ETH zürich Research Collection



Conference Paper

Distributed Hypothesis Testing Over a Noisy Channel

Author(s):
Salehkalaibar, Sadaf; Wigger, Michèle

Publication Date:
2018-02

Permanent Link:
https://doi.org/10.3929/ethz-b-000245046 →

Rights / License:
In Copyright - Non-Commercial Use Permitted →

This page was generated automatically upon download from the ETH Zurich Research Collection. For more information please consult the Terms of use.

ETH Library

Distributed Hypothesis Testing Over a Noisy Channel

[1]Sadaf Salehkalaibar and [2]Michèle Wigger
[1]ECE Department, College of Engineering, University of Tehran, Tehran, Iran, s.saleh@ut.ac.ir
[2]LTCI, Telecom ParisTech, Université Paris-Saclay, 75013 Paris, France, michele.wigger@telecom-paristech.fr

***Abstract*—A coding and testing scheme is presented for the distributed hypothesis testing problem over a noisy channel. The coding scheme combines the Shimokawa-Han-Amari hypothesis testing scheme with Borade's unequal error protection (UEP) channel coding. The type-II error exponent of our scheme consists of three competing error exponents: two of them coincide with the exponents found by Shimokawa-Han-Amari for distributed hypothesis testing over a noiseless link (with the rate be replaced by the mutual information between channel input and output), and the third includes Borade's miss-detection exponent for UEP over a noisy channel. Depending on the problem setup, any of the three exponents can be active. When testing against conditional independence, only the two Shimokawa-Han-Amari exponents are active, and the scheme achieves the optimal type-II error exponent found by Sreekuma and Gündüz.**

I. Introduction

Consider a distributed hypothesis testing problem where a sensor describes its collected information to a remote decision center over a noisy channel. The decision center decides on a binary hypothesis ($\mathcal{H}=0$ or $\mathcal{H}=1$) that determines the joint probability distribution underlying its own observation and the information observed at the sensor. The goal of the communication is to maximize the type-II error (deciding $\hat{\mathcal{H}}=0$ when $\mathcal{H}=1$) exponent under a constrained type-I error (deciding $\hat{\mathcal{H}}=1$ when $\mathcal{H}=0$).

The special case of this problem where communication takes place over a noiseless link was studied in [1]–[4]. These works present achievable type-II error exponents for general joint probability distributions underlying the two hypotheses and the optimal type-II error exponent for the special case called "testing against conditional independence" [4]. Distributed hypothesis testing problems over noiseless networks with multiple sensors or decision centers or with relays have been considered in [4]–[8]. The work most closely related to this paper is by Sreekumar and Gündüz [9]. It proves that the optimal type-II error exponent for "testing against conditional independence" over a noisy channel, coincides with the optimal type-II error exponent of the same test over a noiseless link of rate equal to the capacity of the noisy channel. Their result is based on a joint hypothesis-testing and channel-coding scheme, see also [9, Remark 6] for a discussion on this.

In this work, we propose a coding scheme for distributed hypothesis testing over a noisy channel with general probability distributions. The coding and testing scheme applies separate hypothesis testing and channel coding by combining the Shimokawa-Han-Amari (SHA) hypothesis-testing scheme [3] with Borade's unequal error protection (UEP) channel coding [12]. The idea is to reinforce the protection of the message that the SHA scheme produces to indicate that the transmitter decides on the alternative hypothesis $\mathcal{H}=1$. Our analysis in general shows three competing error exponents, two of them coincide with the two competing error exponents obtained for testing over a noiseless link [3] when the communication rate is replaced by the mutual information between input and output of the channel. The third error exponent depends again on this mutual information, and on Borade's *miss-detection exponent* [12] for channel coding with UEP. In the special case of "testing against conditional independence", recover the optimal exponent by Sreekuma and Gündüz [9]. In this case, our third error exponent is never active and the overall type-II error exponent depends on the noisy channel only through its capacity.

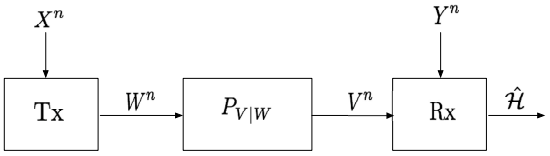


Fig. 1. Hypothesis testing over a noisy channel

Notation: We mostly follow the notation in [10]. Moreover, we use $\text{tp}(\cdot)$ to denote the *joint type* of a tuple. For a joint type π_{AB} over alphabets $\mathcal{A}\times\mathcal{B}$, we denote by $I_{\pi_{AB}}(A;B)$ the mutual information of a pair of random variables (A,B) with probability mass function (pmf) π_{AB}. Similarly for entropy, conditional entropy, and conditional mutual information. When it is unambiguous, we may abbreviate π_{AB} by π. We also abbreviate *independent and identically distributed* by i.i.d.

II. System Model

Consider the distributed hypothesis testing problem in Fig. 1, where a transmitter observes source sequence X^n and a receiver source sequence Y^n. Under the null hypothesis:

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

and under the alternative hypothesis:

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

for two given pmfs P_{XY} and Q_{XY}. The transmitter can communicate with the receiver over n uses of a discrete

memory channel $(\mathcal{W}, \mathcal{V}, P_{V|W})$ where $\mathcal{W}$ denotes the finite channel input alphabet and $\mathcal{V}$ the finite channel output alphabet. Specifically, the transmitter feeds inputs

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

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

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

Based on the sequence of channel outputs V^n and the source sequence Y^n, the receiver decides on the hypothesis $\mathcal{H}$. That means, it produces the guess

$$\hat{\mathcal{H}} = g^{(n)}(V^n, Y^n), \tag{5}$$

by means of a decoding function

$$g^{(n)} \colon \mathcal{V}^n \times \mathcal{Y}^n \to \{0, 1\}. \tag{6}$$

Definition 1: For each $\epsilon \in (0, 1)$, an exponent θ is said ϵ-achievable, if for each sufficiently large blocklength n, there exist encoding and decoding functions $(f^{(n)}, g^{(n)})$ such that the corresponding type-I and type-II error probabilities at the receiver

$$\alpha_n \triangleq \Pr[\hat{\mathcal{H}} = 1 | \mathcal{H} = 0], \tag{7}$$

$$\beta_n \triangleq \Pr[\hat{\mathcal{H}} = 0 | \mathcal{H} = 1], \tag{8}$$

satisfy

$$\alpha_n \leq \epsilon, \tag{9}$$

and

$$-\varlimsup_{n\to\infty} \frac{1}{n} \log \beta_n \geq \theta. \tag{10}$$

The goal is to maximize the type-II error exponent θ.

III. Coding and Testing Scheme

We describe a coding and testing scheme for the general distributed hypothesis testing problem over a noisy channel. The analysis of the scheme is postponed to Section V.

Preparations: Choose a large positive integer n, an auxiliary distribution P_T over $\mathcal{W}$, a conditional channel input distribution $P_{W|T}$, and a conditional source distribution $P_{S|X}$ over a finite auxiliary alphabet $\mathcal{S}$ so that

$$I(S; X) < I(S; Y) + I(V; W|T), \tag{11}$$

where the mutual informations in (11) are calculated according to the following joint distribution

$$P_{SXYWVT} = P_{S|X} \cdot P_{XY} \cdot P_T \cdot P_{W|T} \cdot P_{V|W}. \tag{12}$$

Then choose a sufficiently small $\mu > 0$ and nonnegative rates (R, R') so that

$$R + R' = I(X; S) + \mu \tag{13}$$

$$R < I(V; W|T) \tag{14}$$

$$R' < I(S; Y). \tag{15}$$

Code Construction: Construct a random codebook

$$\mathcal{C}_S = \big\{S^n(m, \ell) \colon m \in \{1, ..., \lfloor 2^{nR} \rfloor\}, \ell \in \{1, ..., \lfloor 2^{nR'} \rfloor\}\big\},$$

by independently drawing all codewords i.i.d. according to $P_S(s) = \sum_{x\in\mathcal{X}} P_X(x) P_{S|X}(s|x)$.

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

$$\mathcal{C}_W = \big\{W^n(m) : m \in \{1, ..., \lfloor 2^{nR} \rfloor\}\big\}$$

superpositioned on T^n where each codeword is drawn independently according to $P_{W|T}$ conditioned on T^n. Reveal the realizations of the codebooks and the sequence T^n to all terminals.

Transmitter: Given that it observes the source sequence $X^n = x^n$, the transmitter looks for a pair (m, ℓ) that satisfies

$$(s^n(m, \ell), x^n) \in \mathcal{T}^n_{\mu/2}(P_{SX}). \tag{16}$$

If successful, it picks one of these pairs uniformly at random and sends the codeword $w^n(m)$ over the channel. Otherwise it sends the sequence of inputs t^n over the channel.

Receiver: Assume that $V^n = v^n$ and $Y^n = y^n$ and that the "time-sharing sequence" $T^n = t^n$. The receiver first looks for an index $m' \in \{1, \ldots, \lfloor 2^{nR} \rfloor\}$ so that

$$(w^n(m'), v^n, t^n) \in \mathcal{T}^n_\mu(P_{WVT}). \tag{17}$$

If it is not successful, it declares $\hat{\mathcal{H}} = 1$. Otherwise, it randomly picks one of the indices ℓ' that satisfy

$$H_{\mathrm{tp}(s^n(m',\ell'),y^n)}(S|Y) = \min_{\tilde{\ell}\in\{1,\ldots,\lfloor 2^{nR'} \rfloor\}} H_{\mathrm{tp}(s^n(m',\tilde{\ell}),y^n)}(S|Y), \tag{18}$$

and checks whether

$$(s^n(m', \ell'), y^n) \in \mathcal{T}^n_\mu(P_{SY}). \tag{19}$$

If successful, it declares $\hat{\mathcal{H}} = 0$. Otherwise, it declares $\hat{\mathcal{H}} = 1$.

IV. An Achievable Error Exponent

The coding and testing scheme described in the previous section allows to establish the following theorem.

Theorem 1: Every error exponent $\theta \geq 0$ that satisfies the following condition (33) is achievable:

$$\theta \leq \max_{\substack{P_{S|X}, P_{TW}\colon \\ I(S;X|Y) \leq I(W;V|T)}} \min\big\{\theta_1, \theta_2, \theta_3\big\}, \tag{20}$$

where

$$\theta_1 = \min_{\substack{\tilde{P}_{SXY}\colon \\ \tilde{P}_{SX}=P_{SX} \\ \tilde{P}_{SY}=P_{SY}}} D(\tilde{P}_{SXY} || Q_{XY} P_{S|X}), \tag{21}$$

$$\theta_2 = \min_{\substack{\tilde{P}_{SXY}\colon \\ \tilde{P}_{SX}=P_{SX} \\ \tilde{P}_{Y}=P_{Y} \\ H(S|Y) \leq H_{\tilde{P}}(S|Y)}} \Big[D(\tilde{P}_{SXY} || P_{S|X} Q_{XY}) + I(V; W|T) - I(S; X|Y)\Big], \tag{22}$$

$$\theta_3 = D(P_Y || Q_Y) + I(V; W|T) - I(S; X|Y)$$

$$+ \sum_{t \in \mathcal{W}} P_T(t) \cdot D(P_{V|T=t} || P_{V|W=t}), \tag{23}$$

and all expressions are calculated with respect to the joint distribution in (12).

Proof: Based on the scheme in Section V. ∎

Lemma 1: It suffices to consider the auxiliary random variable S over an alphabet $\mathcal{S}$ that is of size $|\mathcal{S}| = |\mathcal{X}| + 2$. For the specical case of $P_Y = Q_Y$, it suffices to consider $|\mathcal{S}| = |\mathcal{X}| + 1$.

Proof: Based on Carathéodory's theorem. Omitted. ∎

Our coding and testing scheme combines the SHA hypothesis testing scheme for a noiseless link [3] with Borade's UEP channel coding that protects the 0-message (which indicates that the transmitter decides on $\mathcal{H} = 1$) better than the other messages [11], [12]. In fact, since here we are only interested in the type-II error exponent, the receiver should decide on $\mathcal{H} = 0$ only if the transmitter also shares this opinion.

The expressions in Theorem 1 show three competing error exponents. In (21) and (22), we recognize the two competing error exponents of the SHA scheme for the noiseless setup: θ_1 is the exponent associated to the event that the receiver reconstructs the correct binned codeword and θ_2 is associated to the event that either the binning or the noisy channel introduces a decoding error. The exponent θ_3 in (23) is new and can be associated to the event that the specially protected 0-message is wrongly decoded. We remark in particular that θ_3 contains the term

$$E_{\text{miss}} := \sum_{t \in \mathcal{W}} P_T(t) \cdot D(P_{V|T=t} || P_{V|W=t}), \tag{24}$$

which represents the largest possible *miss-detection exponent* for a single specially protected message at a given rate $I(W;V|T)$ [12, Th. 34].

Which of the three exponents $\theta_1, \theta_2, \theta_3$ is smallest depends on the source and channel parameters and the choice of $P_{S|X}$ and P_W. Notice that the third error exponent θ_3 is inactive for channels with large miss-detection exponent (24), such as binary symmetric channels with small cross-over probability, or for sources where

$$\min_{\substack{\tilde{P}_{SXY}: \\ \tilde{P}_{SX}=P_{SX} \\ \tilde{P}_Y=P_Y}} D(\tilde{P}_{SXY} || P_{S|X} Q_{XY}) = D(P_Y || Q_Y), \tag{25}$$

This is the case for example when "testing against conditional independence" [4] where both terms are 0.

Corollary 1 (Lemma 5 in [9]): Consider the "testing against independence" setup where

$$Y = (\bar{Y}, Z), \tag{26}$$

and $Q_{X\bar{Y}Z}$ decomposes as

$$Q_{X\bar{Y}Z} = P_{XZ} \cdot P_{\bar{Y}|Z}. \tag{27}$$

Error exponent $\theta \geq 0$ is achievable if,

$$\theta \leq \max_{\substack{P_{S|X},\ P_W: \\ I(S;X|Z) \leq I(W;V)}} I(S; \bar{Y}|Z), \tag{28}$$

where mutual informations are calculated with respect to the joint law $P_{X\bar{Y}Z} P_{S|X} P_W P_{V|W}$.

Proof: Fix independent random variables T and W and a random variable S so that

$$I(S;X|Z) \leq I(W;V|T) = I(W;V). \tag{29}$$

Then, Theorem 1 specializes to:

$$\begin{aligned}
\theta_1 &= \min_{\substack{\tilde{P}_{SX\bar{Y}Z}: \\ \tilde{P}_{SX}=P_{SX} \\ \tilde{P}_{S\bar{Y}Z}=P_{S\bar{Y}Z}}} D(\tilde{P}_{SX\bar{Y}Z} || Q_{X\bar{Y}Z} P_{S|X}) \\
&= \min_{\substack{\tilde{P}_{SX\bar{Y}Z}: \\ \tilde{P}_{SX}=P_{SX} \\ \tilde{P}_{S\bar{Y}Z}=P_{S\bar{Y}Z}}} D(\tilde{P}_{SX\bar{Y}Z} || P_{XZ} P_{\bar{Y}|Z} P_{S|X}) \\
&= D(P_{S\bar{Y}Z} || P_Z P_{\bar{Y}|Z} P_{S|Z}) \\
&= I(S; \bar{Y}|Z).
\end{aligned}$$

Moreover, exponents θ_2 and θ_3 cannot be smaller than $I(S;\bar{Y}|Z)$ because of the nonnegativity of the KL-divergence and the mutual information and because

$$\begin{aligned}
&I(V;W) - I(S;X) + I(S;\bar{Y},Z) \\
&= I(V;W) - I(S;X|Z) + I(S;\bar{Y}|Z) \\
&\geq I(S;\bar{Y}|Z),
\end{aligned} \tag{30}$$

where the inequality holds by (29). ∎

Notice that the error exponent in Corollary 1 is optimal [9].

We now present an example and evaluate the largest type-II error exponents attained by our scheme. We also show that depending on the choice of the model parameters, a different error exponent θ_1, θ_2, or θ_3 is active.

Example 1: Let under the null hypothesis

$$\begin{aligned}
\mathcal{H} = 0\colon \quad & X \sim \text{Bern}(p_0), \quad Y = X \oplus N_0, \\
& N_0 \sim \text{Bern}(q_0),
\end{aligned} \tag{31}$$

for N_0 independent of X. Under the alternative hypothesis:

$$\mathcal{H} = 1\colon \quad X \sim \text{Bern}(p_1), \quad Y \sim \text{Bern}(p_0 \star q_0), \tag{32}$$

with X and Y independent. Assume that $P_{V|W}$ is a binary symmetric channel (BSC) with cross-over probability $r \in [0, 1/2]$.

For this example, $P_Y = Q_Y$ and Theorem 1 simplifies to:

$$\theta \leq \max_{\substack{P_{S|X}, P_{TW}: \\ I(S;X|Y) \leq I(W;V|T)}} \min\left\{\theta_1, \theta_2, \theta_3\right\}, \tag{33}$$

where

$$\theta_1 \leq D(P_X || Q_X) + I(S;Y), \tag{34}$$

$$\theta_2 \leq D(P_X || Q_X) + I(V;W|T) + I(S;Y) - I(S;X), \tag{35}$$

$$\theta_3 \leq \sum_{t\in\mathcal{W}} P_T(t) D(P_{V|T=t}||P_{V|W=t}) + I(V;W|T) + I(S;Y) - I(S;X). \quad (36)$$

Depending on the parameters of the setup and the choice of the auxiliary distributions, either of the exponents θ_1, θ_2, or θ_3 is active. For example, when the cross-over probability of the BSC is large, $r \geq 0.4325$,

$$D(P_X||Q_X) \geq \sum_{t\in\mathcal{W}} P_T(t) D(P_{V|T=t}||P_{V|W=t}) + I(V;W|T), \quad (37)$$

and irrespective of the choice of the random variables S, T, W the exponent θ_3 is smaller than θ_1 and θ_2. It is then optimal to choose S constant and (T, W) so as to maximize the sum $\sum_{t\in\mathcal{W}} P_T(t) D(P_{V|T=t}||P_{V|W=t}) + I(V;W|T)$. In particular, for a scenario with parameters $p_0 = 0.1, q_0 = 0.25, p_1 = 0.2$ and $r = \frac{4}{9}$ one obtains numerically that the optimal error exponent achieved by our scheme is $\theta = 0.0358$.

In contrast, when the cross-over probability of the BSC is small, the miss-detection exponent (24) is large and the exponent θ_3 is never active irrespective of the choice of the auxiliary random variable S. The overall exponent is then determined by the smaller of θ_1 and θ_2, and in particular by a choice S, X, W that makes the two equal. In this case, for a scenario with parameters $p_0 = 0.2, q_0 = 0.3, p_1 = 0.4$, and $r = 0.1$, the largest exponent achieved by our scheme is $\theta = 0.19$.

V. Proof of Theorem 1

The proof of the theorem is based on the scheme in Section III. Before analyzing this scheme, notice that by the functional representation lemma, there exists a function γ over appropriate domains and for each time $t \in \{1, \ldots, n\}$ a random variable ϕ_t over a finite alphabet Φ so that the time-t channel input and output satisfy:

$$V_t = \xi(W_t, \phi_t). \quad (38)$$

Let $\mathcal{P}^n$ be the set of all types over the product alphabets $\mathcal{S}^n \times \mathcal{S}^n \times \mathcal{W}^n \times \mathcal{W}^n \times \mathcal{W}^n \times \mathcal{V}^n \times \Phi^n \times \mathcal{X}^n \times \mathcal{Y}^n$, and let $\mathcal{P}^n_\mu$ be the subset of types $\pi_{SS'TWW'V\phi XY} \in \mathcal{P}^n$ that simultaneously satisfy the following conditions:

$$|\pi_{SX} - P_{SX}| \leq \mu/2, \quad (39a)$$
$$|\pi_{S'Y} - P_{SY}| \leq \mu, \quad (39b)$$
$$|\pi_{TW'V} - P_{TWV}| \leq \mu, \quad (39c)$$
$$\pi_{V|\phi TW} = \mathbb{1}\{V = \xi(\phi, T, W)\}, \quad (39d)$$
$$H_{\pi_{S'Y}}(S|Y) \leq H_{\pi_{SY}}(S|Y). \quad (39e)$$

We first analyze the type-I error probability averaged over the random code construction. Let (M, L) be the indices of the codeword chosen at the transmitter, if they exist, and define the following events:

$$\mathcal{E}_{\text{Tx}} \colon \{\nexists (m,\ell) \colon (S^n(m,\ell), X^n) \in \mathcal{T}^n_{\mu/2}(P_{SX})\} \quad (40)$$
$$\mathcal{E}^{(1)}_{\text{Rx}} \colon \{(S^n(M,L), Y^n) \notin \mathcal{T}^n_\mu(P_{SY})\} \quad (41)$$
$$\mathcal{E}^{(2)}_{\text{Rx}} \colon \{\exists m' \neq M \colon (T^n, W^n(m'), V^n) \in \mathcal{T}^n_\mu(P_{TW}P_{V|W})\} \quad (42)$$
$$\mathcal{E}^{(3)}_{\text{Rx}} \colon \{\exists \ell' \neq L \colon H_{\text{tp}(s^n(M,\ell'),y^n)}(S|Y) = \min_{\tilde{\ell}} H_{\text{tp}(s^n(M,\tilde{\ell}),y^n)}(S|Y)\}. \quad (43)$$

With these definitions, we obtain for all sufficiently small values of μ and sufficiently large blocklengths n:

$$\alpha_n \leq \Pr[\mathcal{E}_{\text{Tx}}] + \Pr[\mathcal{E}^{(1)}_{\text{Rx}}|\mathcal{E}^c_{\text{Tx}}] + \Pr[\mathcal{E}^{(2)}_{\text{Rx}}|\mathcal{E}^c_{\text{Tx}}, \mathcal{E}^{(1)c}_{\text{Rx}}] + \Pr[\mathcal{E}^{(3)}_{\text{Rx}}|\mathcal{E}^{(1)c}_{\text{Rx}}, \mathcal{E}^c_{\text{Tx}}] \quad (44)$$
$$\leq \epsilon/4 + \epsilon/4 + \epsilon/4 + \epsilon/4 = \epsilon, \quad (45)$$

where the first summand of (44) can be upper bounded by means of the covering lemma [10] and the rate constraint (15); the second by means of the Markov lemma [10]; the third by means of the packing lemma [10] and the rate constraint (14); and the fourth by following similar steps as in analysis of the type-I error probability in [5, Appendix H].

Now, consider the type-II error probability. Let $\mathcal{P}^n_{\mu,0}$ be the subset of types $\pi_{S'TW'\phi VXY}$ over the alphabets $\mathcal{S}^n \times \mathcal{W}^n \times \mathcal{W}^n \times \Phi^n \times \mathcal{V}^n \times \mathcal{X}^n \times \mathcal{Y}^n$ that satisfy (39b), (39c), and

$$\pi_{V|\phi T} = \mathbb{1}\{V = \xi(T, \phi)\}. \quad (46)$$

Define for each pair $(m, m') \in \{1, \ldots, \lfloor 2^{nR} \rfloor\}^2$ and $(\ell, \ell') \in \{1, \ldots, \lfloor 2^{nR'} \rfloor\}^2$ the set:

$$\mathcal{A}(m, m', \ell, \ell') := \Big\{(\varphi^n, x^n, y^n) : \text{tp}\big(S^n(m,\ell), S^n(m',\ell'), W^n(m), W^n(m'), \varphi^n, \xi^n(W^n(m), \varphi^n), x^n, y^n\big) \in \mathcal{P}^n_\mu\Big\};$$

and for each $m' \in \{1, \ldots, \lfloor 2^{nR} \rfloor\}$ and $\ell' \in \{1, \ldots, \lfloor 2^{nR'} \rfloor\}$ the set:

$$\mathcal{A}(0, m', \ell') := \Big\{(\varphi^n, x^n, y^n) : \text{tp}\big(S^n(m',\ell'), T^n, W^n(m'), \varphi^n, \xi^n(T^n, \varphi^n), x^n, y^n\big) \in \mathcal{P}^n_{\mu,0}\Big\}. \quad (47)$$

By $\xi^n(W^n(m), \varphi^n)$, here we mean the component-wise application of the function $\xi(.,.)$ defined in (38) to the n-length sequences $W^n(m)$ and φ^n.

Define the region $\mathcal{A}_{\text{Rx},n} \subseteq \Phi^n \times \mathcal{X}^n \times \mathcal{Y}^n$

$$\mathcal{A}_{\text{Rx},n} \triangleq \bigcup_{m,m'} \bigcup_{\ell,\ell'} \mathcal{A}(m, m', \ell, \ell') \cup \bigcup_{m',\ell'} \mathcal{A}(0, m', \ell'), \quad (48)$$

where m and m' take value in $\{1, \ldots, \lfloor 2^{nR} \rfloor\}$ and ℓ and ℓ' in $\{1, \ldots, \lfloor 2^{nR'} \rfloor\}$. Notice that $\mathcal{A}_{\text{Rx},n}$ is deterministic for a given codebook, but random in the analysis here.

Since $\mathcal{A}_{\text{Rx},n}$ includes the acceptance region at the receiver, the average (over the random codebooks) type-II error probability is upper bounded as:

$$\mathbb{E}_{\mathcal{C}}[\beta_n] \leq \Pr\big[(\phi^n, X^n, Y^n) \in \mathcal{A}_{\text{Rx},n}|\mathcal{H} = 1\big]. \quad (49)$$

We can then write:

$$\mathbb{E}_{\mathcal{C}}[\beta_n] \leq \Pr\Big[(\phi^n, X^n, Y^n) \in$$

$$\bigcup_{m,m'}\bigcup_{\ell,\ell'}\mathcal{A}(m,m',\ell,\ell')\cup\bigcup_{m',\ell'}\mathcal{A}(0,m',\ell')|\mathcal{H}=1\Big]$$

$$\leq \Pr\Big[(\phi^n,X^n,Y^n)\in\bigcup_{(m,\ell)}\mathcal{A}(m,m,\ell,\ell)|\mathcal{H}=1\Big]$$

$$+\Pr\Big[(\phi^n,X^n,Y^n)\in\bigcup_{(m,\ell)\neq(m',\ell')}\mathcal{A}(m,m',\ell,\ell')|\mathcal{H}=1\Big]$$

$$+\Pr\Big[(\phi^n,X^n,Y^n)\in\bigcup_{m',\ell'}\mathcal{A}(0,m',\ell')|\mathcal{H}=1\Big]. \quad (50)$$

In a similar way as in [5], it can be shown that for sufficiently large blocklengths n, the first probability in (50) is upper bounded as:

$$\Pr\Big[(\phi^n,X^n,Y^n)\in\bigcup_{m}\bigcup_{\ell}\mathcal{A}(m,m,\ell,\ell)|\mathcal{H}=1\Big]\leq 2^{-n\theta_{1,\mu}}, \quad (51)$$

where

$$\theta_{1,\mu}:=min_{\pi_{SXY}\in\mathcal{P}^n_\mu}D(\pi_{SXY}||P_{S|X}Q_{XY})-\delta(\mu) \quad (52)$$

for a function $\delta(\mu)$ that goes to zero as $\mu\to 0$. Moreover, for sufficiently large n, the second probability in (50) is upper bounded as:

$$\Pr\Big[(\phi^n,X^n,Y^n)\in\bigcup_{(m,\ell)\neq(m',\ell')}\mathcal{A}(m,m',\ell,\ell')\Big]\leq 2^{-n\theta_{2,\mu}}, \quad (53)$$

where

$$\theta_{2,\mu}:=\min_{\pi_{SS'WW'VXY}\in\mathcal{P}^n_\mu}D(\pi_{SXY}||P_{S|X}Q_{XY})$$
$$+I(S;Y)+I(V;W|T)-I(S;X)-\delta'(\mu), \quad (54)$$

for a function $\delta'(\mu)$ that goes to zero as $\mu\to 0$.

The last term in (50) is upper bounded for sufficiently large blocklength n:

$$\Pr\Big[(\phi^n,X^n,Y^n)\in\bigcup_{m',\ell'}\mathcal{A}(0,m',\ell')|\mathcal{H}=1\Big]$$

$$\leq\sum_{m',\ell'}\Pr\Big[(\phi^n,X^n,Y^n)\in\mathcal{A}(0,m',\ell')|\mathcal{H}=1\Big]$$

$$\leq\sum_{m',\ell'}\sum_{\pi_{S'TW'\phi VXY}\in\mathcal{P}^n_{\mu,0}}$$
$$\Pr\Big[\text{tp}\Big(S^n(m',\ell'),T^n,W^n(m'),\phi^n,\xi^n(T^n,\phi^n),X^n,Y^n\Big)$$
$$=\pi_{S'TW'\phi VXY}\Big|\mathcal{H}=1\Big]$$

$$\leq\sum_{m',\ell'}\sum_{\pi_{S'TW'\phi VXY}\in\mathcal{P}^n_{\mu,0}}2^{-nD\left(\pi_{S'TW'\phi VXY}\middle\|P_SP_{TW}P_\phi\pi_{V|\phi T}Q_{XY}\right)}$$

where the last inequality holds by the way the random codebooks are generated and because given $\mathcal{H}=1$, the sources X^n,Y^n are i.i.d. $\sim Q_{XY}$. Define now

$$\tilde{\theta}_{3,\mu}:=\min_{\substack{\pi_{S'TW'\phi VXY}\\ \in\mathcal{P}^n_{\mu,0}}}D(\pi_{S'TW'\phi VXY}||P_SP_{TW}P_\phi\pi_{V|\phi T}Q_{XY})$$
$$-R-R'-\mu \quad (55)$$

and notice that there exist functions $\delta''(\mu)$ that $\to 0$ as $\mu\to 0$ and so that the following inequalities hold:

$$\begin{aligned}
\tilde{\theta}_{3,\mu}&\overset{(a)}{\geq}\min_{\pi_{S'TW'\phi VXY}\in\mathcal{P}^n_{\mu,0}}\Big[D(\pi_{TW'\phi V}||P_{TW}P_\phi\pi_{V|\phi T})\\
&\quad+D(\pi_{XY}||Q_{XY})+\mathbb{E}_{\pi_{XY}}[D(\pi_{S'|XY}||P_S)]\Big]\\
&\quad-I(S;X)-2\mu\\
&\overset{(b)}{\geq}\min_{\pi_{S'TW'VXY}\in\mathcal{P}^n_{\mu,0}}\Big[D(\pi_{TW'V}||P_{TW}P_{V|W=T})\Big]\\
&\quad+D(P_Y||Q_Y)+I(S;Y)-I(S;X)-\delta''(\mu)\\
&\overset{(c)}{\geq}\mathbb{E}_{P_{TW}}[D(P_{V|W}||P_{V|W=T})]\\
&\quad+D(P_Y||Q_Y)+I(S;Y)-I(S;X)-\delta''(\mu)\\
&=D(P_Y||Q_Y)+I(V;W|T)+I(S;Y)-I(S;X)\\
&\quad+\sum_t P_T(t)\cdot D(P_{V|T=t}||P_{V|W=t})-\delta''(\mu)\\
&:=\theta_{3,\mu}. \quad (56)
\end{aligned}$$

All three inequalities are based on the data processing inequality for KL-divergences; (a) also uses (13); and (b) and (c) also use the continuity of KL-divergences and that all types in $\mathcal{P}^n_{0,\mu}$ satisfy (39b), (39c), and (46). Thus, for sufficiently large n:

$$\Pr\Big[(\phi^n,X^n,Y^n)\in\bigcup_{m',\ell'}\mathcal{A}(0,m',\ell')|\mathcal{H}=1\Big]\leq 2^{-n\theta_{3,\mu}}. \quad (57)$$

Combining (50), (51), (53), and (57), taking $\mu\to 0$ and $n\to\infty$, the proof can be established by standard arguments.

REFERENCES

[1] A. Ahlswede and I. Csiszar, "Hypothesis testing with communication constraints," *IEEE Trans. on Info. Theory*, vol. 32, no. 4, pp. 533–542, Jul. 1986.

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

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

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

[5] S. Salehkalaibar, M. Wigger and L. Wang, "Hypothesis testing over multi-hop networks," Available at: https://arxiv.org/abs/1708.05198.

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

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

[8] M. Wigger and R. Timo, "Testing against independence with multiple decision centers," in *Proc. of SPCOM*, Bangalore, India, Jun. 2016.

[9] S. Sreekuma and D. Gunduz, "Distributed hypothesis testing over noisy channels," available at: https://arxiv.org/abs/1704.01535.

[10] A. El Gamal and Y. H. Kim, *Network information theory*, Cambridge Univ. Press, 2011.

[11] D. Wang, V. Chandar, S. Y. Chung and G. W. Wornell, "On reliability functions for single-message unequal error protection," in *Proc. IEEE Int. Symp. on Info. Theory*, MIT, pp. 2934–2938, 2012.

[12] S. P. Borade, "When all information is not created equal," Thesis, Massachusetts Institute of Technology, 2008.

[13] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley, 1991.