Title	An Information-Spectrum Approach to Distributed Hypothesis Testing for General Sources
Author(s)	Adamou, Ismaila Salihou; Dupraz, Elsa; Matsumoto, Tad
Citation	Proceedings of 2024 International Zurich Seminar on Information and Communication: 144
Issue Date	2024-03-05
Type	Conference Paper
Text version	author
URL	http://hdl.handle.net/10119/18824
Rights	Copyright (c) 2024 The Authors. Ismaila Salihou Adamou, Elsa Dupraz, and Tad Matsumoto. Proceedings of 2024 International Zurich Seminar on Information and Communication, 144 (2024).
Description	2024 International Zurich Seminar on Information and Communication, March 6–8, 2024, Zurich, Switzerland

An Information-Spectrum Approach to Distributed Hypothesis Testing for General Sources

Ismaila Salihou Adamou[1], Elsa Dupraz[1], and Tad Matsumoto[1,2]
[1] IMT Atlantique, CNRS UMR 6285, Lab-STICC, Brest, France
[2] JAIST and University of Oulu (Emeritus)

***Abstract*—This paper investigates Distributed Hypothesis testing (DHT), in which a source $\mathbf{X}$ is encoded given that side information $\mathbf{Y}$ is available at the decoder only. Based on the received coded data, the receiver aims to decide on the two hypotheses H_0 or H_1 related to the joint distribution of $\mathbf{X}$ and $\mathbf{Y}$. While most existing contributions in the literature on DHT consider i.i.d. assumptions, this paper assumes more generic, non-i.i.d., non-stationary, and non-ergodic sources models. It relies on information-spectrum tools to provide general formulas on the achievable Type-II error exponent under a constraint on the Type-I error. The achievability proof is based on a quantize-and-binning scheme. It is shown that with the quantize-and-binning approach, the error exponent boils down to a trade-off between a binning error and a decision error, as already observed for the i.i.d. sources. The last part of the paper provides error exponents for particular source models, *e.g.*, Gaussian, stationary, and ergodic models.**

I. Introduction

In distributed communication networks, data is gathered from various remote nodes and then sent to a server for further processing. Often, the primary objective of the server is not to reconstruct the data, but instead to make a decision based on the collected data. This type of setup is known as distributed hypothesis testing (DHT), and it was first investigated from an information-theoretic perspective in [1], [2].

In DHT, a source $\mathbf{X}$ is encoded using side information $\mathbf{Y}$ available only to the decoder, as shown in Figure 1. The receiver aims to make a decision between two hypotheses: H_0, where the joint probability distribution of $(\mathbf{X}, \mathbf{Y})$ is $P_{\mathbf{XY}}$, and H_1, where the joint distribution is $P_{\overline{\mathbf{XY}}}$. Hypothesis testing involves two types of errors, called the Type-I error and the Type-II error [3]. The information-theoretic analysis of DHT aims to determine the achievable error exponent for the Type-II error while keeping the Type-I error below a fixed threshold [1], [2].

Previous contributions on DHT typically assume that the sources $\mathbf{X}$ and $\mathbf{Y}$ generate independent and identically distributed (i.i.d.) pairs of symbols (X_t, Y_t) [4]–[8]. For example, [7] and [8] provide the error exponent achieved by a quantize-and-binning scheme for i.i.d. sources. Some more complex source models have been investigated in [9], [10], which assume that the sources $\mathbf{X}$ and $\mathbf{Y}$ generate pairs of Gaussian vectors $(\mathbf{X}_t^M, \mathbf{Y}_t^N)$ with auto-correlations in each vector $\mathbf{X}_t^M$ and $\mathbf{Y}_t^N$, as well as cross-correlation between them. However, the models of [9], [10] are block-i.i.d. in the sense that the successive pairs $(\mathbf{X}_t^M, \mathbf{Y}_t^N)$ are assumed to be i.i.d. with t.

Nevertheless, i.i.d. and block-i.i.d. models are often inadequate for capturing the statistics of signals like time series or videos, which cannot be decomposed into fixed-length independent blocks and are frequently non-stationary and/or non-ergodic. As a result, the objective of this paper is to consider a more general source model that is non-i.i.d. and can account for non-stationary and non-ergodic signals, while still encompassing the previous models as particular instances. To investigate DHT under these conditions, we utilize information spectrum tools, which were first introduced in [11] and generally provide information theory results that are applicable to a broad range of source models. It should be noted that information spectrum has been previously used for hypothesis testing in [12], but only for the encoding of a source $\mathbf{X}$ alone, without the use of side information $\mathbf{Y}$.

In this paper, we investigate DHT using general source models for $\mathbf{X}$ and $\mathbf{Y}$ and provide an achievability scheme that yields a general expression for the Type-II error exponent. Our approach to the achievability scheme builds upon the quantize-and-binning techniques presented in [8], while taking into account the use of side information for more complex source models. As in [8], the resulting error exponent consists of two terms: one for the binning error and the other for the decision error. We then specialize our error-exponent to source models of interest, including (i) i.i.d. sources, for which we recover the error exponent reported in [8]; (ii) non-i.i.d. stationary and ergodic sources in general; and (iii) non-i.i.d. Gaussian stationary and ergodic sources.

The outline of the paper is as follows. Section II describes the general sources model and restates the DHT problem. Section III provides the achievable error exponent for general sources, and Section IV derives the proof. Section V considers some examples of source models.

II. Problem statement

In what follows, $[\![1, n]\!]$ denotes the set of integers between 1 and n. We also use upper-case letters, *e.g*, X, to denote random variables (RVs) and lower-case letters, *e.g*, x, to denote their realizations. Random sequences of length n are denoted $\mathbf{X}^n = (X_1, X_2, \cdots, X_n)$.

This work has received a French government support granted to the Cominlabs excellence laboratory and managed by the National Research Agency in the "Investing for the Future" program under reference ANR-10-LABX-07-01.

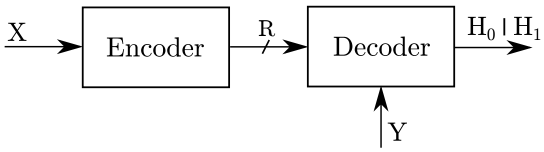

Figure 1. Distributed Hypothesis Testing coding scheme

A. General Sources

In the DHT problem shown in Fig.1, the encoder observes a source sequence $\mathbf{X}$, and the decoder receives a coded version of $\mathbf{X}$ as well as a side information sequence $\mathbf{Y}$, where $\mathbf{X}$ and $\mathbf{Y}$ are correlated. We consider that the sequences $\mathbf{X}$ and $\mathbf{Y}$ are produced from two general sources which are not necessarily i.i.d., and not even stationary or ergodic. As in [12], we define general sources $\mathbf{X}$ and $\mathbf{Y}$ as two infinite sequences :

$$\mathbf{X} = \{\mathbf{X}^n = (X_1, X_2, \cdots, X_n)\}_{n=1}^{\infty}, \quad \mathbf{Y} = \{\mathbf{Y}^n = (Y_1, Y_2, \cdots, Y_n)\}_{n=1}^{\infty} \tag{1}$$

of n-dimensional random variables $\mathbf{X}^n, \mathbf{Y}^n$, respectively. Each component random variable $X_i, Y_i,\ i \in [\![1,n]\!]$, takes values in a finite source alphabet $\mathcal{X}, \mathcal{Y}$, respectively. Next, $P_{\mathbf{X}^n}$ is the probability distribution of the length-n vector $\mathbf{X}^n$, and $P_{\mathbf{X}} = \{P_{\mathbf{X}^n}\}_{n=1}^{\infty}$ is the collection of all probability distributions $P_{\mathbf{X}^n}$. The same holds for the source $\mathbf{Y}$.

We now describe two particular cases of (1). The first one consists of a scalar i.i.d. model in which the sequences $\mathbf{X}^n$ and $\mathbf{Y}^n$ come from two i.i.d. sources, *i.e.*, the successive pairs of symbols (X_n, Y_n) are independent and distributed according to the same joint distribution P_{XY}. This model was considered for DHT in [7], [8]. The second case still relies on an i.i.d. model but for source vectors. In this case, the source sequences $\mathbf{X}^n$ and $\mathbf{Y}^n$ are defined as

$$\mathbf{X}^n = \left\{\mathbf{X}_t^M\right\}_{t=1}^n,\ \mathbf{Y}^n = \left\{\mathbf{Y}_t^M\right\}_{t=1}^n, \tag{2}$$

where $\{\mathbf{X}_t^M\}_{t=1}^n$ and $\{\mathbf{Y}_t^M\}_{t=1}^n$ are sequences of i.i.d. M-dimensional random vectors and the successive pairs $\left(\mathbf{X}_t^M, \mathbf{Y}_t^M\right)$ are distributed according to the same joint distribution $P_{\mathbf{X}^M\mathbf{Y}^M}$. The i.i.d. property of the successive M-length vectors simplifies the DHT analysis by allowing for an orthogonal transform to be applied onto the successive independent blocks $\mathbf{X}_t^M$ and $\mathbf{Y}_t^M$ [9], [10]. Our model described in (1) is more general since it considers infinite sequences without the i.i.d. assumption.

B. Distributed Hypothesis Testing

In what follows, we consider that the joint distribution of the sequence pair $\{(\mathbf{X}^n, \mathbf{Y}^n)\}_{n=1}^{\infty}$ depends on the underlying hypotheses H_0 and H_1 defined as

$$H_0 : (\mathbf{X}^n, \mathbf{Y}^n) \sim P_{\mathbf{X}^n\mathbf{Y}^n}, \tag{3}$$

$$H_1 : (\mathbf{X}^n, \mathbf{Y}^n) \sim P_{\overline{\mathbf{X}}^n\overline{\mathbf{Y}}^n}. \tag{4}$$

where the marginal probability distributions $P_{\mathbf{X}^n}$ and $P_{\mathbf{Y}^n}$ do not depend on the hypothesis.

We consider the following usual coding scheme defined in the literature on DHT [1], [8].

Definition 1: The encoding function $f^{(n)}$ and decoding function $g^{(n)}$ are defined as

$$f^{(n)} : \mathcal{X}^n \longrightarrow \mathcal{M}_n = [\![1, \mathrm{M}_2]\!], \tag{5}$$

$$g^{(n)} : \mathcal{M}_n \times \mathcal{Y}^n \longrightarrow \mathcal{H} = \{H_0, H_1\}, \tag{6}$$

such that $\limsup_{n\to\infty} \frac{1}{n}\log \mathrm{M}_2 \le \mathrm{R}$, where R is the rate and M_2 is the cardinality of the alphabet set $\mathcal{M}_n$.

Definition 2: The Type-I and Type-II error probabilities α_n and β_n are defined as

$$\alpha_n = \mathbb{P}\left[g^{(n)}\left(f^{(n)}(\mathbf{X}^n), \mathbf{Y}^n\right) = H_1 \mid H_0 \text{ is true}\right], \tag{7}$$

$$\beta_n = \mathbb{P}\left[g^{(n)}\left(f^{(n)}(\mathbf{X}^n), \mathbf{Y}^n\right) = H_0 \mid H_1 \text{ is true}\right]. \tag{8}$$

Definition 3: For given The Type-II error exponent θ is said to be achievable for a given rate R, if for large blocklength n, there exists encoding and decoding functions $\left(f^{(n)}, g^{(n)}\right)$ such that the Type-I and Type-II error probabilities α_n and β_n satisfy

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

and

$$\limsup_{n\to\infty} \frac{1}{n}\log\frac{1}{\beta_n} \ge \theta \tag{10}$$

for any $\epsilon > 0$.
In the following, we aim to determine the achievable Type-II error exponent θ for general sources.

III. Main result: Error exponent

A. Definitions

We first provide some definitions which will be useful to express our main result. The $\limsup$ and $\liminf$ in probability of a sequence $\{Z_n\}_{n=1}^{\infty}$ are, respectively, defined as [11]

$$\mathrm{p} - \limsup_{n\to\infty} Z_n = \inf\left\{\alpha \mid \lim_{n\to+\infty} \mathbb{P}(Z_n > \alpha) = 0\right\}, \tag{11}$$

$$\mathrm{p} - \liminf_{n\to\infty} Z_n = \sup\left\{\alpha \mid \lim_{n\to+\infty} \mathbb{P}(Z_n < \alpha) = 0\right\}. \tag{12}$$

The spectral sup-mutual information $\overline{I}(\mathbf{X};\mathbf{U})$, the spectral inf-mutual information $\underline{I}(\mathbf{U};\mathbf{Y})$, the spectral inf-divergence rate $\underline{D}\left(P_{\mathbf{UY}} \| P_{\overline{\mathbf{UY}}}\right)$, and the spectral sup-divergence rate $\overline{D}\left(P_{\mathbf{UY}} \| P_{\overline{\mathbf{UY}}}\right)$ are, respectively, defined as [11]

$$\overline{I}(\mathbf{X};\mathbf{U}) = \mathrm{p} - \limsup_{n\to\infty} \frac{1}{n}\log\frac{P_{\mathbf{U}^n|\mathbf{X}^n}(\mathbf{U}^n \mid \mathbf{X}^n)}{P_{\mathbf{U}^n}(\mathbf{U}^n)}, \tag{13}$$

$$\underline{I}(\mathbf{U};\mathbf{Y}) = \mathrm{p} - \liminf_{n\to\infty} \frac{1}{n}\log\frac{P_{\mathbf{U}^n|\mathbf{Y}^n}(\mathbf{U}^n \mid \mathbf{Y}^n)}{P_{\mathbf{U}^n}(\mathbf{U}^n)}, \tag{14}$$

$$\underline{D}\left(P_{\mathbf{UY}} \| P_{\overline{\mathbf{UY}}}\right) = \mathrm{p} - \liminf_{n\to\infty} \frac{1}{n}\log\frac{P_{\mathbf{U}^n\mathbf{Y}^n}(\mathbf{U}^n, \mathbf{Y}^n)}{P_{\overline{\mathbf{U}}^n\overline{\mathbf{Y}}^n}(\mathbf{U}^n, \mathbf{Y}^n)}, \tag{15}$$

$$\overline{D}\left(P_{\mathbf{UY}} \| P_{\overline{\mathbf{UY}}}\right) = \mathrm{p} - \limsup_{n\to\infty} \frac{1}{n}\log\frac{P_{\mathbf{U}^n\mathbf{Y}^n}(\mathbf{U}^n, \mathbf{Y}^n)}{P_{\overline{\mathbf{U}}^n\overline{\mathbf{Y}}^n}(\mathbf{U}^n, \mathbf{Y}^n)}. \tag{16}$$

B. Achievable error-exponent for general sources

Theorem 1: For the coding scheme of Definition 1, the following error exponent θ is achievable for general sources defined by (1):

$$\theta \leq \min \Big\{ r - \left(\overline{I}(\mathbf{X};\mathbf{U}) - \underline{I}(\mathbf{U};\mathbf{Y}) \right), \underline{D}\left(P_{\mathbf{UY}} \| P_{\overline{\mathbf{UY}}} \right) + \left(\underline{I}(\mathbf{X};\mathbf{U}) - \overline{I}(\mathbf{X};\mathbf{U}) \right) \Big\}, \quad (17)$$

where $\mathbf{U}$ is an auxiliary random variable with same conditional distribution $P_{\mathbf{U}|\mathbf{X}} = P_{\overline{\mathbf{U}}|\overline{\mathbf{X}}}$ under H_0 and H_1 and such that the Markov chain $\mathbf{U} \to \mathbf{X} \to \mathbf{Y}$ is satisfied under both H_0 and H_1. In addition, $P_{\mathbf{UY}}$, and $P_{\overline{\mathbf{UY}}}$ are the joint distributions of $(\mathbf{U}^n, \mathbf{Y}^n)$ under H_0 and H_1, respectively, and $r \leq \mathrm{R}$.

As expected, we find that our error exponent is consistent with that shown in [8] for the i.i.d. case. The error exponent (17) is the result of a trade-off between the binning error and the decision error, as in the i.i.d. case [7], [8]. The binning strategy introduces a new type of error event that does not appear in the DHT scheme without binning for general sources of [12]. In addition, the decision error, *e.g.*, the second term in (17), not only contains a divergence term that appears in [7], [8] and related works, but also the difference $\underline{I}(\mathbf{X};\mathbf{U}) - \overline{I}(\mathbf{X};\mathbf{U})$ between the spectral inf-mutual information and the spectral sup-mutual information of $\mathbf{X}$ and $\mathbf{U}$. Especially, if the term $\frac{1}{n} \log \frac{P_{\mathbf{U}^n|\mathbf{X}^n}(\mathbf{U}^n|\mathbf{X}^n)}{P_{\mathbf{U}^n}(\mathbf{U}^n)}$ does not converge in probability, then the two mutual information terms differ, inducing a penalty in the error exponent. For stationary and ergodic sources, this term converges and there is no such penalty.

IV. Proof of Theorem 1

We first restate the following lemma from [13], which will be useful in the proof.

Lemma 1 ([13]): Let $\mathbf{Z}^n, \mathbf{X}^n$, $\mathbf{U}^n$, be random sequences which take values in finite sets $\mathcal{Z}^n, \mathcal{X}^n, \mathcal{U}^n$, respectively, and satisfy the Markov condition $\mathbf{U}^n \to \mathbf{X}^n \to \mathbf{Z}^n$. Let $\{\Psi_n\}_{n=1}^{\infty}$ be a sequence of mappings such that $\Psi_n : \mathcal{Z}^n \times \mathcal{U}^n \to \{0,1\}$, and

$$\lim_{n\to\infty} \mathbb{P}\left(\Psi_n(\mathbf{Z}^n, \mathbf{U}^n) = 1\right) = 0. \quad (18)$$

Then, $\forall \varepsilon > 0$, there exists a sequence $\{f_n\}_{n=1}^{\infty}$ of mappings $f_n : \mathcal{X}^n \to \{\mathbf{u}_i^n\}_{i=1}^{\mathrm{M}} \subset \mathcal{U}^n$ such that $\mathrm{M} = \lceil e^{n(\overline{I}(\mathbf{U};\mathbf{X})+\varepsilon)} \rceil$ and

$$\lim_{n\to\infty} \mathbb{P}\left(\Psi_n(\mathbf{Z}^n, f_n(\mathbf{X}^n)) = 1\right) = 0. \quad (19)$$

A. Coding scheme

Random codebook generation: Generate $\mathrm{M}_1 = e^{n\overline{r}_0}$ sequences $\mathbf{u}_i^n$ randomly according to a fixed distribution $P_{\mathbf{U}^n|\mathbf{X}^n}$. Assign randomly each $\mathbf{u}_i^n$ to one of $\mathrm{M}_2 = e^{nr}$ bins according to a uniform distribution over $[\![1, \mathrm{M}_2]\!]$. Let $\boldsymbol{B}(\mathbf{u}_i^n) \in [\![1, \mathrm{M}_2]\!]$ denote the index of the bin to which $\mathbf{u}_i^n$ belongs to.

Encoder : Given the sequence $\mathbf{x}^n$, the encoder uses a pre-defined mapping $f_n : \mathcal{X}^n \to \{\mathbf{u}_i^n\}_{i=1}^{\mathrm{M}_1}$ to output a certain sequence $\mathbf{u}_i^n = f_n(\mathbf{x}^n)$ and checks if the condition $(\mathbf{x}^n, \mathbf{u_i^n}) \in T_n^{(1)}$ is satisfied, where

$$T_n^{(1)} = \left\{ (\mathbf{x}^n, \mathbf{u}^n) \text{ s.t. } \underline{r}_0 - \epsilon < \frac{1}{n} \log \frac{P_{\mathbf{U}^n|\mathbf{X}^n}(\mathbf{u}^n \mid \mathbf{x}^n)}{P_{\mathbf{U}^n}(\mathbf{u}^n)} < \overline{r}_0 + \epsilon \right\} \quad (20)$$

where $\underline{r}_0, \overline{r}_0 \in \mathbb{R}$. If such a sequence is found, the encoder sends the bin index $\boldsymbol{B}(\mathbf{u}_i^n)$. Otherwise, it sends an error message.

Decoder : The decoder first looks for a sequence in the bin according to the joint distribution $P_{\mathbf{U}^n\mathbf{Y}^n}$ under H_0. Given the received bin index and the side information $\mathbf{y}^n$, going over the sequences $\mathbf{u}^n$ in the bin one by one, the decoder checks whether $(\mathbf{y}^n, \mathbf{u}^n) \in T_n^{(2)}$ with

$$T_n^{(2)} = \left\{ (\mathbf{y}^n, \mathbf{u}^n) \text{ s.t. } \frac{1}{n} \log \frac{P_{\mathbf{U}^n|\mathbf{Y}^n}(\mathbf{u}^n \mid \mathbf{y}^n)}{P_{\mathbf{U}^n}(\mathbf{u}^n)} > r' - \epsilon \right\}, \quad (21)$$

with $r' \in \mathbb{R}$. The decoder declares H_1 if no such sequence is found in the bin or if it receives an error message from the encoder. Otherwise, it declares H_0 if the sequence $\mathbf{u}^n$ extracted from the bin belongs to the acceptance region $\mathcal{A}_n$ defined as

$$\mathcal{A}_n = \left\{ (\mathbf{y}^n, \mathbf{u}^n) \text{ s.t. } \frac{1}{n} \log \frac{P_{\mathbf{U}^n\mathbf{Y}^n}(\mathbf{u}^n, \mathbf{y}^n)}{P_{\overline{\mathbf{U}}^n\overline{\mathbf{Y}}^n}(\mathbf{u}^n, \mathbf{y}^n)} > S - \epsilon \right\}, \quad (22)$$

where $S \in \mathbb{R}$ is the decision threshold; if otherwise, it declares H_1. The sets $T_n^{(1)}$, $T_n^{(2)}$, and $\mathcal{A}_n$ can be seen as decision regions depending on threshold values $\overline{r}_0, \overline{r}_0, r'$ and S. Those parameters will be chosen such that $\alpha_n \leq \epsilon$, for any $\epsilon > 0$.

B. Error probability analysis

Type-I error α_n : The error events with which the decoder declares H_1 under H_0 are as follows:

$$E_{11} = \Big\{ \nexists \mathbf{u}^n \text{ s.t. } (\mathbf{X}^n, \mathbf{u}^n) \in T_n^{(1)}, (\mathbf{Y}^n, \mathbf{u}^n) \in T_n^{(2)}, (\mathbf{Y}^n, \mathbf{u}^n) \in \mathcal{A}_n \Big\}, \quad (23)$$

$$E_{12} = \Big\{ \exists \mathbf{u'}^n \neq \mathbf{u}^n \text{ s.t. } \boldsymbol{B}(\mathbf{u'}^n) = \boldsymbol{B}(\mathbf{u}^n), \left(\mathbf{Y}^n, \mathbf{u'}^n\right) \in T_n^{(2)}, \text{ but } (\mathbf{Y}^n, \mathbf{u'}^n) \notin \mathcal{A}_n \Big\}. \quad (24)$$

The first event E_{11} is when there is an error either in the encoding, during debinning, or when taking the decision. The second event E_{12} corresponds to a debinning error, where a wrong sequence is extracted from the bin. By the union-bound, the Type-I error probability α_n can be upper bounded as

$$\alpha_n \leq \mathbb{P}(E_{11}) + \mathbb{P}(E_{12}). \quad (25)$$

Regarding the first error event, for $\underline{r}_0 = \underline{I}(\mathbf{X};\mathbf{U})$, $\overline{r}_0 = \overline{I}(\mathbf{X};\mathbf{U})$, and from the definitions of $\underline{I}(\mathbf{X};\mathbf{U})$ and $\overline{I}(\mathbf{X};\mathbf{U})$ in (13) and (14), we have

$$\lim_{n\to\infty} \mathbb{P}\left((\mathbf{X}^n, \mathbf{U}^n) \notin T_n^{(1)} \right) = 0.$$

In addition, according to the definition of $\underline{I}(\mathbf{Y};\mathbf{U})$ in (14), and setting $r^{'} = \underline{I}(\mathbf{Y};\mathbf{U})$, we also have

$$\lim_{n\to\infty} \mathbb{P}\left((\mathbf{Y}^n, \mathbf{U}^n) \notin T_n^{(2)} \right) = 0. \tag{26}$$

Finally, when $S = \underline{D}\left(P_{\mathbf{UY}} \| P_{\overline{\mathbf{UY}}}\right)$ and from the definition of $\underline{D}\left(P_{\mathbf{UY}} \| P_{\overline{\mathbf{UY}}}\right)$, we have

$$\lim_{n\to\infty} \mathbb{P}\left((\mathbf{Y}^n, \mathbf{U}^n) \notin \mathcal{A}_n \right) = 0.$$

Thus, by defining

$$\Psi_n(\mathbf{x}^n, \mathbf{y}^n, \mathbf{u}^n) = \tag{27}$$
$$\begin{cases} 0, & \text{if } (\mathbf{x}^n, \mathbf{u}^n) \in T_n^{(1)}, (\mathbf{y}^n, \mathbf{u}^n) \in T_n^{(2)} \text{ and} \\ & (\mathbf{y}^n, \mathbf{u}^n) \in \mathcal{A}_n, \\ 1, & \text{otherwise.} \end{cases}$$

we get that $\mathbb{P}(\Psi_n(\mathbf{X}^n, \mathbf{Y}^n, \mathbf{U}^n) = 1) \to 0$ as $n \to \infty$. Then, given that $\mathbf{U}^n \to \mathbf{X}^n \to \mathbf{Y}^n$ forms a Markov chain, applying Lemma 1 allows to show that there exists a sequence of functions f_n such that $\mathbb{P}(E_{11}) \to 0$ as $n \to \infty$.

Then, the error probability $\mathbb{P}(E_{12})$ can be expressed as

$$\begin{aligned} \mathbb{P}(E_{12}) &\le \sum_{\mathbf{y}^n} P_{\mathbf{Y}^n}(\mathbf{y}^n) \sum_{\substack{\mathbf{u'}^n : \mathbf{u'}^n \neq \mathbf{u}^n \\ (\mathbf{y}^n, \mathbf{u'}^n) \in \mathcal{T}_n^{(2)} \cap \overline{\mathcal{A}_n}}} \mathbb{P}\left(\boldsymbol{B}(\mathbf{u'}^n) = \boldsymbol{B}(\mathbf{u}^n) \right) \\ &\le \sum_{\mathbf{y}^n} P_{\mathbf{Y}^n}(\mathbf{y}^n) \sum_{\substack{\mathbf{u'}^n : \mathbf{u'}^n \neq \mathbf{u}^n \\ (\mathbf{y}^n, \mathbf{u'}^n) \in \mathcal{T}_n^{(2)}}} e^{-nr} \end{aligned} \tag{28}$$

From (21), for $\left(\mathbf{y}^n, \mathbf{u'}^n\right) \in T_n^{(2)}$ we get

$$P_{\mathbf{Y}^n}(\mathbf{y}^n) < P_{\mathbf{Y}^n|\mathbf{U}^n}\left(\mathbf{y}^n \mid \mathbf{u'}^n\right) e^{-n(r'-\epsilon)},$$

which allows us to write

$$\begin{aligned} \mathbb{P}(E_{12}) &\le \sum_{\mathbf{u'}^n} \sum_{\mathbf{y}^n : (\mathbf{y}^n, \mathbf{u'}^n) \in \mathcal{T}_n^{(2)}} P_{\mathbf{Y}^n|\mathbf{U}^n}\left(\mathbf{y}^n \mid \mathbf{u'}^n\right) e^{-n(r+r'-\epsilon)} \\ &\le e^{-n(r+r'-\overline{r}_0-\epsilon)} \end{aligned} \tag{29}$$

where $e^{n\overline{r}_0}$ is the number of sequences $\mathbf{u}^n$ in the codebook. Therefore, from the condition $r \ge \overline{r}_0 - r' + \epsilon = \overline{I}(\mathbf{X};\mathbf{U}) - \underline{I}(\mathbf{Y};\mathbf{U}) + \epsilon$, we get that $\mathbb{P}(E_{21}) \to 0$ as $n \to \infty$.

Type-II error β_n : A Type-II error occurs when the decoder declares H_0 although H_1 is the true hypothesis. The corresponding error events are:

$$\begin{aligned} E_{21} &= \Big\{ \exists \tilde{\mathbf{u}}^n \neq \mathbf{u}^n : \boldsymbol{B}(\tilde{\mathbf{u}}^n) = \boldsymbol{B}(\mathbf{u}^n), \left(\overline{\mathbf{Y}}^n, \tilde{\mathbf{u}}^n\right) \in T_n^{(2)}, \\ &\qquad \text{and } \left(\overline{\mathbf{Y}}^n, \tilde{\mathbf{u}}^n\right) \in \mathcal{A}_n \Big\}, \\ E_{22} &= \left\{ (\overline{\mathbf{Y}}^n, \mathbf{u}^n) \in T_n^{(2)}, (\overline{\mathbf{Y}}^n, \mathbf{u}^n) \in \mathcal{A}_n \right\}. \end{aligned} \tag{30}$$

The first event E_{21} is a debinning error and the second event E_{22} is the testing error. By the union bound, we get

$$\beta_n \le \mathbb{P}(E_{21}) + \mathbb{P}(E_{22}). \tag{31}$$

Since the marginal probability distribution $P_{\mathbf{Y}^n}$ does not depend on the hypothesis, the probability $\mathbb{P}(E_{21})$ can be expressed by following the same steps as for $\mathbb{P}(E_{12})$. Given that $\overline{r}_0 = \overline{I}(\mathbf{X};\mathbf{U})$ and $r^{'} = \underline{I}(\mathbf{Y};\mathbf{U})$, we get

$$\mathbb{P}(E_{21}) \le e^{-n\left(r - \left(\overline{I}(\mathbf{X};\mathbf{U}) - \underline{I}(\mathbf{Y};\mathbf{U})\right) - \varepsilon\right)}. \tag{32}$$

Next, the probability $\mathbb{P}(E_{22})$ can be expressed as

$$\begin{aligned} \mathbb{P}(E_{22}) &\le \sum_{(\mathbf{x}^n, \mathbf{y}^n)} P_{\overline{\mathbf{X}}^n \overline{\mathbf{Y}}^n}(\mathbf{x}^n, \mathbf{y}^n) \sum_{\substack{\mathbf{u}^n \in [\![1, \mathrm{M}_1]\!], \\ (\mathbf{x}^n, \mathbf{u}^n) \in \mathcal{T}_n^{(1)}}} \mathbb{P}\left((\mathbf{y}^n, \mathbf{u}^n) \in \mathcal{A}_n \right) \\ &\le e^{n\overline{r}_0} \sum_{(\mathbf{x}^n, \mathbf{y}^n)} P_{\overline{\mathbf{X}}^n \overline{\mathbf{Y}}^n}(\mathbf{x}^n, \mathbf{y}^n) \sum_{\substack{\mathbf{u}^n : \\ (\mathbf{x}^n, \mathbf{u}^n) \in \mathcal{T}_n^{(1)} \\ (\mathbf{y}^n, \mathbf{u}^n) \in \mathcal{A}_n}} P_{\mathbf{U}^n}(\mathbf{u}^n) \end{aligned}$$

Since $(\mathbf{x}^n, \mathbf{u}^n) \in \mathcal{T}_n^{(1)}$,

$$P_{\mathbf{U}^n}(\mathbf{u}^n) < P_{\mathbf{U}^n|\mathbf{X}^n}(\mathbf{u}^n \mid \mathbf{x}^n) e^{-n(\underline{r}_0 - \epsilon)}.$$

In addition, the conditional distributions $P_{\mathbf{U}^n|\mathbf{X}^n}$ and $P_{\overline{\mathbf{U}}^n|\overline{\mathbf{X}}^n}$ are the same, and the Markov chain $\mathbf{U}^n \to \mathbf{X}^n \to \mathbf{Y}^n$ is satisfied. Thus, $P_{\mathbf{U}^n|\mathbf{X}^n} = P_{\overline{\mathbf{U}}^n|\overline{\mathbf{X}}^n, \overline{Y}^n}$, and

$$\mathbb{P}(E_{22}) \le e^{n(\overline{r}_0 - \underline{r_0} + \epsilon)} \sum_{\mathbf{u}^n : (\mathbf{y}^n, \mathbf{u}^n) \in \mathcal{A}_n} P_{\overline{\mathbf{U}}^n \overline{\mathbf{Y}}^n}(\mathbf{u}^n, \mathbf{y}^n). \tag{33}$$

For $(\mathbf{y}^n, \mathbf{u}^n) \in \mathcal{A}_n$, we have

$$P_{\overline{\mathbf{U}}^n \overline{\mathbf{Y}}^n}(\mathbf{u}^n, \mathbf{y}^n) < P_{\mathbf{U}^n \mathbf{Y}^n}(\mathbf{u}^n, \mathbf{y}^n) e^{-n(S-\epsilon)}. \tag{34}$$

Combining this with (33) gives that

$$\mathbb{P}(E_{22}) \le e^{-n(\underline{r}_0 - \overline{r_0} + S - 2\epsilon)} \tag{35}$$

Now, substituting (32) and (35) into (31), with $S = \underline{D}\left(P_{\mathbf{UY}} \| P_{\overline{\mathbf{UY}}}\right)$, the Type-II error is upper-bounded as

$$\begin{aligned} \beta_n \le\, & e^{-n\left(r - \left(\overline{I}(\mathbf{X};\mathbf{U}) - \underline{I}(\mathbf{Y};\mathbf{U})\right) - \epsilon\right)} \\ & + e^{-n\left(\underline{I}(\mathbf{X};\mathbf{U}) - \overline{I}(\mathbf{X};\mathbf{U}) + \underline{D}\left(P_{\mathbf{UY}} \| P_{\overline{\mathbf{UY}}}\right) - 2\epsilon\right)}. \end{aligned} \tag{36}$$

Finally, from the definition of the error exponent θ given by (10), we show that (17) is achievable, which proves Theorem 1.

V. Examples

A. Stationary and ergodic sources

We now apply Theorem 1 to sources which are stationary and ergodic, but not necessarily i.i.d.

Proposition 1: If the sources $\mathbf{X}^n$ and $\mathbf{Y}^n$ are stationary and ergodic under both H_0 and H_1, the error exponent (17) becomes :

$$\begin{aligned} \theta \le \min \Bigg\{ & \lim_{n\to\infty} r - \left[\frac{1}{n} h(\mathbf{U}^n \mid \mathbf{Y}^n) - \frac{1}{n} h(\mathbf{U}^n \mid \mathbf{X}^n) \right], \\ & \lim_{n\to\infty} \frac{1}{n} D\left(P_{\mathbf{U}^n \mathbf{Y}^n} \| P_{\overline{\mathbf{U}}^n \overline{\mathbf{Y}}^n} \right) \Bigg\}. \end{aligned} \tag{37}$$

This proposition is due to the *strong converse property* [11].

B. Stationary and ergodic Gaussian sources

Let $\mathbf{X}$ and $\mathbf{Y}$ be two stationary and ergodic sources distributed according to Gaussian distributions $\mathcal{N}\left(\mu_{\mathbf{X}}, \mathbf{K}_{\mathbf{X}}\right)$ and $\mathcal{N}\left(\mu_{\mathbf{Y}}, \mathbf{K}_{\mathbf{Y}}\right)$, with covariance matrices $\mathbf{K}_{\mathbf{X}}$ and $\mathbf{K}_{\mathbf{Y}}$, respectively. The two hypotheses are formulated as

$$H_0 : \left(\begin{array}{c} \mathbf{X}^n \\ \mathbf{Y}^n \end{array} \right) \sim \mathcal{N}(\mu_{\mathbf{XY}}, \mathbf{K}), \tag{38}$$

$$H_1 : \left(\begin{array}{c} \mathbf{X}^n \\ \mathbf{Y}^n \end{array} \right) \sim \mathcal{N}(\overline{\mu}_{\mathbf{XY}}, \overline{\mathbf{K}}). \tag{39}$$

In the expressions (38) and (39), $\mu_{\mathbf{XY}}$ is defined as a block vector $\left[\mu_{\mathbf{X}}, \mu_{\mathbf{Y}}\right]^T$. In addition, $\mathbf{K}$ and $\overline{\mathbf{K}}$ are the joint covariance matrices of $\mathbf{X}$ and $\mathbf{Y}$ defined as

$$\mathbf{K} = \left[\begin{array}{cc} \mathbf{K}_{\mathbf{X}} & \mathbf{K}_{\mathbf{XY}} \\ \mathbf{K}_{\mathbf{YX}} & \mathbf{K}_{\mathbf{Y}} \end{array} \right], \overline{\mathbf{K}} = \left[\begin{array}{cc} \mathbf{K}_{\mathbf{X}} & \overline{\mathbf{K}}_{\mathbf{XY}} \\ \overline{\mathbf{K}}_{\mathbf{YX}} & \mathbf{K}_{\mathbf{Y}} \end{array} \right], \tag{40}$$

We assume that all the matrices $\mathbf{K}_{\mathbf{X}}$, $\mathbf{K}_{\mathbf{Y}}$, $\overline{\mathbf{K}}_{\mathbf{Y}}$, $\mathbf{K}_{\mathbf{XY}}$, and $\overline{\mathbf{K}}_{\mathbf{XY}}$ are positive-definite. We also denote the conditional covariance matrix of $\mathbf{X}^n$ given $\mathbf{Y}^n$ by

$$\mathbf{K}_{\mathbf{X}|\mathbf{Y}} = \mathbf{K}_{\mathbf{X}} - \mathbf{K}_{\mathbf{XY}} \mathbf{K}_{\mathbf{Y}}^{-1} \mathbf{K}_{\mathbf{XY}}. \tag{41}$$

The eigenvalues of $\mathbf{K}_{\mathbf{X}|\mathbf{Y}}$ are further denoted by $\lambda_i^{(X|Y)}$.

Proposition 2: If the sources $\mathbf{X}$ and $\mathbf{Y}$ are Gaussian, stationary, and ergodic, under both H_0 and H_1, the terms in (37) reduce to

$$\lim_{n\to\infty} \frac{1}{n} h\left(\mathbf{U}^n \mid \mathbf{Y}^n\right) - \lim_{n\to\infty} \frac{1}{n} h\left(\mathbf{U}^n \mid \mathbf{X}^n\right) = \lim_{n\to\infty} \frac{1}{2n} \sum_{i=1}^{n} \log \frac{\lambda_i^{(X|Y)} + \kappa}{\kappa}, \tag{42}$$

and

$$\lim_{n\to\infty} \frac{1}{n} D\left(P_{\mathbf{U}^n\mathbf{Y}^n} \| P_{\overline{\mathbf{U}}^n\overline{\mathbf{Y}}^n}\right\} = \lim_{n\to\infty} \frac{1}{2n} \left[\log \frac{|\overline{\mathbf{\Sigma}}|}{|\mathbf{\Sigma}|} - 2n + \left(\overline{\mu}_{\mathbf{UY}} - \mu_{\mathbf{UY}}\right)^T \overline{\mathbf{\Sigma}}^{-1} \left(\overline{\mu}_{\mathbf{UY}} - \mu_{\mathbf{UY}}\right) + \operatorname{tr}\left\{ \overline{\mathbf{\Sigma}}^{-1} \mathbf{\Sigma} \right\} \right], \tag{43}$$

where $\mathbf{\Sigma}$ and $\overline{\mathbf{\Sigma}}$ are the joint covariance matrices of $\mathbf{U}$ and $\mathbf{Y}$ under H_0 and H_1, respectively.

The terms given by (42) and (43) are obtained by considering that the source $\mathbf{U}$ is Gaussian such that $\mathbf{U} = \mathbf{X} + \mathbf{Z}$, where $\mathbf{Z} \sim \mathcal{N}\left(0, \kappa \mathbf{I}_n\right)$ is independent of $\mathbf{X}$, and $\mathbf{I}_n$ is the identity matrix of dimension $n \times n$. The covariance matrices $\mathbf{\Sigma}$ and $\overline{\mathbf{\Sigma}}$ are then defined as

$$\mathbf{\Sigma} = \left[\begin{array}{cc} \mathbf{K}_{\mathbf{U}} & \mathbf{K}_{\mathbf{UY}} \\ \mathbf{K}_{\mathbf{YU}} & \mathrm{K}_{\mathbf{Y}} \end{array} \right], \overline{\mathbf{\Sigma}} = \left[\begin{array}{cc} \mathbf{K}_{\mathbf{U}} & \overline{\mathbf{K}}_{\mathbf{UY}} \\ \overline{\mathbf{K}}_{\mathbf{YU}} & \mathrm{K}_{\mathbf{Y}} \end{array} \right]. \tag{44}$$

We now consider the case where the pair $(\mathbf{U}, \mathbf{Y})$ has different covariance matrices, $\mathbf{\Sigma}$ under H_0 and $\overline{\mathbf{\Sigma}}$ under H_1. We also assume that all the Gaussian vectors are zero-centered. We then define H_0 and H_1 as

$$H_0 : \left(\begin{array}{c} \mathbf{X}^n \\ \mathbf{Y}^n \end{array} \right) \sim \mathcal{N}(\mathbf{0}, \mathbf{K}), \ H_1 : \left(\begin{array}{c} \mathbf{X}^n \\ \mathbf{Y}^n \end{array} \right) \sim \mathcal{N}(\mathbf{0}, \overline{\mathbf{K}}).$$

In this case, it can be shown that the expression (42) remains the same, while the expression (43) reduces to

$$\lim_{n\to\infty} \frac{1}{n} D\left(P_{\mathbf{U}^n\mathbf{Y}^n} \| P_{\overline{\mathbf{U}}^n\overline{\mathbf{Y}}^n}\right\} = \lim_{n\to\infty} \frac{1}{2n} \left[\log \frac{|\overline{\mathbf{\Sigma}}|}{|\mathbf{\Sigma}|} - 2n + \operatorname{tr}\left\{ \overline{\mathbf{\Sigma}}^{-1} \mathbf{\Sigma} \right\} \right]. \tag{45}$$

The matrices $\mathbf{\Sigma}$ and $\overline{\mathbf{\Sigma}}$ are of length $2n \times 2n$, where n tends to infinity. Therefore, for some specific Gaussian sources, one needs to study the convergence of the determinants $|\mathbf{\Sigma}|$ and $|\overline{\mathbf{\Sigma}}|$, and also of the trace $\operatorname{tr}\left\{ \overline{\mathbf{\Sigma}}^{-1} \mathbf{\Sigma} \right\}$.

VI. Conclusion

We provided an information-spectrum approach to DHT for general non-i.i.d., non-stationary, and non-ergodic sources. The derived error exponent is achieved from a quantize-and-binning scheme, which, we found, boils down to a trade-off between a binning error and a decision error. Future works will focus on comparing our error exponent to state-of-the-art ones obtained very recently for the i.i.d. case [14]. Other future works will include designing practical coding schemes, as well as considering other applications of DHT such as synchronism identification in spread spectrum signal detectors [15].

References

[1] T. S. Han, "Hypothesis Testing with Multiterminal Data Compression," *IEEE Trans. Inf. Theory*, vol. 33, no. 6, pp. 759–772, 1987.

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

[3] E. L. Lehmann, J. P. Romano, and G. Casella, *Testing statistical hypotheses*. Springer, 2005, vol. 3.

[4] G. Katz, P. Piantanida, and M. Debbah, "Distributed Binary Detection with Lossy Data Compression," *IEEE Trans. Inf. Theory*, vol. 63, no. 8, pp. 5207–5227, 2017.

[5] S. Salehkalaibar and M. Wigger, "Distributed hypothesis testing over multi-access channels," in *IEEE Global Communications Conference (Globecom)*, 2018, pp. 1–6.

[6] S. Sreekumar and D. Gunduz, "Distributed Hypothesis Testing over Discrete Memoryless Channels," *IEEE Trans. Inf. Theory*, vol. 66, no. 4, pp. 2044–2066, 2020.

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

[8] G. Katz, P. Piantanida, R. Couillet, and M. Debbah, "On the necessity of binning for the distributed hypothesis testing problem," *IEEE Int. Symp. Inf. Theory - Proc.*, pp. 2797–2801, 2015.

[9] M. S. Rahman and A. B. Wagner, "Vector gaussian hypothesis testing and lossy one-helper problem," *IEEE Int. Symp. Inf. Theory - Proc.*, pp. 968–972, 2009.

[10] P. Escamilla, A. Zaidi, and M. Wigger, "Some Results on the Vector Gaussian Hypothesis Testing Problem," *IEEE Int. Symp. Inf. Theory - Proc.*, no. May, pp. 2421–2425, 2020.

[11] T. S. Han, *Information-Spectrum Methods in Information Theory*, Baifukan, Tokyo, 1998.

[12] ——, "Hypothesis testing with the general source," *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2415–2427, 2000.

[13] K.-i. Iwata and J. Muramatsu, "An information-spectrum approach to rate-distortion function with side information," *IEICE transactions on fundamentals of electronics, communications and computer sciences*, vol. 85, no. 6, pp. 1387–1395, 2002.

[14] Y. Kochman and L. Wang, "Improved random-binning exponent for distributed hypothesis testing," *arXiv preprint arXiv:2306.14499*, 2023.

[15] J. Arribas, C. Fernandez-Prades, and P. Closas, "Antenna array based gnss signal acquisition for interference mitigation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 49, no. 1, pp. 223–243, 2013.