



HAL
open science

Collaborative distributed hypothesis testing with general hypotheses

Gil Katz, Pablo Piantanida, Merouane Debbah

► **To cite this version:**

Gil Katz, Pablo Piantanida, Merouane Debbah. Collaborative distributed hypothesis testing with general hypotheses. 2016 IEEE International Symposium on Information Theory (ISIT), Jul 2016, Barcelona, Spain. 10.1109/ISIT.2016.7541590 . hal-01436846

HAL Id: hal-01436846

<https://hal.science/hal-01436846>

Submitted on 23 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Collaborative Distributed Hypothesis Testing with General Hypotheses

Gil Katz*, Pablo Piantanida[†] and Mérouane Debbah*[‡]

*Large Systems and Networks Group (LANEAS), CentraleSupélec-CNRS-Université Paris-Sud, Gif-sur-Yvette, France

[†]Laboratoire des Signaux et Systèmes (L2S), CentraleSupélec-CNRS-Université Paris-Sud, Gif-sur-Yvette, France

[‡]Mathematical and Algorithmic Sciences Lab, Huawei France R&D, Paris, France

{gil.katz,pablo.piantanida,merouane.debbah}@centralesupelec.fr

Abstract—The problem of collaborative distributed hypothesis testing is investigated. In this setting, a binary decision is required about the joint distribution of two arbitrary dependent memoryless processes that are sampled at different physical locations (nodes) in the system. Interactive rate-limited communication is allowed between these nodes. Defining two types of error events, the error exponent for an error of the second type is investigated, under a prescribed probability of error of the first type. A general achievable error exponent, as a function of the total available communication resources, is proposed, for the case of two general hypotheses. The special case of *testing against independence* is revisited for which it is shown that optimality can be attained, as a special case of the general achievable exponent, provided the constraint over the error probability of the first type goes to zero.

I. INTRODUCTION

The field of hypothesis testing (HT) is composed of problems where a statistician is required to declare the joint probability distribution of a random process based on observed samples that are available at the decision end. When binary HT is concerned the choice of the statistician is between two probability distributions (or hypotheses), referred to as H_0 and H_1 . The performance of the test depends on the probabilities of two error events: An error of the first type, with probability α_n , occurs when H_1 is declared while H_0 is the true hypothesis; while an error of the second type, with probability β_n , is defined to be the opposite event. Typically, the trade-off between these probabilities is investigated through the exponential rate of decay with n (sample length) by which β_n vanishes to 0, while α_n is constrained to be below a given $\varepsilon > 0$.

The optimal exponential rate of decay for $\beta_n(\varepsilon)$ in the case of single-variable HT is given by *Stein's Lemma* [1] to be:

$$\theta_n(\varepsilon) \triangleq - \lim_{n \rightarrow \infty} \frac{1}{n} \log \beta_n^*(\varepsilon) = \mathcal{D}(P_{0,X} \| P_{1,X}), \quad (1)$$

where $\mathcal{D}(\cdot \| \cdot)$ is the *Kullback-Leiber* (KL) divergence, and $P_{0,X}, P_{1,X}$ are the distributions implied by hypotheses H_0 and H_1 , respectively. As a matter of fact, in this case the optimal error-exponent $\mathcal{D}(P_{0,X} \| P_{1,X})$ for β_n^* is not dependent on the specific constraint imposed over $\alpha_n \leq \varepsilon$.

When dealing with distributed HT scenarios with constraints due to communication between the different locations in the system, the problem naturally requires an

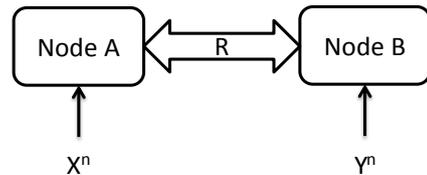


Fig. 1: Two-node detection model with a bidirectional error-free communication link.

interdisciplinary approach between information-theoretic and statistical tools. In [2], [3], a two-node system is considered, much like the one depicted in Fig. 1, where only unidirectional communication is permitted, from node A to node B . Here, node A observes n i.i.d. realizations of a random variable (RV) X , while node B has access to n i.i.d. realizations of the RV Y . After receiving information from node A about its observations, node B declares its decision based on the received message as well as its own observations. While optimal results remain allusive for the general case, an interesting achievable scheme was proposed in [3] which is based on a type-by-type analysis [4].

Recent results in [5], [6] build upon the seminal works [2], [3]. Random binning is proposed as a method of improving on previous achievable schemes. This approach, which was first identified in [7], turns out to be in general beneficial. Indeed, [8] investigates the case of *testing against conditional independence* where under H_1 the observations X and Y are assumed to be statistically independent given some side information Z , and shows that in this case binning is optimal. The joint problem of detection and source reconstruction, where after making a decision node B is required to also reconstruct the source from A (subject to an average fidelity criterion), was also studied in [6], [9].

In this paper, we allow both nodes to communicate with each other, under a constraint over the total amount of information passed through the error-free link, which is now assumed to be bidirectional. We then study the exponent-rate region for the case of two general hypotheses (not necessarily testing against independence) and derive a general achievable region that assume multiple rounds of communication before making the decision. It is worth mentioning that a similar scenario was first studied in [10] for the case of testing against independence, where it was

assumed that $P_{1,XY}(x, y) = P_{0,X}(x)P_{0,Y}(y)$, i.e., the alternative hypothesis H_1 implies a product distribution. The exponent-rate region was derived assuming that only one round of communication was allowed.

The remainder of this paper is organized as follows. In section II the system model is presented. Section III presents our main result, namely an achievable error exponent for the interactive case with general hypotheses. This result is then proved in Section IV, before the case of testing against independence is revisited in Section V. Concluding remarks are discussed in Section VI.

II. SYSTEM MODEL AND PRELIMINARIES

A. Notation

Random variables are denoted by upper-case letters, while lower-case letters denote their realizations. Vectors are denoted by bold-face letters, with their length as a superscript. Sets, including alphabets of RVs, are denoted by calligraphic letters. $P_X \in \mathcal{P}(\mathcal{X})$ denotes a probability distribution over the RV X , out of the set of all possible probability distributions over \mathcal{X} . We consider only RVs on finite alphabets.

For a given vector of realizations \mathbf{x} , $Q_{\mathbf{x}}$ represents its empirical distribution, or *type*. The set of all vectors $\tilde{\mathbf{x}} \in \mathcal{X}^n$ that share the type $Q_{\mathbf{x}}$ is denoted by $\mathcal{T}(Q_{\mathbf{x}})$. The set of all δ -typical sequences of length n for the RV $X \sim P_X(x)$ is denoted by $\mathcal{T}_{[X]_{\delta}}^n$, where δ -typicality refers to strong typicality (see e.g., [3], [11]). $X^{(n)}$ is used to refer to the *type variable* of a vector \mathbf{x} , i.e., a variable whose distribution is defined by the empirical distribution of \mathbf{x} . We use $H(P_X) = H(X) = \mathbb{E}[-\log P_X(x)]$ to denote the entropy of X , $I(X; Y)$ to denote the mutual information between X and Y , and $\mathcal{D}(P_X \| Q_X) = \sum_{x \in \mathcal{X}} P_X(x) \log \frac{P_X(x)}{Q_X(x)}$ to denote the KL-divergence between two probability distributions in $\mathcal{P}(\mathcal{X})$. $X \circlearrowleft Y \circlearrowleft Z$ is used to denote that the RVs X, Y and Z form a Markov chain. Finally, logarithms and exponents are assumed to be of base 2.

B. System Model

We consider a two-node system, as depicted in Fig. 1. Node A observes n realizations of a RV X , while node B sees n realizations of Y . The two RVs are jointly distributed according to one of two possible probability distributions, $P_{0,XY}(x, y) = P_{XY}(x, y)$ and $P_{1,XY}(x, y) = P_{\bar{X}\bar{Y}}(x, y)$, referred to as hypothesis 0 (H_0) and hypothesis 1 (H_1), respectively. We further assume that the marginal distribution of each of the RVs is identical under each of the hypotheses, i.e., $P_{0,X}(x) = P_{1,X}(x)$ and $P_{0,Y}(y) = P_{1,Y}(y)$, where $P_{i,X}(x) = \sum_{y \in \mathcal{Y}} P_{i,XY}(x, y)$ is the marginal distribution of X under hypothesis i , and $P_{i,Y}$ is defined in the same manner. In addition, X and Y are assumed to be jointly independent and identically distributed (i.i.d) in time, under each of the possible hypotheses.

Node A and node B are connected by a bidirectional link with rate $R \left[\frac{\text{bits}}{\text{symbol}} \right]$. Thus, if each node sees n realizations, it is assumed that 2^{nR} bits can be passed on the link. The link is assumed to be perfect, and thus does not introduce errors, as long as the rate limit is respected. In addition, it is assumed that the channel is oblivious

to the direction in which it is used. The purpose of the participants is to declare which of the two hypotheses is the correct one. It is assumed, however, that the location at which the decision is made is unimportant. Notice that this assumption is not restrictive, as the decision can always be transmitted at no cost (zero rate). We take the conventional definition in literature of the two possible error events discussed above (see e.g., [2], [3], [6], [10]):

$$\begin{aligned} \alpha_n &= \Pr(H_1 | XY \sim P_{0,XY}) , \\ \beta_n &= \Pr(H_0 | XY \sim P_{1,XY}) . \end{aligned} \quad (2)$$

In the remainder of this paper, we study the exponential rate with which β_n vanishes to 0, under a fixed constraint over the probability of error of the first type, $\alpha_n \leq \varepsilon$, $\varepsilon > 0$.

III. MAIN RESULT

We now present our main result, followed by an outline of the proof in the next section. We choose to concentrate in this paper on the case where only one round of communication is allowed between the nodes, with node A starting the interaction. The case where *any finite number* of communication rounds is allowed, as well as a comprehensive proof for both cases, can be found in [11].

Proposition 1 (Achievable error exponent with one round of communication). *Define the following sets of auxiliary RVs over finite alphabets:*

$$\begin{aligned} \mathcal{S}(R) &\triangleq \{(UV) : I(U; X) + I(V; Y|U) \leq R, \\ &U \circlearrowleft X \circlearrowleft Y, V \circlearrowleft (U, Y) \circlearrowleft X\} , \end{aligned} \quad (3)$$

$$\begin{aligned} \mathcal{L}(U, V) &\triangleq \{(\tilde{U}\tilde{V}\tilde{X}\tilde{Y}) : P_{\tilde{U}\tilde{V}\tilde{X}}(u, v, x) = P_{UVX}(u, v, x), \\ &P_{\tilde{U}\tilde{V}\tilde{Y}}(u, v, y) = P_{UVY}(u, v, y), \\ &\forall (u, v, x, y) \in \mathcal{U} \times \mathcal{V} \times \mathcal{X} \times \mathcal{Y}\} . \end{aligned} \quad (4)$$

An achievable error exponent for the error probability of the second type, when the error probability of the first type is subject to satisfy $\alpha_n \leq \varepsilon$, is given by:

$$\begin{aligned} - \lim_{n \rightarrow \infty} \frac{1}{n} \log \beta_n &\geq \\ &\max_{U, V \in \mathcal{S}(R)} \min_{\tilde{U}\tilde{V}\tilde{X}\tilde{Y} \in \mathcal{L}(U, V)} \mathcal{D}(P_{\tilde{U}\tilde{V}\tilde{X}\tilde{Y}} \| P_{\tilde{U}\tilde{V}\tilde{X}\tilde{Y}}) . \end{aligned} \quad (5)$$

The strategy that leads to this achievable region is inspired by the one proposed in [10] for the case of testing against independence, while analysis relies on a type-by-type approach, as was done in [3] for the case of general hypotheses with unidirectional communication. Note that the constraint over the rate in expression (3) implies that the total available rate can be divided freely between the messages. It is clear, that by choosing $I(U; X) = R$, $V = \phi$, this result includes the unidirectional case in [3].

IV. PROOF OUTLINE

We now give an outline to the proof of Proposition 1. We start by describing the codebook construction, as well as encoding and decoding strategies. This is followed by an analysis of the probability of error.

A. Codebook Construction

Without loss of generality, we assume that node A is the first to communicate. Let the conditional probability distribution $P_{UV|XY}(u, v|x, y) = P_{U|X}(u|x) \times P_{V|UY}(v|u, y)$ be the one that attains the maximum in Proposition 1. Let $P_U(u) = \sum_{x \in \mathcal{X}} P_{U|X}(u|x)P_X(x)$ and $P_{V|U}(v|u) = \sum_{y \in \mathcal{Y}} P_{V|UY}(v|u, y)P_Y(y)$. Based on these RVs, define $I(U; X) + \epsilon(\delta) = R_U$ and $I(V; Y|U) + \epsilon(\delta') = R_V$, with $\epsilon(\delta) \rightarrow 0$ as $\delta \rightarrow 0$. By definition of the set $\mathcal{S}(R)$, it is clear that $R_U + R_V \leq R + \epsilon(\delta) + \epsilon(\delta')$. Randomly and independently draw 2^{nR_U} sequences from U , each according to the PD $\prod_{i=1}^n P_U(u_i)$. Index these sequences $\mathbf{u}(m_U)$ with $m_U \in [1, M_U = 2^{nR_U}]$. As a second step, for each word in the codebook of U , build a codebook $\mathcal{C}_{\mathbf{u}(m_U)}$ by randomly and independently drawing 2^{nR_V} sequences from V , each according to the PD $\prod_{i=1}^n P_{V|U}(v_i|u_i(m_U))$. Index them by $\mathbf{v}(m_U, m_V)$ with $m_V \in [1, M_V = 2^{nR_V}]$.

B. Encoding and Decoding Strategies

Given a sequence \mathbf{x} , node A searches in the codebook of U for an index m_U such that $(\mathbf{u}(m_U), \mathbf{x}) \in \mathcal{T}_{[UX]}^n$. If no such index is found, node A declares H_1 . If more than one sequence is found, node A chooses one at random. Node A then sends the chosen index m_U to node B , using a portion R_U of the available sum-rate. Upon receiving the index m_U , node B checks if $(\mathbf{u}(m_U), \mathbf{y}) \in \mathcal{T}_{[UY]}^n$. If not, node B declares H_1 . If the received sequence \mathbf{u} and the observed sequence at node B , \mathbf{y} , are jointly typical, node B looks in the specific codebook constructed for message $\mathbf{u}(m_U)$, $\mathcal{C}_{\mathbf{u}(m_U)}$, for an index m_V such that $(\mathbf{u}(m_U), \mathbf{v}(m_U, m_V), \mathbf{y}) \in \mathcal{T}_{[UVY]}^n$. If such an index is not found, node B declares H_1 . If node B finds more than one such index, it chooses one of them at random. Node B then sends the chosen index m_V back to node A . Upon reception of the index m_V , node A checks if $(\mathbf{u}(m_U), \mathbf{v}(m_U, m_V), \mathbf{x}) \in \mathcal{T}_{[UVX]}^n$. If so, it declares H_0 , else it declares H_1 . δ , δ' and δ'' are connected to each other through the generalized Markov lemma (see e.g., [6]). It is important to emphasize that $\delta'(\delta) \rightarrow 0$ when $\delta \rightarrow 0$, and $\delta''(\delta') \rightarrow 0$ when $\delta' \rightarrow 0$.

C. Analysis of α_n

The analysis of α_n is identical to the one proposed in [10], for the case of testing against independence. For sake of clarity, we give here a short summary of this analysis. Assuming that the distribution that controls X and Y is $P_{0,XY}$, and denoting the chosen indices at nodes A and B by m_U and m_V , respectively, the error probability of the first type can be expressed as follows:

$$\begin{aligned} \alpha_n &\equiv \Pr(\mathcal{E}_1 \cup \mathcal{E}_2 \cup \mathcal{E}_3) \\ &\leq \Pr(\mathcal{E}_1) + \Pr(\mathcal{E}_1^c \cap \mathcal{E}_2) + \Pr(\mathcal{E}_1^c \cap \mathcal{E}_2^c \cap \mathcal{E}_3), \end{aligned} \quad (6)$$

where $\mathcal{E}_1, \mathcal{E}_2$ and \mathcal{E}_3 represent the following error events:

$$\begin{aligned} \mathcal{E}_1 &\equiv \{(\mathbf{U}^n(m_U), \mathbf{X}^n) \notin \mathcal{T}_{[UX]}^n, \forall m_U \in [1, M_U]\}, \\ \mathcal{E}_2 &\equiv \{(\mathbf{V}^n(m_U, m_V), \mathbf{U}^n(m_U), \mathbf{Y}^n) \notin \mathcal{T}_{[UVY]}^n, \\ &\quad \forall m_V \in [1, M_V] \text{ and the specific } m_U \text{ sent from node } A\}, \\ \mathcal{E}_3 &\equiv \{(\mathbf{V}^n(m_U, m_V), \mathbf{U}^n(m_U), \mathbf{X}^n) \notin \mathcal{T}_{[UVX]}^n, \\ &\quad \text{for the specific } m_U \text{ and } m_V \text{ previously chosen}\}. \end{aligned} \quad (7)$$

Analyzing each of the probabilities in (6) separately, $\Pr(\mathcal{E}_1)$ tends to 0 as $n \rightarrow \infty$ by the covering lemma [12], if $R_U \geq I(U; X) + \epsilon(\delta)$, with $\epsilon(\delta) \rightarrow 0$ as $\delta \rightarrow 0$. $\Pr(\mathcal{E}_1^c \cap \mathcal{E}_2) \rightarrow 0$ when $n \rightarrow \infty$ by the conditional typicality lemma [12], in addition to the covering lemma, if $R_V \geq I(V; Y|U) + \epsilon(\delta')$. Finally, the third term in (6) can be shown to tend to zero through the use of the Markov lemma (see e.g., [6, Lemma 5]), along with the fact that

$$P_{X^n}(\mathbf{x}) = \exp\left\{-n \left(H(X^{(n)}) + \mathcal{D}(X^{(n)} \| P_X)\right)\right\}, \quad (8)$$

where $X^{(n)}$ is the type variable (or empirical distribution) of the vector \mathbf{x} . Thus, as all three components tend to zero with large n , we may conclude that $\alpha_n \leq \varepsilon$ for an arbitrary desired constraint $\varepsilon > 0$ and n large enough.

D. Analysis of β_n

The probability of error of the second type is defined to be $\beta_n \equiv \Pr(H_0 | XY \sim P_{1,XY})$. Thus, we assume that $P_{1,XY}$ controls the distribution of the observed RVs throughout this analysis. We follow a similar approach to the one used in [3]. As the region proposed in Proposition 1 calls for a maximization over $\mathcal{S}(R)$, it is enough to show that $\min_{\mathcal{L}(U,V)} \mathcal{D}(P_{\bar{U}\bar{V}\bar{X}\bar{Y}} \| P_{UVXY})$ is an achievable error exponent for any choice of $(U, V) \in \mathcal{S}(R)$.

For a given pair of sequences (\mathbf{x}, \mathbf{y}) with type variables $(X^{(n)}Y^{(n)})$, we count all possible events that lead to an error. We notice first, that given a pair of vectors (\mathbf{x}, \mathbf{y}) the probability that these vectors will be the result of n draws, according to the distribution implied by H_1 , can be calculated through [3, Lemma 3] to be:

$$\begin{aligned} \Pr\{\bar{X}\bar{Y} = (\mathbf{x}, \mathbf{y})\} \\ = \exp\left[-n \left(H(X^{(n)}Y^{(n)}) + \mathcal{D}(X^{(n)}Y^{(n)} \| \bar{X}\bar{Y})\right)\right], \end{aligned} \quad (9)$$

where $X^{(n)}Y^{(n)}$ are the type variables of the realizations (\mathbf{x}, \mathbf{y}) . For each pair of codewords in the codebooks \mathbf{u}_i and \mathbf{v}_j , we define the set:

$$S_{i,j}(\mathbf{x}) \equiv \{\mathbf{u}_i\} \times \{\mathbf{v}_j\} \times \mathcal{G}_i \times \{\mathbf{x}\}, \quad (10)$$

where \mathcal{G}_i is the set of all vectors \mathbf{y} that, given the received message \mathbf{u}_i , will result in the message \mathbf{v}_j being transmitted back to node A . Denoting by $K_{i,j}(\mathbf{x})$ the number of elements $(\mathbf{u}_i, \mathbf{v}_j, \mathbf{x}, \mathbf{y}) \in S_{i,j}(\mathbf{x})$ whose type variables coincide with $(U^{(n)}V^{(n)}X^{(n)}Y^{(n)})$, this term can be bounded by:

$$K_{i,j}(\mathbf{x}) \leq \exp\left[nH(Y^{(n)}|U^{(n)}V^{(n)}X^{(n)})\right]. \quad (11)$$

Letting $K(U^{(n)}V^{(n)}X^{(n)}Y^{(n)})$ to denote the number of all elements:

$$(\mathbf{u}, \mathbf{v}, \mathbf{x}, \mathbf{y}) \in \mathcal{S}_n \equiv \bigcup_{i=1}^{M_U} \bigcup_{j=1}^{M_V} \bigcup_{\mathbf{x} \in \mathcal{T}_{[X|\mathbf{u}_i \mathbf{v}_j]}^n} S_{i,j}(\mathbf{x}) \quad (12)$$

that have $U^{(n)}V^{(n)}X^{(n)}Y^{(n)}$ as their type variable. It is not difficult to check that (12) can be bounded as:

$$\begin{aligned} & K(U^{(n)}V^{(n)}X^{(n)}Y^{(n)}) \\ & \leq \sum_{i=1}^{M_U} \sum_{j=1}^{M_V} \exp \left[n H(Y^{(n)}|U^{(n)}V^{(n)}X^{(n)}) \right] |\mathcal{T}_{[X|u_j v_i] \delta''}^n| \\ & \leq \exp \left[n \left(H(Y^{(n)}|U^{(n)}V^{(n)}X^{(n)}) \right. \right. \\ & \quad \left. \left. + I(U; X) + I(V; Y|U) + H(X|UV) + \mu_n \right) \right], \end{aligned} \quad (13)$$

where M_U and M_V are the sizes of the codebooks of both messages, which are dictated by the respective rates. The first and second additional terms in the final expression come from the size of the codebooks and the third is a bound over the size of the delta-typical set. μ_n is a function of $\delta, \delta', \delta''$ that complies with $\mu_n \rightarrow 0$ when $n \rightarrow \infty$. The probability of error of the second type can thus be bounded by:

$$\beta_n \leq \sum \exp \left[-n \left(k(U^{(n)}V^{(n)}X^{(n)}Y^{(n)}) - \mu_n \right) \right], \quad (14)$$

where the function $k(U^{(n)}V^{(n)}X^{(n)}Y^{(n)})$ defined as

$$\begin{aligned} & k(U^{(n)}V^{(n)}X^{(n)}Y^{(n)}) \\ & = H(X^{(n)}Y^{(n)}) + \mathcal{D}(X^{(n)}Y^{(n)} \|\bar{X}\bar{Y}) \\ & \quad - H(Y^{(n)}|U^{(n)}V^{(n)}X^{(n)}) - H(X|UV) \\ & \quad - I(U; X) - I(V; Y|U), \end{aligned} \quad (15)$$

and the sum is taken over all possible type-variables $U^{(n)}V^{(n)}X^{(n)}Y^{(n)}$ of elements $(\mathbf{u}, \mathbf{v}, \mathbf{x}, \mathbf{y})$ in \mathcal{S}_n .

Through considerations of type counting (i.e., the number of possible types grows polynomially, and not exponentially, in n) and continuity similar to the ones in [3], the sum in (14) can be replaced by a maximum operator over the set $\mathcal{L}(U, V)$ in Proposition 1. Considering the error exponent itself, we can thus conclude that

$$-\lim_{n \rightarrow \infty} \frac{1}{n} \log \beta_n \geq \min_{\mathcal{L}(U, V)} k(\tilde{U}, \tilde{V}, \tilde{X}, \tilde{Y}), \quad (16)$$

with

$$\begin{aligned} & k(\tilde{U}, \tilde{V}, \tilde{X}, \tilde{Y}) = H(\tilde{X}\tilde{Y}) + \mathcal{D}(\tilde{X}\tilde{Y} \|\bar{X}\bar{Y}) \\ & \quad - H(\tilde{Y}|\tilde{U}\tilde{V}\tilde{X}) - H(\tilde{X}|\tilde{U}\tilde{V}) - I(\tilde{U}; \tilde{X}) - I(\tilde{V}; \tilde{Y}|\tilde{U}). \end{aligned} \quad (17)$$

Analyzing this expression, it can be shown that

$$\begin{aligned} & k(\tilde{U}, \tilde{V}, \tilde{X}, \tilde{Y}) = \mathcal{D}(P_{\tilde{U}\tilde{X}\tilde{Y}} \| P_{\tilde{U}\tilde{X}\tilde{Y}}) + I(\tilde{X}; \tilde{V}|\tilde{U}\tilde{Y}) \\ & \quad = \mathcal{D}(P_{\tilde{U}\tilde{V}\tilde{X}\tilde{Y}} \| P_{\tilde{U}\tilde{V}\tilde{X}\tilde{Y}}), \end{aligned} \quad (18)$$

which completes the proof.

V. REVISITING THE CASE OF TESTING AGAINST INDEPENDENCE

We now revisit the case of testing against independence, which was studied in [10]. In this case, it is assumed that under hypothesis 1, the RVs X and Y are distributed according to a product distribution:

$$\begin{cases} H_0 : & P_{0,XY}(x, y) = P_{XY}(x, y), \\ H_1 : & P_{1,XY}(x, y) = P_{\bar{X}\bar{Y}}(x, y) = P_X(x)P_Y(y). \end{cases} \quad (19)$$

The *optimal* error exponent for an error of the second type in this case, when the constraint over the error of the first type is very small, $\epsilon \rightarrow 0$, is given by [10] to be

$$-\lim_{n \rightarrow \infty} \frac{1}{n} \log \beta_n^*(\epsilon) = \max_{(U, V) \in \mathcal{S}(R)} I(U; Y) + I(V; X|U), \quad (20)$$

where $\mathcal{S}(R)$ is defined as in Proposition 1. While the achievability proof in [10] inspired our approach taken in this paper for the case of general hypotheses, unfortunately, the auxiliary RVs identified in the converse proof do not match the required Markov chains (3) (see [13] for further details). In order to overcome this issue, we next provide a converse proof by carefully selecting those RVs.

First, we show that this result is contained by the achievable region of Proposition 1, when testing against independence is considered. We then discuss the main steps in the proof of the converse part, which can be found in [11].

A. Proof of Achievability

In order to prove the achievability of the region in (20) through the general region in Proposition 1, it is convenient to start by integrating the expression given in (18) for $k(\tilde{U}, \tilde{V}, \tilde{X}, \tilde{Y})$ into (16):

$$\begin{aligned} & -\lim_{n \rightarrow \infty} \frac{1}{n} \log \beta_n \\ & \geq \max_{\mathcal{S}(R)} \min_{\mathcal{L}(U, V)} \left[\mathcal{D}(P_{\tilde{U}\tilde{X}\tilde{Y}} \| P_{\tilde{U}\tilde{X}\tilde{Y}}) + I(\tilde{X}; \tilde{V}|\tilde{U}\tilde{Y}) \right]. \end{aligned} \quad (21)$$

We analyze each of these components separately:

$$\begin{aligned} & \mathcal{D}(P_{\tilde{U}\tilde{X}\tilde{Y}} \| P_{\tilde{U}\tilde{X}\tilde{Y}}) \\ & \stackrel{(a)}{=} \mathcal{D}(P_{\tilde{U}\tilde{Y}} \| P_{\tilde{U}\tilde{Y}}) + \mathcal{D}(P_{\tilde{X}|\tilde{U}\tilde{Y}} \| P_{\tilde{X}|\tilde{U}\tilde{Y}} | P_{\tilde{U}\tilde{Y}}) \\ & \stackrel{(b)}{=} I(U; Y) + \mathcal{D}(P_{\tilde{X}|\tilde{U}\tilde{Y}} \| P_{\tilde{X}|\tilde{U}} | P_{\tilde{U}\tilde{Y}}) \\ & \stackrel{(c)}{=} I(U; Y) + \mathcal{D}(P_{\tilde{X}|\tilde{U}\tilde{Y}} \| P_{\tilde{X}|\tilde{U}} | P_{\tilde{U}\tilde{Y}}), \end{aligned} \quad (22)$$

where (a) is due to the chain rule, $\mathcal{D}(P_{\tilde{X}|\tilde{U}\tilde{Y}} \| P_{\tilde{X}|\tilde{U}\tilde{Y}} | P_{\tilde{U}\tilde{Y}})$ is the conditional KL-divergence; (b) stems from the assumption of testing against independence, as well as the Markov chain $\tilde{U} \text{---} \tilde{X} \text{---} \tilde{Y}$ and the fact that $P_{\tilde{U}\tilde{Y}} = P_{UY}$; and (c) is due to the fact that $P_{\tilde{U}\tilde{X}} = P_{UX} = P_{\tilde{U}\tilde{X}}$. This is true because of the definition of the set $\mathcal{L}(U, V)$, in addition to the fact that the marginal distribution of X , $P_X(x)$ is identical for both hypotheses, and the encoding at node A is done without prior knowledge of the true hypothesis (thus $P_{\tilde{U}|\tilde{X}}(u|x) = P_{U|X}(u|x)$). By using similar considerations it is straight-forward to show that

$$\mathcal{D}(P_{\tilde{U}\tilde{X}\tilde{Y}} \| P_{\tilde{U}\tilde{X}\tilde{Y}}) = I(U; Y) + I(\tilde{X}; \tilde{Y}|\tilde{U}). \quad (23)$$

As for the second term in (21), we express it as follows:

$$\begin{aligned} & I(\tilde{V}; \tilde{X}|\tilde{U}\tilde{Y}) = I(\tilde{V}\tilde{Y}; \tilde{X}|\tilde{U}) - I(\tilde{X}; \tilde{Y}|\tilde{U}) \\ & \geq I(\tilde{V}; \tilde{X}|\tilde{U}) - I(\tilde{X}; \tilde{Y}|\tilde{U}). \end{aligned} \quad (24)$$

This allows us to conclude through (21) that

$$\begin{aligned}
& - \lim_{n \rightarrow \infty} \frac{1}{n} \log \beta_n \\
& \geq \max_{(U,V) \in \mathcal{S}(R)} \min_{\tilde{U} \tilde{V} \tilde{X} \tilde{Y} \in \mathcal{L}(U,V)} \left[I(U; Y) + I(\tilde{V}; \tilde{X} | \tilde{U}) \right] \\
& = \max_{(U,V) \in \mathcal{S}(R)} \left[I(U; Y) + I(V; X | U) \right], \tag{25}
\end{aligned}$$

where the final equality is due to the fact that in the set $\mathcal{L}(U, V)$, $P_{\tilde{U} \tilde{V} \tilde{X}}(u, v, x) = P_{UVX}(u, v, x)$. This completes the proof of achievability through the result for general hypotheses.

B. Proof of Converse

Starting from the multi-letter characterization of the exponent, which constitutes an upper bound to the error exponent of the second type when the constraint over the error probability of first type vanishes to zero, it follows that:

$$\begin{aligned}
& - \lim_{n \rightarrow \infty} \log \beta_n \leq I(I_A; Y^n) + I(I_B; X^n | I_A) \\
& = \sum_{i=1}^n \left[I(I_A; Y_i | Y_{i+1}^n) + I(I_B; X_i | I_A X^{i-1}) \right] \\
& = \sum_{i=1}^n \left[I(I_A Y_{i+1}^n; Y_i) + I(I_B Y_{i+1}^n; X_i | I_A X^{i-1}) \right. \\
& \quad \left. - I(Y_{i+1}^n; X_i | I_A I_B X^{i-1}) \right] \\
& = \sum_{i=1}^n \left[I(I_A X^{i-1} Y_{i+1}^n; Y_i) - I(X^{i-1}; Y_i | I_A Y_{i+1}^n) \right. \\
& \quad \left. + I(Y_{i+1}^n; X_i | I_A X^{i-1}) + I(I_B; X_i | I_A X^{i-1} Y_{i+1}^n) \right. \\
& \quad \left. - I(Y_{i+1}^n; X_i | I_A I_B X^{i-1}) \right] \\
& = \sum_{i=1}^n \left[I(\hat{U}_i; Y_i) + I(V_i; X_i | \hat{U}_i) - I(Y_{i+1}^n; X_i | I_A I_B X^{i-1}) \right]. \tag{26}
\end{aligned}$$

Here, the final equality is due to the Csiszár Sum Identity [12, Chapter 2], and by setting the auxiliary RVs:

$$\begin{cases} \hat{U}_i \equiv (I_A, X^{i-1}, Y_{i+1}^n), \\ V_i \equiv I_B, \end{cases} \tag{27}$$

for all $i = [1 : n]$. Note that under this definition both required Markov chains in (3) are verified (see e.g., [13] for further details). By introducing a time-sharing RV Q and letting $U \equiv (\hat{U}_Q, Q)$, this bound can be rewritten as:

$$\begin{aligned}
& - \lim_{n \rightarrow \infty} \frac{1}{n} \log \beta_n \leq \left[I(\hat{U}_Q; Y_Q | Q) + I(V_Q; X_Q | \hat{U}_Q, Q) \right] \\
& \quad - \frac{1}{n} \sum_{i=1}^n I(Y_{i+1}^n; X_i | I_A I_B X^{i-1}) \\
& \quad \equiv I(U; Y) + I(V; X | U) - t, \tag{28}
\end{aligned}$$

with $t \geq 0$. It can be shown in a similar manner that

$$R \geq I(U; X) + I(V; Y | U) - t. \tag{29}$$

All that is left is to show that the combination of the constraints in (28) and (29) is equivalent to the region

$$\begin{aligned}
& - \lim_{n \rightarrow \infty} \frac{1}{n} \log \beta_n \leq I(U; Y) + I(V; X | U), \\
& R \geq I(U; X) + I(V; Y | U), \tag{30}
\end{aligned}$$

which can be done through Fourier-Motzkin elimination [14], as both regions are convex. Thus, the converse part of the proof is complete.

VI. CONCLUDING REMARKS

We investigated the problem of collaborative distributed hypothesis testing with general hypotheses. While concentrating on the case of a single round of communication in this paper, broadening results to multiple rounds is straight-forward, and can be found in an extended version of this work [11]. A new achievable exponent-rate region for the case of general hypotheses was proposed. The case of testing against independence was then revisited. It was shown that the proposed achievable region is in fact optimal for this case, by first showing that it achieves the previously known result, and then revising the proof of the converse part.

REFERENCES

- [1] E. Lehmann and J. Romano, *Testing Statistical Hypotheses*, ser. Springer Texts in Statistics.
- [2] R. Ahlswede and I. Csiszár, "Hypothesis testing with communication constraints," *Information Theory, IEEE Transactions on*, vol. 32, no. 4, pp. 533–542, Jul 1986.
- [3] T. Han, "Hypothesis testing with multiterminal data compression," *Information Theory, IEEE Transactions on*, vol. 33, no. 6, pp. 759–772, Nov 1987.
- [4] I. Csiszár, "The method of types," *Information Theory, IEEE Transactions on*, vol. 44, no. 6, pp. 2505–2523, Oct 1998.
- [5] G. Katz, P. Piantanida, R. Couillet, and M. Debbah, "On the necessity of binning for the distributed hypothesis testing problem," in *Inf. Theory, 2015 IEEE International Symposium on (ISIT)*, June 2015, pp. 2797–2801.
- [6] G. Katz, P. Piantanida, and M. Debbah, "Distributed binary detection with lossy data compression," *ArXiv e-prints*, Jan. 2016, submitted to *Information Theory, IEEE Trans. on*. [Online]. Available: <http://arxiv.org/abs/1601.01152>
- [7] H. Shimokawa, T. Han, and S.-I. Amari, "Error bound of hypothesis testing with data compression," in *Inf. Theory, 1994 IEEE International Symposium on (ISIT)*, Jun 1994, p. 114.
- [8] S. Rahman and A. Wagner, "On the optimality of binning for distributed hypothesis testing," *Information Theory, IEEE Transactions on*, vol. 58, no. 10, pp. 6282–6303, Oct 2012.
- [9] G. Katz, P. Piantanida, R. Couillet, and M. Debbah, "Joint estimation and detection against independence," in *Communication, Control, and Computing (Allerton), 52nd Annual Allerton Conference on*, 2014.
- [10] Y. Xiang and Y.-H. Kim, "Interactive hypothesis testing with communication constraints," in *Communication, Control, and Computing (Allerton), 2012 50th Annual Allerton Conference on*, Oct 2012, pp. 1065–1072.
- [11] G. Katz, P. Piantanida, and M. Debbah, "Collaborative distributed hypothesis testing," *ArXiv e-prints*, Apr. 2016, submitted to *The Annals of Statistics*. [Online]. Available: <http://arxiv.org/abs/1604.01292>
- [12] A. El Gamal and Y.-H. Kim, *Network information theory*. Cambridge University Press, 2011.
- [13] L. R. Vega, P. Piantanida, and A. O. Hero, "The three-terminal interactive lossy source coding problem," *Information Theory, IEEE Transactions on*, 2015, (revised). [Online]. Available: <http://arxiv.org/abs/1502.01359>
- [14] A. Schrijver, *Theory of linear and integer programming*. John Wiley & Sons, 1998.