Independence Testing under Zero-Rate Communication: Beyond the Stein-Regime

Nicolas Le Gouic
LTCI Telecom Paris
91120 Palaiseau, France
nicolas.legouic@ip-paris.fr

Michèle Wigger
LTCI Telecom Paris
91120 Palaiseau, France
michele.wigger@telecom-paris.fr

***Abstract*—We study the two-terminal distributed "testing against independence" problem when the remote sensor can only send k noise-free bits to the decision center or communicate over k uses of a discrete memoryless channel, for k sublinear in the observation length n. We show that under these conditions the smallest type-II error probability only decays exponentially in the number of transmitted bits/channel uses k but not in the number of observations n. Moreover, we provide a general lower bound on the exponent of the decay-rate and a matching upper bound when the type-I error probability is required to vanish for growing n.**

For the noise-free bits scenario, we also consider a generalization of "testing against independence" where under the alternative hypothesis the observations at the two terminals follow an arbitrary product distribution. Previous results showed that in this case the type-II error probability decays exponentially in n. In this work, we characterize a lower bound on the second order decay-rate in k when the number of bits is larger than $\sqrt{n}$. We have a matching converse result when the type-I error probability is required to tend to 0 sufficiently fast in n.

***Index Terms*—Testing against independence, communication constraint, zero-rate, Stein-exponent, contraction coefficient.**

I. Introduction

Consider the distributed binary hypothesis testing problem (see also Fig. 1) where a decision center aims to guess between two hypotheses that govern the joint distribution of two independent and identically distributed (IID) sources, one that is directly observed at the decision center and the other observed at a remote sensor. The sensor communicates to the decision center either over a noise-free link or over a noisy communication channel. Based on this communication and its own observations, the decision center attempts to guess the correct hypothesis.

For the noise-free link setup, the seminal paper [1] characterized the *Stein-exponent*, i.e., the fastest exponential decay-rate of the type-II error probability in the observation length n when a threshold is imposed on the type-I error. While generally only in multi-letter form, a computable single-letter form was obtained in the special case of *testing against independence*, where under the alternative hypothesis the joint distribution is the product of the marginals under the null hypothesis. A single-letter formulation for Stein's exponent was subsequently obtained for the more general case of *testing against conditional independence* [2]. Here, the scheme in [3] is not sufficient but the binning scheme in [4] was shown to be optimal. This scheme has recently been improved in [5], [6].

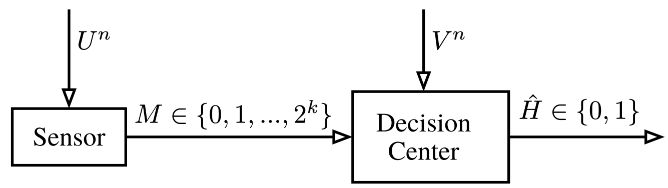

Fig. 1. Hypothesis Testing under One-Sided Compression Constraint.

Numerous extensions of this problem have been considered, including multiple decision centers [7], two-hop relay networks [8], [9], interactive communications [10], and security constraints [11].

For the setup with a noisy link, lower bounds on the Stein-exponent where proposed in [12]–[15]. Moreover, weak and strong converse results were proved for testing against independence and conditional independence [12], [14], [16].

All above results concern scenarios where the number of transmitted noise-free bits scales linearly in the observation length n or the noisy channel can be used a linear number of times. In this work, we focus on the highly communication-limited regime where the noise-free link or noisy channel can be used only k times, for k a function that scales sublinearly in n. Under this assumption, [3], [17] determined the Stein-exponent for the noise-free link setup and showed that it neither depends on the growth rate of k—in fact, $k = 1$ is sufficient to achieve the exponent—nor on the allowed type-I error probability threshold ε. This result has recently been extended to noisy channels in [18].

For many sources, the Stein-exponents found in [3], [17], [18] are strictly positive, implying that the type-II error probability can vanish exponentially in the observation length n. However, for certain sources, as in testing against independence, the Stein-exponent is 0. As a consequence, the best scaling of the type-II error probability remains open in such cases. For testing against independence, it has been conjectured in [19, Chap. 16.5] that the smallest type-II error probability decays exponentially in k with an exponent equals to the contraction coefficient of the transition probability $P_{V|U}$ from the sensor's source U to the decision center's source V under the null hypothesis, with for input P_U the marginal distribution of U (the result was hinted through a limiting argument from the result in [1], were an interchange of limits lacks proper justification). The contraction coefficient is the smallest factor that strengthen the data-processing inequality (DPI) for the according channel and input into a strong DPI.

In this work, we prove the conjecture in [19] for vanishing type-I errors. More generally, we propose a coding scheme for all sources where under the alternative hypothesis the observations at the sensor and the decision center are independent (but not necessarily of same marginal distributions as under the null hypothesis). The type-II error probability of our scheme decays as $2^{-n\theta^*(\varepsilon)-k\eta+o(k)}$ for $\theta^*(\varepsilon)$ the Stein-exponent determined in [3], [17] and η the mentioned contraction coefficient. We show that this result is optimal whenever the type-I error probability is required to vanish faster than k/n while k is faster than $\sqrt{n}$.

In our last contribution, we assume that the sensor communicates to the decision center over k (sublinear in n) uses of a discrete memoryless channel (DMC). We focus on testing against independence where we show that when the type-I error probability is required to vanish, the optimal type-II error probability is $2^{-k\eta C+o(k)}$, for again η the contraction coefficient and C the capacity of the DMC.

This article is organized as follows. The setup is described in Section II along with the main results for the noise-free link setup. Section III describes the noisy-channel scenario, for which we restrict to testing against independence only. The proofs are provided in the following sections IV to VII.

II. Problem Statement and Main Result

A. Problem Statement

In this work, the tester aims to decide whether the n IID samples (U^n, V^n) have been generated according to the null or alternative hypothesis as

$$H_0 : (U^n, V^n) \overset{\text{IID}}{\sim} P_{UV},$$
$$H_1 : (U^n, V^n) \overset{\text{IID}}{\sim} Q_U Q_V,$$

for P_{UV} a probability mass function (PMF) over the finite alphabet $\mathcal{U} \times \mathcal{V}$, Q_U a PMF over $\mathcal{U}$ and Q_V a PMF over $\mathcal{V}$. We will assume that $Q_U(u) = 0$ and $Q_V(v) = 0$ only when $P_U(u) = 0$ and $P_V(v) = 0$, respectively.

As opposed to classical binary hypothesis testing, the decision center doesn't have complete access to the samples, but is subject to a compression constraint on the observation of U^n while it gets full access to V^n. As depicted in Fig. 1, U^n is observed at a remote sensor that can send a message $M \in \{0, ..., 2^k\}$—roughly a k-bits message—that it obtains by means of an encoding function $f_n : \mathcal{U}^n \to \{0, ..., 2^k\}$ as $M = f_n(U^n)$.

We consider the case where k grows with n, but at most sublinearly, so that $k = k(n)$ with

$$\lim_{n\to\infty} k(n) = \infty, \tag{1.1}$$

$$\lim_{n\to\infty} \frac{k(n)}{n} = 0. \tag{1.2}$$

Based on its observations (M, V^n), the decision center produces a guess $\hat{H} \in \{0, 1\}$ according to a deterministic hypothesis test $g_n : \{0, ..., M\} \times \mathcal{V}^n \to \{0, 1\}$.

The performance criteria associated with a pair (f_n, g_n) are its type-I error probability

$$\alpha_n := \mathbb{P}\Big[\hat{H} = 1 | H_0\Big], \tag{2}$$

and its type-II error probability

$$\beta_n := \mathbb{P}\Big[\hat{H} = 0 | H_1\Big]. \tag{3}$$

Definition 1. Let $\varepsilon \in [0, 1)$. A Stein-exponent θ is called ε-achievable if there exists a sequence (in n) of encodings and decision functions (f_n, g_n) so that

$$\limsup_{n\longrightarrow\infty} \alpha_n \le \varepsilon, \tag{4}$$

and

$$\liminf_{n\longrightarrow\infty} -\frac{1}{n} \log \beta_n \ge \theta. \tag{5}$$

The largest ε-achievable exponent is called the Stein-exponent and denoted $\theta^*(\varepsilon)$.

The Stein-exponent $\theta^*(\varepsilon)$ of our setup has been characterized in [3], [17]:

$$\theta^*(\varepsilon) = D(P_U \parallel Q_U) + D(P_V \parallel Q_V). \tag{6}$$

Interestingly, this exponent doesn't depend neither on the threshold ε nor on the growth-rate of $k(n)$, and in particular can be achieved even with a single bit of communication $k(n) = 1$. Note that when $P_U = Q_U$ and $P_V = Q_V$ the resulting exponent is $\theta^*(\varepsilon) = 0$.

The focus in this article is on the "second-order" asymptotic of the error probability, that is the additional decay of type-II error that can be obtained by leveraging the growth rate $k(n)$.

Definition 2. Let $\varepsilon \in [0, 1)$. A second-order exponent η is said ε-achievable if there exists a sequence (f_n, g_n) so that (4) holds and moreover

$$\liminf_{n\to\infty} \frac{1}{k} \cdot (-\log(\beta_n) - n \cdot \theta^*(\varepsilon)) \ge \eta. \tag{7}$$

The largest ε-achievable second-order exponent is denoted $\eta^*(\varepsilon)$.

Even thought we restrict to deterministic encoding and tests, our converse results apply equally to randomized procedures.

B. Main Results

Our results include a lower bound (achievability) for $\eta^*(\varepsilon)$ for $\varepsilon \in [0, 1)$ when $k(n)/\sqrt{n} \longrightarrow \infty$ or $(P_U, Q_U) = (P_V, Q_V)$, a weak converse for $\varepsilon = 0$ when $(P_U, Q_U) = (P_V, Q_V)$, and a "very weak" converse for general Q_U and Q_V where we require that α_n be such that even $\alpha_n \cdot n/k \longrightarrow 0$.

Theorem 1. *[Achievability]* When one of the following two conditions is satisfied

1) $P_U = Q_U$ and $P_V = Q_V$,
2) k grows fast enough as

$$\lim_{n\to\infty} \frac{k(n)}{\sqrt{n}} = \infty, \tag{8}$$

we have for any $\varepsilon \in [0, 1)$

$$\eta^*(\varepsilon) \ge \eta(P_U, P_{V|U}), \tag{9}$$

where $\eta(P_U, P_{V|U})$ is the contraction coefficient [19, Def. 33.10] expressed as

$$\eta(P_U, P_{V|U}) = \sup_{\substack{P_{W|U} \\ W-U-V}} \frac{\mathrm{I}(W;V)}{\mathrm{I}(W;U)}. \tag{10}$$

This result is proved in Section IV.

Theorem 2. *[Weak converse]* When $P_U = Q_U$ and $P_V = Q_V$:

$$\eta^*(0) \le \eta(P_U, P_{V|U}). \tag{11}$$

The proof is based on the forthcoming Theorem 3, as explained in the concluding remark of Section VII.

Equipped with Theorem 1 and Theorem 2, we have an exact characterization of $\eta^*(0)$ when $P_U = Q_U$ and $P_V = Q_V$ as

$$\eta^*(0) = \eta(P_U, P_{V|U}). \tag{12}$$

So far, it is not clear whether a strong converse holds for the setup under consideration due to the normalization factor k in the definition of η in (7) instead of n as used in the definition of θ in (6).

For arbitrary Q_U and Q_V (not just for when $Q_U = P_U$ and $Q_V = P_V$) we can obtain a converse result under more restrictive conditions on α_n.

Theorem 3. *[Very weak converse]* If requirement (4) is strengthened to

$$\limsup_{n\longrightarrow\infty} \left(\alpha_n \cdot \frac{n}{k(n)} \right) = 0, \tag{13}$$

which in particular implies $\alpha_n \to 0$, it holds that

$$\eta^*(0) \le \eta(P_U, P_{V|U}). \tag{14}$$

The proof is delayed to Section VII.

By Theorems 1 and 3, under (8) and (13) it holds that

$$\eta^*(0) = \eta(P_U, P_{V|U}). \tag{15}$$

III. Communication over a DMC

In this section, we consider the same setup as described in Section II, except that the sensor communicates to the decision center over $k(n)$ uses of a DMC $\Gamma_{Y|X}$ with capacity $C > 0$. We focus on the case of testing against independence were the alternative hypothesis is $P_U P_V$.

Stein-exponent and second-order exponent are defined analogously to before, scaling the second-order by channel uses k.

Theorem 4. *[Achievability]* When $P_U = Q_U$ and $P_V = Q_V$, for any $\varepsilon \in [0,1)$

$$\eta^*(\varepsilon) \ge C \cdot \eta(P_U, P_{V|U}). \tag{16}$$

The proof is given in Section V.

Theorem 5. *[Weak converse]* When $P_U = Q_U$ and $P_V = Q_V$,

$$\eta^*(0) \le C \cdot \eta(P_U, P_{V|U}). \tag{17}$$

Proof in Section VI.

These two results allow us to conclude that

$$\eta^*(0) = C \cdot \eta(P_U, P_{V|U}), \tag{18}$$

similarly to the noiseless setting.

IV. Proof of Theorem 1

Assume $P_U \neq Q_U$ and $P_V \neq Q_V$, other cases are treated at the end of the section. Let $\mathbb{V}_U > 0$ and $\mathbb{T}_U$ denote the respective second and third centered moments of $\log(P_U/Q_U)$, and identically $\mathbb{V}_V > 0$ and $\mathbb{T}_V$ for $\log(P_V/Q_V)$.

We use the definitions and δ-convention from [20]. Fix a sufficiently large n and small constants $\mu, \nu > 0$. Pick a conditional $P_{W|U}$ on a finite auxiliary alphabet $\mathcal{W}$ and set

$$R = \mathrm{I}(W;U) + \nu. \tag{19}$$

This and all following mutual informations are calculated according to the joint PMF $P_{WUV} = P_{W|U} P_{UV}$.

Let u^n and v^n be the realizations of U^n and V^n. Let

$$\tau := k/R, \tag{20}$$

and notice that $\tau < n$ for n large enough (1).

Draw a length-τ random codebook IID according to P_W with rate R to get $\mathcal{C} = \{w^\tau(1), ..., w^\tau(2^k)\}$. Consider local tests $\varphi_1 : \mathcal{U}^n \to \{0,1\}$ and $\varphi_2 : \mathcal{V}^n \to \{0,1\}$ specified later.

Encoder: Looks for an index m such that $w^\tau(m)$ and u^τ are jointly typical, $(w^\tau(m), u^\tau) \in \mathcal{T}_\mu^\tau(P_{WU})$. If successful, and moreover $\varphi_1(u^n) = 0$, it sends $M = m$, else it sends $M = 0$.

Decoder: If $M \neq 0$, $(w^\tau(M), v^\tau) \in \mathcal{T}_\mu^\tau(P_{WV})$, and $\varphi_2(v^n) = 0$, set $\hat{H} = 0$. Otherwise, set $\hat{H} = 1$.

Analysis: We start by analyzing the type-I error probability α_n of our scheme, which can be upper-bounded as

$$\begin{aligned}
\alpha_n &= \mathbb{P}[\varphi_1(U^n) = 1 \text{ or } \varphi_2(V^n) = 1 \\
&\qquad \text{or } \forall m : (W^\tau(m), U^\tau) \notin \mathcal{T}_\mu^\tau(P_{WU}) \\
&\qquad \text{or } (W^\tau(M), V^\tau) \notin \mathcal{T}_\mu^\tau(P_{WV}) | H_0] &(21.1)\\
&\le \mathbb{P}[\varphi_1(U^n) = 1 | H_0] + \mathbb{P}[\varphi_2(V^n) = 1 | H_0] \\
&\quad + \mathbb{P}[\forall m : (W^\tau(m), U^\tau) \notin \mathcal{T}_\mu^\tau(P_{WU}) \\
&\qquad \text{or } (W^\tau(M), V^\tau) \notin \mathcal{T}_\mu^\tau(P_{WV}) | H_0] &(21.2)\\
&= \mathbb{P}[\varphi_1(U^n) = 1 | H_0] + \mathbb{P}[\varphi_2(V^n) = 1 | H_0] \\
&\quad + \mathbb{P}[\forall m : (W^\tau(m), U^\tau) \notin \mathcal{T}_\mu^\tau(P_{WU}) | H_0] \\
&\quad + \mathbb{P}[(W^\tau(M), V^\tau) \notin \mathcal{T}_\mu^\tau(P_{WV}) \\
&\qquad | \exists m : (W^\tau(m), U^\tau) \in \mathcal{T}_\mu^\tau(P_{WU}), H_0]. &(21.3)
\end{aligned}$$

We now analyze the behavior of each summand in (21.3). Take sequences $(\alpha_{1,n})$ and $(\alpha_{2,n})$ such that

$$\lim_{n\to\infty} \alpha_{1,n} = \lim_{n\to\infty} \alpha_{2,n} = \varepsilon/2. \tag{22}$$

For the first summand, we take φ_1, the local test at sensor between P_U and Q_U, as a likelihood ratio test with threshold

$$\gamma_1 = n\,\mathrm{D}(P_U \,\|\, Q_U) - \sqrt{n\mathbb{V}_U}\,\mathcal{Q}^{-1}\left(\alpha_{1,n} - \frac{6\mathbb{T}_U}{\sqrt{n}\mathbb{V}_U^{3/2}} \right), \tag{23}$$

for $\mathcal{Q}^{-1}$ the inverse of the $\mathcal{Q}$-function. Thus, $\varphi_1(u^n) = 0$ when

$$\log\left(\frac{P_{U^n}(u^n)}{Q_{U^n}(u^n)} \right) \ge \gamma_1. \tag{24}$$

Berry-Esseen theorem allows to upper bound the probability of $\varphi_1(u^n) = 1$ under H_0 by $\alpha_{1,n}$ [21, eq. A.11]. The second summand is upper bounded by $\alpha_{2,n}$ the same way by choosing φ_2 as a likelihood ratio test with threshold

$$\gamma_2 = n\,\mathrm{D}(P_V \,\|\, Q_V) - \sqrt{n\mathbb{V}_V}\mathcal{Q}^{-1}\left(\alpha_{2,n} - \frac{6\mathbb{T}_V}{\sqrt{n}\mathbb{V}_V^{3/2}}\right). \quad (25)$$

Remaining terms vanish: the third summand by $R > I(W;V)$ and covering lemma [19, Cor. 25.6], while the fourth by Markov lemma [19, Prop. 25.7]. Overall, we obtain

$$\lim_{n\to\infty} \alpha_n \leq \varepsilon. \quad (26)$$

For type-II error, see that when $\varphi_1(u^n) = 0$, (24) gives

$$Q_{U^n}(u^n) \leq P_{U^n}(u^n) \cdot 2^{-\gamma_1}. \quad (27)$$

Same goes for φ_2, and we get for realizations such that $\hat{H} = 0$

$$Q_{U^n}(u^n)Q_{V^n}(v^n) \leq P_{U^n}(u^n)P_{V^n}(v^n) \cdot 2^{-\gamma_1-\gamma_2}. \quad (28)$$

Using (28) and denoting H_2 for $(U^n, V^n) \overset{\text{IID}}{\sim} P_U P_V$,

$$\begin{aligned}\beta_n &= \mathbb{P}\left[\hat{H} = 0 | H_1\right] && (29.1)\\ &\leq \mathbb{P}\left[\hat{H} = 0 | H_2\right] \cdot 2^{-\gamma_1-\gamma_2} && (29.2)\\ &\leq \mathbb{P}\left[(W^\tau(M), V^\tau) \in \mathcal{T}_\mu^\tau(P_{WV}) | M \neq 0, H_2\right]\\ &\quad \cdot 2^{-\gamma_1-\gamma_2}, && (29.3)\end{aligned}$$

and by law of total probability

$$\begin{aligned}&\mathbb{P}\left[(W^\tau(M), V^\tau) \in \mathcal{T}_\mu^\tau(P_{WV}) | M \neq 0, H_2\right]\\ &= \sum_{j=1}^{2^k} \underbrace{\mathbb{P}\left[(W^\tau(j), V^\tau) \in \mathcal{T}_\mu^\tau(P_{WV}) | M = j, H_2\right]}_{=(\star)}\\ &\quad \cdot \mathbb{P}[M = j | M \neq 0, H_2]. \quad (30)\end{aligned}$$

Note that V^τ is independent $W^\tau(j)$ for any $j = 1, ..., 2^k$ under H_2. Thus, probability of success of the typicality check considered in $(\star)$ is upper-bounded by $2^{-\tau(\mathrm{I}(W;V)-\nu)}$ provided n is large enough [22, Cor. 4.8]. We obtain

$$\beta_n \leq 2^{-\tau(\mathrm{I}(W;V)-\nu)-\gamma_1-\gamma_2}. \quad (31)$$

Consequently, we obtain the lower bound

$$\begin{aligned}-\log(\beta_n) &\geq n\theta^*(\varepsilon) + \tau(\mathrm{I}(W;V) - \nu) && (32.1)\\ &\quad -\sqrt{n\mathbb{V}_U}\mathcal{Q}^{-1}\left(\alpha_{1,n} - \frac{6\mathbb{T}_U}{\sqrt{n}\mathbb{V}_U^{3/2}}\right) && (32.2)\\ &\quad -\sqrt{n\mathbb{V}_V}\mathcal{Q}^{-1}\left(\alpha_{2,n} - \frac{6\mathbb{T}_V}{\sqrt{n}\mathbb{V}_V^{3/2}}\right). && (32.3)\end{aligned}$$

Assuming $k(n)$ verifies (8), this yields

$$\begin{aligned}\lim_{n\to\infty} -\frac{1}{k}(\log(\beta_n) - n\theta^*(\varepsilon)) &\geq \frac{\tau}{k}(I(W;V) - \nu) && (33.1)\\ &= \frac{I(W;V) - \nu}{I(W;U) + \nu}. && (33.2)\end{aligned}$$

Letting ν tend to 0, then optimizing over $P_{W|U}$ gives the result claimed in Theorem 1.

When $P_U = Q_U$ or $P_V = Q_V$, use same scheme without irrelevant local tests. (8) isn't needed when both are equals.

V. Proof of Theorem 4

Choose a large positive integer n, a channel input distribution P_X, and a conditional source distribution $P_{W|U}$ over a finite alphabet $\mathcal{W}$. All following mutual informations are calculated according to the joint PMF $P_{UVWXY} = P_{UV}P_{W|U}P_X\Gamma_{Y|X}$. Then, take small $\delta, \mu, \nu > 0$ and set

$$\tau = k\left(\frac{I(X;Y)}{R} - \delta\right), \quad (34)$$

for $R = I(W;U) + \nu$.

Code Construction: Construct a random source codebook

$$\mathcal{C}_W = \{W^\tau(m) : m \in \{1, ..., 2^{\tau R}\}\}, \quad (35)$$

by independently drawing all codewords IID according to P_W.

Construct a random channel codebook

$$\mathcal{C}_X = \{X^k(m) : m \in \{1, ..., 2^{\tau R}\}\}, \quad (36)$$

by independently drawing all codewords IID according to P_X.

Encoding: Given it observes the source sequence u^n, the transmitter looks for an index m that satisfies

$$(w^\tau(m), u^\tau) \in \mathcal{T}_\mu^\tau(P_{WU}). \quad (37)$$

If successful, it picks one of these indices arbitrarily and sends the codeword $x^k(m)$ over the channel.

Decoding: For realizations (y^k, v^n) of (Y^k, V^n), the decision center first looks for an index $m \in \{1, ...2^{kR}\}$ so that

$$\left(x^k(m), y^k\right) \in \mathcal{T}_\mu^k(P_{XY}). \quad (38)$$

If this is not successful, it declares $\hat{H} = 1$. Otherwise, it checks whether

$$(w^\tau(m), v^\tau) \in \mathcal{T}_\mu^\tau(P_{WV}). \quad (39)$$

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

Analysis: We only provide a very rough sketch. A precise analysis requires carefully aligning several type considerations and can be obtained by following the steps in [13, Thm 2].

For type-I error, we have the upper bound

$$\begin{aligned}\alpha_n \leq\ &\mathbb{P}[\forall m : (W^\tau(m), U^\tau) \notin \mathcal{T}_\mu^\tau(P_{WU}) | H_0]\\ &+\mathbb{P}[(W^\tau(M), V^\tau) \notin \mathcal{T}_\mu^\tau(P_{WV})\\ &\qquad |\exists m : (W^\tau(m), U^\tau) \in \mathcal{T}_\mu^\tau(P_{WU})\\ &\qquad ,\exists m : \left(X^k(m), Y^k\right) \in \mathcal{T}_\mu^k(P_{XY}), H_0]\\ &+\mathbb{P}[\forall m : \left(X^k(m), Y^k\right) \notin \mathcal{T}_\mu^k(P_{XY}) | H_0]. \quad (40)\end{aligned}$$

First two terms vanishes by similar analysis as in Section IV, while the remaining term also decays to zero by noticing that

$$\tau R < kI(X;Y). \quad (41)$$

For type-II error,

$$\begin{aligned}\beta_n \leq\ &\mathbb{P}\big[(W^\tau(M), V^\tau) \in \mathcal{T}_\mu^\tau(P_{WV})\\ &\qquad |\exists m : (W^\tau(m), U^\tau) \in \mathcal{T}_\mu^\tau(P_{WU})\\ &\qquad ,\exists m : \left(X^k(m), Y^k\right) \in \mathcal{T}_\mu^k(P_{XY}), H_1\big], \quad (42)\end{aligned}$$

which can be treated as (30) to get for n large enough

$$\beta_n \leq 2^{-\tau(I(W;V)-\nu)}. \quad (43)$$

This establishes the exponent

$$-\frac{1}{k}\log\beta_n = \frac{\tau}{k}I(W;V) = \frac{I(W;V)}{R}I(X;Y) - \delta' - \nu, \quad (44)$$

for $\delta' = \delta I(W;V)$. Letting $n \longrightarrow \infty$, then $\delta, \nu \to 0$ and optimizing over the compressor $P_{W|U}$ and over the channel input PMF P_X yields the desired result.

VI. Proof of Theorem 5

Take an ε-achievable second order exponent η. Accordingly, consider a sequence (g_n, f_n) so that (4) and (7) hold. Notice that $\theta^*(\varepsilon) = 0$ here, due to (6).

Fix n. The PMF of (Y^k, V^n) accessible at the decision center is denoted $P_{Y^kV^n}$ under H_0 and $P_{Y^k}P_{V^n}$ under H_1.

The DPI gives

$$d(\alpha_n \| \beta_n) \leq \mathrm{D}(P_{Y^kV^n} \,\|\, P_{Y^k}P_{V^n}), \quad (45)$$

where $d(\cdot \,\|\, \cdot)$ denotes the binary divergence. Expending the left-hand side yields, for $h(\cdot)$ the binary entropy,

$$\overbrace{(1-\alpha_n)\log(1-\alpha_n) + \alpha_n\log(\alpha_n)}^{=-h(\alpha_n)} \; \overbrace{-\alpha_n\log(1-\beta_n)}^{\geq 0}$$
$$-(1-\alpha_n)\log(\beta_n) \leq \mathrm{D}(P_{Y^kV^n} \,\|\, P_{Y^k}P_{V^n}). \quad (46)$$

Therefore,

$$-(1-\alpha_n)\log(\beta_n) \leq \mathrm{D}(P_{Y^kV^n} \,\|\, P_{Y^k}P_{V^n}) + h(\alpha_n) \quad (47.1)$$
$$\leq \mathrm{D}(P_{Y^kV^n} \,\|\, P_{Y^k}P_{V^n}) + 1 \quad (47.2)$$
$$= \mathrm{I}\big(Y^k; V^n\big) + 1. \quad (47.3)$$

Now, we use the strong DPI on $Y^k - U^n - V^n$, stated as

$$\mathrm{I}\big(Y^k; V^n\big) \leq \eta(P_{U^n}, P_{V^n|U^n})\,\mathrm{I}\big(Y^k; U^n\big), \quad (48)$$

to obtain

$$-(1-\alpha_n)\log(\beta_n) \leq \eta(P_{U^n}, P_{V^n|U^n})\,\mathrm{I}\big(Y^k; U^n\big) + 1 \quad (49.1)$$
$$= \eta(P_U, P_{V|U})\,\mathrm{I}\big(Y^k; U^n\big) + 1, \quad (49.2)$$

where (49.2) holds by the tensorization of the contraction coefficient [19, Prop. 33.11]. We further have by DPI

$$\mathrm{I}\big(Y^k; U^n\big) \leq \mathrm{I}\big(Y^k; X^k\big) \leq k \cdot C. \quad (50)$$

Combining (49.2) with (50),

$$-\frac{1}{k}\log(\beta_n) \leq \frac{C \cdot \eta(P_U, P_{V|U})}{1-\alpha_n} + \frac{1}{(1-\alpha_n)k}. \quad (51)$$

And finally taking the limit on $n \longrightarrow \infty$,

$$\eta \leq \frac{C \cdot \eta(P_U, P_{V|U})}{1-\varepsilon}. \quad (52)$$

Since (52) is valid for any ε-achievable η, we obtain

$$\eta^*(\varepsilon) \leq \frac{C \cdot \eta(P_U, P_{V|U})}{1-\varepsilon}, \quad (53)$$

and the desired converse result for $\varepsilon = 0$.

VII. Proof of Theorem 3

Take an ε-achievable second order exponent η and a sequence (g_n, f_n) so that (4) and (7) hold. Fix n, let P_{MV^n} and $Q_MQ_{V^n}$ be the PMFs of (M, V^n) under H_0 and H_1 respectively.

Mimicking the steps of Section VI up to (47.2) for a noiseless channel, meaning take $Y^k = M$,

$$-(1-\alpha_n)\log(\beta_n) \leq \mathrm{D}(P_{MV^n} \,\|\, Q_MQ_{V^n}) + 1. \quad (54)$$

By multiplying and dividing by $P_M(.)P_{V^n}(.)$ in the logarithm,

$$\mathrm{D}(P_{MV^n} \,\|\, Q_MQ_{V^n}) = \mathrm{D}(P_{MV^n} \,\|\, P_MP_{V^n}) + \mathrm{D}(P_M \,\|\, Q_M) + \mathrm{D}(P_{V^n} \,\|\, Q_{V^n}), \quad (55)$$

and by DPI

$$\mathrm{D}(P_M \,\|\, Q_M) \leq \mathrm{D}(P_{U^n} \,\|\, Q_{U^n}). \quad (56)$$

Based on (55) and (56), it holds that

$$\mathrm{D}(P_{MV^n} \,\|\, Q_MQ_{V^n})$$
$$\leq \mathrm{D}(P_{MV^n} \,\|\, P_MP_{V^n}) + \mathrm{D}(P_{U^n} \,\|\, Q_{U^n}) \quad (57.1)$$
$$+ \mathrm{D}(P_{V^n} \,\|\, Q_{V^n}) \quad (57.2)$$
$$= \mathrm{I}(M; V^n) + n \cdot (\mathrm{D}(P_U \,\|\, Q_U) + \mathrm{D}(P_V \,\|\, Q_V)) \quad (57.3)$$
$$= \mathrm{I}(M; V^n) + n \cdot \theta^*(\varepsilon). \quad (57.4)$$

Then, applying the strong DPI to (57.4) as in (49),

$$\mathrm{D}(P_{MV^n} \,\|\, Q_MQ_{V^n})$$
$$\leq \eta(P_U, P_{V|U})\,\mathrm{I}(M; U^n) + n \cdot \theta^*(\varepsilon) \quad (58.1)$$
$$\leq \eta(P_U, P_{V|U}) \cdot (k+1) + n \cdot \theta^*(\varepsilon), \quad (58.2)$$

where in (58.2) we used $\mathrm{I}(M; U^n) \leq \mathrm{H}(M) \leq k+1$.

Going back to (54) equipped with (58.2), we obtain

$$-(1-\alpha_n)\log(\beta_n)$$
$$\leq \eta(P_U, P_{V|U}) \cdot (k+1) + n \cdot \theta^*(\varepsilon) + 1. \quad (59)$$

Thus,

$$-\log(\beta_n) - n \cdot \theta^*(\varepsilon) \leq \frac{\eta(P_U, P_{V|U}) \cdot (k+1)}{1-\alpha_n} + n \cdot \frac{\alpha_n \cdot \theta^*(\varepsilon)}{1-\alpha_n} + \frac{1}{1-\alpha_n}, \quad (60)$$

and

$$\frac{1}{k}(-\log(\beta_n) - n \cdot \theta^*(\varepsilon)) \leq \frac{\eta(P_U, P_{V|U}) \cdot (k+1)}{k(1-\alpha_n)} + \frac{n}{k} \cdot \frac{\alpha_n}{1-\alpha_n} \cdot \theta^*(\varepsilon) + \frac{1}{k(1-\alpha_n)}. \quad (61)$$

Now, under the assumption

$$\limsup_{n \longrightarrow \infty}\left(\alpha_n \cdot \frac{n}{k}\right) = 0, \quad (62)$$

which notably implies that $\varepsilon = 0$, letting $n \longrightarrow \infty$ gives

$$\eta \leq \eta(P_U, P_{V|U}). \quad (63)$$

Finally, as claimed we get under (62)

$$\eta^*(0) \leq \eta(P_U, P_{V|U}). \quad (64)$$

To prove Theorem 2, note that in this scenario $\theta^*(\varepsilon) = 0$. Hence, we can directly obtain from (61) that

$$\eta^*(\varepsilon) \leq \frac{\eta(P_U, P_{V|U})}{1-\varepsilon}, \quad (65)$$

yielding the desired result when specifying for $\varepsilon = 0$.

Acknowledgements

The authors acknowledge funding from the ERC under Grant Agreement 101125691.

References

[1] I. Csiszár and R. Ahlswede, "Hypothesis testing with communication constraints," *IEEE Trans. Inf. Theory*, vol. IT-32, no. 4, pp. 533–542, 1986.

[2] M. S. Rahman and A. Wagner, "On the optimality of binning for distributed hypothesis testing," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 6282–6303, 2012.

[3] T. S. Han, "Hypothesis testing with multiterminal data compression," *IEEE Trans. Inf. Theory*, vol. IT-36, no. 6, pp. 759–772, 1987.

[4] H. Shimokawa, T. S. Han, and S.-i. Amari, "Error bound of hypothesis testing with data compression," in *Int. Symp. Inf. Theory (ISIT)*, 1994, p. 114–.

[5] Y. Kochman and L. Wang, "Improved random-binning exponent for distributed hypothesis testing," *IEEE Trans. Inf. Theory*, vol. 71, no. 11, pp. 8217–8222, 2025.

[6] S. Watanabe, "On sub-optimality of random binning for distributed hypothesis testing," in *IEEE Int. Symp. Inf. Theory (ISIT)*, IEEE, 2022, pp. 2708–2713.

[7] S. Salehkalaibar, M. Wigger, and R. Timo, "On hypothesis testing against conditional independence with multiple decision centers," *IEEE Trans. Com.*, vol. 66, no. 6, pp. 2409–2420, 2018.

[8] S. Salehkalaibar, M. Wigger, and L. Wang, "Hypothesis testing over the two-hop relay network," *IEEE Trans. Inf. Theory*, vol. 65, no. 7, pp. 4411–4433, 2019.

[9] D. Cao, L. Zhou, and V. Tan, "Strong converse for hypothesis testing against independence over a two-hop network," in *IEEE Int. Symp. Inf. Theory (ISIT)*, IEEE, 2019, pp. 1994–1998.

[10] Y. Xiang and Y. Kim, "Interactive hypothesis testing against independence," in *IEEE Int. Symp. Inf. Theory (ISIT)*, IEEE, 2013, pp. 2840–2844.

[11] A. Gilani, B. Amor, and S. Salehkalaibar, "Distributed hypothesis testing with privacy constraints," *Entropy*, vol. 21, pp. 1–27, 2019.

[12] S. Sreekumar and D. Gündüz, "Distributed hypothesis testing over noisy channels," in *IEEE Int. Symp. Inf. Theory (ISIT)*, IEEE, 2017, pp. 983–987.

[13] S. Salehkalaibar and M. Wigger, "Distributed hypothesis testing over a noisy channel," in *International Zurich Seminar on Communications*, 2018, pp. 943–947.

[14] S. Salehkalaibar and M. Wigger, "Distributed hypothesis testing based on unequal-error protection codes," *IEEE Trans. Inf. Theory*, vol. 66, no. 7, pp. 4150–4182, 2020.

[15] N. Weinberger and Y. Kochman, "Exponent trade-off for hypothesis testing over noisy channels," in *IEEE Int. Symp. Inf. Theory (ISIT)*, IEEE, 2019, pp. 1852–1856.

[16] S. Skreekumar and D. Gündüz, "Strong converse for testing against independence over a noisy channel," in *IEEE Int. Symp. Inf. Theory (ISIT)*, 2020, pp. 1283–1288.

[17] H. Shalaby and A. Papamarcou, "Multiterminal detection with zero-rate data compression," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 759–772, 1992.

[18] A. Bounhar, M. Sarkiss, and M. Wigger, "A dichotomy for distributed detection with limited communication," in *IEEE Inf. Theory Work.*, 2025.

[19] Y. Polyanskiy and Y. Wu, *Information theory: From coding to learning*. Cambridge University Press, 2025.

[20] I. Csiszár and J. Körner, *Information theory: Coding theorems for discrete memoryless systems*. Cambridge University Press, 2nd ed., 2011.

[21] Y. Polyanskyi, "Channel coding: Non-asymptotic fundamental limits," Ph.D. Thesis, Princeton University, 2010.

[22] S. M. Moser, *Advanced topics in information theory (lecture notes)*, 6th ed. 2025. [Online]. Available: https://moser-isi.ethz.ch/scripts.html