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 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", our third error exponent is not active and our overall type-II error exponent depends on the noisy channel only through its capacity. We thus recover the optimal exponent by Sreekuma and Gündüz [9].

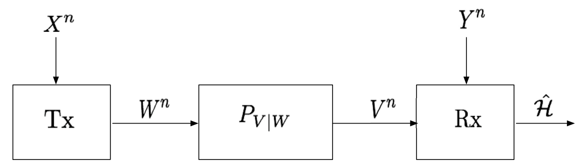

Fig. 1. Hypothesis testing over a noisy channel

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.

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:

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

and under the alternative hypothesis:

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

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

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

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.

Preparations: Choose $\mu > 0$, a sufficiently large positive integer n, an auxiliary distribution P_Q over $\mathcal{W}$, a conditional channel input distribution $P_{W|Q}$, and a conditional source distribution $P_{S|X}$ over a finite auxiliary alphabet $\mathcal{S}$ so that

$$I(S;X) \leq I(S;Y) + I(V;W|Q). \tag{11}$$

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

$$P_{SXYWVQ} = P_{S|X} \cdot P_{XY} \cdot P_Q \cdot P_{W|Q} \cdot P_{V|W}. \tag{12}$$

Further set the positive rates

$$R = I(V;W|Q) - \mu, \tag{13}$$

$$R' = I(S;X) + \mu - R. \tag{14}$$

Code Construction: Construct a random codebook

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

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 Q^n i.i.d. according to P_Q. Construct a random codebook

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

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

$$(s^n(m,\ell), x^n) \in \mathcal{T}^n_{\mu/2}(P_{SX}). \tag{15}$$

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 q^n over the channel.

Receiver: Assume that the receiver observes the sequences $V^n = v^n$ and $Y^n = y^n$ and that the "time-sharing sequence" $Q^n = q^n$. It first looks for an index $m' \in \{1, \ldots, \lfloor 2^{nR}\rfloor\}$ so that

$$(w^n(m'), v^n, q^n) \in \mathcal{T}^n_{\mu}(P_{WVQ}). \tag{16}$$

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

$$H_{\text{tp}(s^n(m',\ell'),y^n)}(S|Y) = \min_{\tilde{\ell}\in\{1,\ldots,\lfloor 2^{nR'}\rfloor\}} H_{\text{tp}(s^n(m',\tilde{\ell}),y^n)}(S|Y), \tag{17}$$

and checks whether

$$(s^n(m',\ell'), y^n) \in \mathcal{T}^n_{\mu}(P_{SY}). \tag{18}$$

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

IV. An Achievable Error Exponent

Theorem 1: Every error exponent $\theta \geq 0$ that satisfies the following condition (31) is achievable:

$$\theta \leq \max_{\substack{P_{S|X}, P_{QW}:\\ I(S;X|Y) \leq I(W;V|Q)}} \min\left\{\theta_1, \theta_2, \theta_3\right\}, \tag{19}$$

where

$$\theta_1 = \min_{\substack{\tilde{P}_{SXY}:\\ \tilde{P}_{SX}=P_{SX}\\ \tilde{P}_{SY}=P_{SY}}} D(\tilde{P}_{SXY}||Q_{XY}P_{S|X}), \tag{20}$$

$$\theta_2 = \min_{\substack{\tilde{P}_{SXY}:\\ \tilde{P}_{SX}=P_{SX}\\ \tilde{P}_{Y}=P_{Y}\\ H(S|Y)\leq H_{\tilde{P}}(S|Y)}} \Big[D(\tilde{P}_{SXY}||P_{S|X}Q_{XY}) + I(V;W|Q) - I(S;X|Y)\Big], \tag{21}$$

$$\theta_3 = D(P_Y||Q_Y) + I(V;W|Q) - I(S;X|Y) + \sum_{q\in\mathcal{W}} P_Q(q) \cdot D(P_{V|Q=q}||P_{V|W=q}), \tag{22}$$

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. ∎

In the next following Section V, we present a coding and testing scheme that combines the SHA hypothesis testing scheme for a noiseless link [3] with Borade's UEP channel coding that protects the 0-message better than the other messages [11], [12]. Recall that in the SHA scheme, the 0-message indicates that the transmitter decides on the alternative hypothesis $\mathcal{H} = 1$. In fact, since here we are only interested in the type-II error exponent, it is important that the receiver decides only on the null hypothesis $\mathcal{H} = 0$ if the transmitter shares this opinion. For this reason the 0-message needs special protection when being sent over the channel.

The expressions in Theorem 1 show three competing error exponents. In (20) and (21), 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 (22) 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

$$E_{\text{miss}} := \sum_{q\in\mathcal{W}} P_Q(q) \cdot D(P_{V|Q=q}||P_{V|W=q}), \tag{23}$$

which represents the largest possible *miss-detection exponent* for a single specially protected message at a given rate $I(W;V|Q)$ [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 (23), such as binary symmetric channels with small cross-over probability, or for sources where

$$\min_{\substack{\tilde{P}_{SXY}:\\ \tilde{P}_{SX}=P_{SX}\\ \tilde{P}_{Y}=P_{Y}}} D(\tilde{P}_{SXY}||P_{S|X}Q_{XY}) = D(P_Y||Q_Y), \tag{24}$$

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

$$Y = (\bar{Y}, Z), \tag{25}$$

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

$$Q_{X\bar{Y}Z} = P_{XZ} \cdot P_{\bar{Y}|Z}. \tag{26}$$

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

$$\theta \leq \max_{\substack{P_{S|X},\ P_W:\\ I(S;X|Z)\leq I(W;V)}} I(S;\bar{Y}|Z), \tag{27}$$

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 Q and W and a random variable S so that $I(S;X|Z) \leq I(W;V|Q) = I(W;V)$. Then, Theorem 1 specializes to:

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

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

$$\begin{aligned}
&I(V;W) - I(S;X) + I(S;\bar{Y},Z)\\
&= I(V;W) - I(S;X|Z) + I(S;\bar{Y}|Z)\\
&\geq I(S;\bar{Y}|Z),
\end{aligned} \tag{28}$$

where the inequality holds because we imposed $I(S;X|Z) \leq I(W;V)$. ∎

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

$$\begin{aligned}
\mathcal{H} = 0: \qquad & X \sim \text{Bern}(p_0), \quad Y = X \oplus N_0,\\
& N_0 \sim \text{Bern}(q_0),
\end{aligned} \tag{29}$$

for N_0 independent of X. Under the alternative hypothesis:

$$\mathcal{H} = 1: \qquad X \sim \text{Bern}(p_1), \quad Y \sim \text{Bern}(p_0 \star q_0), \tag{30}$$

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, Theorem 1 simplifies to:

$$\theta \leq \max_{\substack{P_{S|X}, P_{QW}:\\ I(S;X|Y)\leq I(W;V|Q)}} \min\{\theta_1, \theta_2, \theta_3\}, \tag{31}$$

where

$$\theta_1 \leq D(P_X||Q_X) + I(S;Y), \tag{32}$$

$$\theta_2 \leq D(P_X||Q_X) + I(V;W|Q) + I(S;Y) - I(S;X), \tag{33}$$

$$\theta_3 \leq \sum_{q\in\mathcal{W}} P_Q(q) D(P_{V|Q=q}||P_{V|W=q}) + I(V;W|Q) + I(S;Y) - I(S;X). \tag{34}$$

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, $r \geq 0.4325$,

$$D(P_X||Q_X) \geq \sum_{q\in\mathcal{W}} P_Q(q) D(P_{V|Q=q}||P_{V|W=q}) + I(V;W|Q), \tag{35}$$

and exponent θ_3 is smaller than θ_1 and θ_2, irrespective of the choice of the random variables S, Q, W. It is then optimal to choose S constant and (Q, W) so as to maximize the sum $\sum_{q\in\mathcal{W}} P_Q(q)D(P_{V|Q=q}||P_{V|W=q}) + I(V;W|Q)$. 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 (23) 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:

$$V_t = \xi(W_t, \phi_t). \tag{36}$$

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

$$|\pi_{SX} - P_{SX}| \leq \mu/2, \tag{37a}$$
$$|\pi_{S'Y} - P_{SY}| \leq \mu, \tag{37b}$$
$$|\pi_{W'V} - P_{WV}| \leq \mu, \tag{37c}$$
$$\pi_{V|\phi W} = \mathbb{1}\{V = \xi(W, \phi)\}, \tag{37d}$$
$$H_{\pi_{S'Y}}(S|Y) \leq H_{\pi_{SY}}(S|Y). \tag{37e}$$

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:

$$\mathcal{E}_{\text{Tx}}\colon \{\nexists(m,\ell)\colon (S^n(m,\ell), X^n) \in \mathcal{T}^n_{\mu/2}(P_{SX})\} \tag{38}$$
$$\mathcal{E}^{(1)}_{\text{Rx}}\colon \{(S^n(M,L), Y^n) \notin \mathcal{T}^n_\mu(P_{SY})\} \tag{39}$$
$$\mathcal{E}^{(2)}_{\text{Rx}}\colon \{(W^n(M), V^n) \notin \mathcal{T}^n_\mu(P_{WV})\} \tag{40}$$
$$\mathcal{E}^{(3)}_{\text{Rx}}\colon \{\exists \ell' \neq L\colon H_{\text{tp}(s^n(M,\ell'),y^n)}(S|Y) = \min_{\tilde{\ell}} H_{\text{tp}(s^n(M,\tilde{\ell}),y^n)}(S|Y)\}. \tag{41}$$

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

$$\alpha_n \leq \Pr[\mathcal{E}_{\text{Tx}}] + \Pr[\mathcal{E}^{(1)}_{\text{Rx}}|\mathcal{E}^c_{\text{Tx}}] + \Pr[\mathcal{E}^{(2)}_{\text{Rx}}|\mathcal{E}^c_{\text{Tx}}, \mathcal{E}^{(1)c}_{\text{Rx}}] + \Pr[\mathcal{E}^{(3)}_{\text{Rx}}|\mathcal{E}^{(1)c}_{\text{Rx}}, \mathcal{E}^{(2)c}_{\text{Rx}}, \mathcal{E}^c_{\text{Tx}}] \tag{42}$$
$$\leq \epsilon/4 + \epsilon/4 + \epsilon/4 + \epsilon/4 = \epsilon, \tag{43}$$

where the first summand of (42) can be upper bounded by means of the covering lemma [10] and the rate constraint (14); the second by means of the Markov lemma [10]; the third by means of the packing lemma [10] and the rate constraint (13); 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 $\pi_{S'QW'V\phi XY}$ over the alphabets $\mathcal{S}^n \times \mathcal{W}^n \times \mathcal{W}^n \times \mathcal{V}^n \times \Phi^n \times \mathcal{X}^n \times \mathcal{Y}^n$ that satisfy (37b), (37c), and

$$\pi_{V|\phi Q} = \mathbb{1}\{V = \xi(Q,\phi)\}. \tag{44a}$$

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:

$$\mathcal{A}(0,m',\ell') := \Big\{(\varphi^n, x^n, y^n) : \text{tp}\big(S^n(m',\ell'), Q^n, W^n(m'), \varphi^n, \xi^n(Q^n,\varphi^n), x^n, y^n\big) \in \mathcal{P}^n_{\mu,0}\Big\}. \tag{45}$$

By $\xi^n(w^n, \varphi^n)$, here we mean the component-wise application of the function $\xi(.,.)$ defined in (36) to the n-length sequences w^n and φ^n.

The average (over the random codebooks) type-II error probability satisfies:

$$\mathbb{E}_{\mathcal{C}}[\beta_n] \leq \Pr\left[(\phi^n, X^n, Y^n) \in \mathcal{A}_{\text{Rx},n} | \mathcal{H} = 1\right], \tag{46}$$

where $\mathcal{A}_{\text{Rx},n} \subseteq \Phi^n \times \mathcal{X}^n \times \mathcal{Y}^n$ includes the acceptance region at the receiver:

$$\mathcal{A}_{\text{Rx},n} \triangleq \bigcup_{m,m'} \bigcup_{\ell,\ell'} \mathcal{A}(m,m',\ell,\ell') \cup \bigcup_{m',\ell'} \mathcal{A}(0,m',\ell'), \tag{47}$$

where m and m' take value in $\{1, \ldots, \lfloor 2^{nR} \rfloor\}$ and ℓ and ℓ' in $\{1, \ldots, \lfloor 2^{nR'} \rfloor\}$. We can then write:

$$\begin{aligned}
&\mathbb{E}_{\mathcal{C}}[\beta_n] \\
&\leq \Pr\Big[(\phi^n, X^n, Y^n) \in \bigcup_{m,m'} \bigcup_{\ell,\ell'} \mathcal{A}(m,m',\ell,\ell') \cup \bigcup_{m',\ell'} \mathcal{A}(0,m',\ell') | \mathcal{H} = 1\Big] \\
&\leq \Pr\Big[(\phi^n, X^n, Y^n) \in \bigcup_{(m,\ell)} \mathcal{A}(m,m,\ell,\ell) | \mathcal{H} = 1\Big] \\
&+ \Pr\Big[(\phi^n, X^n, Y^n) \in \bigcup_{(m,\ell)\neq(m',\ell')} \mathcal{A}(m,m',\ell,\ell') | \mathcal{H} = 1\Big]
\end{aligned}$$

$$+ \Pr\Big[(\phi^n, X^n, Y^n) \in \bigcup_{m',\ell'} \mathcal{A}(0, m', \ell')|\mathcal{H}=1\Big]. \tag{48}$$

Consider the last term in (48). By the code construction and Sanov's theorem, for sufficiently large n:

$$\begin{aligned}
&\Pr\Big[(\phi^n, X^n, Y^n) \in \bigcup_{m',\ell'} \mathcal{A}(0, m', \ell')|\mathcal{H}=1\Big] \\
&\leq \sum_{m',\ell'} \Pr\Big[(\phi^n, X^n, Y^n) \in \mathcal{A}(0, m', \ell')|\mathcal{H}=1\Big] \\
&\leq 2^{-n\tilde{\theta}_{3,\mu}},
\end{aligned} \tag{49}$$

where we define for some function $\delta(\mu)$ that $\to 0$ as $\mu \to 0$:

$$\begin{aligned}
\tilde{\theta}_{3,\mu} &:= \min_{\substack{\pi_{S'QW'V\phi XY} \\ \in \mathcal{P}^n_{\mu,0}}} D(\pi_{S'QXY\phi W'}||P_S P_{WQ} P_\phi Q_{XY}) \\
&\quad - R - R' - \mu \\
&= \min_{\pi_{S'QW'V\phi XY} \in \mathcal{P}^n_{\mu,0}} \Big[D(\pi_{Q\phi W'}||P_{WQ}P_\phi) \\
&\quad + D(\pi_{XY}||Q_{XY}) + \mathbb{E}_{\pi_{XY}}[D(\pi_{S'|XY}||P_S)]\Big] \\
&\quad - I(S;X) - \mu \\
&\overset{(a)}{\geq} \min_{\pi_{S'QW'V\phi XY} \in \mathcal{P}^n_{\mu,0}} \Big[D(\pi_{QVW'}||P_{WQ}P_{V|W=Q}) \\
&\quad + D(\pi_Y||Q_Y) + I_\pi(S';YX)\Big] - I(S;X) - \mu \\
&\overset{(b)}{\geq} \mathbb{E}_{WQ}[D(P_{V|W}||P_{V|Q}) + D(P_{V|Q}||P_{V|W=Q})] \\
&\quad + D(P_Y||Q_Y) + I(S;Y) - I(S;X) - \delta(\mu) \\
&= D(P_Y||Q_Y) + I(V;W|Q) \\
&\quad + \sum_q P_Q(q) \cdot D(P_{V|Q=q}||P_{V|W=q}) \\
&\quad + I(S;Y) - I(S;X) - \delta(\mu) \\
&:= \theta_{3,\mu}.
\end{aligned} \tag{50}$$

Here, (a) follows by defining $P_{V|\phi Q} = \mathbb{1}\{V = \xi(Q,\phi)\}$ and by the data processing inequality for KL-divergences and (b) follows by the definition of the type class $\mathcal{P}^n_{0,\mu}$ and because $I_\pi(S';XY) \geq I_\pi(S';Y)$.

Consider now the first and second probabilities in (48). Following similar steps to above, one finds:

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

and

$$\Pr\Big[(\phi^n, X^n, Y^n) \in \bigcup_{(m,\ell)\neq(m',\ell')} \mathcal{A}(m,m',\ell,\ell')\Big] \leq 2^{-n\tilde{\theta}_{2,\mu}}, \tag{52}$$

where

$$\theta_{1,\mu} := \min_{\pi_{SXY} \in \mathcal{P}^n_\mu} D(\pi_{SXY}||Q_{XY}P_{S|X}) - \delta'(\mu), \tag{53}$$

$$\begin{aligned}
\theta_{2,\mu} &:= \min_{\pi_{SXY} \in \mathcal{P}^n_\mu} D(\pi_{SXY}||Q_{XY}P_{S|X}) + I(V;W|Q) \\
&\quad + I(S;Y) - I(S;X) - \delta''(\mu),
\end{aligned} \tag{54}$$

for functions $\delta'(\mu)$ and $\delta''(\mu)$ that go to zero as $\mu \to 0$. These exponents are derived in detail in [5, Appendix H, Proof of Thm 4] for a noiseless communication link.

Combining (48)–(54), proves that for sufficiently large blocklengths n, the average type-II error probability satisfies

$$\mathbb{E}_\mathcal{C}[\beta_n] \leq \max\Big\{2^{-n\theta_{1,\mu}},\ 2^{-n\theta_{2,\mu}},\ 2^{-n\theta_{3,\mu}}\Big\}. \tag{55}$$

By standard arguments and successively eliminating the worst half codebooks with respect to the exponents $\theta_{1,\mu}$, $\theta_{2,\mu}$, and $\theta_{3,\mu}$, it can be shown that there exists at least one codebook for which

$$\beta_n \leq 8 \cdot \max\Big\{2^{-n\theta_{1,\mu}},\ 2^{-n\theta_{2,\mu}},\ 2^{-n\theta_{3,\mu}}\Big\}. \tag{56}$$

Letting $\mu \to 0$ and $n \to \infty$, we get $\theta_{1,\mu} \to \theta_1$, $\theta_{2,\mu} \to \theta_2$, $\theta_{3,\mu} \to \theta_3$, which establishes achievability of Theorem 1.

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.