Submitted to the Annals of Statistics

COLLABORATIVE DISTRIBUTED HYPOTHESIS TESTING

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

CentraleSupélec and Huawei France

Abstract

A collaborative distributed binary decision problem is considered. Two statisticians are required to declare the correct probability measure of two jointly distributed memoryless process, denoted by $X^n = (X_1, \ldots, X_n)$ and $Y^n = (Y_1, \ldots, Y_n)$, out of two possible probability measures on finite alphabets, namely P_{XY} and $P_{\bar{X}\bar{Y}}$. The marginal samples given by X^n and Y^n are assumed to be available at different locations. The statisticians are allowed to exchange limited amount of data over multiple rounds of interactions, which differs from previous work that deals mainly with unidirectional communication. A single round of interaction is considered before the result is generalized to any finite number of communication rounds. A feasibility result is shown, guaranteeing the feasibility of an error exponent for general hypotheses, through information-theoretic methods. The special case of testing against independence is revisited as being an instance of this result for which also an unfeasibility result is proven. A second special case is studied where zero-rate communication is imposed (data exchanges grow sub-exponentially with n) for which it is shown that interaction does not improve asymptotic performance.

1. Introduction. The field of hypothesis testing (HT) is comprised of different problems, in which the goal is to determine the probability measure (PM) of one or more random variables (RVs), based on a number of available observations. Considering binary HT problems, it is assumed that this choice is made out of two possible hypotheses, denoted the null hypothesis H_0 and the alternative hypothesis H_1. In this setting, two error events may occur: An error of Type I, with probability α_n (dependent on the number

[*]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

MSC 2010 subject classifications: Primary 94A24,94A15; secondary 94A13,68P30

Keywords and phrases: Binary hypothesis testing, Type I and II error rates, Shannon information, Neyman-Pearson lemma, Coding theorem, Distributed processing, Interactive statistical computing, Exchange rate, Converse

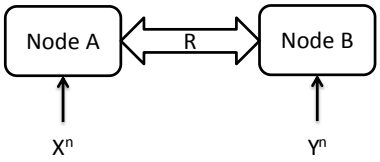

FIGURE 1. *Collaborative Distributed Hypothesis Testing model.*

of observations n), occurs when the alternative hypothesis H_1 is declared while H_0 is true. Conversely, an error of Type II with probability β_n, occurs when H_0 is declared despite H_1 being true. Often, for fixed $0 < \epsilon < 1$, the goal is to find the optimal error exponent:

$$E(\epsilon) \coloneqq \liminf_{n\to\infty} -\frac{1}{n}\log \beta_n(\epsilon) \ , \tag{1}$$

for a constrained error probability of Type I: $\alpha_n \leq \epsilon$.

Let $\{X_i\}_{i=1}^{\infty}$ be an independent and identically distributed (i.i.d) process, commonly refereed to as a *memoryless process*, taking values in a countably finite alphabet $\mathcal{X}$ equipped with probability measures P_0 or P_1 defined on the measurable space $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$, where $\mathcal{B}_{\mathcal{X}} = 2^{\mathcal{X}}$. Denote $X^n = (X_1, \ldots, X_n)$ the finite block of the process following the product measures P_0^n or P_1^n on $(\mathcal{X}^n, \mathcal{B}_{\mathcal{X}^n})$. Let us denote by $\mathcal{P}(\mathcal{X})$ the family of probability measures in $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$, where for every $\mu \in \mathcal{P}(\mathcal{X})$, $f_\mu(x) \coloneqq \frac{d\mu}{d\lambda}(x) = \mu(\{x\})$ is a short-hand for its probability mass function (pmf). The optimal error exponent for the Type II error probability of the binary HT problem is well-known and given by *Stein's Lemma* (see e.g., [15, 7]) to be:

$$E(\epsilon) = \mathcal{D}(P_0||P_1) \ , \ \ \forall\, 0 < \epsilon < 1 \tag{2}$$

where P_0 and P_1 are the probability measures implied by hypotheses H_0 and H_1, respectively, and $\mathcal{D}(\cdot||\cdot)$ is the *Kullback-Leiber divergence* satisfying $P_0 \ll P_1$. The optimal exponential rate of decay of the error probability of Type II does not depend on the specific constraint over the error probability of Type I. This property is referred to as *strong unfeasibility*.

In many scenarios, the realizations of different parts of a random process are available at different physical locations (with different statisticians) in the system (see Fig. 1). Assuming that exchanging data between the statisticians is possible but costly, a new question arises –for a given constraint over the total amount of data exchange between the nodes, what is the optimal

error exponent to the error probability of Type II, under a fixed constraint over the error probability of Type I? In this paper, we compose together two stories. One is from statistics concerning binary HT originating in the works of Wald [24, 25]. The other story is from information theory concerning the case of *unidirectional* data exchanges where only one statistician can share information with the other one due to [2, 11]. We focus on *bidirectional collaborative* binary HT problem. It is assumed that the available resources for interaction can be divided between the statisticians in any way that would benefit performance, and that without loss of generality no importance is given to the location at which the decision is made – as the decision can always be transmitted with sub-exponential resources. First, we concentrate on a special case where only one "round of interaction" (only a query and its reply) is allowed between the statisticians, i.e., a decision is made after each statistician communicates one statistics, which will be commonly referred to as a message. This scenario was first studied in [26] for a special case called *testing against independence*. While the scenario studied in this paper borrows ideas from [26], the mathematical tools are fundamentally different since these rely on the *method of types* [8], as it was the case to deal with general hypothesis in [11]. We then extend our result for any finite number of interaction rounds, before showing that this new result for general hypotheses implies the special case of testing against independence, for which optimality is proven via an *unfeasibility* property.

The remainder of this paper is organized as follows. We finish this introduction with a short summary of related results, before presenting the considered statistical model in Section 2. In Section 3, we present and prove our first result, being a feasible error exponent for the case of general hypotheses and interactive exchanges, under the assumption of a single communication round. Section 4 extends this result to any *finite number* of interaction rounds. In Section 5, we revisit the special case of testing against independence and show that the known exponent for this case is indeed feasible through our general exponent result. Then, we show an unfeasibility property (thus proving optimality, at least in a "weak" sense) for the case of a single communication round. In Section 6, we give the optimal error exponent when communication is constrained to be of *zero rate*, meaning that the sizes of the codebooks grows sub-exponentially with the number of observations n.

1.1. *Summary of related works.* Some of the first contributions on binary HT are due to Wald [24, 25] where an optimal course of action is given by which a sequential probability ration test (SPRT) is used. It was shown that

the expected number of observations required to reach a conclusion is lower than by any other approach, when a similar constraint over the probabilities of error is enforced. Stein's Lemma takes an information-theoretic form since by considering the limit where the number of observations $n \to \infty$, it is shown that the optimal error exponent for the error probability of Type II, under any fixed constraint over the error probability of Type I, is given by the KL divergence. Later [5] proves an important property by which when $\alpha_n \equiv \exp(-nc) \to 0$ as $n \to \infty$, then $\beta_n \to 0$ or $\beta_n \to 1$, exponentially depending on the rate of decay $c > 0$.

Among the first works that started enforcing constraints on the basic HT problem, which are independent from the statistical nature of the data, are references [6, 12]. The single-variable HT is considered, and the enforced constraint is related to the *memory* of the system, rather than to communication between different locations. It is assumed that a realistic system cannot hold a large number of observations for future use, and thus at each step a function must be used that would best encapsulate the "knowledge" gained from the new observation, combined with the compressed representation of previous observations. This problem was then revisited in [27, 4], which are motivated by new scenarios in which memory efficiency is an important aspect, such as satellite communication systems. [4] focuses on the case where both probabilities of error simultaneously decay to zero.

Distributed HT with communication constraints was the focus of the seminal works [2, 11]. Both of them investigated binary decisions in presence of a helper, i.e., unidirectional communication, and propose a feasible error exponent for β_n while enforcing a strict constraint over α_n. Although both of these approaches achieve optimality for the case of testing against independence, where it is assumed that under the alternative hypothesis H_1 the samples from (X, Y) are independent with the same marginal measures implied by H_0, optimal results for the case of general hypotheses remain allusive until this day. Improving these results by using further randomization of the codebooks, referred to "random binning", was first briefly suggested in [22] and analyzed thoroughly in [14]. In [1] a similar scenario is considered for parameter estimation with unidirectional communication. This is a generalization of the binary HT problem where the mean square-error loss was considered instead of exponential decay of the error probability.

A special case referred to as HT under "complete data compression" was studied in [11]. In this case, it is assumed that node A is allowed to communicate with node B by sending only one bit of information. A feasible scheme was proposed and its optimality proved. The much broader scenario, by which codebooks are allowed to grow with n, but not exponentially fast,

was studied in [21]. Interestingly, it was shown that this scenario does not offer any advantage, with relation to complete data compression. This setting, referred to as zero-rate communication, was recently revisited in [28] where both α_n and β_n are required to decrease exponentially with n.

Interactive communication was considered for the problem of distributed binary HT within the framework of testing against independence in [26]. In the present paper, we further study this problem in the framework of general hypotheses, as well as revisit the special case of testing against independence via a strong unfeasibility proof. Other works in recent years evolve the problem of HT in many different directions. Two interesting examples are [17] (see references therein), which assumes a tighter control by the statistician throughout the process, allowing him to choose and evaluate the testing procedure through past information, and [18] which investigates HT in the framework of quantum statistical models.

2. Statistical Model and Preliminaries.

2.1. *Notation.* We use upper-case letters to denote random variables (RVs) and lower-case letters to denote realizations of RVs. Vectors are denoted by boldface letters, with their length as a superscript, emitted when it is clear from the context. Sets, including alphabets of RVs, are denoted by calligraphic letters. Throughout this paper we assume all RVs have an alphabet of finite cardinality. $P_X \in \mathcal{P}(\mathcal{X})$ denotes a probability measure (PM) for the RV $X \in \mathcal{P}(\mathcal{X})$ defined on the measurable space $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$, that belongs to the set of all possible PMs over $\mathcal{X}$; $X \multimap Y \multimap Z$ denotes that X, Y and Z form a Markov chain. We shall use tools from information theory. Notations generally comply with the ones introduced in [8]. Thus, for a RV X, distributed by $X \sim P_X(x)$, the *entropy* is defined to be $H(X) = H(P) \coloneqq -\sum_{x\in\mathcal{X}} P_X(x) \log P_X(x)$. Similarly, the *conditional entropy*:

$$H(Y|X) = H(V|P) \coloneqq -\sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{X}} P_X(x)V(y|x) \log V(y|x)$$

for a stochastic mapping $V : \mathcal{X} \mapsto \mathcal{P}(\mathcal{Y})$. The *conditional Kullback-Leiber (KL) divergence* between two stochastic mappings $P_{Y|X} : \mathcal{X} \mapsto \mathcal{P}(\mathcal{Y})$ and $Q_{Y|X} : \mathcal{X} \mapsto \mathcal{P}(\mathcal{Y})$, is:

$$\mathcal{D}(P_{Y|X}\|Q_{Y|X}|P_X) \coloneqq \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} P_X(x)P_{Y|X}(y|x) \log \frac{P_{Y|X}(y|x)}{Q_{Y|X}(y|x)} \ , \tag{3}$$

satisfying that $P_{Y|X} \ll Q_{Y|X}$ *a.e.* wrt P_X. For any two RVs, X and Y, whose measure is controlled by $XY \sim P_{XY}(x, y) = P_X(x)P_{Y|X}(y|x)$, the

following is defined to be the *mutual information* between them: $I(X;Y) \coloneqq \mathcal{D}(P_{XY}\|P_XP_Y)$. Given a vector $\mathbf{x} = (x_1, \ldots, x_n) \in \mathcal{X}^n$, let $N(a|\mathbf{x})$ be the *counting measure*, i.e., the number of times the letter $a \in \mathcal{X}$ appears in the vector X. The *type* of the vector $\mathbf{x}$, denoted by $Q_{\mathbf{x}}$, is defined through its *empirical measure*: $Q_{\mathbf{x}}(a) = n^{-1}N(a|\mathbf{x})$ with $a \in \mathcal{X}$. $\mathcal{P}_n(\mathcal{X})$ denotes the set of all possible types (or empirical measures) of length n over $\mathcal{X}$. We use type variables of the form $X^{(n)} \in \mathcal{P}_n(\mathcal{X})$ to denote a RV with a probability measure identical to the empirical measure induced by $\mathbf{x}$. The set of all vectors $\mathbf{x}$ that share this type is denoted by $\mathcal{T}(Q_{\mathbf{x}}) = \mathcal{T}_{[Q_{\mathbf{x}}]}$. Main definitions of *$\delta$-typical sets* and some of their properties, are given in Appendix A. All exponents and logarithms are assumed to be of base 2.

2.2. *Statistical model and problem statement.* In a system comprising two statisticians, as depicted in Fig. 1, each of them is assumed to observe the i.i.d. realizations of one random variable. Let $X^nY^n = (X_1, Y_1), \ldots, (X_n, Y_n)$ be independent random variables in $(\mathcal{X}^n \times \mathcal{Y}^n, \mathcal{B}_{\mathcal{X}^n \times \mathcal{Y}^n})$ that are jointly distributed in one of two ways, denoted by hypothesis 0 and 1, with probability measures as follows:

$$\begin{cases} H_0: & P_{XY}(x,y)\ , \forall\,(x,y) \in \mathcal{X} \times \mathcal{Y}\ , \\ H_1: & P_{\bar{X}\bar{Y}}(x,y)\ , \forall\,(x,y) \in \mathcal{X} \times \mathcal{Y}\ . \end{cases} \tag{4}$$

Communication between the two statisticians is assumed to be done in rounds, with node A starting the interaction. These interactions are limited, however, by a total (exponential) rate R bits per symbol. That is, if each of the nodes sees n realizations, the total amount of bits allowed to exchange data between the nodes before the decision is made is $\exp(nR)$. The data exchange is assumed to be *perfect*, meaning that within the rate limit no errors are introduced by the communication. It is assumed that the total rate can be distributed by the two statisticians in any way that is beneficial to performance. Moreover, we assume that it does not matter *where* the decision is finally made, as its transmission can be done at no cost.

As is the case in the standard centralized HT problem, we consider two error events. An error of the Type I, with probability α_n, occurs when H_1 is declared despite H_0 being true, while an error event of Type II, with probability β_n, is the opposite error event. The goal is to find the exponential rate: $-\frac{1}{n}\log\beta_n$ (n being the number of samples) s.t. $\beta_n \to 0$ as $n \to \infty$, while fixed constraints are enforced on α_n and the total exchange rate R.

Definition 1 (K-round collaborative HT). *A K-round decision code for the two node collaborative hypothesis testing system, when each of the*

statisticians is allowed to observe X^n and Y^n realizations of X and Y, respectively, is defined by a sequence of encoders and a decision mapping:

$$f_{[k]} : \mathcal{X}^n \times \prod_{i=1}^{k-1} \{1, \ldots, |g_{[i]}|\} \longrightarrow \{1, \ldots, |f_{[k]}|\} \ , \ k = [1 : K] \tag{5}$$

$$g_{[k]} : \mathcal{Y}^n \times \prod_{i=1}^{k} \{1, \ldots, |f_{[i]}|\} \longrightarrow \{1, \ldots, |g_{[k]}|\} \ , \ k = [1 : K] \tag{6}$$

$$\phi : \mathcal{X}^n \times \prod_{i=1}^{K} \{1, \ldots, |g_{[i]}|\} \longrightarrow \{0, 1\} \ , \tag{7}$$

where $f_{[k]}$ and $g_{[k]}$ are encoder mappings with image sizes satisfying $\log |f_{[i]}| \equiv \mathcal{O}(n)$ and $\log |g_{[i]}| \equiv \mathcal{O}(n)$, respectively, while ϕ is the decision mapping. The corresponding Type I and II error probabilities are given by

$$\alpha_n(R\,|K) \coloneqq \Pr\left[\phi\big(X^n, g_{[1:K]}\big) = 1 \,|\, X^n Y^n \sim P_{XY}\right] \ , \tag{8}$$

$$\beta_n(R\,|K) \coloneqq \Pr\left[\phi\big(X^n, g_{[1:K]}\big) = 0 \,|\, X^n Y^n \sim P_{\bar{X}\bar{Y}}\right] \ . \tag{9}$$

An exponent E to the error probability of Type II, constrained to an error probability of Type I to be below $\epsilon > 0$ and a total exchange rate R, is said to be feasible, if for any $\varepsilon > 0$ there exists a code satisfying:

$$-\frac{1}{n} \log \beta_n(R, \epsilon\,|K) \geq E - \varepsilon \ , \tag{10}$$

$$\frac{1}{n} \sum_{k=1}^{K} \log \big(|g_{[k]}||f_{[k]}|\big) \leq R + \varepsilon \ , \ \alpha_n(R\,|K) \leq \epsilon \ , \tag{11}$$

provided that n is large enough. The supremum of all feasible exponents for given (R, ϵ) is defined to be the optimal error exponent.

3. Collaborative Hypothesis Testing with One Round. In this section, we present and prove a *feasible error exponent* $-\frac{1}{n} \log \beta_n(R, \epsilon\,|K = 1)$ to the error probability of Type II, under any fixed constraint $\epsilon > 0$ on the error probability of Type I for a total exchange rate R. Here, we only consider one round of exchange whereby each of the nodes exchanges one statistics (or message) before a decision is made. The extension to the case with multiple exchanging rounds is relegated to the next section.

Proposition 1 (Sufficient conditions for one round of interaction). *Let $\mathscr{S}(R) \subset \mathcal{P}(\mathcal{U} \times \mathcal{V})$ and $\mathscr{L}(U, V) \subset \mathcal{P}(\mathcal{U} \times \mathcal{V} \times \mathcal{X} \times \mathcal{Y})$ denote the sets of*

probability measures defined in terms of corresponding RVs:

$$\mathscr{S}(R) \coloneqq \big\{UV:\ I(U;X)+I(V;Y|U)\le R \\ U \multimap X \multimap Y\ ,\ V \multimap (U,Y) \multimap X\ ,\ |\mathcal{U}|,|\mathcal{V}|<+\infty\big\}\ , \tag{12}$$

$$\mathscr{L}(U,V) \coloneqq \big\{\tilde{U}\tilde{V}\tilde{X}\tilde{Y}\ :\ P_{\tilde{U}\tilde{V}\tilde{X}}=P_{UVX}\ ,\ P_{\tilde{U}\tilde{V}\tilde{Y}}=P_{UVY}\big\}\ . \tag{13}$$

A feasible error exponent to the error probability of Type II, when the total exchange rate is R (bits per sample), is given by

$$\liminf_{n\to\infty}\ \lim_{\epsilon\to 0} -\frac{1}{n}\log\beta_n(R,\epsilon\,|K=1) \ge \\ \max_{UV\in\mathscr{S}(R)}\ \min_{\tilde{U}\tilde{V}\tilde{X}\tilde{Y}\in\mathscr{L}(U,V)} \mathcal{D}\big(P_{\tilde{U}\tilde{V}\tilde{X}\tilde{Y}}||P_{\bar{U}\bar{V}\bar{X}\bar{Y}}\big)\ . \tag{14}$$

Proof. We start by describing the random construction of codebooks, as well as encoding and decision functions. By analyzing the asymptotic properties of such decision systems, we aim at implying a *feasibility (existence) result* of interactive functions and decision regions that satisfy, for any given $\epsilon,\varepsilon>0$, the following inequalities:

$$\frac{1}{n}\log\big(|f_{[1]}||g_{[1]}|\big)\le I(U;X)+I(V;Y|U)+\varepsilon\ ,\ \alpha_n(R\,|K=1)\le\epsilon\ , \tag{15}$$

$$-\frac{1}{n}\log\beta_n(R,\epsilon\,|K=1)\ge \min_{\tilde{U}\tilde{V}\tilde{X}\tilde{Y}\in\mathscr{L}(U,V)} \mathcal{D}\big(P_{\tilde{U}\tilde{V}\tilde{X}\tilde{Y}}||P_{\bar{U}\bar{V}\bar{X}\bar{Y}}\big)-\varepsilon\ , \tag{16}$$

provided that n is large enough and for any given pair of random variables $(U,V)\in\mathscr{S}(R)$, where $|f_{[1]}|$ and $|g_{[1]}|$ denote the number of codewords in the codebooks[1] used for interaction.

Codebook generation. Without loss of generality, we assume that node A is the first to communicate. Fix a conditional probability $P_{UV|XY}(u,v|x,y)=P_{U|X}(u|x)P_{V|UY}(v|u,y)$ that attains the maximum in Proposition 1. Let

$$P_U(u)\equiv\sum_{x\in\mathcal{X}}P_{U|X}(u|x)P_X(x)\ ,\ P_{V|U}(v|u)\equiv\sum_{y\in\mathcal{Y}}P_{V|UY}(v|u,y)P_Y(y).$$

For this choice of RVs, set the rates (R_U,R_V) to be

$$I(U;X)+\epsilon(\delta)\coloneqq R_U\ ,\ I(V;Y|U)+\epsilon(\delta')\coloneqq R_V$$

with $\epsilon(\delta)\to 0$ as $\delta\to 0$. By the definition of the set $\mathscr{S}(R)$, it is clear that $R_U+R_V\le R+\epsilon(\delta)+\epsilon(\delta')$. Randomly and independently draw 2^{nR_U}

[1]Note that *feasibility* is defined in the information-theoretic sense which implies the *random existence* of interactive and decision functions with desired properties.

sequences $\mathbf{u} = (u_1, \ldots, u_n)$, each according to $\prod_{i=1}^{n} P_U(u_i)$. Index these sequences by $m_U \in [1 : M_U \coloneqq 2^{nR_U}]$ to form the random codebook $\mathcal{C}_{\mathbf{u}} \coloneqq \{\mathbf{u}(m_U) : m_U \in [1 : M_U]\}$. As a second step, for each word $\mathbf{u} \in \mathcal{C}_{\mathbf{u}}$, build a codebook $\mathcal{C}_{\mathbf{v}}(m_U)$ by randomly and independently drawing 2^{nR_V} sequences $\mathbf{v}$, each according to $\prod_{i=1}^{n} P_{V|U}(v_i|u_i(m_U))$. Index these sequences by $m_V \in [1 : M_V \coloneqq 2^{nR_V}]$ to form the collection of codebooks $\mathcal{C}_{\mathbf{v}}(m_U) \coloneqq \{\mathbf{v}(m_U, m_V) : m_V \in [1 : M_V]\}$ for $m_U \in [1 : M_U]$.

Encoding and decision mappings. Given a sequence $\mathbf{x}$, node A searches in the codebook $\mathcal{C}_{\mathbf{u}}$ for an index m_U such that $(\mathbf{u}(m_U), \mathbf{x}) \in \mathcal{T}^n_{[UX]_\delta}$ (note that this notation denotes the δ-typical set with relation to the probability measure implied by H_0). 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 communicates the chosen index m_U to node B, using a portion R_U bits of the available exchange rate. Upon receiving the index m_U, node B checks if $(\mathbf{u}(m_U), \mathbf{y}) \in \mathcal{T}^n_{[UY]_{\delta'}}$. If not, node B declares H_1. If the received sequence $\mathbf{u}$ and $\mathbf{y}$ (the observed sequence at node B) are jointly typical, node B looks in the specific codebook $\mathcal{C}_{\mathbf{v}}(m_U)$, for an index m_V such that $\big(\mathbf{u}(m_U), \mathbf{v}(m_U, m_V), \mathbf{y}\big) \in \mathcal{T}^n_{[UVY]_{\delta'}}$. 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 transmits the chosen index m_V to node A. Upon reception of the index m_V, node A checks if $\big(\mathbf{u}(m_U), \mathbf{v}(m_U, m_V), \mathbf{x}\big) \in \mathcal{T}^n_{[UVX]_{\delta''}}$. If so, it declares H_0 and otherwise, it declares H_1. The relation between δ, δ' and δ'' can be deducted from Lemma 5. It is, however, important to emphasize that $\delta'(\delta) \to 0$ as $\delta \to 0$, and $\delta''(\delta') \to 0$ as $\delta' \to 0$ with $n \to \infty$.

Analysis of α_n (Type I). The analysis of α_n is identical to the one proposed in [26], for the case of testing against independence. We give here a short summary of the analysis available in [26]. Assuming that the measure that controls X and Y is P_{XY}, and denoting the chosen indices at nodes A and B by m_U and m_V respectively, the error probability of the Type I can be expressed as follows

$$\alpha_n \equiv \Pr(\mathcal{E}_1 \cup \mathcal{E}_2 \cup \mathcal{E}_3) \leq \Pr(\mathcal{E}_1) + \Pr(\mathcal{E}_1^c \cap \mathcal{E}_2) + \Pr(\mathcal{E}_1^c \cap \mathcal{E}_2^c \cap \mathcal{E}_3) \ , \tag{17}$$

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

$$\mathcal{E}_1 \equiv \big\{(\mathbf{U}(m_U), \mathbf{X}) \notin \mathcal{T}^n_{[UX]\delta} \, \forall \, m_U \in [1 : M_U]\big\} \ , \tag{18}$$

$$\mathcal{E}_2 \equiv \big\{(\mathbf{V}(m_U, m_V), \mathbf{U}(m_U), \mathbf{Y}) \notin \mathcal{T}^n_{[VUY]_{\delta'}} \, \forall \, m_V \in [1 : M_V] \text{ and the specific } m_U \text{ selected at node } A\big\} \ , \tag{19}$$

$$\mathcal{E}_3 \equiv \big\{(\mathbf{V}(m_U, m_V), \mathbf{U}(m_U), \mathbf{X}) \notin \mathcal{T}^n_{[VUX]_{\delta''}}, \text{ for the specific } m_U \text{ and } m_V \text{ previously chosen}\big\} \ . \tag{20}$$

Analyzing each of the probabilities in (17) separately, $\Pr(\mathcal{E}_1) \to 0$ as $n \to \infty$ by the *covering lemma* [10], provided that $R_U \geq I(U;X) + \epsilon(\delta)$, with $\epsilon(\delta) \to 0$ as $\delta \to 0$. $\Pr(\mathcal{E}_1^c \cap \mathcal{E}_2) \to 0$ when $n \to \infty$ by the *conditional typicality lemma* [10], in addition to the covering lemma, provided that $R_V \geq I(V;Y|U)+\epsilon(\delta')$. Finally, the third term in (17) can be shown to tend to zero through the use of the Markov lemma (see Lemma 6), as well as Lemma 4 and Lemma 5 in Appendix A. Thus, as all three components tend to zero with large n, we may conclude that $\alpha_n \leq \epsilon$ for any constraint $0 < \epsilon < 1$ and n large enough.

Analysis of β_n (Type II). The error probability of Type II is defined by

$$\beta_n(R, \epsilon \,|K = 1) \equiv \Pr\left(\text{decide } H_0 | XY \sim P_{\bar{X}\bar{Y}}\right) . \tag{21}$$

Thus, we assume that $P_{\bar{X}\bar{Y}}$ controls the measure of the observed RVs throughout this analysis. We use similar methods to what was done in [11], although we choose to work with random codebooks. The influence of this choice is on the analysis of α_n only, as seen above, and not on β_n.

For a given pair of sequences $(\mathbf{x}, \mathbf{y})$ with type variables $X^{(n)}Y^{(n)} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y})$, we count all possible events that lead to an error. We notice first, that given a pair of vectors $(\mathbf{x}, \mathbf{y}) \in \mathcal{X}^n \times \mathcal{Y}^n$ the probability that these vectors will be the result of n i.i.d. draws, according to the measure implied by H_1, is given by Lemma 4 to be:

$$\Pr\{\bar{X}^n \bar{Y}^n = (\mathbf{x}, \mathbf{y})\} = \exp\left[-n\left(H(X^{(n)}Y^{(n)}) + \mathcal{D}(X^{(n)}Y^{(n)} || \bar{X}\bar{Y})\right)\right] , \tag{22}$$

where $X^{(n)}Y^{(n)} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y})$ are the type variables of the realizations $(\mathbf{x}, \mathbf{y})$ (see Appendix A). For each pair of codewords $\mathbf{u}_i \in \mathcal{C}_\mathbf{u}$ and $\mathbf{v}_{ij} \in \mathcal{C}_\mathbf{v}(i)$, we define the set:

$$\mathcal{S}_{ij}(\mathbf{x}) \coloneqq \{\mathbf{u}_i\} \times \{\mathbf{v}_{ij}\} \times \mathcal{G}_{ij} \times \{\mathbf{x}\} , \tag{23}$$

where $\mathcal{G}_{ij} \subseteq \mathcal{Y}^n$ is the set of all vectors $\mathbf{y}$ that, given the received message $\mathbf{u}_i$, will result in the message $\mathbf{v}_{ij}$ being transmitted back to node A. Denoting by $K_{ij}(\mathbf{x})$ the number of elements $(\mathbf{u}_i, \mathbf{v}_{ij}, \mathbf{x}, \mathbf{y}) \in \mathcal{S}_{ij}(\mathbf{x})$ whose type variables coincide with $U^{(n)}V^{(n)}X^{(n)}Y^{(n)}$, we have by Lemma 3 that:

$$K_{ij}(\mathbf{x}) \leq \exp\left[nH(Y^{(n)}|U^{(n)}V^{(n)}X^{(n)})\right] . \tag{24}$$

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

$$(\mathbf{u}, \mathbf{v}, \mathbf{x}, \mathbf{y}) \in \mathscr{S}_n \coloneqq \bigcup_{i=1}^{M_U} \bigcup_{j=1}^{M_V} \bigcup_{\mathbf{x} \in \mathcal{T}^n_{[X|\mathbf{u}_i\mathbf{v}_{ij}]_{\delta''}}} \mathcal{S}_{ij}(\mathbf{x})$$

that have type variable $U^{(n)}V^{(n)}X^{(n)}Y^{(n)} \in \mathcal{P}_n(\mathcal{U}\times\mathcal{V}\times\mathcal{X}\times\mathcal{Y})$, then

(25)
$$\begin{aligned} K(U^{(n)}V^{(n)}X^{(n)}Y^{(n)}) &\leq \sum_{i=1}^{M_U}\sum_{j=1}^{M_V} \exp\left[nH(Y^{(n)}|U^{(n)}V^{(n)}X^{(n)})\right] \left|\mathcal{T}^n_{[X|\mathbf{u}_i\mathbf{v}_{i,j}]_{\delta''}}\right| \\ &\leq \exp\Big[n\Big(H(Y^{(n)}|U^{(n)}V^{(n)}X^{(n)}) \\ &\qquad + I(U;X) + I(V;Y|U) + H(X|UV) + \mu_n\Big)\Big] \ , \end{aligned}$$

where M_U and M_V are the sizes of the codebooks $\mathcal{C}_{\mathbf{u}}$ and $\mathcal{C}_{\mathbf{v}}(\cdot)$. 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 (see Lemma 7). The resulting sequence μ_n is a function of $\delta, \delta', \delta''$ that complies with $\mu_n \to 0$ as $n \to \infty$. The error probability of Type II satisfies:

(26)
$$\beta_n(R,\epsilon\,|K=1) \leq \sum_{U^{(n)}V^{(n)}X^{(n)}Y^{(n)}\in\mathscr{S}_n} \exp\left[-n\left(k(U^{(n)}V^{(n)}X^{(n)}Y^{(n)}) - \mu_n\right)\right] \ ,$$

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

(27)
$$\begin{aligned} k(U^{(n)}V^{(n)}X^{(n)}Y^{(n)}) \coloneqq\ & H(X^{(n)}Y^{(n)}) + \mathcal{D}(X^{(n)}Y^{(n)}||\bar{X}\bar{Y}) \\ & - H(Y^{(n)}|U^{(n)}V^{(n)}X^{(n)}) - H(X|UV) \\ & - I(U;X) - I(V;Y|U) \ . \end{aligned}$$

We deliberately made an abuse of notation in (26) to indicate that the sum is taken over all possible type-variables $U^{(n)}V^{(n)}X^{(n)}Y^{(n)} \in \mathcal{P}_n(\mathcal{U}\times\mathcal{V}\times\mathcal{X}\times\mathcal{Y})$ formed by empirical probability measures from elements $(\mathbf{u},\mathbf{v},\mathbf{x},\mathbf{y}) \in \mathscr{S}_n$.

From the construction of $\mathscr{S}_n$, it is clear that if $(\mathbf{u},\mathbf{v},\mathbf{x},\mathbf{y}) \in \mathscr{S}_n$, then at least $(\mathbf{u},\mathbf{v},\mathbf{x}) \in \mathcal{T}^n_{[UVX]_{\delta''}}$ and $(\mathbf{u},\mathbf{v},\mathbf{y}) \in \mathcal{T}^n_{[UVY]_{\delta'}}$. Thus, the summation in (26) is only over all types satisfying:

(28)
$$\begin{aligned} |P_{U^{(n)}V^{(n)}X^{(n)}}(u,v,x) - P_{UVX}(u,v,x)| &\leq \delta'' \ , \\ |P_{U^{(n)}V^{(n)}Y^{(n)}}(u,v,y) - P_{UVY}(u,v,y)| &\leq \delta' \ , \end{aligned}$$

for all $(u,v,x) \in \mathrm{supp}(P_{UVX})$ and $(u,v,y) \in \mathrm{supp}(P_{UVY})$. In addition, it follows by Lemma 2 from the total number of types of length n that:

(29)
$$\begin{aligned} \beta_n(R,\epsilon\,|K=1) &\leq (n+1)^{|\mathcal{U}||\mathcal{V}||\mathcal{X}||\mathcal{Y}|} \\ &\quad\times \max_{U^{(n)}V^{(n)}X^{(n)}Y^{(n)}\in\mathscr{S}_n} \exp\left[-n\left(k(U^{(n)}V^{(n)}X^{(n)}Y^{(n)}) - \mu_n\right)\right] \ . \end{aligned}$$

By (28) and the continuity of the entropy function as well as the KL divergence, we can conclude that

$$k(U^{(n)}V^{(n)}X^{(n)}Y^{(n)}) = H(\tilde{X}\tilde{Y}) + \mathcal{D}(\tilde{X}\tilde{Y}||\bar{X}\bar{Y}) - H(\tilde{Y}|\tilde{U}\tilde{V}\tilde{X}) - H(\tilde{X}|\tilde{U}\tilde{V}) - I(\tilde{U};\tilde{X}) - I(\tilde{V};\tilde{Y}|\tilde{U}) + \mu'_n \ , \tag{30}$$

with $\tilde{U}\tilde{V}\tilde{X}\tilde{Y} \in \mathscr{L}(U,V)$ and $\mu'_n \to 0$ when $n \to \infty$. We can further simplify the expression of $k(U^{(n)}V^{(n)}X^{(n)}Y^{(n)})$ by observing that:

$$\begin{aligned}
k(U^{(n)}V^{(n)}X^{(n)}Y^{(n)}) &= H(\tilde{X}\tilde{Y}) + \mathcal{D}(\tilde{X}\tilde{Y}||\bar{X}\bar{Y}) - H(\tilde{Y}|\tilde{U}\tilde{V}\tilde{X}) \\
&\quad - H(\tilde{X}|\tilde{U}\tilde{V}) - I(\tilde{U};\tilde{X}) - I(\tilde{V};\tilde{Y}|\tilde{U}) + \mu'_n \\
&= H(\tilde{X}\tilde{Y}) + \mathcal{D}(\tilde{X}\tilde{Y}||\bar{X}\bar{Y}) - H(\tilde{X}\tilde{Y}|\tilde{U}\tilde{V}) - I(\tilde{U};\tilde{X}) - I(\tilde{V};\tilde{Y}|\tilde{U}) + \mu'_n \\
&= I(\tilde{X}\tilde{Y};\tilde{U}\tilde{V}) + \mathcal{D}(\tilde{X}\tilde{Y}||\bar{X}\bar{Y}) - I(\tilde{U};\tilde{X}) - I(\tilde{V};\tilde{Y}|\tilde{U}) + \mu'_n \\
&= I(\tilde{X}\tilde{Y};\tilde{U}) + I(\tilde{X}\tilde{Y};\tilde{V}|\tilde{U}) + \mathcal{D}(\tilde{X}\tilde{Y}||\bar{X}\bar{Y}) - I(\tilde{U};\tilde{X}) - I(\tilde{V};\tilde{Y}|\tilde{U}) + \mu'_n \\
&\overset{(a)}{=} \mathcal{D}(\tilde{U}\tilde{X}\tilde{Y}||\bar{U}\bar{X}\bar{Y}) + I(\tilde{X}\tilde{Y};\tilde{V}|\tilde{U}) - I(\tilde{Y};\tilde{V}|\tilde{U}) + \mu'_n \\
&= \mathcal{D}(\tilde{U}\tilde{X}\tilde{Y}||\bar{U}\bar{X}\bar{Y}) + I(\tilde{X};\tilde{V}|\tilde{U}\tilde{Y}) + \mu'_n \ ,
\end{aligned} \tag{31}$$

where equality (a) stems from the identity [11]:

$$\begin{aligned}
I(\tilde{X}\tilde{Y};\tilde{U}) + \mathcal{D}(\tilde{X}\tilde{Y}||\bar{X}\bar{Y}) - I(\tilde{U};\tilde{X}) &= I(\tilde{U};\tilde{Y}|\tilde{X}) + \mathcal{D}(\tilde{X}\tilde{Y}||\bar{X}\bar{Y}) \\
&= \mathcal{D}(\tilde{U}\tilde{X}\tilde{Y}||\bar{U}\bar{X}\bar{Y}) \ ,
\end{aligned} \tag{32}$$

which holds the case on unidirectional communication. Note that the following Markov chain: $X \multimap (U,Y) \multimap V$ holds under both hypotheses (i.e., the same chain can be written with a bar over all variables), *but not* for the auxiliary RVs, marked with a tilde.

Finally, we conclude our development of $k(U^{(n)}V^{(n)}X^{(n)}Y^{(n)})$ as follows:

$$\begin{aligned}
k(U^{(n)}V^{(n)}X^{(n)}Y^{(n)}) &= \mathcal{D}(\tilde{U}\tilde{X}\tilde{Y}||\bar{U}\bar{X}\bar{Y}) + I(\tilde{X};\tilde{V}|\tilde{U}\tilde{Y}) + \mu'_n \\
&= \sum_{\forall(u,v,x,y)} P_{\tilde{U}\tilde{V}\tilde{X}\tilde{Y}}(u,v,x,y) \times \\
&\qquad \times \log\left(\frac{P_{\tilde{U}\tilde{X}\tilde{Y}}(u,x,y)}{P_{\bar{U}\bar{X}\bar{Y}}(u,x,y)} \frac{P_{\tilde{X}\tilde{V}|\tilde{U}\tilde{Y}}(x,v|u,y)}{P_{\tilde{X}|\tilde{U}\tilde{Y}}(x|u,y) P_{\tilde{V}|\tilde{U}\tilde{Y}}(v|u,y)}\right) + \mu'_n \\
&\overset{(b)}{=} \sum_{\forall(u,v,x,y)} P_{\tilde{U}\tilde{V}\tilde{X}\tilde{Y}}(u,v,x,y) \log\left(\frac{P_{\tilde{U}\tilde{V}\tilde{X}\tilde{Y}}(u,v,x,y)}{P_{\bar{U}\bar{X}\bar{Y}}(u,x,y) P_{\bar{V}|\bar{U}\bar{Y}}(v|u,y)}\right) + \mu'_n
\end{aligned} \tag{33}$$

$$= \sum_{\forall(u,v,x,y)} P_{\tilde{U}\tilde{V}\tilde{X}\tilde{Y}}(u,v,x,y) \log\left(\frac{P_{\tilde{U}\tilde{V}\tilde{X}\tilde{Y}}(u,v,x,y)}{P_{\bar{U}\bar{V}\bar{X}\bar{Y}}(u,v,x,y)}\right) + \mu_n'$$
$$= \mathcal{D}(\tilde{U}\tilde{V}\tilde{X}\tilde{Y}||\bar{U}\bar{V}\bar{X}\bar{Y}) + \mu_n' \ ,$$

where the sums are over the $\text{supp}(P_{\tilde{U}\tilde{V}\tilde{X}\tilde{Y}})$; and (b) is due to the definition of the set $\mathscr{L}(U,V)$ that implies $P_{\tilde{V}|\tilde{U}\tilde{Y}}(v|u,y) = P_{V|UY}(v|u,y)$. In addition, as coding (at each side) is performed before a decision is made, it is clear it is done in the same way under both hypotheses. Thus, while $P_{UVY}(u,v,y) \neq P_{\bar{U}\bar{V}\bar{Y}}(u,v,y)$, it is true that $P_{\bar{V}|\bar{U}\bar{Y}}(v|u,y) = P_{V|UY}(v|u,y) = P_{\tilde{V}|\tilde{U}\tilde{Y}}(v|u,y)$. As μ_n, μ_n' are arbitrarily small, as a function of the choices of δ and δ' provided that n is large enough, this concludes the proof of Proposition 1. □

4. Collaborative Hypothesis Testing with Multiple Rounds. We now allow the statisticians to exchange data over an arbitrary but *finite* number of exchange rounds, and investigate the extension of Proposition 1 to this more general case. The corresponding result is stated below.

PROPOSITION 2 (Sufficient conditions for K-rounds of interaction). *Let $\mathscr{S}(R)$ and $\mathscr{L}\big(U_{[1:K]}, V_{[1:K]}\big)$ denote the sets of probability measures defined in terms of corresponding RVs:*

$$\mathscr{S}(R) \coloneqq \Big\{ U_{[1:K]}V_{[1:K]} : R \geq \sum_{k=1}^{K} \big[I(X; U_{[k]}|U_{[1:k-1]}V_{[1:k-1]}) + I(Y; V_{[k]}|U_{[1:k-1]}V_{[1:k-2]}) \big] \ ,$$
$$U_{[k]} \multimap \big(X, U_{[1:k-1]}, V_{[1:k-1]}\big) \multimap Y \ , \ |\mathcal{U}_{[k]}| < +\infty \ ,$$
$$V_{[k]} \multimap \big(Y, U_{[1:k]}, V_{[1:k-1]}\big) \multimap X \ , \ |\mathcal{V}_{[k]}| < +\infty \ , \forall\, k \in [1:K] \Big\} \ , \tag{34}$$

$$\mathscr{L}\big(U_{[1:K]}, V_{[1:K]}\big) \coloneqq \Big\{ \tilde{U}_{[1:K]}\tilde{V}_{[1:K]}\tilde{X}\tilde{Y} : P_{\tilde{U}_{[1:K]}\tilde{V}_{[1:K]}\tilde{X}} = P_{U_{[1:K]}V_{[1:K]}X} \ , \ P_{\tilde{U}_{[1:K]}\tilde{V}_{[1:K]}\tilde{Y}} = P_{U_{[1:K]}V_{[1:K]}Y} \Big\} \ , \tag{35}$$

where $U_{[1:k]} \coloneqq (U_{[1]}, \ldots, U_{[k]})$ and $V_{[1:k]} \coloneqq (V_{[1]}, \ldots, V_{[k]})$ represent the exchanged data between nodes A and B until round k. A feasible error exponent to the error probability of Type II, when the total (over K-rounds) exchange rate is R (bits per sample), is given by

$$\liminf_{n\to\infty} \lim_{\epsilon\to 0} -\frac{1}{n} \log \beta_n(R, \epsilon, |K) \geq \max_{\mathscr{S}(R)} \min_{\mathscr{L}\left(U_{[1:K]}, V_{[1:K]}\right)} \mathcal{D}\Big(P_{\tilde{U}_{[1:K]}\tilde{V}_{[1:K]}\tilde{X}\tilde{Y}} \big|\big| P_{\bar{U}_{[1:K]}\bar{V}_{[1:K]}\bar{X}\bar{Y}}\Big) \ . \tag{36}$$

This proposition is very clearly an extension of Proposition 1 to allow multiple rounds of interaction. The implication of this result is as follows. Given a limited budget of rate R for data exchange, which the nodes can divide as they choose into any finite number of K exchange rounds, the gain of interaction attained through the different characteristics of the underlying Markov process between the RVs comes at no cost in terms of the form of the expression for the error exponent.

Proof of Proposition 2. The proof of this proposition is very similar to the one presented above for Proposition 1. Codebook construction, as well as encoding and decision mappings remain similar. At each round, a codebook is built based on any possible combination of the previous messages. Given previous messages, each node chooses a message in the relevant codebook and communicates its index to the other statistician. The process continues until a message cannot be found, which is jointly typical with all previous messages as well as the observed sequence, in which case H_1 is declared. Otherwise, until the end of round K in which case H_0 is declared, provided that all the messages are jointly typical with the observed sequence. We next provide a sketch of the proof to this simple extension.

The analysis of α_n applies similarly to the previous case, as long as a finite number of rounds is considered. Regarding the analysis of β_n, the following important changes are needed:

- The set $\mathcal{S}_{\mathbf{ij}}(\mathbf{x})$ is now defined by using all exchanged messages:
(37)
$$\mathcal{S}_{\mathbf{ij}}(\mathbf{x}) \coloneqq \{\mathbf{u}_{[1],i_1}\}\times\{\mathbf{v}_{[1],i_1j_1}\}\times\cdots\times\{\mathbf{u}_{[K],i_K}\}\times\{\mathbf{v}_{[K],i_Kj_K}\}\times\mathcal{G}_{\mathbf{ij}}\times\{\mathbf{x}\}\ ,$$
where $(\mathbf{i},\mathbf{j}) \coloneqq (i_1,j_1),\ldots,(i_K,j_K)$ and $\mathbf{u}_{[k],i_k}$ is the i_k-th message in the codebook $\mathcal{C}_{\mathbf{u}_{[k]}}$, similarly for the other random variables.
- Similarly, $\mathscr{S}_n$ is now defined by the union over the codewords of *all* auxiliary RVs.
- The bound over $K_{\mathbf{ij}}$ (analogues to expression (24) before) writes:
$$K_{\mathbf{ij}}(\mathbf{x}) \leq \exp\left[nH\big(Y^{(n)}|U^{(n)}_{[1:K]}V^{(n)}_{[1:K]}X^{(n)}\big)\right]\ . \tag{38}$$
- Finally, $K\big(U^{(n)}_{[1:K]}V^{(n)}_{[1:K]}X^{(n)}Y^{(n)}\big)$, i.e., see (25), is now calculated through the summation over the codebooks of all messages, considering the cardinality of the conditional set: $\big|\mathcal{T}^n_{[X|\mathbf{u}_{[1:K]},\mathbf{i}\mathbf{v}_{[1:K],\mathbf{ij}}]_\delta}\big|$.
- As more steps are performed, each of which requires encoding, we also need to define new δ's for each of these steps. We refrain from this for the sake of readability, as all of these δ's go to 0 together, as was seen in the case of a single round.

Considering these differences, after k rounds of interactions, $k\big(U_{[1:k]}^{(n)}V_{[1:k]}^{(n)}\big)$ can be shown to be equal to (e.g. see (27)):

$$
\begin{aligned}
(39)\quad k\big(U_{[1:k]}^{(n)}V_{[1:k]}^{(n)}\big) &= \mathcal{D}\big(P_{\tilde{U}_{[1:k-1]}\tilde{V}_{[1:k-1]}\tilde{X}\tilde{Y}}||P_{\bar{U}_{[1:k-1]}\bar{V}_{[1:k-1]}\bar{X}\bar{Y}}\big)\\
&\quad + I(\tilde{Y};\tilde{U}_{[k]}|\tilde{U}_{[1:k-1]}\tilde{V}_{[1:k-1]}\tilde{X}) + I(\tilde{X};\tilde{V}_{[k]}|\tilde{U}_{[1:k]}\tilde{V}_{[1:k-1]}\tilde{Y}) + \mu_n' \ .
\end{aligned}
$$

By continuing in the same manner as in (33), we show:

$$
\begin{aligned}
&k\big(U_{[1:k]}^{(n)}V_{[1:k]}^{(n)}\big) - \mu_n' = \sum_{\forall} P_{\tilde{U}_{[1:k-1]}\tilde{V}_{[1:k-1]}\tilde{X}\tilde{Y}} \log \frac{P_{\tilde{U}_{[1:k-1]}\tilde{V}_{[1:k-1]}\tilde{X}\tilde{Y}}}{P_{\bar{U}_{[1:k-1]}\bar{V}_{[1:k-1]}\bar{X}\bar{Y}}}\\
&+ \sum_{\forall} P_{\tilde{U}_{[1:k]}\tilde{V}_{[1:k-1]}\tilde{X}\tilde{Y}} \log \frac{P_{\tilde{U}_{[k]}\tilde{Y}|\tilde{U}_{[1:k-1]}\tilde{V}_{[1:k-1]}\tilde{X}}}{P_{\tilde{U}_{[k]}|\tilde{U}_{[1:k-1]}\tilde{V}_{[1:k-1]}\tilde{X}}P_{\tilde{Y}|\tilde{U}_{[1:k-1]}\tilde{V}_{[1:k-1]}\tilde{X}}}\\
&+ \sum_{\forall} P_{\tilde{U}_{[1:k]}\tilde{V}_{[1:k]}\tilde{X}\tilde{Y}} \log \frac{P_{\tilde{V}_{[k]}\tilde{X}|\tilde{U}_{[1:k]}\tilde{V}_{[1:k-1]}\tilde{Y}}}{P_{\tilde{V}_{[k]}|\tilde{U}_{[1:k]}\tilde{V}_{[1:k-1]}\tilde{Y}}P_{\tilde{X}|\tilde{U}_{[1:k]}\tilde{V}_{[1:k-1]}\tilde{Y}}}\\
&= \sum_{\forall} P_{\tilde{U}_{[1:k]}\tilde{V}_{[1:k]}\tilde{X}\tilde{Y}} \log \left[\frac{P_{\tilde{U}_{[1:k]}\tilde{V}_{[1:k]}\tilde{X}\tilde{Y}}}{P_{\bar{U}_{[1:k-1]}\bar{V}_{[1:k-1]}\bar{X}\bar{Y}}P_{\tilde{U}_{[k]}|\tilde{U}_{[1:k-1]}\tilde{V}_{[1:k-1]}\tilde{X}}P_{\tilde{V}_{[k]}|\tilde{U}_{[1:k]}\tilde{V}_{[1:k-1]}\tilde{Y}}}\right]\\
&\overset{(c)}{=} \sum_{\forall} P_{\tilde{U}_{[1:k]}\tilde{V}_{[1:k]}\tilde{X}\tilde{Y}} \log \left[\frac{P_{\tilde{U}_{[1:k]}\tilde{V}_{[1:k]}\tilde{X}\tilde{Y}}}{P_{\bar{U}_{[1:k-1]}\bar{V}_{[1:k-1]}\bar{X}\bar{Y}}P_{\bar{U}_{[k]}|\bar{U}_{[1:k-1]}\bar{V}_{[1:k-1]}\bar{X}}P_{\bar{V}_{[k]}|\bar{U}_{[1:k]}\bar{V}_{[1:k-1]}\bar{Y}}}\right]\\
&= \sum_{\forall} P_{\tilde{U}_{[1:k]}\tilde{V}_{[1:k]}\tilde{X}\tilde{Y}} \log \left[\frac{P_{\tilde{U}_{[1:k]}\tilde{V}_{[1:k]}\tilde{X}\tilde{Y}}}{P_{\bar{U}_{[1:k]}\bar{V}_{[1:k]}\bar{X}\bar{Y}}}\right]\\
&= \mathcal{D}\big(P_{\tilde{U}_{[1:k]}\tilde{V}_{[1:k]}\tilde{X}\tilde{Y}}||P_{\bar{U}_{[1:k]}\bar{V}_{[1:k]}\bar{X}\bar{Y}}\big) \ ,
\end{aligned}
$$

where all sums are over all the alphabets of the relevant RVs. Here, (c), much like in the case of single-round exchange above, is due to the definition of the set $\mathscr{L}(U_{[1:k]},V_{[1:k]})$ and to the fact that encoding occurs without knowledge of the PM controlling the RVs, and thus behaves the same under each of the hypotheses. Thus,

$$
P_{\tilde{U}_{[k]}|\tilde{U}_{[1:k-1]}\tilde{V}_{[1:k-1]}\tilde{X}} = P_{U_{[k]}|U_{[1:k-1]}V_{[1:k-1]}X} = P_{\bar{U}_{[k]}|\bar{U}_{[1:k-1]}\bar{V}_{[1:k-1]}\bar{X}},
$$

and similarly for the messages $V_{[k]}$ at node B. Pursuing this until round K, the proposition is proved. □

REMARK 1. *For reasons of brevity and clarity, we chose in this paper to concentrate on scenarios where the interactions begins and ends at node A. However, it is easy to see that this does not necessarily need to be the case. The process could start or end at node B, implying that the final round of exchange is in fact only half of a round, without any significant changes to the theory or our proofs.*

5. Collaborative Testing Against Independence. We now concentrate on the special problem of testing against independence, where it is assumed that under H_1 the n observed samples of the RVs (X, Y) defined on $(\mathcal{X} \times \mathcal{Y}, \mathcal{B}_{\mathcal{X}\times\mathcal{Y}})$ are distributed according to a product measure:

$$\begin{cases} H_0: & P_{XY}(x,y) \ , \forall\, (x,y) \in \mathcal{X} \times \mathcal{Y} \ , \\ H_1: & P_{\bar{X}\bar{Y}}(x,y) = P_X(x) P_Y(y) \ , \forall\, (x,y) \in \mathcal{X} \times \mathcal{Y} \ , \end{cases} \tag{40}$$

where $P_X(x)$ and $P_Y(y)$ are the marginal probability measures implied by $P_{XY}(x,y)$. Testing against independence was first studied, for a unidirectional communication link [11] (see also [2]). It was shown that the *optimal* rate of exponential decay to the error probability of Type II is:

$$\liminf_{n\to\infty} -\frac{1}{n} \log \beta_n(R, \epsilon, |K = 1/2) = \max_{\substack{P_{U|X}\,:\,\mathcal{X} \mapsto \mathcal{P}(\mathcal{U}) \\ \text{s.t. } I(U;X) \le R}} I(U;Y) \ , \ \ \forall\, 0 < \epsilon < 1 \ , \tag{41}$$

where R is the available exchange rate from node A to node B. Note that much like the case of centralized HT, the *optimal* error exponent does not depend on ϵ and thus a *strong unfeasibility* (converse) result holds.

Testing against independence in a cooperative scenario was first studied in [26], for the case of a single round of interaction. It was shown that a feasible error exponent to the error probability of Type II is given by

$$\liminf_{n\to\infty} \lim_{\epsilon\to 0} -\frac{1}{n} \log \beta_n(R, \epsilon, |K = 1) \ge E(R) \tag{42}$$

subject to a total available exchange rate R, where:

$$E(R) := \max_{\substack{P_{U|X}\,:\,\mathcal{X} \mapsto \mathcal{P}(\mathcal{U}) \\ P_{V|UY}\,:\,\mathcal{U}\times\mathcal{Y} \mapsto \mathcal{P}(\mathcal{V}) \\ \text{s.t. } I(U;X) + I(V;Y|U) \le R}} \big[I(U;Y) + I(V;X|U) \big] \ . \tag{43}$$

While the proof of feasibility inspired the approach taken in Proposition 1 for general hypotheses, unfortunately, the auxiliary RVs identified in the *weak* unfeasibility proof in [26] do not match the required Markov chains to lead to a feasible exponent (the reader may refer to [23, 13] for further details).

In this section, we revisit the problem of characterizing the reverse inequality in (42). We prove a *weak unfeasibility* result, determining necessary

and sufficient conditions to the optimality of the error exponent (43) satisfying $\alpha_n \leq \epsilon$ for *any* $0 < \epsilon < 1$ (i.e., we prove that the exponent in (43) is optimal in the case where we constrain α_n to go to 0 with n). We first show that Proposition 1 implies the feasibility part, i.e., inequality (42), and then follow with a new proof for the unfeasibility (for ϵ arbitrarily small) of any higher exponent.

THEOREM 3 (Necessary and sufficient conditions for testing against independence with $K = 1$). *The* optimal *error exponent to the error probability of Type II for testing against independence is given by*

$$\liminf_{n\to\infty} \lim_{\epsilon\to 0} -\frac{1}{n} \log \beta_n(R, \epsilon, |K = 1) \coloneqq E(R) \ , \quad \forall\ 0 < \epsilon < 1 \ , \tag{44}$$

where $E(R)$ is defined in (43), and R denotes the available rate of interaction between the statisticians and ϵ is the error probability of Type I.

REMARK 2. *In a similar manner to Theorem 3, a* feasible *error exponent to the error probability of Type II with K rounds is given by*

$$\liminf_{n\to\infty} \lim_{\epsilon\to 0} -\frac{1}{n} \log \beta_n(R, \epsilon, |K) \geq \tag{45}$$
$$\max_{U_{[1:K]}V_{[1:K]}\in\mathscr{S}(R)} \sum_{k=1}^{K} \left[I\big(U_{[k]}; Y|U_{[1:k-1]}V_{[1:k-1]}\big) + I\big(V_{[k]}; X|U_{[1:k]}V_{[1:k-1]}\big)\right] \ .$$

The proof of the feasibility of (45) follows largely the same path as the one for the feasibility part provided below for Theorem 3. However, for $K > 1$ our unfeasibility proof does not hold and this feasible exponent result may not longer be optimal.

5.1. *Proof of Theorem 3.* We first enunciate and prove some preliminary results from which the proof of Theorem 3 will easily follow.

LEMMA 1 (Multi-letter representation for testing against independence with $K = 1$ [26]). *The error exponent to the error probability of Type II for testing against independence with one round satisfies:*

$$\limsup_{n\to\infty} \lim_{\epsilon\to 0} -\frac{1}{n} \log \beta_n(R, \epsilon\,|K = 1) \leq \frac{1}{n}\big[I(I_A; Y^n) + I(I_B; X^n|I_A)\big] \ , \tag{46}$$

$$R \geq \frac{1}{n}\big[I(I_A; X^n) + I(I_B; Y^n|I_A)\big] \ , \tag{47}$$

where $I_A \coloneqq f_1(X^n)$ and $I_B \coloneqq g_1\big(f_1(X^n), Y^n\big)$ for any mappings (f_1, g_1), as given in Defintion 1.

PROOF. The proof follows [2, 26] and is given in Appendix C. □

PROOF OF THEOREM 3. We start showing the feasibility, followed by a proof of the unfeasibility part.

Feasibility. In order to show the feasibility to the exponent (43) through the general result stated in Proposition 1, it is convenient to use the form of the last expression in (31):

$$\liminf_{n\to\infty} -\frac{1}{n}\log\beta_n(R,\epsilon,|K=1) \geq \max_{UV\in\mathscr{S}(R)} \min_{\tilde{U}\tilde{V}\tilde{X}\tilde{Y}\in\mathscr{L}(U,V)} \left[\mathcal{D}(P_{\tilde{U}\tilde{X}\tilde{Y}}||P_{\bar{U}\bar{X}\bar{Y}}) + I(\tilde{X};\tilde{V}|\tilde{U}\tilde{Y})\right] . \tag{48}$$

We analyze each of these components separately:

$$\begin{aligned}\mathcal{D}(P_{\tilde{U}\tilde{X}\tilde{Y}}||P_{\bar{U}\bar{X}\bar{Y}}) &\overset{(d)}{=} \mathcal{D}(P_{\tilde{U}\tilde{Y}}||P_{\bar{U}\bar{Y}}) + \mathcal{D}(P_{\tilde{X}|\tilde{U}\tilde{Y}}||P_{\bar{X}|\bar{U}\bar{Y}}|P_{\tilde{U}\tilde{Y}}) \\ &\overset{(e)}{=} I(U;Y) + \mathcal{D}(P_{\tilde{X}|\tilde{U}\tilde{Y}}||P_{\bar{X}|\bar{U}}|P_{\tilde{U}\tilde{Y}}) \\ &= I(U;Y) + \mathcal{D}(P_{\tilde{X}|\tilde{U}\tilde{Y}}||P_{\tilde{X}|\tilde{U}}|P_{\tilde{U}\tilde{Y}}) + \mathcal{D}(P_{\tilde{X}|\tilde{U}}||P_{\bar{X}|\bar{U}}|P_{\tilde{U}}) \\ &\overset{(f)}{\geq} I(U;Y) + \mathcal{D}(P_{\tilde{X}|\tilde{U}\tilde{Y}}||P_{\tilde{X}|\tilde{U}}|P_{\tilde{U}\tilde{Y}}) ,\end{aligned} \tag{49}$$

where (d) is due to the chain rule and $\mathcal{D}(P_{\tilde{X}|\tilde{U}\tilde{Y}}||P_{\bar{X}|\bar{U}\bar{Y}}|P_{\tilde{U}\tilde{Y}})$ is the conditional KL-divergence; (e) stems from the assumption of testing against independence, as well as the Markov chain $\bar{U} \multimap\!\circ\!\!- \bar{X} \multimap\!\circ\!\!- \bar{Y}$ and the fact that $P_{\tilde{U}\tilde{Y}} = P_{UY}$; and (f) is due to the fact that the KL-divergence is non-negative. To conclude the analysis, we note that:

$$\begin{aligned}&\mathcal{D}(P_{\tilde{U}\tilde{X}\tilde{Y}}||P_{\bar{U}\bar{X}\bar{Y}}) \geq \\ &\quad I(U;Y) + \sum_{(u,x,y)\in\mathcal{U}\times\mathcal{X}\times\mathcal{Y}} P_{\tilde{U}\tilde{X}\tilde{Y}}(u,x,y)\log\left(\frac{P_{\tilde{X}|\tilde{U}\tilde{Y}}(x|u,y)}{P_{\tilde{X}|\tilde{U}}(x|u)}\right) \\ &= I(U;Y) + \sum_{(u,x,y)\in\mathcal{U}\times\mathcal{X}\times\mathcal{Y}} P_{\tilde{U}\tilde{X}\tilde{Y}}(u,x,y)\log\left(\frac{P_{\tilde{X}\tilde{Y}|\tilde{U}}(x,y|u)}{P_{\tilde{X}|\tilde{U}}(x|u)P_{\tilde{Y}|\tilde{U}}(y|u)}\right) \\ &= I(U;Y) + I(\tilde{X};\tilde{Y}|\tilde{U}) .\end{aligned} \tag{50}$$

As for the second term in (48), we express it as follows:

$$I(\tilde{V};\tilde{X}|\tilde{U}\tilde{Y}) = I(\tilde{V}\tilde{Y};\tilde{X}|\tilde{U}) - I(\tilde{X};\tilde{Y}|\tilde{U}) \geq I(\tilde{V};\tilde{X}|\tilde{U}) - I(\tilde{X};\tilde{Y}|\tilde{U}) . \tag{51}$$

This allows us to conclude through (48) that

(52)
$$\begin{aligned}\liminf_{n\to\infty} -\frac{1}{n}\log\beta_n(R,\epsilon\,|K=1) &\geq \max_{UV\in\mathscr{S}(R)}\min_{\tilde{U}\tilde{V}\tilde{X}\tilde{Y}\in\mathscr{L}(U,V)}\Big[I(U;Y)+I(\tilde{V};\tilde{X}|\tilde{U})\Big]\\ &= \max_{UV\in\mathscr{S}(R)}\left[I(U;Y)+I(V;X|U)\right]\ ,\end{aligned}$$

which completes the proof of feasibility through Proposition 1.

Weak unfeasibility. We are now ready to complete the proof of weak unfeasibility (converse) to Theorem 3. From Lemma 1, it follows that:

$$\begin{aligned}(53)\quad &\limsup_{n\to\infty}\lim_{\epsilon\to 0} -\frac{1}{n}\log\beta_n(R,\epsilon\,|K=1)\\ &\qquad\leq \limsup_{n\to\infty}\frac{1}{n}[I(I_A;Y^n)+I(I_B;X^n|I_A)] \coloneqq \limsup_{n\to\infty}\Delta_n\ ,\end{aligned}$$

where I_A is the message sent from node A while I_B is its reply from node B. In order to derive a single-letter expression, we expand (53) as follows:

(54)
$$\begin{aligned}\Delta_n &\overset{(g)}{=} \frac{1}{n}\sum_{i=1}^{n}\left[I(I_A;Y_i|Y_{i+1}^n)+I(I_B;X_i|I_AX^{i-1})\right]\\ &\overset{(h)}{=} \frac{1}{n}\sum_{i=1}^{n}\left[I(I_AY_{i+1}^n;Y_i)+I(I_BY_{i+1}^n;X_i|I_AX^{i-1})-I(Y_{i+1}^n;X_i|I_AI_BX^{i-1})\right]\\ &= \frac{1}{n}\sum_{i=1}^{n}\big[I(I_AX^{i-1}Y_{i+1}^n;Y_i)-I(X^{i-1};Y_i|I_AY_{i+1}^n)+I(Y_{i+1}^n;X_i|I_AX^{i-1})\\ &\qquad +I(I_B;X_i|I_AX^{i-1}Y_{i+1}^n)-I(Y_{i+1}^n;X_i|I_AI_BX^{i-1})\big]\\ &\overset{(i)}{=} \frac{1}{n}\sum_{i=1}^{n}\Big[I(\hat{U}_i;Y_i)+I(V_i;X_i|\hat{U}_i)-I(Y_{i+1}^n;X_i|I_AI_BX^{i-1})\Big]\ ,\end{aligned}$$

where X^i denotes the first i samples and $X_i^n=(X_i,\ldots,X_n)$; (g) stems from the chain rule and (h) from the assumed i.i.d. nature of the sources. In (i), the following identity is used [9]:

$$(55)\qquad \sum_{i=1}^{n} I(\mathbf{A}^{i-1};B_i|C,\mathbf{B}_{i+1}^n)=\sum_{i=1}^{n} I(\mathbf{B}_{i+1}^n;A_i|C,\mathbf{A}^{i+1})\ ,$$

where C can be arbitrarily dependent to the vectors $\mathbf{A}$ and $\mathbf{B}$, as long as it does not change with i, and the following auxiliary RVs are defined on measurable spaces $(\mathcal{U}_i\times\mathcal{V}_i,\mathcal{B}_{\mathcal{U}_i\times\mathcal{V}_i})$ by setting:

$$(56)\qquad \hat{U}_i\coloneqq(I_A,X^{i-1},Y_{i+1}^n)\quad\text{and}\quad V_i\coloneqq I_B\ ,\ \forall\, i=[1:n]\ .$$

It is important to emphasize that the required Markov chains in (43) are verified for each $i = [1 : n]$ (see Appendix B). Let Q be a RV uniformly distributed over $[1 : n]$, then:

$$\begin{aligned}\Delta_n &\leq I(\hat{U}_Q; Y_Q|Q) + I(V_Q; X_Q|\hat{U}_Q, Q) - \frac{1}{n}\sum_{i=1}^{n} I(Y_{i+1}^n; X_i|I_A I_B X^{i-1}) \\ &= I(U;Y) + I(V;X|U) - T \ ,\end{aligned} \tag{57}$$

where $U \coloneqq (\hat{U}_Q, Q)$. We now bound the required rate, from the size of the mappings, we have

$$nR \geq I(I_A; X^n) + I(I_B; Y^n I_A) \geq I(I_A; X^n) + I(I_B; Y^n|I_A) \ . \tag{58}$$

For convenience, we analyze each of these terms separately:

$$\begin{aligned}I(I_A; X^n) &\overset{(j)}{=} \sum_{i=1}^{n} I(I_A X^{i-1}; X_i) \\ &= \sum_{i=1}^{n} \left[I(I_A X^{i-1} Y_{i+1^n}; X_i) - I(Y_{i+1}^n; X_i|I_A X^{i-1})\right] \ ,\end{aligned} \tag{59}$$

where (j) is due to the i.i.d nature of samples. The second term writes as:

$$\begin{aligned}I(I_B; Y^n|I_A) &= \sum_{i=1}^{n} \left[I(I_B X^{i-1}; Y_i|I_A Y_{i+1}^n) - I(X^{i-1}; Y_i|I_A I_B Y_{i+1}^n)\right] \\ &= \sum_{i=1}^{n} \left[I(X^{i-1}; Y_i|I_A Y_{i+1}^n) + I(I_B; Y_i|I_A X^{i-1} Y_{i+1}^n) - I(X^{i-1}; Y_i|I_A I_B Y_{i+1}^n)\right] \\ &= \sum_{i=1}^{n} \left[I(I_B; Y_i|I_A X^{i-1} Y_{i+1}^n) + I(X_i; Y_{i+1}^n|I_A X^{i-1}) - I(X^{i-1}; Y_i|I_A I_B Y_{i+1}^n)\right] \ ,\end{aligned} \tag{60}$$

where the final step is due to identity (55). These inequalities lead to

$$\begin{aligned}nR \geq \sum_{i=1}^{n} \big[&I(I_A X^{i-1} Y_{i+1^n}; X_i) + I(I_B; Y_i|I_A X^{i-1} Y_{i+1}^n) \\ &- I(X^{i-1}; Y_i|I_A I_B Y_{i+1}^n)\big] \ .\end{aligned} \tag{61}$$

Using the same definitions for the auxiliary RVs as above, this result can be expressed as follows:

$$R \geq I(\hat{U}_Q; X_Q|Q) + I(V_Q; Y_Q|\hat{U}_Q, Q) - T \ , \tag{62}$$

and thus, the following region is an outer bound:

$$\begin{cases} \Delta_n \leq I(U;Y) + I(V;X|U) - T \ , \\ R \geq I(U;X) + I(V;Y|U) - T \ , \end{cases} \tag{63}$$

where (U,V) are auxiliary RVs that respect the required Markov chains in (43). It is left to show that (63) is equivalent or stricter than:

$$\begin{cases} \Delta_n \leq I(U;Y) + I(V;X|U) \ , \\ R \geq I(U;X) + I(V;Y|U) \ . \end{cases} \tag{64}$$

That is, all pairs (R,Δ_n) that are forbidden in the region in (63) are also forbidden in (64). In order to do so we use *Fourier-Motzkin* elimination [20] over $T \geq 0$. By removing T, we get:

$$\begin{cases} \Delta_n \leq I(U;Y) + I(V;X|U) \ , \\ R \geq I(U;X) + I(V;Y|U) - I(U;Y) - I(V;X|U) + \Delta_n \ , \end{cases} \tag{65}$$

and using the Markovian relations between the different RVs we obtain:

$$\begin{cases} \Delta_n \leq I(U;Y) + I(V;X|U) \ , \\ R \geq I(U;X|Y) + I(V;Y|UX) + \Delta_n \ . \end{cases} \tag{66}$$

In order to show the equivalence between the two regions, we need to check the extremal points. The point where $\Delta_n = 0$ is trivial, as $R = 0$ is optimal under both regions. When checking $\Delta_n = I(U;Y) + I(V;X|U)$ we have:

$$\begin{aligned} R &\geq I(U;X|Y) + I(V;Y|UX) + I(U;Y) + I(V;X|U) \\ &= I(U;X) + I(V;Y|U) \ , \end{aligned} \tag{67}$$

which completes the proof of the weak unfeasibility. □

Remark 3. *We conjecture that in contrast to the unidirectional testing problem [2], the strong unfeasibility property –implying that the error exponent does not depend on ϵ– does not hold for the collaborative hypothesis testing problem. A possible reason for this failure is that such a property heavily relies on the Blowing Up lemma (see Lemma 9) which does not hold conditioned on arbitrary probability events (e.g. the corresponding event induced from the first information layer).*

6. Collaborative Hypothesis Testing with Zero Rate. We now consider another special case of Proposition 2, whereby testing is done over two general hypotheses, but the total exchange rate is zero. It is worth mention that zero-rate does not mean that *no information exchange* is possible, but rather that the size of the codebook grows slower than exponentially with the blocklength n, as stated in the following proposition.

THEOREM 4 (Necessary and sufficient conditions under zero-rate). *Let P_{XY} and $P_{\bar{X}\bar{Y}}$ be any probability measures such that $supp(P_{\bar{X}\bar{Y}}) = supp(P_{XY}) = \mathcal{X} \times \mathcal{Y}$. Assume the total exchange rate $R = 0$, that is:*

$$\sum_{k=1}^{K} \log |f_{[k]}| + \sum_{k=1}^{K} \log |g_{[k]}| \equiv o(n) \ , \tag{68}$$

the optimal error exponent to the probability of Type II is given by

$$\lim_{n\to\infty} -\frac{1}{n} \log \beta_n(R = 0, \epsilon \,|K) = \min_{\tilde{X}\tilde{Y} \in \mathscr{L}_0(X,Y)} \mathcal{D}(P_{\tilde{X}\tilde{Y}} \| P_{\bar{X}\bar{Y}}) \coloneqq E(R = 0) \ , \ \ \forall\, 0 < \epsilon < 1 \ , \tag{69}$$

where $\mathscr{L}_0(X,Y) \coloneqq \left\{\tilde{X}\tilde{Y} : P_{\tilde{X}} = P_X \ , \ P_{\tilde{Y}} = P_Y\right\}$.

It is worth mentioning that the same expression (69) was proven in [11] to be feasible based on *unidirectional one bit exchange*, i.e., $|f_{[1]}| = 2$, $|g_{[1]}| = 0$. This observation implies that when zero-rate is enforced, not only data exchanges do not help, but only one bit of exchange is enough. In addition, note that this is a *strong unfeasability* result, as the optimal exponent for β_n is not dependent on the constraint ϵ over the error probability of Type I.

PROOF OF THEOREM 4. From the expression of the error exponent in (69), it is clear that it is enough to show the result for $K = 1$, since it is feasible with one round and the extension of the unfeasibility proof is straightforward. We start by proving the feasibility of the error exponent in (69) and then, we prove the unfeasibility result using methods similar to the ones in [21] for the case of a unidirectional exchanges.

Feasibility. As the error exponent in (69) is feasible with single-side exchange, we use Proposition 1 setting $V = \phi$. Thus, a feasible error exponent for zero-rate, as defined in Theorem 4:

$$\liminf_{n\to\infty} -\frac{1}{n} \log \beta_n(R = 0, \epsilon \,|K) \geq \max_{\mathscr{S}(R=0)} \min_{\mathscr{L}(U,X,Y)} \mathcal{D}(P_{\tilde{U}\tilde{X}\tilde{Y}} || P_{\bar{U}\bar{X}\bar{Y}}) \ , \tag{70}$$

where $\mathscr{S}$ and $\mathscr{L}$ are the sets defined in Proposition 1. Using the chain rule for KL divergence, this exponent can be bounded as follows:

(71)
$$\begin{aligned}
&\max_{\mathscr{S}(R=0)} \min_{\mathscr{L}(U,X,Y)} \mathcal{D}(P_{\tilde{U}\tilde{X}\tilde{Y}}||P_{\bar{U}\bar{X}\bar{Y}}) \\
&\quad = \max_{\mathscr{S}(R=0)} \min_{\mathscr{L}(U,X,Y)} \left[\mathcal{D}(P_{\tilde{X}\tilde{Y}}||P_{\bar{X}\bar{Y}}) + \mathcal{D}(P_{\tilde{U}|\tilde{X}\tilde{Y}}||P_{\bar{U}|\bar{X}\bar{Y}}|P_{\tilde{X}\tilde{Y}})\right] \\
&\quad = \max_{\mathscr{S}(R=0)} \min_{\mathscr{L}_0(X,Y)} \left[\mathcal{D}(P_{\tilde{X}\tilde{Y}}||P_{\bar{X}\bar{Y}}) + \min_{P_{\tilde{U}|\tilde{X}\tilde{Y}}} \mathcal{D}(P_{\tilde{U}|\tilde{X}\tilde{Y}}||P_{\bar{U}|\bar{X}\bar{Y}}|P_{\tilde{X}\tilde{Y}})\right] \\
&\quad \geq \min_{\mathscr{L}_0(X,Y)} \mathcal{D}(P_{\tilde{X}\tilde{Y}}||P_{\bar{X}\bar{Y}}) \ .
\end{aligned}$$

Here, the minimum over $P_{\tilde{U}|\tilde{X}\tilde{Y}}$ is such that $\tilde{U}\tilde{X}\tilde{Y} \in \mathscr{L}(U,X,Y)$, and the final inequality is due to the non-negativity of the KL divergence.

Strong unfeasibility. We now prove the optimality of Theorem 4, by showing that the error exponent of $\beta_n(R=0,\epsilon)$ does not depend on $\epsilon \in (0,1)$, and that (69) cannot be beaten. We follow a similar approach to [21], which addressed this proof for the case of unidirectional exchanges.

Let $f_{[1]} : \mathcal{X}^n \to \{1,\dots,|f_{[1]}|\}$ and $g_{[1]} : \mathcal{Y}^n \times \{1,\dots,|f_{[1]}|\} \to \{1,\dots,|g_{[1]}|\}$ be the encoding functions at node A and B, respectively, and let $\phi\big(X^n, g_{[1]}(Y^n, f_{[1]}(X^n))\big) \in \{0,1\}$ be the decoding function at node A. Define sets:

$$\mathcal{C}_{ij} \coloneqq \big\{\mathbf{x} \in \mathcal{X}^n : f_{[1]}(\mathbf{x}) = i \ \text{ and } \phi(\mathbf{x},j) = 0\big\} \ , \ \mathcal{C}_i \coloneqq \bigcup_{i=1}^{|f_{[1]}|} \mathcal{C}_{ij} \ ,$$
$$\mathcal{F}_{ij} \coloneqq \big\{\mathbf{y} \in \mathcal{Y}^n : g_{[1]}(\mathbf{y},i) = j\big\} \ , \ (i,j) \in \{1,\dots,|f_{[1]}|\} \times \{1,\dots,|g_{[1]}|\} \ .$$

Note that $\mathcal{C}_{ij}$ (respectively, $\mathcal{F}_{ij}$) cannot be said to be pairwise disjoint in $\mathcal{X}^n$ (respectively, $\mathcal{Y}^n$) while the sets $\mathcal{C}_i$ are pairwise disjoint. Similarly, for each index i_0, the sets $\mathcal{F}_{i_0 j}$ are disjoint. The acceptance set of H_0 can be expressed by

$$\mathcal{A}_n \coloneqq \bigcup_{i=1}^{|f_{[1]}|} \bigcup_{j=1}^{|g_{[1]}|} \mathcal{C}_{ij} \times \mathcal{F}_{ij} \ . \tag{72}$$

That is, if $(\mathbf{x},\mathbf{y}) \in \mathcal{A}_n$, $\phi\big(\mathbf{x}, g_{[1]}(\mathbf{y}, f_{[1]}(\mathbf{x}))\big) = 0$ and otherwise, the result is H_1. By the definition, $P_{XY}^n(\mathcal{A}_n^c) \leq \epsilon$, or equivalently

$$P_{XY}^n(\mathcal{A}_n) = P_{XY}^n\left(\bigcup_{i=1}^{|f_{[1]}|} \bigcup_{j=1}^{|g_{[1]}|} \mathcal{C}_{ij} \times \mathcal{F}_{ij}\right) > 1-\epsilon \ . \tag{73}$$

Since the sets $\mathcal{B}_i \coloneqq \bigcup_{j=1}^{|g_{[1]}|} \mathcal{C}_{ij} \times \mathcal{F}_{ij}$ are disjoint, by relying on (73) and on the size $|f_{[1]}|$, there exists an index i_0 such that

$$P_{XY}^n \left(\bigcup_{j=1}^{|g_{[1]}|} \mathcal{C}_{i_0 j} \times \mathcal{F}_{i_0 j} \right) \geq \frac{1-\epsilon}{|f_{[1]}|} \ . \tag{74}$$

As the sets $F_{i_0 j}$ are disjoint, there exists an index j_0 such that

$$P_{XY}^n (\mathcal{C}_{i_0 j_0} \times \mathcal{F}_{i_0 j_0}) \geq \frac{1-\epsilon}{|f_{[1]}||g_{[1]}|} \ . \tag{75}$$

Letting $\mathcal{C} \equiv \mathcal{C}_{i_0 j_0}$ and $\mathcal{F} \equiv \mathcal{F}_{i_0 j_0}$, we rewrite this as:

$$P_{XY}^n (\mathcal{C} \times \mathcal{F}) \geq \frac{1-\epsilon}{|f_{[1]}||g_{[1]}|} \equiv \exp(-n\delta_n) \ , \tag{76}$$

with $\delta_n \equiv \frac{1}{n} \log \left(|f_{[1]}||g_{[1]}| \right) - \frac{1}{n} \log(1-\epsilon)$. As the log-function is monotonic and both $|f_{[1]}|$ and $|g_{[1]}|$ are non-negative, expression (68) implies that $\log |f_{[1]}| = o(n)$ and $\log |g_{[1]}| = o(n)$ and thus $\delta_n = o(1)$.

Having shown that there exist sets $\mathcal{C}$ and $\mathcal{F}$, such that $\mathcal{C} \times \mathcal{F} \in \mathcal{A}_n$, and the probability $P_{XY}(\mathcal{C} \times \mathcal{F})$ does not approach 0 exponentially with n, the rest of the proof follows along the lines in [21]. For the sake of completeness, this proof is completed in Appendix D. □

APPENDIX A: TECHNICAL DEFINITIONS AND LEMMAS

In this appendix, we revise fundamental notions and properties of *method of types* [8], which are extensively used through this paper.

DEFINITION 2 (Types [9]). *The type of a sequence $\mathbf{x} \in \mathcal{X}^n$ is the measure $\hat{P}_X$ on $\mathcal{X}$ defined by $\hat{P}_X(a) \coloneqq \frac{1}{n} N(a|\mathbf{x})$, $\forall a \in \mathcal{X}$, where $N(a|\mathbf{x})$ is the counting measure of the letter a in $\mathbf{x}$. The* joint type *of a pair $(\mathbf{x}, \mathbf{y}) \in \mathcal{X}^n \times \mathcal{Y}^n$ is the empirical measure $\hat{P}_{XY}$ on $\mathcal{X} \times \mathcal{Y}$ such that*

$$\hat{P}_{XY}(a,b) \coloneqq \frac{1}{n} N(a,b|\mathbf{x},\mathbf{y}) \ , \quad \forall (a,b) \in \mathcal{X} \times \mathcal{Y} \ , \tag{77}$$

where $N(a,b|\mathbf{x},\mathbf{y})$ is the joint counting measure of the pair (a,b) in $(\mathbf{x},\mathbf{y})$.

DEFINITION 3 (Conditional Types [9]). *The vector $\mathbf{y} \in \mathcal{Y}^n$ is said to have* conditional type $V : \mathcal{X} \mapsto \mathcal{P}_n(\mathcal{Y})$ *given $\mathbf{x} \in \mathcal{X}^n$ if*

$$N(a,b|\mathbf{x},\mathbf{y}) = N(a|\mathbf{x}) V(b|a) \ , \quad \forall (a,b) \in \mathcal{X} \times \mathcal{Y} \ , \tag{78}$$

where V is a stochastic mapping.

LEMMA 2 (Type Counting). *Let $\mathcal{P}_n(\mathcal{X})$ be the set of all possible types of sequences in $\mathcal{X}^n$. Then, $|\mathcal{P}_n(\mathcal{X})| \leq (n+1)^{|\mathcal{X}|}$.*

PROOF. Refer to reference [9, Lemma 2.2]. □

LEMMA 3. *For any type $\hat{P} \in \mathcal{P}_n(\mathcal{X})$ of sequences in $\mathcal{X}^n$, denote by $\mathcal{T}_{[\hat{P}]}$ the set of all sequences with this type. Then,*

$$(n+1)^{-|\mathcal{X}|} \exp\left[nH(\hat{P})\right] \leq |\mathcal{T}_{[\hat{P}]}| \leq \exp\left[nH(\hat{P})\right] . \tag{79}$$

In a similar fashion, for every $\mathbf{x} \in \mathcal{X}^n$ and stochastic mapping $V : \mathcal{X} \mapsto \mathcal{P}_n(\mathcal{Y})$, let $\mathcal{T}_{[V]}(\mathbf{x})$ be the set of all sequences $\mathbf{y} \in \mathcal{Y}^n$ with the conditional type V given $\mathbf{x}$. Then,

$$(n+1)^{-|\mathcal{X}||\mathcal{Y}|} \exp\left[nH(V|\hat{P})\right] \leq |\mathcal{T}_{[V]}(\mathbf{x})| \leq \exp\left[nH(V|\hat{P})\right] , \tag{80}$$

where $H(V|\hat{P})$ is the conditional entropy function,

$$H(V|\hat{P}) = \sum_{x \in \mathcal{X}} \hat{P}(x) H(V(\cdot|x)) . \tag{81}$$

PROOF. Refer to reference [9, Lemma 2.3, Lemma 2.5]. □

LEMMA 4 (Inaccuracy). *Let $\hat{P} \in \mathcal{P}_n(\mathcal{X})$ be the type of $\mathbf{x} \in \mathcal{X}^n$ ($X^{(n)} \sim \hat{P}$ is referred to as the* type variable*). Then, for any RV X on $(\mathcal{X}, \mathcal{B}_{\mathcal{X}}, P_X)$,*

$$P_X^n(X^n = \mathbf{x}) = \exp\Big\{-n\left[H(\hat{P}) + \mathcal{D}(\hat{P}\|P_X)\right]\Big\} . \tag{82}$$

PROOF. Refer to reference [11, Lemma 3],[9, Lemma 2.6]. □

DEFINITION 4 (δ-Typicality [11]). *Let $\delta > 0$, an n-sequence $\mathbf{x}$ is called δ-typical, denoted by $\mathcal{T}_{[X]_\delta}$, if $|N(a|\mathbf{x}) - nP_X(a)| \leq \mathcal{O}(\delta), \quad \forall a \in \mathcal{X}$, and $\hat{P}_X \ll P_X$. Jointly δ-typical $\mathcal{T}_{[XY]_\delta}$ and conditionally δ-typical sequences $\mathcal{T}_{[Y|X]_\delta}(\mathbf{x})$ are defined in a similar manner.*

LEMMA 5. *Let $\mathcal{T}_{[X]_\delta}$, $\mathcal{T}_{[XY]_\delta}$ and $\mathcal{T}_{[Y|X]_\delta}$ denote the sets of typical, jointly typical and conditionally typical sequences, respectively. For any $\mathbf{x} \in \mathcal{T}_{[X]_\delta}$ and $\mathbf{y} \in \mathcal{T}_{[Y|X]_{\delta'}}$, then $(\mathbf{x}, \mathbf{y}) \in \mathcal{T}_{[XY]_{\delta+\delta'}}$. Moreover, $\mathbf{y} \in \mathcal{T}_{[Y]_{\delta''}}$, with $\delta'' := (\delta + \delta')|\mathcal{X}|$.*

PROOF. Refer to reference [9]. □

LEMMA 6 (Generalized Markov Lemma). *Let $p_{UXY} \in \mathcal{P}(\mathcal{U} \times \mathcal{X} \times \mathcal{Y})$ be a probability measure that satisfies: $U \multimap X \multimap Y$. Consider $(\mathbf{x}, \mathbf{y}) \in \mathcal{T}^n_{[XY]_{\epsilon'}}$ and random vectors U^n generated according to:*

$$\Pr\left\{U^n = \mathbf{u} \middle| U^n \in \mathcal{T}^n_{[U|X]_{\epsilon''}}(\mathbf{x}), \mathbf{x}, \mathbf{y}\right\} = \frac{\mathbb{1}\left\{u^n \in \mathcal{T}^n_{[U|X]_{\epsilon''}}(\mathbf{x})\right\}}{\left|\mathcal{T}^n_{[U|X]_{\epsilon''}}(\mathbf{x})\right|} \ . \tag{83}$$

For sufficiently small $\epsilon, \epsilon', \epsilon'' > 0$,

$$\Pr\left\{U^n \notin \mathcal{T}^n_{[U|XY]_{\epsilon}}(\mathbf{x}, \mathbf{y}) \middle| U^n \in \mathcal{T}^n_{[U|X]_{\epsilon''}}(\mathbf{x}), \mathbf{x}, \mathbf{y}\right\} \equiv \mathcal{O}\left(c^{-n}\right) \tag{84}$$

holds uniformly on $(\mathbf{x}, \mathbf{y}) \in \mathcal{T}^n_{[XY]_{\epsilon'}}$ where $c > 1$.

PROOF. Refer to reference [19]. □

LEMMA 7. *For every probability measure $P_X \in \mathcal{P}(\mathcal{X})$ and stochastic mapping $W : \mathcal{X} \mapsto \mathcal{P}(\mathcal{Y})$, there exist sequences $(\varepsilon_n)_{n\in\mathbb{N}_+}, (\varepsilon'_n)_{n\in\mathbb{N}_+} \to 0$ as $n \to \infty$ satisfying:*

$$\left|\frac{1}{n}\log|\mathcal{T}_{[X]_\delta}| - H(X)\right| \leq \varepsilon_n \ , \ \left|\frac{1}{n}\log|\mathcal{T}_{[Y|X]_\delta}(\mathbf{x})| - H(Y|X)\right| \leq \varepsilon_n \ , \tag{85}$$

for each $\mathbf{x} \in \mathcal{T}_{[X]_\delta}$ where $\varepsilon_n \equiv \mathcal{O}(n^{-1}\log n)$, and

$$P_X^n\left(\mathcal{T}_{[X]_\delta}\right) \geq 1 - \varepsilon'_n \ , \ W^n\left(\mathcal{T}_{[Y|X]_\delta}(\mathbf{x})|X^n = \mathbf{x}\right) \geq 1 - \varepsilon'_n \ , \tag{86}$$

for all $\mathbf{x} \in \mathcal{X}^n$ where $\varepsilon'_n \equiv \mathcal{O}\left(\frac{1}{n\delta^2}\right)$, provided that n is sufficiently large.

PROOF. Refer to reference [9, Lemma 2.13]. □

APPENDIX B

As a part of the weak unfeasibility part of the proof of Theorem 3, two Markov chains are necessary:

$$\begin{cases} \hat{U}_i \multimap X_i \multimap Y_i \ , \ \forall\, i = [1:n] \\ V_i \multimap (\hat{U}_i, Y_i) \multimap X_i \ , \ \forall\, i = [1:n]. \end{cases} \tag{87}$$

Using the chosen RVs from (56), these Markov chains are represented by

$$\begin{cases} (I_A, X^{i-1}, Y_{i+1}^n) \multimap X_i \multimap Y_i \ , \ \forall\, i = [1:n] \\ I_B \multimap (I_A, X^{i-1}, Y_i^n) \multimap X_i \ , \ \forall\, i = [1:n]. \end{cases} \tag{88}$$

In order to check this, we use the next lemma.

LEMMA 8. *Let A_1, A_2, B_1, B_2 be RVs with joint probability measure $P_{A_1A_2B_1B_2} = P_{A_1B_1}P_{A_2B_2}$ and assume that $\{f^i\}_{i=1}^k, \{g^i\}_{i=1}^k$ are any collection of P-measurable mappings with domain structure given by:*

$$f^1(A_1, A_2); f^2(A_1, A_2, g^1); \ldots; f^k(A_1, A_2, g^1, \ldots, g^{k-1}) \ , \tag{89}$$

$$g^1(B_1, B_2, f^1); g^2(B_1, B_2, f^1, f^2); \ldots; g^k(B_1, B_2, f^1, \ldots, f^k) \ . \tag{90}$$

Then,

$$I(A_2; B_1 | f^1, f^2, \ldots, f^k, g^1, g^2, \ldots, g^k, A_1, B_2) = 0 \ . \tag{91}$$

PROOF. Refer to reference [13, Lemma 1]. □

In order to prove the first Markov chain, we simply let:

$$\begin{cases} A_1 \coloneqq X_i, \quad B_1 \coloneqq Y_i \ , \\ A_2 \coloneqq (X^{i-1}, X_{i+1}^n, Y_{i+1}^n) \ , \quad B_2 \coloneqq Y^{i-1} \ . \end{cases} \tag{92}$$

It can be easily verified that $P_{A_1A_2B_1B_2} = P_{A_1B_1}P_{A_2B_2}$, which stems directly from the i.i.d. nature of the samples. Thus, according to Lemma 8:

$$\begin{aligned} 0 &= I(X^{i-1}X_{i+1}^n Y_{i+1}^n; Y_i | X_i Y^{i-1}) \\ &= I(X^{i-1}X_{i+1}^n Y^{i-1} Y_{i+1}^n; Y_i | X_i) - I(Y^{i-1}; Y_i | X_i) \ , \end{aligned} \tag{93}$$

which shows the Markov chain:

$$(X^{i-1}, X_{i+1}^n, Y^{i-1}, Y_{i+1}^n) \multimap X_i \multimap Y_i \ , \ \forall\, i = [1:n]. \tag{94}$$

As $I_A \coloneqq f_{[1]}(X^n)$, the following Markov chain is also true:

$$(I_A, X^{i-1}, Y_{i+1}^n) \multimap X_i \multimap Y_i \ , \ \forall\, i = [1:n] \tag{95}$$

which proves the first Markov chain in (88). As for the second one, we let:

$$\begin{cases} A_1 \coloneqq X^{i-1} \ , \quad B_1 \coloneqq Y^{i-1} \ , \\ A_2 \coloneqq (X_i, X_{i+1}^n) \ , \quad B_2 \coloneqq (Y_i, Y_{i+1}^n) \ . \end{cases} \tag{96}$$

Under this choice, $I_A \coloneqq f_{[1]}(A_1, A_2)$ and thus,

$$I(X_i X_{i+1}^n; Y^{i-1} | I_A X^{i-1} Y_i Y_{i+1}^n) = 0 \ , \ \forall\, i = [1:n]. \tag{97}$$

The later identity proves the following Markov chain:

$$(X_i, X_{i+1}^n) \multimap (I_A, X^{i-1}, Y_i, Y_{i+1}^n) \multimap Y^{i-1} \ , \ \forall\, i = [1:n]. \tag{98}$$

As $I_B \coloneqq g_{[1]}(I_A, Y^n)$, it also holds that:

$$X_i \multimap (I_A, X^{i-1}, Y_i^n) \multimap I_B \ , \ \forall\, i = [1:n] \tag{99}$$

which yields the desired Markov chain.

APPENDIX C

PROOF OF LEMMA 1. For block-length n, given a code characterized by the encoding mappings $f_{[1]}, g_{[1]}$ at nodes A and B respectively, and a decoding mapping ϕ at node A. Let the acceptance region be denoted by

(100)
$$\mathcal{A}_n \coloneqq \left\{(\mathbf{x}, j) \in \mathcal{X}^n \times \{1, \ldots, |g_{[1]}|\} : g_{[1]}\big(\mathbf{y}, f_{[1]}(\mathbf{x})\big) = j,\ \mathbf{y} \in \mathcal{Y}^n,\ \phi\big(\mathbf{x}, j\big) = 0\right\}.$$

Let P and Q denote the probabilities measures on $\mathcal{X}^n \times \{1, \ldots, |g_{[1]}|\}$ induced by H_0 and H_1, respectively. From the *log-sum inequality* [9], we have:

$$\begin{aligned}(101)\qquad &\mathcal{D}\left(P_{X^n I_A I_B} \| Q_{X^n I_A I_B}\right) = \mathcal{D}\left(P_{X^n I_B} \| Q_{X^n I_B}\right) \\ &\geq (1-\alpha_n) \log \frac{1-\alpha_n}{\beta_n(R, \epsilon\,|K=1)} + \alpha_n \log \frac{\alpha_n}{1-\beta_n(R, \epsilon\,|K=1)}\ ,\end{aligned}$$

where $I_A \coloneqq f_{[1]}(X^n)$, $I_B \coloneqq g_{[1]}(I_A, Y^n)$, $\alpha_n(R|K=1) \coloneqq P(\mathcal{A}_n^c) \leq \epsilon$ and $\beta_n(R, \epsilon\,|K=1) \coloneqq Q(\mathcal{A}_n)$. Through some algebra this yields:

$$(102)\qquad \mathcal{D}\left(P_{X^n I_A I_B} \| Q_{X^n I_A I_B}\right) \geq (1-\alpha_n) \log \frac{1}{\beta_n(R, \epsilon\,|K=1)} - h_2(\alpha_n)\ ,$$

where $h_2(p) \coloneqq -p \log p - (1-p)\log(1-p)$ is the *binary entropy* function. By assumption $\epsilon \to 0$ as $n \to \infty$, one conclude that for n large enough

$$(103)\qquad -\frac{1}{n} \log \beta_n(R, \epsilon\,|K=1) \leq \frac{1}{n} \mathcal{D}\left(P_{X^n I_A I_B} \| Q_{X^n I_A I_B}\right) - \delta_n\ ,$$

with $\delta_n \to 0$ as $n \to \infty$. Using the chain rule, we continue to get:

$$\begin{aligned}\mathcal{D}\big(P_{X^n I_A I_B} \| Q_{X^n I_A I_B}\big) &\overset{(k)}{=} I(I_B; X^n | I_A) + \mathcal{D}\left(P_{I_B|I_A} \| Q_{I_B|I_A} | P_{I_A}\right) \\ &\overset{(l)}{\leq} I(I_B; X^n | I_A) + \mathcal{D}\left(P_{Y^n I_A I_B} \| Q_{Y^n I_A I_B}\right) \\ &\overset{(m)}{=} I(I_B; X^n | I_A) + \mathcal{D}\left(P_{Y^n I_A} \| P_Y^n P_{I_A}\right) \\ &= I(I_B; X^n | I_A) + I(I_A; Y^n)\ .\end{aligned}$$

Here, (k) and (l) stem from the chain rule for the KL-divergence, and (m) is due to the fact that we consider the case of testing against independence. With this, the *weak unfeasibility* proof is completed. □

APPENDIX D

COMPLEMENTARY PROOF OF THEOREM 4. We now complete the proof of the strong unfeasibility to Theorem 4. To this end, we recall that we showed there exist sets $\mathcal{C} \subset \mathcal{X}^n$ and $\mathcal{F} \subset \mathcal{Y}^n$ such that $\mathcal{C} \times \mathcal{F} \in \mathcal{A}_n$, and $P_{XY}^n(\mathcal{C} \times \mathcal{F}) \geq \exp(-n\delta_n)$, with $\delta_n \to 0$ as $n \to \infty$. We now evoke the "Blowing-Up" Lemma:

LEMMA 9 (Blowing-up Lemma). *Let $Y^n = (Y_1, \ldots, Y_n)$ be independent random variables in $(\mathcal{Y}^n, \mathcal{B}_{\mathcal{Y}^n})$ distributed according to $W^n(Y^n|X^n = \mathbf{x})$ for some fixed vector $\mathbf{x} \in \mathcal{X}^n$ and a stochastic mapping $W : \mathcal{X} \mapsto \mathcal{P}(\mathcal{Y})$ and let $\delta_n \to 0$ be a given sequence. There exist sequences $k_n \equiv o(n)$ and $\gamma_n \equiv o(1)$, such that for every subset $\mathcal{A}_n \subset \mathcal{Y}^n$:*

$$W^n(\mathcal{A}_n|X^n = \mathbf{x}) \geq \exp(-n\delta_n) \text{ implies } W^n\left(\Gamma^{k_n}\mathcal{A}_n|X^n = \mathbf{x}\right) \geq 1-\gamma_n \tag{104}$$

where $\Gamma^{k_n}\mathcal{A}_n$ denotes the Γ^{k_n}-neighborhood of the set $\mathcal{A}_n$ defined by

$$\Gamma^{k_n}\mathcal{A}_n \coloneqq \left\{\hat{\mathbf{y}} \in \mathcal{Y}^n : \min_{\mathbf{y}\in\mathcal{A}_n} \rho_n(\hat{\mathbf{y}}, \mathbf{y}) \leq k_n\right\} , \tag{105}$$

where $\rho_n(\hat{\mathbf{y}}, \mathbf{y}) \coloneqq \sum_{i=1}^{n} \mathbb{1}\{\hat{y}_i \neq y_i\}$ and $\mathbb{1}\{\hat{y} \neq y\} = 1$ if $\hat{y} \neq y$ or $= 0$ otherwise.

PROOF. Refer to references [16, 3]. □

The rest of the proof follows closely the steps taken in [21]. As $P^n_{XY}(\mathcal{C} \times \mathcal{F}) \geq \exp(-n\delta_n)$, clearly $P^n_X(\mathcal{C}) \geq \exp(-n\delta_n)$ and $P^n_Y(\mathcal{F}) \geq \exp(-n\delta_n)$. Using the non-conditional version of Lemma 9, there exist sequences $k_n = o(n)$ and $\gamma_n = o(1)$ s.t.:

$$P^n_X\left(\Gamma^{k_n}\mathcal{C}\right) \geq 1 - \gamma_n , \quad P^n_Y\left(\Gamma^{k_n}\mathcal{F}\right) \geq 1 - \gamma_n , \tag{106}$$

where k_n, γ_n only depend on $|\mathcal{X}|, |\mathcal{Y}|$ and δ_n, but not on P_{XY}. Equation (106) holds true if we change P_X to $P_{\tilde{X}}$ and P_Y to $P_{\tilde{Y}}$, for some $\tilde{X}\tilde{Y} \in \mathscr{L}_0$. As we wish to analyze the error probability for fixed n, during most of this proof we take the liberty to dismiss the subscript n from k_n, for the sake of readability.

Using the fact $\Pr(A \cap B) \geq \Pr(A) + \Pr(B) - 1$ and (106), we obtain:

$$P^n_{\tilde{X}\tilde{Y}}\left(\Gamma^k\mathcal{C} \times \Gamma^k\mathcal{F}\right) \geq P^n_{\tilde{X}}\left(\Gamma^k\mathcal{C}\right) + P^n_{\tilde{Y}}\left(\Gamma^k\mathcal{F}\right) - 1 \geq 1 - 2\gamma_n . \tag{107}$$

Consider the set of η-typical sequences defined by $P_{\tilde{X}\tilde{Y}}$. By Lemma 7,

$$P^n_{\tilde{X}\tilde{Y}}(\mathcal{T}_{[\tilde{X}\tilde{Y}]_\eta}) \geq 1 - \mathcal{O}\left(\frac{1}{n\eta^2}\right) = 1 - \mathcal{O}\left(n^{-\frac{1}{3}}\right) , \tag{108}$$

where the last equality is a result of the choice $\eta \equiv \eta_n \coloneqq n^{-\frac{1}{3}}$. Combining (107) and (108), it is clear that for sufficiently large n,

$$P^n_{\tilde{X}\tilde{Y}}\left(\Gamma^k\mathcal{C} \times \Gamma^k\mathcal{F}\right) \cap \mathcal{T}_{[\tilde{X}\tilde{Y}]_\eta}\big) \geq \frac{1}{2} . \tag{109}$$

By the definition of the η-typical set (see Definition 4), we have:

$$\mathcal{T}_{[\tilde{X}\tilde{Y}]_\eta} = \bigcup_{\substack{P_{\hat{X}\hat{Y}} \in \mathcal{P}_n(\mathcal{X}\times\mathcal{Y}) \\ |P_{\hat{X}\hat{Y}} - P_{\tilde{X}\tilde{Y}}| \leq \eta \ , \ P_{\hat{X}\hat{Y}} \ll P_{\tilde{X}\tilde{Y}}}} \mathcal{T}_{[\hat{X}\hat{Y}]} \ , \tag{110}$$

where $|P_{\hat{X}\hat{Y}} - P_{\tilde{X}\tilde{Y}}| \leq \eta$ refers to the maximum over all the arguments in $\mathcal{X} \times \mathcal{Y}$. As all elements of $\mathcal{T}_{[\hat{X}\hat{Y}]}$ are equiprobable under an i.i.d measure, (109) can be rewritten as

$$\sum_{\substack{P_{\hat{X}\hat{Y}} \in \mathcal{P}_n(\mathcal{X}\times\mathcal{Y}) \\ |P_{\hat{X}\hat{Y}} - P_{\tilde{X}\tilde{Y}}| \leq \eta \ , \ P_{\hat{X}\hat{Y}} \ll P_{\tilde{X}\tilde{Y}}}} P^n_{\tilde{X}\tilde{Y}}\big(\mathcal{T}_{[\hat{X}\hat{Y}]}\big) \frac{|(\Gamma^k\mathcal{C} \times \Gamma^k\mathcal{F}) \cap \mathcal{T}_{[\hat{X}\hat{Y}]_\eta}|}{|\mathcal{T}_{[\hat{X}\hat{Y}]_\eta}|} \geq \frac{1}{2} \ . \tag{111}$$

As $P^n_{\tilde{X}\tilde{Y}}(\mathcal{T}_{[\hat{X}\hat{Y}]}) \leq 1$, by using the bound over the size of the set $\mathcal{P}_n(\mathcal{X} \times \mathcal{Y})$ in Lemma 2, there must be *at least one type* $\mathcal{T}_{[\hat{X}\hat{Y}]}$, for which

$$\frac{|(\Gamma^k\mathcal{C} \times \Gamma^k\mathcal{F}) \cap \mathcal{T}_{[\hat{X}\hat{Y}]_\eta}|}{|\mathcal{T}_{[\hat{X}\hat{Y}]_\eta}|} \geq \frac{1}{2}(n+1)^{-|\mathcal{X}|\mathcal{Y}|} = \frac{1}{2}\exp(-n\epsilon_n) \ , \tag{112}$$

with $\epsilon_n = \mathcal{O}(n^{-1}\log(n+1)) \to 0$ as $n \to \infty$. The equiprobability property is also true for the probability measure implied by H_1, that is $P_{\bar{X}\bar{Y}}$. Thus,

$$\begin{aligned} P^n_{\bar{X}\bar{Y}}\big(\Gamma^k\mathcal{C} \times \Gamma^k\mathcal{F}\big) &\geq P^n_{\bar{X}\bar{Y}}\big(\Gamma^k\mathcal{C} \times \Gamma^k\mathcal{F}\big) \cap \mathcal{T}_{\hat{X}\hat{Y}}) \\ &= P^n_{\bar{X}\bar{Y}}(\mathcal{T}_{\hat{X}\hat{Y}}) \frac{|(\Gamma^k\mathcal{C} \times \Gamma^k\mathcal{F}) \cap \mathcal{T}_{\hat{X}\hat{Y}}|}{|\mathcal{T}_{\hat{X}\hat{Y}}|} \\ &\geq \frac{1}{2}\exp(-n\epsilon_n) P^n_{\bar{X}\bar{Y}}(\mathcal{T}_{\hat{X}\hat{Y}}) \ , \end{aligned} \tag{113}$$

where the final inequality stems from (112).

Consider now an arbitrary element $(\mathbf{u}, \mathbf{v}) \in \Gamma^k\mathcal{C} \times \Gamma^k\mathcal{F}$. By definition, there exist an element $(\mathbf{x}, \mathbf{y}) \in \mathcal{C} \times \mathcal{F}$, such that $(u_i, v_i) \neq (x_i, y_i)$ at most in $2k$ locations. Thus,

$$P^n_{\bar{X}\bar{Y}}(\mathbf{u}, \mathbf{v}) = \prod_{i=1}^n P_{\bar{X}\bar{Y}}(u_i, v_i) \leq \rho^{-2k} \prod_{i=1}^n P_{\bar{X}\bar{Y}}(x_i, y_i) = \rho^{-2k} P^n_{\bar{X}\bar{Y}}(\mathbf{x}, \mathbf{y}) \ , \tag{114}$$

with $\rho = \min\limits_{(x,y)\in\mathcal{X}\times\mathcal{Y}} P_{\bar{X}\bar{Y}}(x, y)$, and we assume that $\rho > 0$ (which complies with the preliminaries of Theorem 4). As $(\mathbf{u}, \mathbf{v})$ range over $\Gamma^k\mathcal{C} \times \Gamma^k\mathcal{F}$,

each element $(\mathbf{x},\mathbf{y}) \in \mathcal{C}\times\mathcal{F}$ will be chosen as the closest neighbor at most $|\Gamma^k(\mathbf{x})| \times |\Gamma^k(\mathbf{y})|$ times. Thus,

$$P^n_{\bar{X}\bar{Y}}\big(\Gamma^k\mathcal{C}\times\Gamma^k\mathcal{F}\big) \le \rho^{-2k}|\Gamma^k(\mathbf{x})| \times |\Gamma^k(\mathbf{y})| P^n_{\bar{X}\bar{Y}}(\mathcal{C}\times\mathcal{F}) \ . \tag{115}$$

From [9, Lemma 5.1] we have:

$$|\Gamma^k_n(\mathbf{x})| \le \exp\left[n\left(h_2\left(\frac{k_n}{n}\right)+\frac{k_n}{n}\log|\mathcal{X}|\right)\right] \equiv \exp(n\zeta'_n) \ , \tag{116}$$

with $h_2(\cdot)$ being the *binary entropy* function and $\zeta'_n \to 0$ as $n\to\infty$. This implies that

$$P^n_{\bar{X}\bar{Y}}\big(\Gamma^k\mathcal{C}\times\Gamma^k\mathcal{F}\big) \le \exp(n\zeta_n) P^n_{\bar{X}\bar{Y}}(\mathcal{C}\times\mathcal{F}) \ , \tag{117}$$

with $\zeta_n \coloneqq 2h_2\left(\frac{k_n}{n}\right)+\frac{k_n}{n}\log(|\mathcal{X}|\mathcal{Y}|)-\frac{2k_n}{n}\log\rho \to 0$ as $n\to\infty$. Combining this with (113), we finally get

$$\begin{aligned} P^n_{\bar{X}\bar{Y}}(\mathcal{C}\times\mathcal{F}) &\ge \exp(-n\zeta_n) P^n_{\bar{X}\bar{Y}}\big(\Gamma^k\mathcal{C}\times\Gamma^k\mathcal{F}\big) \\ &\ge \frac{1}{2}\exp\left[-n(\zeta_n+\epsilon_n)\right] P^n_{\bar{X}\bar{Y}}(\mathcal{T}_{\hat{X}\hat{Y}}) \\ &\ge \frac{(n+1)^{|\mathcal{X}|\mathcal{Y}|}}{2}\exp\left[-n\big(\mathcal{D}(P_{\hat{X}\hat{Y}}\|P_{\bar{X}\bar{Y}})+\zeta_n+\epsilon_n\big)\right] \\ &\ge \exp\left[-n\big(\mathcal{D}(P_{\hat{X}\hat{Y}}\|P_{\bar{X}\bar{Y}})+\mu_n\big)\right] \ , \end{aligned} \tag{118}$$

and $\mu_n \equiv \mu_n(\rho,\epsilon,M_n,N_n,|\mathcal{X}|,|\mathcal{Y}|) \to 0$ as $n\to\infty$.

The previous conclusion is true for *some type* $P_{\hat{X}\hat{Y}}$ over the range of all types that are η-typical for the measure $P_{\tilde{X}\tilde{Y}}$. As the divergence functional $\mathcal{D}(\cdot\|\cdot)$ is convex and bounded, it is also uniformly continuous. It follows that we can find a sequence $\mu'_n \equiv \mu'_n(\rho,|\mathcal{X}|,|\mathcal{Y}|)$ such that $|P_{\hat{X}\hat{Y}}-P_{\tilde{X}\tilde{Y}}| \le \eta = o(n^{-\frac{1}{3}})$ implies that $|\mathcal{D}(P_{\hat{X}\hat{Y}}\|P_{\bar{X}\bar{Y}})-\mathcal{D}(P_{\tilde{X}\tilde{Y}}\|P_{\bar{X}\bar{Y}})| \le \mu'_n$. Hence

$$P^n_{\bar{X}\bar{Y}}(\mathcal{C}\times\mathcal{F}) \ge \exp\left[-n\big(\mathcal{D}(P_{\tilde{X}\tilde{Y}}\|P_{\bar{X}\bar{Y}})+\mu_n+\mu'_n\big)\right] \ , \tag{119}$$

and consequently

$$\begin{aligned} -\liminf_{n\to\infty}\frac{1}{n}\log P^n_{\bar{X}\bar{Y}}(\mathcal{A}_n) &= -\lim_{n\to\infty}\frac{1}{n}\log\beta_n(R=0,\epsilon\,|K=1) \\ &\le \mathcal{D}(P_{\tilde{X}\tilde{Y}}\|P_{\bar{X}\bar{Y}}) \ , \end{aligned} \tag{120}$$

and the RVs $\tilde{X}\tilde{Y}$ are chosen from the set $\mathscr{L}_0$, which concludes the proof. □

REFERENCES

[1] Ahlswede, R. and Burnashev, M. (1990). On minimax estimation in the presence of side information about remote data. *The Annals of Statistics* **18** 141–171.

[2] Ahlswede, R. and Csiszar, I. (1986). Hypothesis testing with communication constraints. *Information Theory, IEEE Transactions on* **32** 533-542.

[3] Ahlswede, R., Gács, P. and Körner, J. (1976). Bounds on conditional probabilities with applications in multi-user communication. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **34** 157–177.

[4] Bucklew, J. and Ney, P. (1991). Asymptotically optimal hypothesis testing with memory constraints. *The Annals of Statistics* **18** 982–998.

[5] Chiyonobu, T. (2001). Hypothesis testing for signal detection problem and large deviations. *Nagoya Mathematical Journal* **162** 187–203.

[6] Cover, T. M. (1969). Hypothesis testing with finite statistics. *The Annals of Mathematical Statistics* **40** 828–835.

[7] Cover, T. M. and Thomas, J. A. (1991). *Elements of information theory*. John Wiley & Sons, New York.

[8] Csiszár, I. (1998). The Method of Types. *Information Theory, IEEE Transactions on* **44** 2505-2523.

[9] Csiszar, I. and Körner, J. (2011). *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press.

[10] El Gamal, A. and Kim, Y.-H. (2011). *Network information theory*. Cambridge University Press.

[11] Han, T. (1987). Hypothesis testing with multiterminal data compression. *Information Theory, IEEE Transactions on* **33** 759-772.

[12] Hellman, M. E. and Cover, T. M. (1970). Learning with finite memory. *The Annals of Mathematical Statistics* **41** 765–782.

[13] Kaspi, A. (1985). Two-way source coding with a fidelity criterion. *Information Theory, IEEE Transactions on* **31** 735-740.

[14] Katz, G., Piantanida, P. and Debbah, M. (2016). Distributed Binary Detection with Lossy Data Compression. *ArXiv e-prints*. Submitted to Information Theory, IEEE Trans. on.

[15] Lehmann, E. L. and Romano, J. P. *Testing Statistical Hypotheses. Springer Texts in Statistics*.

[16] Margulis, G. A. (1974). Probabilistic characteristics of graphs with large connectivity. *Problemy Peredači Informacii* **10** 101–108.

[17] Naghshvar, M. and Javidi, T. (2013). Active sequential hypothesis testing. *The Annals of Statistics* **41** 2703–2738.

[18] Nussbaum, M. and Szkoła, A. (2009). The Chernoff lower bound for symmetric quantum hypothesis testing. *The Annals of Statistics* **37** 1040–1057.

[19] Piantanida, P., Rey Vega, L. and Hero, A. (2014). A Proof of the Generalized Markov Lemma with Countable Infinite Sources. In *Information Theory Proceedings (ISIT), 2014 IEEE International Symposium on*.

[20] Schrijver, A. (1998). *Theory of linear and integer programming*. John Wiley & Sons.

[21] Shalaby, H. M. H. and Papamarcou, A. (1992). Multiterminal detection with zero-rate data compression. *Information Theory, IEEE Transactions on* **38** 254-267.

[22] Shimokawa, H., Han, T. and Amari, S. I. (1994). Error Bound of Hypothesis Testing with Data Compression. In *Inf. Theory, 1994 IEEE International Symposium on (ISIT)* 114.

[23] Vega, L. R., Piantanida, P. and Hero, A. O. (2015). The Three-Terminal Interactive Lossy Source Coding Problem. *Information Theory, IEEE Trans. on.* (revised).

[24] Wald, A. (1945). Sequential tests of statistical hypotheses. *The Annals of Mathematical Statistics* **16** 117–186.

[25] Wald, A. and Wolfowitz, J. (1948). Optimum character of the sequential probability ratio test. *The Annals of Mathematical Statistics* **19** 326–339.

[26] Xiang, Y. and Kim, Y.-H. (2012). Interactive hypothesis testing with communication constraints. In *Communication, Control, and Computing (Allerton), 2012 50th Annual Allerton Conference on* 1065-1072.

[27] Yakowitz, S. (1974). Multiple hypothesis testing by finite memory algorithms. *The Annals of Statistics* **2** 323–336.

[28] Zhao, W. and Lai, L. (2015). Distributed testing with zero-rate compression. In *Inf. Theory, 2015 IEEE International Symposium on (ISIT)* 2792-2796.

E-mail: gil.katz@CentraleSupelec.fr; pablo.piantanida@CentraleSupelec.fr; merouane.debbah@CentraleSupelec.fr