Error-Exponent of Distributed Hypothesis Testing for Gilbert-Elliot Source Models

Ismaila Salihou Adamou[1], Elsa Dupraz[1], Amin Zribi[1,2], and Tad Matsumoto[1,3]
[1] IMT Atlantique, CNRS UMR 6285, Lab-STICC, Brest, France
[2] ICT Department, Higher Institute of Communication Technologies (ISETCom), Ariana, Tunisia
[3] JAIST and University of Oulu (Emeritus)

***Abstract*—This paper considers Distributed Hypothesis Testing (DHT), in which a source X is encoded given that a side information Y is available only to the decoder. The decoder's task is to make a decision between two hypothesis $\mathcal{H}_0$ and $\mathcal{H}_1$ related to the joint probability distribution of X and Y. While most works on DHT have adopted an information-theoretic approach by providing generic error exponents for the decision error, we focus on a specific and practical source model: the Gilbert-Elliot (GE) model. The later is a two-states hidden Markov model which describes time-varying correlation between X and Y. Starting from the generic error exponents, we develop a method for numerically estimating the error exponent for the GE model. We provide numerical results to evaluate the impact of the model parameters on the error exponents and explore the tradeoff between two types of error events, namely the testing error and the binning error.**

I. Introduction

Emerging Internet of Things (IoT) networks rely on multi-sensors architectures, where the destination primary objective is often to make accurate decisions. In order to minimize both latency and energy consumption, the decision-making process can directly use the received data, without need for reconstructing the original information. This scenario, known as Distributed Hypothesis Testing (DHT), was initially investigated in [1], [2], and has gained significant interest in the field of information theory. It falls within the field of goal-oriented communications, a growing area of study currently under investigation from both theoretical and practical perspectives [3].

A widely used DHT model assumes that a source $\mathbf{X}$ is encoded given that a side information $\mathbf{Y}$ is available only to the decoder. The receiver's task is to decide between two hypotheses, $\mathcal{H}_0$ and $\mathcal{H}_1$, related to the joint probability distribution $P(\mathbf{X}, \mathbf{Y})$. As in standard hypothesis testing [4], the objective of DHT is to minimize the Type-II error probability, defined as the probability of deciding on $\mathcal{H}_0$ when $\mathcal{H}_1$ is true, under a constraint on the Type-I error probability, which is the probability of deciding on $\mathcal{H}_1$ when $\mathcal{H}_0$ is true. Consequently, the information-theoretic analysis of DHT provides Type-II error exponents which can be achieved under rate-limited communication links [5], [6]. These error exponents are often expressed as a tradeoff between a binning error and a testing error.

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.

Several variations of this problem have been considered in the literature. For instance, [7] considers a joint setup where the receiver objective encompasses both data reconstruction and hypothesis testing, and identifies a tradeoff between the two objectives. Additionally, [8] analyzes the impact of a noisy point-to-point channel on the error exponent, and investigates joint versus separate source/channel coding in the context of DHT. In these previous works, the sources $\mathbf{X}$ and $\mathbf{Y}$ are assumed to generate independent and identically distributed (i.i.d.) symbol sequences. To alleviate this assumption, [9] derives an achievable error-exponent for general sources, not necessarily stationary or ergodic, by employing information-spectrum tools [10].

The generic information-theoretic error exponents have been specified for Gaussian models in [5], [9], [11] and for binary i.i.d. models in [6], [12]. To progress towards the development of practical DHT coding schemes, it is crucial to investigate a broader range of source models of interest. Therefore, the objective of this work is to study the Gilbert-Elliot (GE) model, which finds applications in many domains such as video coding [13], link quality estimation [14], or packet loss analysis [15], and has not been previously investigated in the context of DHT. The GE model is a non-i.i.d. time-varying binary model which involves a good state (G) and a bad state (B) [16]. The transition between these two states is modeled by a Markov chain, thus accounting for memory in the successive state values. For instance, in sensor networks, state G (respectively state B) represents high correlation (respectively low correlation) between sensors measurements.

In this paper, we aim to analyze the error-exponent of DHT for GE models representing the correlation between $\mathbf{X}$ and $\mathbf{Y}$. We assume that each hypothesis corresponds to a distinct set of parameters for the GE model. To evaluate the error exponent, we build upon our previous work [9], where the error-exponent is expressed for general sources as a tradeoff between a binning error probability and a testing error probability. Since there does not exist closed-form expressions of the error exponents related to each of these two error probabilities, we begin by proposing estimators for these error exponents. Subsequently, we provide an efficient numerical method to evaluate these estimators, using forward recursions proposed for Hidden Markov Models (HMM) in [17]. Although the focus of this paper is specifically on the GE model, the proposed approach is generic and can be easily extended to

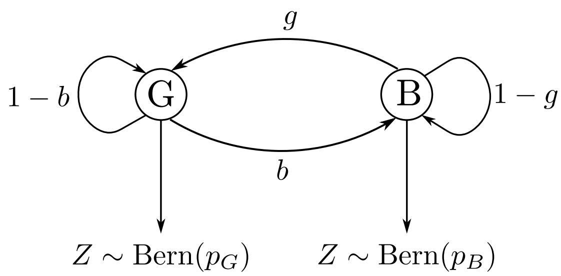

Figure 1: Gilbert-Elliot model for the correlation noise Z under hypothesis $\mathcal{H}_0$. Under hypothesis $\mathcal{H}_1$, the crossover probabilities become $\overline{p}_G$ and $\overline{p}_B$, while the parameters g and b remain the same.

handle more complex source models with memory. To provide valuable insights for the design of practical coding schemes for DHT, we present numerical results that explore: (i) the impact of different GE model parameters on the error-exponent, and (ii) the tradeoff between the binning error and the testing error.

The outline of the paper is as follows. Section II introduces our notation as well as the considered GE model. Section III provides the DHT error-exponent for general sources. Section IV presents a new method to numerically evaluate the terms involved in the error-exponent. Section V shows numerical results.

II. Source models

A. Notation

Throughout the paper, we use $[\![1, M]\!]$ to denote the set of integers between 1 and M. Sources are denoted by bold capital letters, *e.g.*, $\mathbf{X}$ and they generate infinite sequences of binary symbols $\{X_k\}_{k=1}^{+\infty}$ where source symbols $X_k \in \{0,1\}$ are denoted using capital letters, and k is the time index. We use $\mathbf{X}^n$ to denote length-n random source vectors, and we use $\mathbf{x}^n$ to denote their realizations. The limit in probability γ of a sequence of random variables $\{A_n\}_{n=1}^{+\infty}$ is denoted by $\gamma = \mathrm{p}-\lim_{n\to\infty} A_n$ and it verifies

$$\lim_{n\to\infty} P\left(|A_n - \gamma| > \epsilon\right) = 0 \tag{1}$$

for all $\epsilon > 0$.

B. Sources definition

We consider two correlated sources, $\mathbf{X}$ and $\mathbf{Y}$, where $\mathbf{X}$ is the source to be encoded, and $\mathbf{Y}$ is the side information available at the decoder. The binary sequences generated by $\mathbf{X}$ and $\mathbf{Y}$ are denoted as $\{X_k\}_{k=1}^{+\infty}$ and $\{Y_k\}_{k=1}^{+\infty}$, respectively. We assume that the source $\mathbf{X}$ is i.i.d., such that for all $k \geq 1$, X_k follows a Bernoulli distribution $\mathrm{Bern}(p)$ with parameter p, constant across the hypotheses. On the other hand, the source $\mathbf{Y}$ is not i.i.d. and its probability distribution depends on the hypotheses $\mathcal{H}_0$ or $\mathcal{H}_1$, as described below:

$$\mathcal{H}_0 : Y_k = X_k \oplus Z_k \tag{2}$$
$$\mathcal{H}_1 : \overline{Y}_k = X_k \oplus \overline{Z}_k. \tag{3}$$

With a slight abuse of notation, we use $\overline{Y}_k$ to denote the side information symbols generated under $\mathcal{H}_1$, to make it clear that they have a different probability distribution than under $\mathcal{H}_0$. The sources $\mathbf{Z}$ and $\overline{\mathbf{Z}}$ which generate sequences of binary symbols $\{Z_k\}_{k=1}^{+\infty}$ and $\{\overline{Z}_k\}_{k=1}^{+\infty}$, respectively, are independent of $\mathbf{X}$ and follow GE models described below.

Under $\mathcal{H}_0$, the source $\mathbf{Z}$ follows a GE model [16] with hidden state $\mathbf{S}$ depicted in Figure 1. The sequence output from the binary hidden states $\{S_k\}_{k=1}^{+\infty}$ is such that $S_k \in \{G, B\}$, and it follows a Markov model described with the following state transition probabilities:

$$P\left(S_k = G \mid S_{k-1} = B\right) = g, \tag{4}$$
$$P\left(S_k = B \mid S_{k-1} = G\right) = b. \tag{5}$$

Due to the Markov property,

$$P(S_k | S_{k-1}, S_{k-2} \cdots S_1) = P(S_k | S_{k-1}). \tag{6}$$

Each symbol Z_k takes value 0 or 1 depending on the hidden state value $S_k = s$ such that

$$P\left(Z_k = 1 \mid S_k = G\right) = p_G, \tag{7}$$
$$P\left(Z_k = 1 \mid S_k = B\right) = p_B. \tag{8}$$

where p_G and p_B are crossover probabilities. We often consider that $p_G < p_B$, so that the state G corresponds to high correlation between $\mathbf{X}$ and $\mathbf{Y}$ while the state B corresponds to low correlation. In addition, the GE model assumes that

$$P(Z_k | Z_1, \cdots Z_n, S_1, \cdots S_n) = P(Z_k | S_k). \tag{9}$$

Under hypothesis $\mathcal{H}_1$, the source $\overline{\mathbf{Z}}_k$ also follows a GE model, with the same hidden state $\mathbf{S}$ described by equations (4) and (5) and same values g and b as under $\mathcal{H}_0$. However, the crossover probabilities differ from those specified by (7) and (8). They are now denoted by $\overline{p}_G$ and $\overline{p}_B$ with

$$P\left(\overline{Z}_k = 1 \mid S_k = G\right) = \overline{p}_G, \tag{10}$$
$$P\left(\overline{Z}_k = 1 \mid S_k = B\right) = \overline{p}_B. \tag{11}$$

III. Error exponent

In this section, we first introduce the DHT coding scheme [5], [6], and then restate our result of [9] which provides an achievable error exponent for general sources.

A. Coding scheme

Definition 1: The DHT coding scheme is defined from an encoding function f_n and a decoding function g_n defined as

$$f_n : \{0,1\}^n \longrightarrow [\![1, M]\!], \tag{12}$$
$$g_n : [\![1, M]\!] \times \{0,1\}^n \longrightarrow \{\mathcal{H}_0, \mathcal{H}_1\}, \tag{13}$$

under a rate constraint $R \in (0,1)$ such that

$$\limsup_{n\to\infty} \frac{1}{n} \log M \leq R. \tag{14}$$

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

$$\alpha_n = P\left(g_n\left(f_n\left(\mathbf{X}^n\right), \mathbf{Y}^n\right) = \mathcal{H}_1 \mid \mathcal{H}_0\right), \quad (15)$$

$$\beta_n = P\left(g_n\left(f_n\left(\mathbf{X}^n\right), \mathbf{Y}^n\right) = \mathcal{H}_0 \mid \mathcal{H}_1\right). \quad (16)$$

Definition 3: For given $\epsilon > 0$ and $R \in (0,1)$, a Type-II error exponent θ is achievable if for large blocklength n, there exists encoding and decoding functions (f_n, g_n) such that the Type-I and Type-II error probabilities α_n and β_n satisfy

$$\alpha_n \leq \epsilon, \quad (17)$$

$$\limsup_{n\to\infty} \frac{1}{n} \log \frac{1}{\beta_n} \geq \theta. \quad (18)$$

B. Information-spectrum terms for ergodic GE models

In [9], we derived general expressions for the error-exponent which are applicable to a large range of sources, neither necessarily stationary nor ergodic. These expressions rely on information-spectrum terms [10] which are defined from limits inferior and limits superior in probability. These information spectrum terms have simplified expressions in our case, given that the underlying Markov chain is ergodic, where ergodic means that it admits a unique stationary distribution as the sequence length n tends to infinity [18]. The GE model described in Section II is ergodic given that the conditions $0 < g < 1$ and $0 < b < 1$ are satisfied [18].

For ergodic GE models for some sources $\mathbf{U}, \mathbf{Y}, \overline{\mathbf{U}}, \overline{\mathbf{Y}}$, combining the information-spectrum definitions from [10] with the convergence proofs from [18] allows us to show that the spectral conditional entropy has expression

$$H_s(\mathbf{U}|\mathbf{Y}) = p\text{-}\lim_{n\to\infty} \frac{1}{n} \log \frac{1}{P(\mathbf{U}^n|\mathbf{Y}^n)}, \quad (19)$$

and the spectral divergence has expression

$$D_s(P_{\mathbf{U},\mathbf{Y}} || P_{\overline{\mathbf{U}},\overline{\mathbf{Y}}}) = p\text{-}\lim_{n\to\infty} \frac{1}{n} \log \frac{P(\mathbf{U}^n, \mathbf{Y}^n)}{P(\overline{\mathbf{U}}^n, \overline{\mathbf{Y}}^n)}. \quad (20)$$

C. Achievable error-exponent

We now apply the general expression of the error exponent of [9] to ergodic GE models. In this case, the error exponent in [9] reduces to

$$\theta \leq \min\left\{\theta_{\text{bin}}, \theta_{\text{test}}\right\}, \quad (21)$$

where

$$\theta_{\text{bin}} = r - \left[H_s\left(\mathbf{U} \mid \mathbf{Y}\right) - H_s\left(\mathbf{U} \mid \mathbf{X}\right)\right] \quad (22)$$

$$\theta_{\text{test}} = D_s\left(P_{\mathbf{UY}} \| P_{\overline{\mathbf{UY}}}\right) \quad (23)$$

and $r \leq R$.

The proof of [9] is based on an achievability scheme with the following main steps. The codebook is built from 2^{nr_0} sequences $\mathbf{u}^n$ randomly generated according to a pre-defined joint distribution $P_{\mathbf{U}^n|\mathbf{X}^n}$. The sequences $\mathbf{u}^n$ are then assigned uniformly and randomly to 2^{nr} bins with $r < r_0$. The encoder seeks for a sequence $\mathbf{u}^n$ that is typical in a certain sense with $\mathbf{x}^n$, and sends the index of the bin to which the sequence $\mathbf{u}^n$ belongs. The decoder extracts a sequence $\hat{\mathbf{u}}^n$ from the bin which is typical in a certain sense with the side information sequence $\mathbf{y}^n$. It then checks whether the pair $(\hat{\mathbf{u}}^n, \mathbf{y}^n)$ belongs to an acceptance region. If so, it declares $\mathcal{H}_0$. Otherwise, or if there was an error during encoding or decoding, it declares $\mathcal{H}_1$. As a result, two types of errors can occur: the binning error, which is represented by the term θ_{bin} in (22), and the testing error, which is represented by the term θ_{test} in (23). The error exponent in (21) results from a tradeoff between these two terms.

D. Auxiliary source

The previous error exponent depends on an auxiliary source $\mathbf{U}$ which has to satisfy the Markov chain $\mathbf{Y} \to \mathbf{X} \to \mathbf{U}$ [9]. In what follows, we assume that this auxiliary source generates a sequence of symbols $\{U_k\}_{k=1}^{+\infty}$ such that

$$U_k = X_k \oplus \phi_k, \quad (24)$$

where $\phi_k \sim \text{Bern}(\delta)$, and X_k and ϕ_k are independent. The parameter δ is key at it addresses the tradeoff between the binning error and the testing error. Indeed, a small value of δ means that the sequence $\mathbf{u}^n$ selected by the encoder will be close to $\mathbf{x}^n$, which reduces the testing error. But this increases the binning error at the same time, since this increases the value of r_0 that ensures that the encoding error vanishes asymptotically, which also increases the number of sequences in each bin in order to satisfy the rate constraint for a given R .

We choose definition (24) for the auxiliary source $\mathbf{U}$ because due to the Markov chain, $\mathbf{U}$ needs to be expressed from $\mathbf{X}$ which is itself Bernoulli and i.i.d. From this choice of $\mathbf{U}$, the term $H_s(\mathbf{U}|\mathbf{X})$ in (22) can be expressed as

$$H_s(\mathbf{U}|\mathbf{X}) = -\delta \log \delta - (1-\delta)\log(1-\delta), \quad (25)$$

which is the conventional entropy of a Bernoulli source. We leave to future work the investigation of if (24) is the optimal choice for the auxiliary source $\mathbf{U}$. However, since there are no known analytical expressions for the terms $H_s\left(\mathbf{U} \mid \mathbf{Y}\right)$ in (22) and $D_s\left(P_{\mathbf{UY}} | P_{\overline{\mathbf{UY}}}\right)$ in (23), we now describe our numerical evaluation procedure for these terms.

IV. Statistical evaluation of the error exponent

This section provides a statistical method to evaluate the error exponent for the GE model described in Section II.

A. Estimators of spectral information-theory terms

Given that the spectral conditional entropy in (19) and the spectral divergence in (20), are both defined from a limit in probability, the terms

$$\widehat{H}_s\left(\mathbf{U} \mid \mathbf{Y}\right) = \frac{1}{n} \log \frac{1}{P(\mathbf{u}^n|\mathbf{y}^n)}, \quad (26)$$

$$\widehat{D}_s(\mathbf{U}, \mathbf{Y} || P_{\overline{\mathbf{U}},\overline{\mathbf{Y}}}) = \frac{1}{n} \log \frac{P(\mathbf{u}^n, \mathbf{y}^n)}{P(\overline{\mathbf{u}}^n, \overline{\mathbf{y}}^n)}, \quad (27)$$

are consistent estimators [19, Section 1.8] of $H_s\left(\mathbf{U} \mid \mathbf{Y}\right)$ and $D_s(P_{\mathbf{U},\mathbf{Y}} || P_{\overline{\mathbf{U}},\overline{\mathbf{Y}}})$, respectively. To evaluate these estimators, we propose to use a large value of n, and to randomly generate samples $(\mathbf{x}^n, \mathbf{y}^n, \mathbf{u}^n)$ according to the Gilbert-Elliot model described in Section II, and to the definition of $\mathbf{U}$ in (24). We then calculate the probability terms $P(\mathbf{u}^n|\mathbf{y}^n)$, $P(\mathbf{u}^n)$, $P(\mathbf{u}^n, \mathbf{y}^n)$, and, $P(\overline{\mathbf{u}}^n, \overline{\mathbf{y}}^n)$ involved in (26) or (27), which allows evaluating the previous estimators, and then the error exponent (21).

A similar methodology was employed in [20] to numerically evaluate the capacity of a Gilbert-Elliot channel, from an estimator defined as the log-probability of certain random vectors. In [20], the estimator was defined from the notion of information-rate, while here, we rely on the definition of information spectrum terms. Moreover, since the capacity of a channel only involves computing the mutual information between the channel input and output, the probability computation in [20] is simpler. In this paper, we evaluate the probability terms involved in (26) and (27) from forward recursions which we now describe.

B. Forward recursions

In this section, we describe the computation of all the probabilities involved in the estimators (26) and (27). We first calculate the joint probability $p(\mathbf{u}^n, \mathbf{y}^n)$ by evaluating for $s \in \{G, B\}$, $\alpha_n^{(u,y)}(s) = P(\mathbf{u}^n, \mathbf{y}^n, S_n = s)$. This probability can be computed efficiently from the HMM forward recursion described in [17]. This recursion is initialized for $s \in \{G, B\}$ as

$$\alpha_1^{(u,y)}(s) = P(S_1 = s)\, P(y_1 \mid S_1 = s)\, P(u_1 \mid y_1, S_1 = s),$$

and then defined for all $k \in [\![2, n-1]\!]$ as

$$\alpha_{k+1}^{(u,y)}(s) = \left[\sum_{s' \in \{G,B\}} \alpha_k^{(u,y)}(s') P(S_{k+1} = s \mid S_k = s') \right] \dots$$
$$P(y_{k+1} \mid S_{k+1} = s)\, P(u_{k+1} \mid y_{k+1}, S_{k+1} = s).$$

We them compute

$$p(\mathbf{u}^n, \mathbf{y}^n) = \sum_{s \in \{G,B\}} \alpha_n^{(u,y)}(s). \tag{28}$$

The marginal probability $p(\mathbf{y}^n)$ can also be evaluated from a forward recursion initialized as

$$\alpha_1^{(y)}(s) = P(S_1 = s)\, P(y_1 \mid S_1 = s),$$

and defined for all $k \in [\![2, n-1]\!]$ as

$$\alpha_{k+1}^{(y)}(s) = \left[\sum_{s' \in \{G,B\}} \alpha_k^{(y)}(s') P(S_{k+1} = s \mid S_k = s') \right] \dots$$
$$P(y_{k+1} \mid S_{k+1} = s).$$

Then,

$$p(\mathbf{y}^n) = \sum_{s \in \{G,B\}} \alpha_n^{(y)}(s). \tag{29}$$

This allows us to calculate the conditional probability $p(\mathbf{u}^n \mid \mathbf{y}^n)$ using the formula

$$p(\mathbf{u}^n \mid \mathbf{y}^n) = \frac{p(\mathbf{u}^n, \mathbf{y}^n)}{p(\mathbf{y}^n)} = \frac{\sum_{s \in \{G,B\}} \alpha_n^{(u,y)}(s)}{\sum_{s \in \{G,B\}} \alpha_n^{(y)}(s)}. \tag{30}$$

We apply a similar recursion to calculate the joint probability $P(\bar{\mathbf{u}}^n \mid \bar{\mathbf{y}}^n)$ which appears in (27). To avoid numerical issues, we compute all the previous terms in $\log$.

V. Numerical results

In this section, we aim to numerically evaluate the DHT error exponents with the GE model, by applying the method described in Section IV. We first investigate the convergence of the proposed estimators with respect to the sequence length n. Figure 2 shows the estimated error-exponents θ_{bin} and θ_{test} as functions of n, for a set of parameters given in the caption of the figure. We consider a maximum value $n = 15000$, and represent two types of curves: one obtained from a single set of source vectors $(\mathbf{x}^n, \mathbf{y}^n, \mathbf{u}^n)$, and the other obtained by averaging over $K = 15$ realizations of the source vectors. It is clear that averaging improves the estimation quality, especially for the testing error.

Next, we consider a large value of $n = 70000$, and investigate the effect of the model parameters onto the error exponent. First of all, Figure 3 shows the two error exponents θ_{bin} and θ_{test} as functions of p_G with p_B values as a parameter, where p_G and p_B are the crossover probabilities under the hypothesis $\mathcal{H}_0$. All other parameters are fixed, and indicated in the caption of the figure. We see that both error exponents decrease as p_G and p_B increase, due to the fact that increasing these parameters makes the source closer to uniform. This, in turn, causes $\mathcal{H}_0$ and $\mathcal{H}_1$ to become more similar, making it challenging for the decoder to accurately distinguish between them. In addition, Figure 4 shows the two error exponents with respect to a parameter $\mu = 1 - b - g$. While the binning error only slightly varies with μ, the testing error varies with large values of μ, especially when b is fixed.

Finally, we aim to investigate the tradeoff between the testing error and the binning error. Figure 5 shows the error exponents θ_{bin} and θ_{test} as function of δ, which is the parameter that defines the auxiliary source $\mathbf{U}$. The other parameters are fixed and indicated in the figure. It is found that for a small value range of δ, the binning error is smaller than the testing error, and therefore is a dominating factor of the Type-II error. However, when δ increases, the testing error tends to become the dominant error event. Finally, as pointed out in [6] for i.i.d. binary sources, the best tradeoff between the two error events is provided by the value δ at which the two curves intersect, *e.g*, $\delta \approx 0.25$ for this set of parameters.

VI. Conclusion

In this work, we have considered the DHT coding scheme and represented the correlation between the sources $\mathbf{X}$ and $\mathbf{Y}$ by a GE model. We developed an efficient method to estimate the error-exponent of the Type-II error for this model, which utilizes the forward recursion of HMMs. Our numerical results have evaluated the effects of the model parameters onto the error-exponent, as well as the tradeoff between the testing error and the binning error. Future works will be dedicated to the analysis of more generic Hidden Markov Models as well as

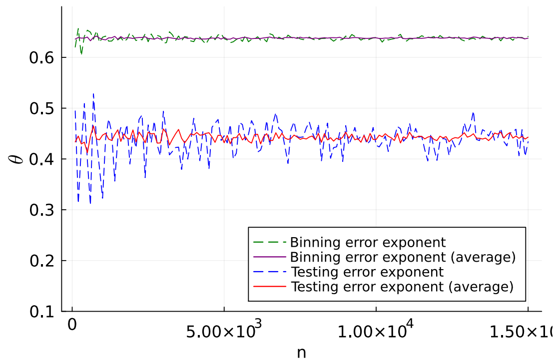

Figure 2: Estimated error exponents as functions of n for $p_G = 0.05$, $p_B = 0.03$, $\bar{p}_G = 0.3$, $\bar{p}_B = 0.5$, $g = 0.001$, $b = 0.002$, $p = 0.2$, $\delta = 0.15$, $R = 0.4$. The red and purple curves are averaged over $K = 15$ sequence realizations.

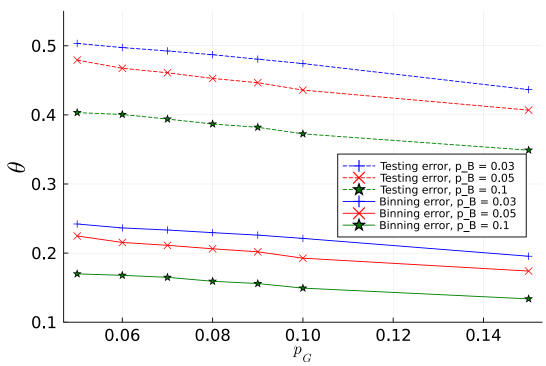

Figure 3: Error exponents as functions of p_g, for various values of p_b and for $\bar{p}_G = 0.3$, $\bar{p}_B = 0.5$, $g = 0.001$, $b = 0.002$, $p = 0.2$, $\delta = 0.001$.

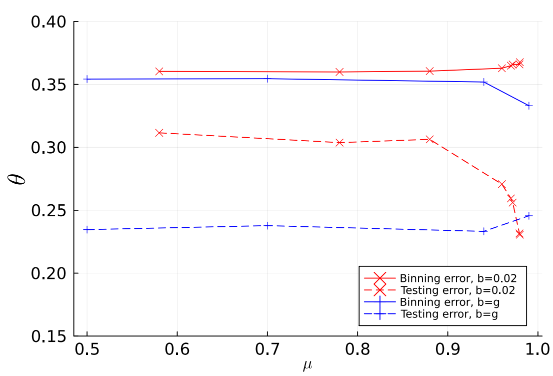

Figure 4: Error exponents as functions of $\mu = 1 - g - b$, for $\bar{p}_G = 0.3$, $\bar{p}_B = 0.5$, $p = 0.2$, $\delta = 0.15$.

Gauss Markov models, and to the design of practical coding schemes for DHT.

REFERENCES

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

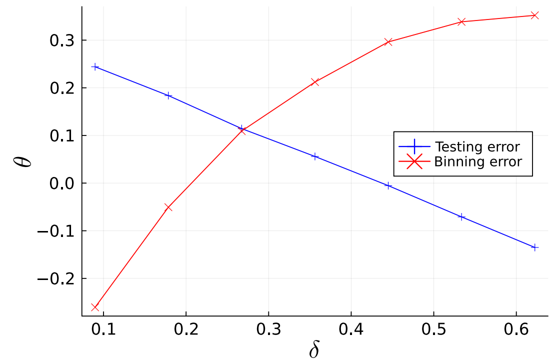

Figure 5: Error exponents as functions of δ for $p_G = 0.5$, $p_B = 0.3$, $\bar{p}_G = 0.05$, $\bar{p}_B = 0.03$, $p = 0.2$, $R = 0.4$

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

[3] E. C. Strinati and S. Barbarossa, "6g networks: Beyond shannon towards semantic and goal-oriented communications," *Computer Networks*, vol. 190, p. 107930, 2021.

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

[5] 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.

[6] 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.*, vol. 2015-June, pp. 2797–2801, 2015.

[7] 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.

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

[9] I. S. Adamou, E. Dupraz, and T. Matsumoto, "An information-spectrum approach to distributed hypothesis testing for general sources," *arXiv preprint arXiv:2305.06887*, 2023.

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

[11] A. Zaidi and I. E. Aguerri, "Optimal Rate-Exponent Region for a Class of Hypothesis Testing Against Conditional Independence Problems," *2019 IEEE Inf. Theory Work. ITW 2019*, no. 2, 2019.

[12] E. Haim and Y. Kochman, "Binary distributed hypothesis testing via körner-marton coding," in *2016 IEEE Information Theory Workshop (ITW)*. IEEE, 2016, pp. 146–150.

[13] V. Toto-Zarasoa, A. Roumy, and C. Guillemot, "Hidden markov model for distributed video coding," in *2010 IEEE International Conference on Image Processing*. IEEE, 2010, pp. 3709–3712.

[14] A. Bildea, O. Alphand, F. Rousseau, and A. Duda, "Link quality estimation with the gilbert-elliot model for wireless sensor networks," in *2015 IEEE 26th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 2015, pp. 2049–2054.

[15] C. A. G. da Silva and C. M. Pedroso, "Packet loss characterization using cross layer information and hmm for wi-fi networks," *Sensors*, vol. 22, no. 22, p. 8592, 2022.

[16] M. Mushkin and I. Bar-David, "Capacity and coding for the gilbert-elliott channels," *IEEE Transactions on Information Theory*, vol. 35, no. 6, pp. 1277–1290, 1989.

[17] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.

[18] S. Biswas, "Various proofs of the fundamental theorem of markov chains," *arXiv preprint arXiv:2204.00784*, 2022.

[19] E. L. Lehmann and G. Casella, *Theory of point estimation*. Springer Science & Business Media, 2006.

[20] M. Rezaeian, "Computation of capacity for gilbert-elliott channels, using a statistical method," in *2005 Australian Communications Theory Workshop*. IEEE, 2005, pp. 56–61.