Testing Against Independence with Multiple Decision Centers

Michèle Wigger[†] and Roy Timo[‡]
[†]LTCI CNRS, Telecom ParisTech, Université Paris-Saclay, Paris, France, 75013.
[‡]Technical University of Munich, Munich, Germany, 80333.
michele.wigger@telecom-paristech.fr, roy.timo@tum.de

***Abstract*—We consider a three-terminal distributed hypothesis testing *against independence* problem. Two terminals, which act as decision centers, are required to decide whether their observed sequences are probabilistically independent of the sequence observed at the third terminal. The third terminal communicates with the two decision centers over three rate-limited noise-free pipes (error free channels): A *common* pipe that is connected to both centers, and two *private* pipes that connect separately to each center. We characterize the optimal exponential decay of the type-II error probabilities at the two decision centers given that the type-I error probabilities vanish for increasing blocklengths. The optimal exponents are determined by a certain information-theoretic optimization problem that depends on the maximum rates allowed over the noise-free communication pipes.**

I. Introduction

This work is motivated by joint problems of communications and distributed hypothesis testing. For example, imagine that many physically separated *decision centers* observe different datasets, and suppose that each center must determine whether or not its data is independent of the other datasets. Naturally the decision centers will need to exchange information to ensure reliable decisions. We would like to understand the fundamental tradeoffs between the hypothesis-testing error probabilities and the communications rates.

This paper will characterize the above tradeoffs for the simplified binary hypothesis testing problem with communications constraints shown in Figure 1. This problem is formulated in the same spirit as the seminal works of Ahlswede, Csiszár and Han [1], [2]: The main aim will be to determine the maximum exponents of the *type-II error probabilities* (the probability of incorrectly choosing the dependent-data hypothesis when the independent-data hypothesis is true) when the *type-I error probabilities* (the probability of incorrectly choosing the independent-data hypothesis when the dependent-data hypothesis is true) are smaller than some constant $\epsilon \in (0, 1)$.

Ahlswede, Csiszár and Han [1], [2] considered a different, but rather canonical, binary hypothesis-testing problem with three terminals. They assumed that two separated terminals observed different random sequences. These terminals communicated information about their observations to a single decision center, and the decision center was required to choose between two hypothesis on the joint distributions of these sequences. Ahlswede, Csisźar and Han established several important upper and lower bounds on maximum exponent of

978-1-5090-1746-1/16/$31.00 © 2016 IEEE

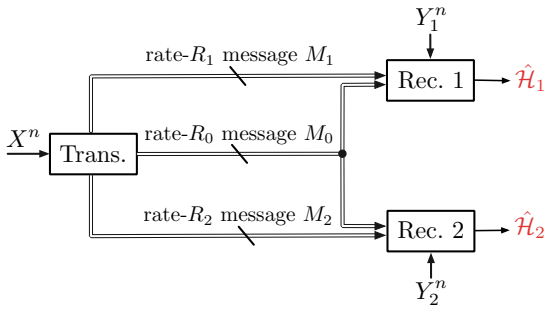

Fig. 1. Distributed hypothesis testing problem with one transmitter (observer) and two receivers (decision centers). The receivers must decide whether or not (X^n, Y_1^n, Y_2^n) is generated in an independent and identically distributed manner using the "dependent data" null hypothesis $(X, Y_1, Y_2) \sim P_{XY_1Y_2}$ or the "independent-data" alternative hypothesis $(X, Y_1, Y_2) \sim P_X P_{Y_1Y_2}$.

the type-II error probabilities as functions of the available communication rates, and they completely resolved the "testing against independence" problem. More work on this distributed hypothesis testing problem can be found in [3], [4], and a comprehensive literature review is given by [5]. Extensions to more involved communication scenarios have been considered in [6], [7], [8]. Zhao and Lai assumed that the observers (i.e., the terminals that are not decision centers) can exchange data. Xiang and Kim [7], [8] and Katz, Piantanida and Debbah [9] allowed interactive communication between a single observer and a single decision center.

In contrast to the above mentioned work, this paper considers a scenario with two decision centers and one observer (that is not a decision center), see Figure 1. We assume that the single observer can communicate to the two decision centers over three bit-pipes of given rates: A common pipe to both decision centers, and two separate private pipes to each center. The main result of this paper is to determine, for this communication scenario, the optimal exponents of the type-II error probabilities under constrained type-I error probabilities.

Outline: The problem setup is formally described in Section II. Our main result is summarized by Theorem 1 in Section III, and it is proved in Section IV.

Notation: Random variables are identified by uppercase letters, e.g. X, their alphabets by matching calligraphic font, e.g. $\mathcal{X}$, and elements of an alphabet by lowercase letters, e.g. $x \in \mathcal{X}$. The n-fold Cartesian product of an alphabet $\mathcal{X}$ is $\mathcal{X}^n$. Finally, $\log(\cdot)$ denotes the base-2 logarithm.

II. Problem Definition

Consider the distributed hypothesis testing problem with a transmitter and two receivers (the decision centers) illustrated in Figure 1. The transmitter observes the sequence

$$X^n := (X_1, X_2, \ldots, X_n)$$

and Receivers 1 and 2 respectively observe

$$Y_1^n := (Y_{1,1}, Y_{1,2}, \ldots, Y_{1,n}) \quad \text{and}$$
$$Y_2^n := (Y_{2,1}, Y_{2,2}, \ldots, Y_{2,n}).$$

The receivers are required to choose between two hypothesis $\mathcal{H} \in \{0, 1\}$. The "dependent data" *null hypothesis* is

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

and the "independent data" *alternative hypothesis* is

$$\mathcal{H} = 1\colon \quad (X^n, Y_1^n, Y_2^n) \sim \text{ i.i.d. } P_X P_{Y_1Y_2}, \tag{1b}$$

Here i.i.d. stands for *independent and identically distributed*; P_X and $P_{Y_1Y_2}$ respectively denote the X and (Y_1, Y_2) marginals of the joint law $P_{XY_1Y_2}$:

$$P_X(x) = \sum_{y_1 \in \mathcal{Y}_1, y_2 \in \mathcal{Y}_2} P_{XY_1Y_2}(x, y_1, y_2) \quad x \in \mathcal{X},$$
$$P_{Y_1Y_2}(y_1, y_2) = \sum_{x \in \mathcal{X}} P_{XY_1Y_2}(x, y_1, y_2), \quad (y_1, y_2) \in \mathcal{Y}_1 \times \mathcal{Y}_2.$$

The transmitter computes three messages

$$(M_0, M_1, M_2) = \phi^{(n)}(X^n), \tag{2}$$

using a (possibly stochastic) encoding function $\phi^{(n)}$ of the form

$$\phi^{(n)}\colon \mathcal{X}^n \to \{0, \ldots, \lfloor 2^{nR_0} \rfloor\} \times \{0, \ldots, \lfloor 2^{nR_1} \rfloor\} \times \{0, \ldots, \lfloor 2^{nR_2} \rfloor\}.$$

The transmitter sends M_0, M_1 and M_2 over the three noise-free bit-pipes as depicted in Figure 1. Receiver 1 obtains messages M_0 and M_1 and Receiver 2 obtains messages M_0 and M_2.

Each receiver $i \in \{1, 2\}$ outputs an estimate $\hat{\mathcal{H}}_i \in \{0, 1\}$ of the actual hypothesis based on its observed sequence Y_i^n and on the messages obtained over the pipes. That is,

$$\hat{\mathcal{H}}_i = g_i^{(n)}(Y_i^n, M_0, M_i), \qquad i \in \{1, 2\}, \tag{3}$$

using some (possibly stochastic) decoding function

$$g_i^{(n)}\colon \mathcal{Y}_i^n \times \{0, \ldots, \lfloor 2^{nR_0} \rfloor\} \times \{0, \ldots, \lfloor 2^{nR_i} \rfloor\} \to \{0, 1\}.$$

The main goal here is to maximize the exponent of the type-II error probabilities under a constant constraint on the type-I error probabilities.

Definition 1: For each $\epsilon \in (0, 1)$, we say that the exponent-rate tuple $(\theta_1, \theta_2, R_0, R_1, R_2)$ is ϵ-achievable if there exists a sequence of encoding and decoding functions $\{(\phi^{(n)}, g_1^{(n)}, g_2^{(n)})\}_n$ such that corresponding sequences of type-I error probabilities

$$\alpha_{1,n} := \mathbb{P}\big[\hat{\mathcal{H}}_1 = 1 \big| \mathcal{H} = 0\big] \tag{4a}$$

$$\alpha_{2,n} := \mathbb{P}\big[\hat{\mathcal{H}}_2 = 1 \big| \mathcal{H} = 0\big] \tag{4b}$$

and type-II error probabilities

$$\beta_{1,n} := \mathbb{P}\big[\hat{\mathcal{H}}_1 = 0 \big| \mathcal{H} = 1\big] \tag{5a}$$

$$\beta_{2,n} := \mathbb{P}\big[\hat{\mathcal{H}}_2 = 0 \big| \mathcal{H} = 1\big] \tag{5b}$$

satisfy (for $i = 1$ and 2)

$$\alpha_{i,n} \le \epsilon \tag{6}$$

for all n and

$$-\varlimsup_{n \to \infty} \frac{1}{n} \log \beta_{i,n} \ge \theta_i \tag{7}$$

Definition 2: For a given rate triple (R_0, R_1, R_2), we define the *exponent region* $\mathcal{E}^\star(R_0, R_1, R_2)$ as the closure of all non-negative exponent pairs (θ_1, θ_2) for which $(\theta_1, \theta_2, R_0, R_1, R_2)$ is ϵ-achievable for every $\epsilon \in (0, 1)$.

Characterizing the *exponents region* $\mathcal{E}^\star(R_0, R_1, R_2)$ is main problem of interest in this paper.

Remark 1: For our envisioned communication scenario, the two receivers cannot exploit differences in the joint conditional laws $P_{Y_1Y_2|X}$ that do not lead to different conditional laws $P_{Y_1|X}$ and $P_{Y_2|X}$ under the two hypothesis. In this sense, the solution to our hypothesis testing problem remains unchanged when

1) we replace the conditional law $P_{Y_1Y_2|X}$ in (1a) by any other joint conditional law with same marginal laws $P_{Y_1|X}$ and $P_{Y_2|X}$; or
2) we replace the joint law $P_{Y_1Y_2}$ in (1b) by any other joint law with same marginals P_{Y_1} and P_{Y_2}.

III. Main Result

Definition 3: For a given rate tuple (R_0, R_1, R_2), let $\mathcal{E}(R_0, R_1, R_2)$ denote the set of all nonnegative pairs (θ_1, θ_2) that satisfy

$$\theta_1 \le I(U_0, U_1; Y_1) \tag{8a}$$

$$\theta_2 \le I(U_0, U_2; Y_2), \tag{8b}$$

for some auxiliary random variables U_0, U_1, U_2 satisfying the Markov chain

$$(U_0, U_1, U_2) \to X \to (Y_1, Y_2), \tag{9}$$

and for which

$$R_0 \ge I(U_0; X)$$
$$R_1 \ge I(U_1; X|U_0)$$
$$R_2 \ge I(U_2; X|U_0). \tag{10}$$

Theorem 1: The exponents region $\mathcal{E}^\star(R_0, R_1, R_2)$ coincides with $\mathcal{E}(R_0, R_1, R_2)$:

$$\mathcal{E}(R_0, R_1, R_2) = \mathcal{E}^\star(R_0, R_1, R_2). \tag{11}$$

Proof: See Section IV. ∎

Remark 2: Theorem 1 readily leads to the following result for the related scenario where communication from the transmitter to the two receivers is over a discrete memoryless

broadcast channel (BC) of transition law $P_{Z_1Z_2|W}$. The exponents region over a discrete memoryless broadcast channel $P_{Z_1Z_2|W}$ includes all pairs (θ_1, θ_2) that satisfy (8) for some auxiliary random variables U_0, U_1, U_2 satisfying the Markov chain (9) and for which the rate triple

$$\big(R_0 = I(U_0;X),\ R_1 = I(U_1;X|U_0),\ R_2 = I(U_2;X|U_0)\big)$$

lies inside the capacity region of the discrete memoryless BC $P_{Z_1Z_2|W}$.

It is further interesting to notice that in the important special case with only a common pipe to both receivers, i.e., when $R_1 = R_2 = 0$, a single auxiliary random variable U suffices to characterize the optimal exponents region.

Definition 4: Given rate $R > 0$, let $\mathcal{E}_{\text{com-pipe}}(R)$ denote the set of all nonnegative pairs (θ_1, θ_2) that satisfy

$$\theta_1 \le I(U;Y_1) \tag{12}$$
$$\theta_2 \le I(U;Y_2), \tag{13}$$

for some auxiliary random variable U satisfying the Markov chain

$$U \to X \to (Y_1, Y_2), \tag{14}$$

and for which

$$R \ge I(U;X). \tag{15}$$

Corollary 1.1: When $R_1 = R_2 = 0$, the exponents region $\mathcal{E}^\star(R_0, R_1, R_2)$ coincides with $\mathcal{E}_{\text{com-pipe}}(R_0)$:

$$\mathcal{E}_{\text{com-pipe}}(R) = \mathcal{E}^\star(R, R_1 = 0, R_2 = 0). \tag{16}$$

It is not too difficult to see that Corollary 1.1 readily extends to arbitrary number of $K \ge 1$ receivers.

IV. Proof of Theorem 1

Feasibility of all pairs in $\mathcal{E}(R_0, R_1, R_2)$, i.e., inclusion

$$\mathcal{E}(R_0, R_1, R_2) \subseteq \mathcal{E}^\star(R_0, R_1, R_2) \tag{17}$$

is proved in Subsection IV-A. Infeasibility of all pairs outside $\mathcal{E}(R_0, R_1, R_2)$, i.e., inclusion

$$\mathcal{E}(R_0, R_1, R_2) \supseteq \mathcal{E}^\star(R_0, R_1, R_2) \tag{18}$$

is proved in Subsection IV-B.

A. Proof of Feasibility Statement (17)

Fix $\epsilon > 0$, an arbitrary large blocklength n, and a joint conditional distribution $P_{U_0U_1U_2|X}$. Consider any nonnegative rate tuple (R_0, R_1, R_2) satisfying

$$R_0 > I(U_0;X) \tag{19a}$$
$$R_1 > I(U_1;X|U_0) \tag{19b}$$
$$R_2 > I(U_2;X|U_0). \tag{19c}$$

Codebook Generation: Randomly generate a U_0-*codebook*

$$\Big\{U_0^n(m_0) = \big(U_{0,1}(m_0), \ldots, U_{0,n}(m_0)\big)\Big\}_{m_0=1}^{\lfloor 2^{nR_0}\rfloor}$$

by selecting each entry of each n-length codeword $U_0^n(m_0)$ in an i.i.d. manner using P_{U_0}. For each index m_0, randomly generate a codebook (U_1, m_0)-*codebook*

$$\Big\{U_1^n(m_0, m_1) = \big(U_{1,1}(m_0, m_1), \ldots, U_{1,n}(m_0, m_1)\big)\Big\}_{m_1=1}^{\lfloor 2^{nR_1}\rfloor}$$

by picking the j-th entry of codeword $U_1^n(m_0, m_1)$ in a memoryless manner using the conditional distribution $P_{U_1|U_0}(\cdot|U_{0,j}(m_0))$. In a similar manner, randomly generate a (U_2, m_0)-*codebook* for each index m_0 using the conditional distribution $P_{U_2|U_0}$ instead of $P_{U_1|U_0}$.

All codebooks are revealed to all terminals. Choose a small $\epsilon' > 0$.

Transmitter: The transmitter looks for a tuple of indices (m_0, m_1, m_2) such that

$$\big(X^n, U_0^n(m_0), U_1^n(m_0, m_1), U_2^n(m_0, m_2)\big) \in \mathcal{T}_{\epsilon'/2}^{(n)}(P_{XU_0U_1U_2}).$$

If successful, the transmitter picks one such tuple uniformly at random and sends the corresponding indices M_0, M_1, M_2 over pipes 0, 1 and 2 respectively. If no such triple exists, the transmitter sends $M_0 = 0, M_1 = 0$ and $M_2 = 0$ over the respective pipes.

Receiver 1: If $M_0 = 0$, then receiver 1 outputs hypothesis $\hat{\mathcal{H}}_1 = 1$. If $M_0 \ne 0$, then it checks whether checks

$$\big(Y_1^n, U_0^n(M_0), U_1^n(M_0, M_1)\big) \in \mathcal{T}_{\epsilon'}^{(n)}(P_{Y_1U_0U_1}). \tag{20}$$

If this check is successful, then it outputs hypothesis $\hat{\mathcal{H}}_1 = 0$; otherwise, it outputs $\hat{\mathcal{H}}_1 = 1$.

Receiver 2: If $M_0 = 0$, then receiver 2 outputs hypothesis $\hat{\mathcal{H}}_2 = 1$. If $M_0 \ne 0$, then it checks whether

$$\big(Y_2^n, U_0^n(m_0), U_2^n(m_0, m_2)\big) \in \mathcal{T}_{\epsilon'}^{(n)}(P_{Y_2U_0U_2}). \tag{21}$$

If this check is successful, then it outputs hypothesis $\hat{\mathcal{H}}_2 = 0$; otherwise, if outputs $\hat{\mathcal{H}}_2 = 1$.

Analysis: We first analyse the two type-I probabilities of error. We thus assume that $\mathcal{H} = 0$ and (X^n, Y_1^n, Y_2^n) i.i.d. $\sim P_{XY_1Y_2}$. Let us first bound the expectation of the type-I error probability averaged over the randomly generated codebooks.

By the *covering lemma* [10, Sec. 3.7] and the rate-constraints (19), the probability that the encoding operation at the transmitter outputs $M_0 = 0$ tends to 0 exponentially fast in n. Notice that if $M_0 \ne 0$, then the tuple

$$\big(X^n, U_0^n(M_0), U_1^n(M_1|M_0), U_2^n(M_2|M_0)\big) \in \mathcal{T}_{\epsilon'/2}^{(n)}(P_{XU_0U_1U_2}). \tag{22}$$

Conditioned on the event (22), the probability that

$$\big(Y_1^n, Y_2^n, X^n, U_0^n(M_0), U_1^n(M_1|M_0), U_2^n(M_2|M_0)\big) \in \mathcal{T}_{\epsilon'}^{(n)}(P_{Y_1Y_2XU_0U_1U_2}) \tag{23}$$

tends to 1 exponentially fast in n by the *Markov and the conditional typicality lemma* [10].

The above discussion implies that both receivers decide for the correct hypothesis $\mathcal{H}=0$ with probability tending to 1 exponentially fast in n. Or, put another way, we have

$$\bar{\alpha}_{i,n} := \mathbb{E}[\alpha_{i,n}] \leq 2^{-an}, \qquad i \in \{1,2\} \tag{24}$$

for some $a > 0$. (Here the expectation is taken over the randomly generated codebooks).

We now analyze the type-II probabilities of error. Assume that $\mathcal{H}=1$ and thus (X^n, Y_1^n, Y_2^n) i.i.d. $\sim P_X \cdot P_{Y_1Y_2}$. We start by bounding the probability of error for any given codebooks by

$$\begin{aligned}
\beta_{i,n} &= \mathbb{P}\big[\hat{\mathcal{H}}_i = 0, M_0 = 0 \big| \mathcal{H}=1\big] \\
&\quad + \mathbb{P}\big[\hat{\mathcal{H}}_i = 0, M_0 \neq 0 \big| \mathcal{H}=1\big] \\
&\overset{\text{a}}{=} \mathbb{P}\big[\hat{\mathcal{H}}_i = 0, M_0 \neq 0 \big| \mathcal{H}=1\big] \\
&\leq \mathbb{P}\big[\hat{\mathcal{H}}_i = 0 \big| M_0 \neq 0, \mathcal{H}=1\big] \\
&\overset{\text{b}}{=} \mathbb{P}\big[\hat{\mathcal{H}}_i = 0 \big| (M_0,M_1,M_2) = (1,1,1), \mathcal{H}=1\big] \\
&= \mathbb{P}\Big[\big(Y_i^n, U_0^n(1), U_i^n(1|1)\big) \in \mathcal{T}_{\epsilon'}^{(n)}(P_{Y_iU_0U_i}) \Big| \\
&\qquad\qquad (M_0,M_1,M_2) = (1,1,1), \mathcal{H}=1\Big]
\end{aligned}$$

where (a) holds because the receivers output $\hat{\mathcal{H}}_i = 1$ whenever $M_0 = 0$, and (b) holds by symmetry in the code construction. It can be shown that the probability that $(U_0^n(1), U_i^n(1|1))$ is jointly typical with the (independently generated) Y_i^n is bounded by

$$\bar{\beta}_{i,n} := \mathbb{E}[\beta_{i,n}] \leq 2^{-n[I(U_0,U_i;Y_i)-\delta(\epsilon')]}, \tag{25}$$

where $\delta(\epsilon')$ is a function that tends to 0 as $\epsilon' \to 0$. (Here the expectation is over the randomly generated codebooks.)

We now show that the expectations (over the randomly chosen codebooks) in (24) and (25) imply that for all sufficiently large blocklengths n there exists at least one codebook for which

$$\alpha_{i,n} \leq \epsilon, \tag{26a}$$

$$\beta_{i,n} \leq 2^{-n[I(U_0,U_i;Y_i)-\delta']}, \tag{26b}$$

for any $\delta > \delta(\epsilon')$. Since ϵ' can be chosen arbitrarily close to 0, this proves the theorem.

Let $\mathcal{C}_n$ denote the (finite) product space representing every possible configuration of the U_0, $\{(U_1, m_0)\}_{m_0}$ and $\{(U_2, m_0)\}_{m_0}$ codebooks with blocklength n (as described above), and let the random variable $C_n \in \mathcal{C}_n$ represent the random codebook construction method. For any given codebook configuration $c \in \mathcal{C}_n$, let $\alpha_{i,n}(c)$ and $\beta_{i,n}(c)$ denote the type-I and type-II error probabilities at receiver i respectively.

Fix $\delta' > \delta(\epsilon')$. Now define a set of *bad codes*

$$\mathcal{B}_n := \mathcal{B}_{\alpha,1,n} \cup \mathcal{B}_{\alpha,2,n} \cup \mathcal{B}_{\beta,1,n} \cup \mathcal{B}_{\beta,2,n}, \tag{27}$$

where

$$\mathcal{B}_{\alpha,i,n} := \Big\{ c \in \mathcal{C}_n : \alpha_{i,n}(c) > \epsilon \Big\} \tag{28}$$

and

$$\mathcal{B}_{\beta,i,n} := \Big\{ c \in \mathcal{C}_n : \beta_{i,n}(c) > 2^{-n[I(U;Y_i)-\delta']} \Big\} \tag{29}$$

The probability that we choose a bad code is bounded by

$$\begin{aligned}
\mathbb{P}\big[C_n \in \mathcal{B}_n\big] &\overset{\text{a}}{\leq} \sum_{i=1}^{2} \Big(\mathbb{P}\big[C_n \in \mathcal{B}_{\alpha,i,n}\big] + \mathbb{P}\big[C_n \in \mathcal{B}_{\beta,i,n}\big] \Big) \\
&\overset{\text{b}}{\leq} \sum_{i=1}^{2} \left(\frac{1}{\epsilon} \bar{\alpha}_{i,n} + \bar{\beta}_{i,n} 2^{n[I(U;Y_i)-\delta']} \right) \\
&\overset{\text{c}}{\leq} 2 \left(\frac{1}{\epsilon} 2^{-na} + 2^{-n[\delta' - \delta(\epsilon')]} \right),
\end{aligned} \tag{30}$$

where (a) follows by the union bound, (b) follows by Markov's inequality, and (c) substitutes (24) and (25). Since $\delta' > \delta(\epsilon')$, we have $\mathbb{P}\big[C_n \notin \mathcal{B}_n\big] \to_n 1$. Therefore, for some sufficiently large n there exists at least one codebook configuration $c \in \mathcal{C}_n$ that satisfies (26).

B. Proof of Infeasibility Statement (18)

Fix a sequence of encoding and decoding functions $\{\phi^{(n)}, g_1^{(n)}, g_2^{(n)}\}_{n=1}^{\infty}$ so that (6) and (7) hold.

Now consider a fixed blocklength n. By the data-processing inequality and the definitions of $\alpha_{i,n}$ and $\beta_{i,n}$,

$$\begin{aligned}
&D\big(P_{M_0M_iY_i^n|\mathcal{H}=0} \big\| P_{M_0M_iY_i^n|\mathcal{H}=1}\big) \\
&\quad \geq D\big(P_{\hat{\mathcal{H}}_i|\mathcal{H}=0} \big\| P_{\hat{\mathcal{H}}_i|\mathcal{H}=1}\big) \\
&\quad = \alpha_{i,n} \log \frac{\alpha_{i,n}}{1-\beta_{i,n}} + (1-\alpha_{i,n}) \log \frac{(1-\alpha_{i,n})}{\beta_{i,n}} \\
&\quad = H_{\text{b}}(\alpha_{i,n}) - \alpha_{i,n} \log(1-\beta_{i,n}) - (1-\alpha_{i,n}) \log \beta_{i,n} \\
&\quad \geq -(1-\alpha_{i,n}) \log \beta_{i,n} \\
&\quad \geq -(1-\epsilon) \log \beta_{i,n},
\end{aligned} \tag{31}$$

where $H_{\text{b}}(\cdot)$ denotes the binary entropy function and in the last inequality we have used that $\alpha_{i,n} \leq \epsilon$.

We have

$$\begin{aligned}
&-\frac{1}{n} \log \beta_{1,n} \\
&\quad \leq \frac{1}{1-\epsilon} \, \frac{1}{n} \, D\big(P_{M_0M_1Y_1^n|\mathcal{H}=0} \big\| P_{M_0M_1Y_1^n|\mathcal{H}=1}\big) \\
&\quad = \frac{1}{1-\epsilon} \, \frac{1}{n} \, I(M_0, M_1; Y_1^n) \\
&\quad = \frac{1}{1-\epsilon} \, \frac{1}{n} \, \sum_{t=1}^{n} I(M_0, M_1; Y_{1,t} | Y_1^{t-1}) \\
&\quad = \frac{1}{1-\epsilon} \, \frac{1}{n} \, \sum_{t=1}^{n} I(M_0, M_1, Y_1^{t-1}; Y_{1,t}), \\
&\quad \leq \frac{1}{1-\epsilon} \, \frac{1}{n} \, \sum_{t=1}^{n} I(M_0, M_1, X^{t-1}; Y_{1,t}) \\
&\quad = \frac{1}{1-\epsilon} I(\tilde{U}_{0,T_n}, U_{1,T_n}; Y_{1,T_n} | T_n) \\
&\quad = \frac{1}{1-\epsilon} I(U_{0,n}, U_{1,n}; Y_{1,n}),
\end{aligned} \tag{32}$$

where T_n denotes a uniform random variable over $\{1, \ldots, n\}$ independent of the tuple $(M_0, M_1, M_2, X^n, Y_1^n, Y_2^n)$, and where we defined $\tilde{U}_{0,t} := (M_0, X^{t-1})$; $U_{0,n} := (\tilde{U}_{0,T_n}, T_n)$; $U_{1,n} := M_1$; $Y_{1,n} := Y_{1,T_n}$. The second equality holds because under hypothesis $\mathcal{H}=1$ the messages M_0 and M_1,

which are functions of X^n, are independent of Y_1^n. The third equality and the last equality hold because Y_1^n is i.i.d. and independent of T_n. The second inequality holds because for each t we have the Markov chain $Y_{1,t} \to (M_0, M_1, X^{t-1}) \to Y_1^{t-1}$.

Following similar steps we also obtain

$$-\frac{1}{n}\log \beta_{2,n} \leq \frac{1}{1-\epsilon} I(U_{0,n}, U_{2,n}; Y_{2,n}), \tag{33}$$

where we defined $U_{2,n} := M_2$ and $Y_{2,n} := Y_{2,T_n}$.

Further, we have for the common rate

$$\begin{aligned} R_0 &\geq \frac{1}{n} H(M_0) \\ &\geq \frac{1}{n} I(M_0; X^n) \\ &\geq \frac{1}{n} \sum_{t=1}^{n} I(M_0; X_t | X^{t-1}) \\ &= \frac{1}{n} \sum_{t=1}^{n} I(M_0, X^{t-1}; X_t) \\ &= \frac{1}{n} \sum_{t=1}^{n} I(\tilde{U}_{0,t}; X_t) \\ &= I(\tilde{U}_{0,T_n}; X_{T_n} | T_n) \\ &= I(U_{0,n}; X_n), \end{aligned} \tag{34}$$

where we defined $X_n := X_{T_n}$. And for the private rates

$$\begin{aligned} R_1 &\geq \frac{1}{n} H(M_1) \\ &\geq \frac{1}{n} H(M_1 | M_0) \\ &\geq \frac{1}{n} I(M_1; X^n | M_0) \\ &\geq \frac{1}{n} \sum_{t=1}^{n} I(M_1; X_t | X^{t-1}, M_0) \\ &= \frac{1}{n} \sum_{t=1}^{n} I(M_1, ; X_t | X^{t-1}, M_0) \\ &= I(U_{1,T_n}; X_{T_n} | U_{0,T_n}, T_n) \\ &= I(U_{1,n}; X_n | U_{0,n}). \end{aligned} \tag{35}$$

Similarly,

$$R_2 \leq I(U_{2,n}; X_n | U_{0,n}), \tag{36}$$

Notice that the Markov chain

$$(U_{0,n}, U_{1,n}, U_{2,n}) \to X_n \to (Y_{1,n}, Y_{2,n}) \tag{37}$$

holds and that

$$(X_n, Y_{1,n}, Y_{2,n}) \sim P_{XY_1Y_2}. \tag{38}$$

From (32)–(38) and by continuity of mutual information we conclude that if the tuple $(\theta_1, \theta_2, R_0, R_1, R_2)$ is ϵ-achievable, then there exist auxiliary random variable U_0, U_1, U_2 forming the Markov chain

$$(U_0, U_1, U_2) \to X \to (Y_1, Y_2)$$

and satisfying

$$\begin{aligned} \theta_1 &\leq \frac{1}{1-\epsilon}\ I(U_0, U_1; Y_1) \\ \theta_2 &\leq \frac{1}{1-\epsilon}\ I(U_0, U_2; Y_2) \\ R_0 &\geq I(U_0; X) \\ R_1 &\geq I(U_1; X | U_0) \\ R_2 &\geq I(U_2; X | U_0). \end{aligned}$$

Taking $\epsilon \to 0$, concludes the proof.

ACKNOWLEDGEMENTS

R. Timo was supported by a fellowship from the Alexander von Humboldt Foundation.

REFERENCES

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

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

[3] W. Zhao and L. Lai, "Distributed testing against independence with multiple terminals," in *proceedings 52nd Annual Allerton Conference Communication, Control, and Computing*, Monticello (IL), USA, pp. 1246–1251, Oct. 2014.

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

[5] T. S. Han and S. I. Amari, "Statistical inference under multiterminal data compression," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2300–2324, June, 1998.

[6] W. Zhao and L. Lai, "Distributed testing against independence with conferencing encoders," in *proceedings IEEE Information Theory Workshop (ITW)*, Jeju Island, Korea, October, 2015.

[7] Y. Xiang and Young-Han Kim, "Interactive hypothesis testing against independence," in *proceedings IEEE International Symposium on Information Theory (ISIT)*, pp. 2840–2844, Istanbul, Turkey, June, 2013

[8] Y. Xiang and Y.-H. Kim, "Interactive hypothesis testing with communication constraints," in *proceedings 50th Annual Allerton Conference on Communication, Control, and Computing*, Monticello (IL), USA, September, 2012.

[9] G. Katz, P. Piantanida and M. Debbah, "Collaborative distributed hypothesis testing," *arXiv*, 1604.01292, April, 2016.

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