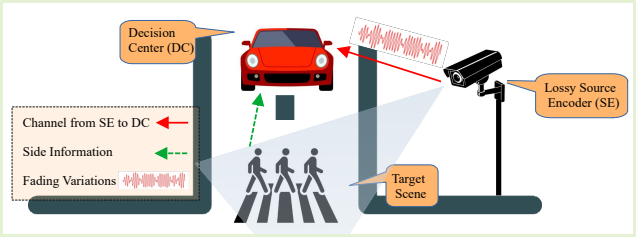
Goal-Oriented Wireless Sensor Networks: Error Exponent and Outage Analyses with Distributed Hypothesis Testing

Fatemeh Khaledian, *Member, IEEE*, Reza Asvadi, *Senior Member, IEEE*, Ismaila Salihou Adamou, Wensheng Lin, *Member, IEEE*, Tad Matsumoto, *Life Fellow, IEEE*

***Abstract*— This paper presents an analytical framework for characterizing and evaluating the rate-threshold (RT) region in distributed hypothesis testing (DHT) against independence, assuming a binary symmetric source (BSS) model. We consider a single sensor–single decision center (DC) setting, where a lossy source encoder (SE) compresses its observations using a binary linear block code and transmits them over a fading wireless channel. The DC, in turn, has perfect side information via an ideal (error-free, non-fading) channel and performs hypothesis testing using the Neyman–Pearson criterion, as shown in the graphical example. In a finite-length regime, we derived lower and upper bounds on the Type-II error probability and determined the sufficient number of samples (SNS) and the corresponding decision threshold, for which the bounds are tight and yield an exponential approximation to the Type-II error probability under an ideal channel. Next, we define the *admissible* RT region by considering the critical relationship between the source coding rate and the decision threshold, an essential factor that influences the achievable Type-II error exponent. The impact of channel fading is then analyzed through the lens of the lossy source–channel separation theorem, leading to an outage probability expression. Numerical results are provided under the Nakagami fading model, illustrating how the outage probability is affected by the fading severity, decision threshold, and the correlation level between the source and side information.**



***Index Terms*— Distributed hypothesis testing (DHT), error exponent, decision threshold, sufficient number of samples, Type-II error approximation, admissible rate-threshold region, outage probability, wireless sensor networks (WSNs).**

I. Introduction

HYPOTHESIS testing (HT) is a fundamental component of statistical inference, significantly contributing to decision-making, health monitoring, security alerts, and many other applications [1]–[4]. In a centralized context, HT draws inferences from the data collected at a single source (or location), providing a robust framework for identifying the underlying distribution of the observations from a specific set of hypotheses. With the growing use of distributed sensing technologies and wireless networks, such as Internet of Things (IoT) systems, wireless sensor networks (WSNs), and cloud radio access networks, an increasingly volume of data is being collected by spatially distributed sensors and/or agents. The sensors transmit their observations to a decision center (DC) while adhering to strict communication constraints [5], within the framework of a distributed hypothesis testing (DHT) problem. The implementation of DHT becomes particularly challenging in sensitive applications such as vehicular networks and local verification systems [6]. Consequently, evaluating decision accuracy under communication constraints, particularly limited transmission rate, emerges as a critical design factor [7].

A fundamental scenario in DHT involves binary hypothesis testing, where the DC aims to identify between a null hypothesis $\mathcal{H}_0$ and an alternative hypothesis $\mathcal{H}_1$. In this context, *testing against independence* arises as a fundamental problem, wherein understanding dependencies between sensor observations becomes essential for both theoretical and practical purposes. In particular, $\mathcal{H}_0$ represents the case of statistically dependent, i.e., correlated observations, while $\mathcal{H}_1$ corresponds to statistical independence, as commonly addressed in a wide range of studies [8]–[13]. The decision process at the DC can result in two types of errors. A Type-I error arises when the DC incorrectly selects $\mathcal{H}_1$ although $\mathcal{H}_0$ is true. Conversely, a Type-II error occurs when the DC selects $\mathcal{H}_0$ despite $\mathcal{H}_1$ being the correct hypothesis.

Due to the high computational complexity involved in deriving exact analytical expressions for Type-I and Type-II error probabilities, the error exponent, which characterizes

F. Khaledian and R. Asvadi are with the Department of Telecommunications, Faculty of Electrical Engineering, Shahid Beheshti University, Tehran 1983963113, Iran (e-mails: fa.khaledian@alumni.sbu.ac.ir; r_asvadi@sbu.ac.ir).

Ismaila Salihou Adamou is with the Department of Communication and Electronic, Télécom Paris, 91120 Palaiseau , France (email: ismaila.salihou@telecom-paris.fr).

W. Lin is with the School of Electronics and Information, Faculty of Northwestern Polytechnical University, Xi'an, Shaanxi 710129, China (e-mail: linwest@nwpu.edu.cn).

T. Matsumoto is Professor Emeritus of Japan Advanced Institute of Science and Technology, Japan, and University of Oulu, Finland. He is also currently serving as a Research Fellow at Institute of Science Tokyo, Japan. (e-mail: matumoto@jaist.ac.jp).

This paper was presented in part at the IEEE Information Theory Workshop (ITW), Shenzhen, China, in November 2024 [28].

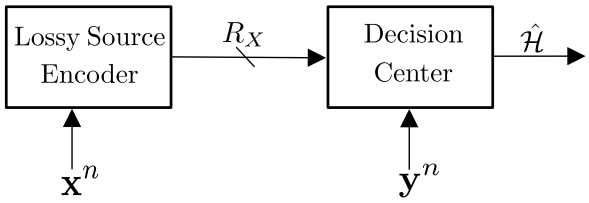


Fig. 1: Single sensor-single DC distributed hypothesis testing problem over a rate-limited error-free channel.

the exponential decay rate of Type-II error probability as the number of observed samples increases, is often used as a key performance metric in the asymptotic regime. Error exponent analyses across a variety of scenarios are reviewed in [14].

While the trade-off between Type-I and Type-II error probabilities can be addressed through various approaches, a common technique focuses on analyzing the error exponent of Type-II error probability while keeping Type-I error probability below a fixed constraint [15], [16]. In the specific scenario of *asymmetric* setup of DHT, a single sensor communicates with a single DC over an error-free, rate-limited channel, while the side information is fully available at the DC, as illustrated in Fig. 1. In this scenario, [15] establishes a general lower bound on Type-II error exponent which is optimal in the case of testing against independence, which can be achieved through a quantization-based scheme while maintaining Type-I error probability smaller than a fixed constraint $\epsilon > 0$. By transmitting both the quantized source and its joint type with the source to the DC, [16] enhanced the quantization scheme and established a tighter lower bound on Type-II error exponent for the *general* DHT problem. To better exploit the correlation between the source and side information, a quantize-binning scheme, known as the Shimokawa-Han-Amari (SHA) scheme, was introduced in [17], achieving a further reduction in the required rate. The optimality of this scheme is shown in [18] for the case of *testing against conditional independence*. By employing the quantize-binning strategy within a layered coding framework, an optimal single-letter characterization of the *rate-error-distortion region* for testing against independence was established in [19]. For a *symmetric* DHT scenario with two sensors and a single DC, both Type-I and Type-II error exponents have been derived for the binary source case. These results were obtained using the Körner-Marton scheme [20] and are presented in [21]. The symmetric setup under a zero-rate compression constraint, where the coding rate of one or both sensors asymptotically approaches zero, has been analyzed. This line of work has primarily focused on developing testing schemes and characterizing the achievable error exponents [16], [22]-[24]. Furthermore, a scenario involving L sensors ($L \geq 1$) and a single decision center (DC) has been examined in the context of testing against conditional independence. The *rate-exponent region* has been characterized for both discrete memoryless sources and memoryless vector Gaussian sources; see [25] and references therein for further details.

Although distributed hypothesis testing (DHT) is grounded in a robust theoretical framework within information theory, it has also attracted significant attention in practical applications. In the context of rate-limited, error-free channels, recent studies, such as [26], have established lower bounds on the Type-II error exponent for general sources. Furthermore, efficient short-length quantization and quantize-binning schemes employing binary linear block codes have been developed specifically for binary symmetric sources (BSS). Analytical expressions for both Type-I and Type-II error probabilities, based on these short-length coding strategies, were derived in [27]. The study of DHT over noisy channels was initiated in [29], where bounds on Type-II error probability were provided for testing against conditional independence. Extensions to multi-access channels (MACs) and hybrid coding schemes have been presented in [30] and [31], incorporating both joint and separate source-channel coding approaches. The SHA strategy, combined with unequal error protection (UEP), has been shown to achieve optimal error exponents in discrete memoryless channel (DMC) and MAC scenarios [32]. Some other works on interactive two-node problems are addressed in [33] and [34], where the optimal error exponent for testing against independence is established in [33]. In networks with multiple decision centers, trade-offs in Type-II error exponent have been explored in the context of broadcast and relay channels [35], [36], where optimal performance was characterized using Markov models and cascade architectures. Further extensions have included cooperative DHT systems with two decision centers [37], and two-hop or multi-hop network configurations under expected rate constraints [38], [39]. Furthermore, DHT has been investigated under privacy constraints [40], [41] and secrecy constraints [42], [43].

Despite extensive research on DHT across a variety of channel models and network topologies, most existing studies assume *non-fading* communication links. In contrast, real-world WSN deployments are often affected by fading, which can substantially degrade system performance. This limitation is especially significant as WSNs are now commonly used in dynamic environments where channel conditions change frequently. In cooperative wireless IoT systems with multi-terminal source coding, performance is typically evaluated in terms of outage probabilities under fading channel conditions. As discussed in [44], many analytical frameworks have been developed for these scenarios, most of which are based on Shannon's source-channel separation theorem [45]. Building on these principles, recent works such as [46] and [47] have proposed transmission schemes that integrate source coding with capacity-achieving channel coding. This approach can also be extended to DHT systems, offering a strong theoretical framework for analyzing trade-offs among compression rate, decision threshold, and accuracy.

The primary objective of this paper is to analyze the performance of a DHT system, with a particular focus on outage probability under fading conditions. We investigate a DHT framework for testing against independence, adopting the BSS model in a single-sensor, single DC configuration. A binary linear block code serves as the lossy source encoder (SE), and the communication channel between the SE and DC is subject to fading. In particular, the SE-DC channel is modeled using block Nakagami fading, where the channel signal-to-noise ratio (SNR) varies from block to block but

TABLE I: Definitions of the Key Variables in the Paper

Notations	Definition(s)
X, Y, U	Source, side information, and quantized random variables
x, y, u	Source, side information and quantized realizations
$\mathcal{X}, \mathcal{Y}, \mathcal{U}$	Finite alphabet sets
$\mathbf{x}^n$, $\mathbf{y}^n$, $\mathbf{u}^n$	Source, side information and quantized vectors with length n
$P_{XY}(x,y), P_{U\|X}(u\|x)$	Joint probability density function (PDF) of X and Y, conditional PDF between X and U
$\mathcal{H}_0$, $\mathcal{H}_1$	Null and alternative hypotheses
α_n, β_n	Type-I and Type-II error probabilities
$\widetilde{\beta}_n$, β_n^*	Finite-length and optimal finite-length approximations of Type-II error probability
E_1, E_1^*	Type-II error exponent, optimal Type-II error exponent
λ_n, λ_r	Deterministic and random decision thresholds
ϵ	Type-I error probability constraint in $(0,1)$
τ	Normalized decision threshold with respect to the number of samples n
$\mathcal{R}(\tau)$, $\mathcal{R}^c(\tau)$	Admissible and inadmissible rate-threshold regions
$d_H(.)$, $D_H(.)$	Hamming and the average Hamming distortion
$\Lambda(\cdot,\cdot)$	Distortion transfer function
$H_b(X)$	Binary entropy function of X
$I(X;Y)$	Mutual information function between X and Y
$D_b(p,q)$	Binary Kullback-Leibler divergence between two Bernoulli parameters p and q
log	base-2 logarithm
$\binom{n}{k}$	Binomial coefficient for integers n and k
$[n,k]$	Binary linear block code of dimension $k \times n$
$\mathbb{E}[\cdot]$	Expectation operator
$[\![a,b]\!]$	Set of integers between a and b

remains constant within each block. Under this setting, the main contributions of the paper are summarized as follows:

- We derive both exponential-form lower and upper bounds on the Type-II error probability, as well as an upper bound on the Type-I error probability. These bounds are obtained using exact analytical error probability expressions, specifically tailored to a practical finite-length coding scheme employing binary linear block codes, as detailed in [27].
- Next, we investigate the exponential approximation of the Type-II error probability in the finite-length regime, establishing explicit conditions on the sufficient number of samples (SNS) and the corresponding decision thresholds that ensure the approximation is reliable. We then assess the accuracy of this approximation by comparing it with the optimal exponential characterization of the Type-II error in the finite-length regime, as discussed in [48]. To this end, we employ a search strategy based on the randomization decision method proposed in [49] within the Neyman-Pearson (NP) framework.
- Additionally, we derive the achievable Type-II error exponent in the asymptotic regime and evaluate system performance in terms of outage probability, assuming hypothesis $\mathcal{H}_0$ holds. In particular, when the decision center (DC) erroneously selects $\mathcal{H}_1$ due to fading, we characterize the admissible *rate-threshold* (RT) region by linking the rate constraint from [15] to a normalized decision threshold, considering the Markov chain among the source encoder (SE), side information, and the quantized source. This relationship is formalized via the distortion transfer function (DTF) introduced in [50]. The analysis leverages the source-channel separation theorem, employing binary linear block codes for source coding and Gaussian codebooks for channel coding. We further illustrate how outage probability behavior varies with fading severity, decision thresholds, and the correlation between SE and DC observations.

The remainder of this paper is structured as follows. Section II details the system and channel models and introduces key definitions pertinent to the DHT framework. Section III presents finite-length results, including the conditions necessary for the exponential approximation of the Type-II error probability. Section IV characterizes the admissible rate-threshold (RT) region and examines outage probability under fading channel conditions. Section V reports numerical results, while Section VI concludes the paper by summarizing the main findings. A comprehensive list of notations is provided in Table I.

II. Preliminaries and System Model

In this section, we present some fundamental definitions associated with the DHT problem in a single sensor–single DC system, focusing on the specific case of testing against independence. We initially assume an error-free SE-DC channel, and subsequently define the channel model used in the outage probability analysis.

A. Distributed Hypothesis Testing against Independence

Consider the system illustrated in Fig. 1, where the SE and DC observe random vectors $\mathbf{x}^n$ and $\mathbf{y}^n$, both with length n, respectively. The components of these vectors take values from finite alphabets $\mathcal{X}$, $\mathcal{Y}$, respectively. The random pairs (X_i, Y_j) are independently and identically distributed (i.i.d.) over the time indices $i, j = 1, ..., n$, according to one of the following

hypotheses

$$\mathcal{H}_0 : (X, Y) \sim P_{XY}(x, y), \tag{1}$$
$$\mathcal{H}_1 : (X, Y) \sim P_X(x)P_Y(y). \tag{2}$$

The hypothesis $\mathcal{H}_1$ represents the case X and Y are independent, where the joint distribution is given by the product of the marginal distributions $P_X(x)$ and $P_Y(y)$ under the hypothesis $\mathcal{H}_0$. This scenario is known as testing against independence. In the case of the BSS model, X and Y take values in $\mathcal{X}, \mathcal{Y} = \{0, 1\}$, with $P_X(x) = P_Y(y) \sim \text{Bern}\left(\frac{1}{2}\right)$. While the marginal distributions are identical, Y may be a noisy version of X, modeled as $Y = X \oplus W$, where $\oplus$ denotes addition over the binary field. The two hypotheses are thus specified as

$$\mathcal{H}_0 : W \sim \text{Bern}(p_0), \tag{3}$$
$$\mathcal{H}_1 : W \sim \text{Bern}(p_1), \tag{4}$$

where $\text{Bern}(p)$ represents Bernoulli distribution with parameter $p \in [0, 1]$. We assume $0 \leq p_0 < \frac{1}{2}$ as the crossover probability between X and Y under $\mathcal{H}_0$, while $\mathcal{H}_1$ represent statistical independent with $p_1 = \frac{1}{2}$.

As shown in Fig. 1, the SE observes $\mathbf{x}^n$, and transmits a compressed representation to the DC at the rate R_X. The DC then makes a decision using the received message and the side information $\mathbf{y}^n$. Following the approach in [15], the lossy SE is modeled as a mapping $f : \{0, 1\}^n \to [\![1, \mathsf{M}]\!]$, where each $\mathsf{m} \in [\![1, \mathsf{M}]\!]$ is given by $\mathsf{m} = f(\mathbf{x}^n)$. The DC is modeled as a function $g : [\![1, \mathsf{M}]\!] \times \{0, 1\}^n \to \hat{\mathcal{H}} \in \{\mathcal{H}_0, \mathcal{H}_1\}$, such that the final decision is denoted by $\hat{\mathcal{H}} = g(f(\mathbf{x}^n), \mathbf{y}^n)$. We consider the lossy SE under a rate constraint $\frac{\log \mathsf{M}}{n} \leq R_X$. For a given pair (f, g), the decision process can result in two types of errors, defined by the following probabilities

$$\alpha_n = \Pr\left(g\left(f\left(\mathbf{x}^n\right), \mathbf{y}^n\right) = \mathcal{H}_1 \mid \mathcal{H}_0\right), \tag{5}$$
$$\beta_n = \Pr\left(g\left(f\left(\mathbf{x}^n\right), \mathbf{y}^n\right) = \mathcal{H}_0 \mid \mathcal{H}_1\right), \tag{6}$$

where α_n and β_n represent Type-I error and Type-II error probabilities, respectively. Given a Type-I error constraint $\alpha_n \leq \epsilon$, with $\epsilon \in (0, 1)$, Type-II error exponent E_1 satisfies [21]

$$E_1 \leq \lim_{n \to \infty} \sup \left\{ -\frac{1}{n} \log \beta_n \;\middle|\; \log \mathsf{M} \leq nR_X \right\} \tag{7}$$

The optimal Type-II error exponent E_1^* is then defined as [15]

$$E_1^* = \max_{\substack{P_{U|X} \\ |\mathcal{U}| \leq |\mathcal{X}|+1 \\ R_X \geq I(U;X)}} I(U; Y), \tag{8}$$

where U is the random variable representing the compressed, and transmitted version of the source X. Moreover, the mutual information between two random variables X and Y is denoted by $I(X; Y)$. For the BSS model, the optimal Type-II error exponent E_1^* becomes [51]

$$E_1^* = 1 - H_b\left(H_b^{-1}(1 - R_X) \star p_0\right), \tag{9}$$

with $\star$ denoting the binary convolution, defined as $x \star y = x(1 - y) + y(1 - x)$ for $x, y \in [0, 1]$. The function $H_b^{-1}(\cdot)$ denotes inverse of $H_b(\cdot)$, defined in the probability range $[0, 0.5]$.

B. Quantization Scheme

In [15], it is shown that for DHT against independence, an information-theoretic quantizer-based scheme minimizes Type-II error probability β_n under the constraint $\alpha_n \leq \epsilon$. According to the practical approach proposed in [27], we adopt the following scheme.

1) Code Construction: To perform binary quantization at the lossy SE, we consider a generator matrix $\mathbf{G}_q$ of size $k \times n$, corresponding to a binary linear block code [52]. Given a source vector $\mathbf{x}^n$, the SE generates a codeword $\mathbf{u}_q^k$ as described in [53]

$$\mathbf{u}_q^k = \underset{\mathbf{z}^k \in \{0,1\}^k}{\arg\min}\; d_{\text{H}}\left(\mathbf{z}^k \mathbf{G}_q, \mathbf{x}^n\right). \tag{10}$$

The codeword $\mathbf{u}_q^k$ is then transmitted to the DC at a code rate of $R_X = \frac{k}{n}$.

2) Decision Criterion: To perform the decision-making task, it is not necessary for the DC to fully reconstruct the source vector $\mathbf{x}^n$. Instead, the DC can apply an NP test [54] using the quantized vector $\mathbf{u}^n$, defined as $\mathbf{u}^n = \mathbf{u}_q^k \mathbf{G}_q$. According to the NP lemma, the following test is used

$$P_1(\mathbf{u}^n, \mathbf{y}^n) \leq \mu P_0(\mathbf{u}^n, \mathbf{y}^n), \tag{11}$$

which minimizes Type-II error probability β_n. The parameter μ is a pre-specified threshold selected to ensure that Type-I error probability satisfies $\alpha_n \leq \epsilon$. In (11), P_0 and P_1 denote the joint probability distributions of $(\mathbf{u}^n, \mathbf{y}^n)$ under hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. As shown in [54], the test defined by (11) is equivalent to the following condition

$$\sum_{i=1}^{n} d_{\text{H}}\left(u_i, y_i\right) < \lambda_n, \tag{12}$$

where $\lambda_n \in \mathbb{N}$ is an integer threshold selected to ensure that $\alpha_n \leq \epsilon$. Notably, if (12) holds, the DC decides $\mathcal{H}_0$; otherwise, it selects $\mathcal{H}_1$. This test depends on the parameter p_0 which influences the choice of λ_n such that the Type-I error constraint is satisfied.

C. Channel Model

We assume the decision center (DC) is mobile, while the source encoder (SE) remains stationary, resulting in fading variations on the SE-DC channel. The severity of the fading can range from severe, as characterized by Rayleigh fading, to moderate, as modeled by Nakagami fading with a shape parameter $m > 0.5$. Rayleigh fading is commonly employed to model random multipath propagation in environments with only non-line-of-sight (NLOS) components. When a line-of-sight (LOS) path is present, however, the fading characteristics change substantially. In such cases, the Rician fading distribution, derived from mathematical propagation models, is typically used to represent channels with a dominant LOS component. Nevertheless, the Nakagami-m distribution is often considered more practical, as it provides a better fit to experimental data than the Rician model. The PDF of the SNR γ_X for the SE-DC channel under Nakagami-m fading is given by

$$P(\gamma_X) = \frac{m^m \gamma_X^{m-1}}{\bar{\gamma}_X^m \Gamma(m)} \exp\left(\frac{-m\gamma_X}{\bar{\gamma}_X}\right), \quad m > \frac{1}{2}, \tag{13}$$

where $\Gamma(\cdot)$ is the complete Gamma function, and m is the shape parameter that quantifies the severity of the fading. The case $m = 1$ corresponds to Rayleigh fading. The relationship between the Nakagami parameter m and the Rician K-factor is expressed as

$$m = \frac{(K+1)^2}{2K+1}, \tag{14}$$

where K denotes the ratio of LOS to average NLOS component power [56], [57]. Throughout this paper, we adopt a block fading model, where the channel gain remains constant within each block and varies independently from block to block.

III. Finite-Length Result

In this section, we first present an upper bound for the Type-I error probability and both lower and upper bounds for the Type-II error probability associated with the quantization-and-decision scheme from Section II. The tightness of these Type-II error bounds determines the SNS and the corresponding decision threshold needed to guarantee that the exponential approximation of the Type-II error probability achieves a specified accuracy, while keeping the Type-I error probability below a set constraint. We then evaluate how closely this exponential approximation aligns with the optimal Type-II error characterization. Finally, we detail a randomization-based method for selecting the decision threshold.

The exact analytical expressions for Type-I and Type-II error probabilities in DHT against independence are given by [27]

$$\alpha_n = 1 - \frac{1}{N}\sum_{\lambda=0}^{\lambda_n}\sum_{i=0}^{d_{\max}}\sum_{u=0}^{\min(\lambda,i)} \Gamma_{\lambda,i,u} p_0^j (1-p_0)^{n-j} N_i, \tag{15}$$

$$\beta_n = \frac{1}{N}\left(\frac{1}{2}\right)^n \sum_{\lambda=0}^{\lambda_n}\sum_{i=0}^{d_{\max}}\sum_{u=0}^{\min(\lambda,i)} \Gamma_{\lambda,i,u} N_i, \tag{16}$$

where $\Gamma_{\lambda,i,u} = \binom{i}{u}\binom{n-i}{\lambda-u}$, and $j = i + \lambda - 2u$. For the set of non-negative integers $\{N_i\}_{i\in[\![0,d_{\max}]\!]}$, N_i denotes the number of vectors $\mathbf{x}^n$ of Hamming distance i that belong to the decision region of all-zero codeword $\mathbf{u}^n = \mathbf{0}^n$. The normalization constant N is given by $N = \sum_{i=0}^{d_{\max}} N_i$.

Proposition 1: Type-I and Type-II error probabilities are bounded by the following exponential expressions

$$\mathrm{LB}_{\beta_n}(n,\lambda_n) \le \beta_n \le \mathrm{UB}_{\beta_n}(n,\lambda_n), \tag{17}$$

$$\alpha_n \le \mathrm{UB}_{\alpha_n}(n,\lambda_n,p_0) \tag{18}$$

where $\mathrm{LB}_{\beta_n}(n,\lambda_n) \triangleq \left(8\lambda_n(1-\frac{\lambda_n}{n})\right)^{-\frac{1}{2}} 2^{-nD_b(1-\frac{\lambda_n}{n}||\frac{1}{2})}$ and $\mathrm{UB}_{\beta_n}(n,\lambda_n) \triangleq 2^{-nD_b(1-\frac{\lambda_n}{n}||\frac{1}{2})}$. Moreover, $\mathrm{UB}_{\alpha_n}(n,\lambda_n,p_0) = \frac{\Delta}{N} 2^{-nD_b(\frac{\lambda_n}{n}+\frac{1}{n}||p_0)}$, with $\Delta = \sum_{i=0}^{d_{\max}} \left(\frac{1}{\kappa}\right)^{\frac{i}{2}} \left(\frac{\kappa^i-1}{\kappa-1}\right) N_i$ and $\kappa = \left(\frac{1-p_0}{p_0}\right)^2$.[1] The threshold λ_n is chosen from the intervals $[\![0, \frac{n}{2}]\!]$ and $[\![np_0 - 1, n-1]\!]$ to ensure that (17) and (18) are satisfied, respectively.

proof: A detailed proof is provided in Appendix.

According to Proposition 1, the Type-II error bounds depend only on the number of samples n and the decision threshold λ_n. This motivates a search for the finite-length regime that achieves the smallest number of samples, SNS, along with the corresponding decision threshold λ_n that jointly satisfy the following conditions:

$$\mathrm{UB}_{\beta_n}(n,\lambda_n) - \mathrm{LB}_{\beta_n}(n,\lambda_n) \le \delta_1, \tag{19}$$

within a desired precision δ_1, while ensuring $\alpha_n \le \epsilon$. When (19) is satisfied, the Type-II error probability can be tightly approximated as $\widetilde{\beta}_n \simeq \mathrm{UB}_{\beta_n}(n,\lambda_n)$, with accuracy guaranteed to be within δ_1. Furthermore, we assess the proximity of $\widetilde{\beta}_n$ to the optimal finite-length benchmark $\widetilde{\beta}_n^* \simeq 2^{-nE_1^*}$.[2]

$$\widetilde{\beta}_n - \widetilde{\beta}_n^* \le \delta_2, \tag{20}$$

within a prescribed precision δ_2.

Because n and λ_n are restricted to integer values, we employ a randomized scheme to facilitate the optimal selection of λ_n [49, Section 7.4]. This approach ensures that, as the number of samples increases, the average Type-I error probability $\bar{\alpha}_n$ converges to ϵ. To formalize this scheme, we define the cumulative distribution function (CDF) for any $\lambda_n \in [\![0, \frac{n}{2}]\!]$ as follows:

$$F_n(\lambda_n) \triangleq \frac{1}{N}\sum_{\lambda=0}^{\lambda_n}\sum_{i=0}^{d_{\max}}\sum_{u=0}^{\min(\lambda,i)} \Gamma_{\lambda,i,u} p_0^j (1-p_0)^{n-j} N_i, \tag{21}$$

it is straightforward to verify that $F_n(\lambda_n) = 1 - \alpha_n$. Based on (21), we define two candidate decision thresholds as follows [49, Eq. (7.62), Eq. (7.63)]:

$$\underline{\lambda}_n \triangleq \underset{\lambda_n\,:\,F_n(\lambda_n)\le 1-\epsilon}{\arg\max} F_n(\lambda_n),$$

$$\bar{\lambda}_n \triangleq \underset{\lambda_n\,:\,F_n(\lambda_n)\ge 1-\epsilon}{\arg\min} F_n(\lambda_n).$$

In general, $\underline{\lambda}_n < \bar{\lambda}_n$. In this case, we define the randomized threshold as

$$\lambda_r \triangleq \begin{cases} \underline{\lambda}_n & \text{with probability } p, \\ \bar{\lambda}_n & \text{with probability } 1-p, \end{cases} \tag{22}$$

where the randomized probability

$$p \triangleq \frac{(1-\epsilon) - F_n(\bar{\lambda}_n)}{F_n(\underline{\lambda}_n) - F_n(\bar{\lambda}_n)},$$

is selected so that the average Type-I error probability, given by $\bar{\alpha}_n = p(1 - F_n(\bar{\lambda}_n)) + (1-p)(1 - F_n(\underline{\lambda}_n))$, is exactly equal to ϵ [49, Eq. (7.75)].

Section IV presents numerical examples that satisfy (19) and (20), illustrating the evaluation of the SNS and λ_n.

[1] Upper bound in (18) is helpful in problems with exponential Type-I error probability constraints [58]. Throughout this paper, we focus on the fixed constraint $\alpha_n \le \epsilon$ according to the NP setting.

[2] Following [48, Section IV-A, item (v)], we emphasize that $\widetilde{\beta}_n^* \simeq 2^{-nE_1^*}$ serves not only as an asymptotic expression but also as an attainable finite-length benchmark for approximating the Type-II error probability.

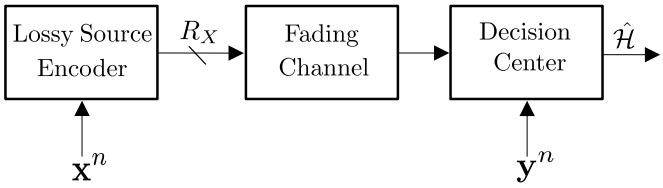


Fig. 2: Single sensor-single DC distributed hypothesis testing problem over a wireless fading channel.

IV. Outage Probability Analysis

In this section, we formalize the admissible RT region to analyze the influence of the normalized decision threshold on system performance. Subsequently, we assess the outage probability under a fading channel model as determined by this RT region.

Corollary 1: According to Proposition 1, for the scheme defined in (10) and (12) utilizing a family of binary linear block codes $[n, k]$ with $k \leq nR_X$, the achievable Type-II error exponent E_1 is given by

$$\begin{aligned} E_1 \;&\leq \lim_{n\to\infty} \sup \left\{ -\frac{1}{n} \log \beta_n \;\middle|\; k \leq nR_X \right\} \\ &= 1 - H_b(\tau), \end{aligned} \tag{23}$$

where $\tau = \lim_{n\to\infty} \frac{\lambda_n}{n}$ and $\tau \in (0, 0.5)$.

The result follows directly from Proposition 1.

From Cor. 1, we observe that the achievable Type-II error exponent, a key measure of the decision process [15], significantly depends on the normalized decision threshold τ. In addition, under the testing against independence setup, the SE rate R_X must satisfy the following inequality [15]

$$R_X \;\geq\; I\,(X;U), \tag{24}$$

where U is an auxiliary random variable representing the compressed version of the source X. In the case of the BSS model, the mutual information in (24) can be expressed as

$$\begin{aligned} R_X \;&\geq\; H_b(U) - H_b\,(U \mid X)\,, \\ &\geq 1 - H_b\,(D_X)\,, \end{aligned} \tag{25}$$

where D_X represents the average Hamming distortion between the source vector $\mathbf{x}^n$ and its quantized version $\mathbf{u}^n = \mathbf{u}^k\mathbf{G}_q$, with $\mathbf{u}^k$ being the codeword generated using the binary linear block code with generator matrix $\mathbf{G}_q$.

Remark 1: We consider a Markov chain $U - X - Y$, where U is generated solely from X with a bit-flipping probability D_X. Under the BSS model, X and Y are connected by a bit-flipping channel specified by the crossover probability p_0. As a result, the overall average Hamming distortion between U and Y is given by the binary convolution $D_X \star p_0$.

To explore the impact of the normalized decision threshold τ on the rate bound derived in (25), we make use of the DTF as follows

Proposition 1 ([50, Prop. 1]): For any values $x, t \in [0, 0.5]$ and $y \in [0, 0.5)$, if $x \star y \leq t$, then

$$x \leq \Lambda(y, t) \triangleq \frac{1}{2}\Big(1 - \frac{2t-1}{2y-1}\Big). \tag{26}$$

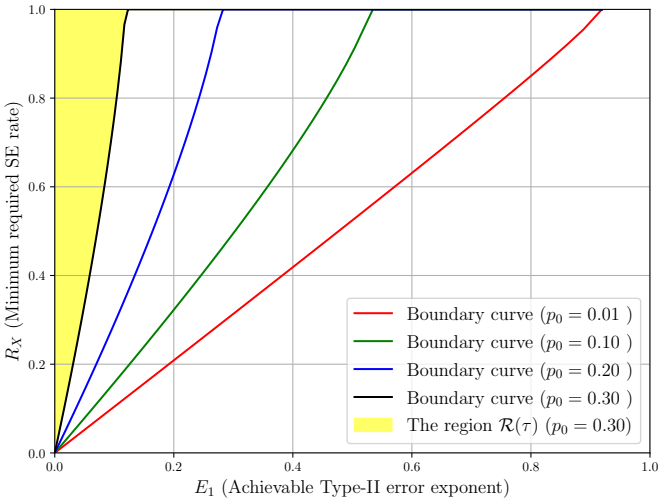


Fig. 3: The trade-off between region $\mathcal{R}(\tau)$ and the achievable Type-II exponent E_1 of the single sensor-single DC distributed hypothesis testing system for various values of p_0 and $\tau \in (0, 0.5)$.

Assuming hypothesis $\mathcal{H}_0$ holds, the condition $D_X \star p_0 \leq \tau$ guarantees a correct decision under the test specified in (12). Utilizing the DTF from (26), we derive the following upper bound for D_X:

$$D_X \;\leq\; \Lambda\,(p_0, \tau)\,. \tag{27}$$

Therefore, considering the inequality above and the corresponding SE rate given in (25), we define the *admissible RT region* $\mathcal{R}(\tau)$ as the set of all source encoder rates R_X that satisfy

$$\mathcal{R}(\tau) = \Big\{ R_X : \lim_{n\to\infty} \mathbb{E}\big[D_H(\mathbf{u}^n, \mathbf{y}^n)\big] \leq \tau + \varepsilon, \quad \forall \varepsilon > 0 \Big\}, \tag{28}$$

Furthermore, based on (25) and (27), the region $\mathcal{R}(\tau)$ is then characterized as follows:

$$\mathcal{R}(\tau) = \Big\{ R_X : R_X \geq 1 - H_b(\Lambda(p_0, \tau)) \Big\}. \tag{29}$$

This region forms the basis for defining the inadmissible RT region, which is used to evaluate outage probability. Figure 3 illustrates the trade-off between the admissible region $\mathcal{R}(\tau)$ and the achievable Type-II error exponent E_1. The boundary curves, derived from (23) and (29), span $\tau \in (0, 0.5)$ and various crossover probabilities p_0. As p_0 decreases and correlation increases, the boundary shifts downward, permitting lower R_X for the same E_1. The shaded area corresponds to $\mathcal{R}(\tau)$ with $p_0 = 0.3$, while the black curve represents the achievable E_1 for this scenario.

A. Outage Probability Derivation

Consider the system model with a fading SE-DC channel depicted in Fig. 2. An outage event occurs when the SE-DC channel fails to meet the rate requirement specified in (29). Consequently, the inadmissible RT region, $\mathcal{R}^c(\tau)$, is defined as

$$\mathcal{R}^c(\tau) = \{ R_X : 0 \leq R_X < 1 - H_b\,(\Lambda\,(p_0, \tau))\}\,. \tag{30}$$

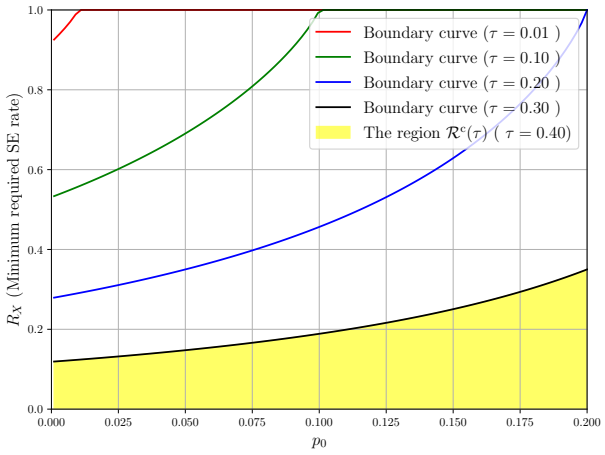


Fig. 4: The inadmissible region $\mathcal{R}^c(\tau)$ for a single-sensor, single-DC distributed hypothesis testing system, displayed for various values of the normalized decision threshold τ.

If $\mathcal{H}_0$ holds and the SE rate R_X lies within the inadmissible region $\mathcal{R}^c(\tau)$, the overall distortion $D_X \star p_0$ will exceed the threshold τ. In this case, the decision center declares $\hat{\mathcal{H}} = \mathcal{H}_1$, resulting in a Type-I error and violating the error trade-off guaranteed by τ. The outage probability is therefore defined as $P_{\text{out}} = \Pr\{R_X \in \mathcal{R}^c(\tau)\}$, representing the average over all transmissions. Fig. 4 depicts the boundary curve, derived from (30), for a range of threshold values. The inadmissible region $\mathcal{R}^c(\tau)$ is highlighted for the specific case where $\tau = 0.4$.

According to Shannon's source–channel separation theorem, the SE rate R_X is constrained by the channel capacity,

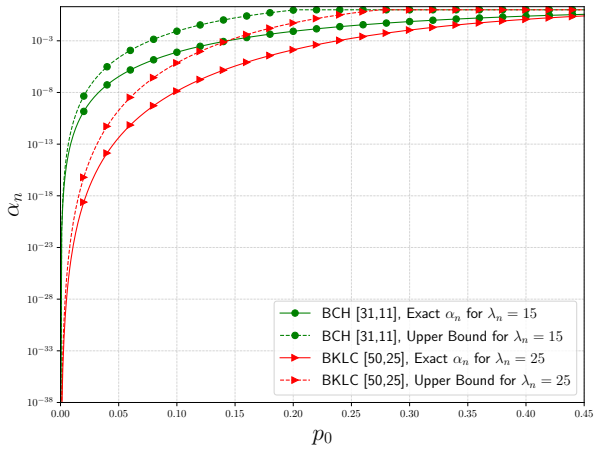
$$R_X \leq \Phi(\gamma_X) \triangleq \frac{C(\gamma_X)}{R_C}, \tag{31}$$

where R_C denotes the normalized spectral efficiency[3] and the function $C(\cdot)$ represents channel's capacity function. For two-dimensional signaling, the channel capacity using a Gaussian codebook with per-block SNR γ_X is given by $C(\gamma_X) = \log(1+\gamma_X)$. Thus, γ_X can be computed using the inverse function Φ^{-1}, given by $\gamma_X = \Phi^{-1}(R_X) = 2^{R_X} - 1$. Clearly, $\Phi^{-1}(0) = 0$. Using (31) the outage probability P_{out} can be characterizes the cases in which the maximum rate supported by the SNR γ_X falls within the inadmissible region $\mathcal{R}^c(\tau)$ as

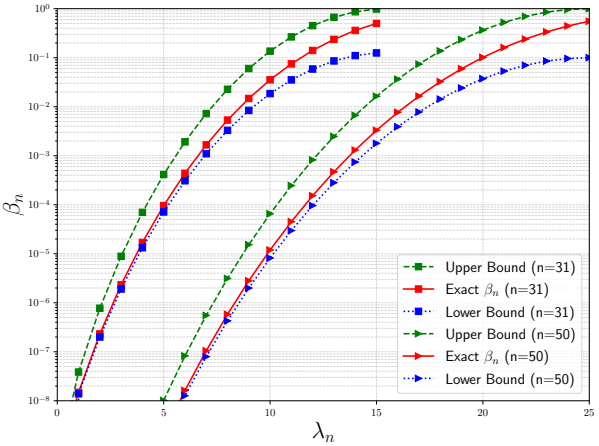
$$\begin{aligned} P_{\text{out}} &= \Pr\{p_0 \in [0, 0.5),\, R_X \in \mathcal{R}^c(\tau)\} \\ &= \Pr\Big\{0 \leq R_X < 1 - H_b(\Lambda(p_0,\tau)),\, p_0 \in [0, 0.5)\Big\} \\ &= \Pr\Big\{\Phi^{-1}(0) \leq \gamma_X \leq \Phi^{-1}(1 - H_b(\Lambda(p_0,\tau))), \\ &\quad p_0 \in [0, 0.5)\Big\}. \end{aligned} \tag{32}$$

The outage probability in (32), under the assumption of block Nakagami-m fading on the SE-DC channel, can be evaluated as follows:

[3]Under the assumption of Shannon's separation theorem, we set $R_C = 1$ for the transmission chain of the SE-DC channel, including both channel coding and modulation multiplicity. This corresponds, for example, to half-rate coding combined with quaternary phase shift keying (QPSK) modulation.



(a) α_n in terms of p_0.



(b) β_n in terms of λ_n for $p_0 = 0.1$.

Fig. 5: Type-I and Type-II error probabilities as functions of p_0 and λ_n.

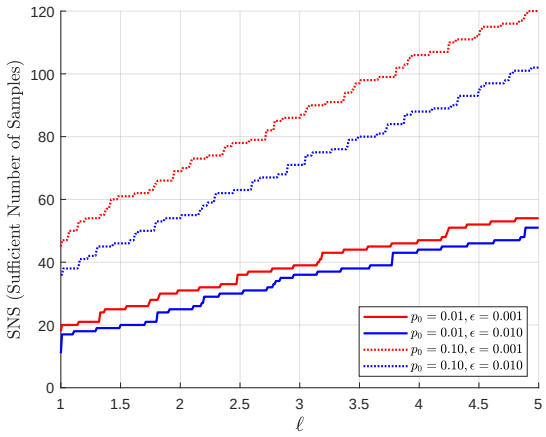
$$\begin{aligned} P_{\text{out}} &= \int_{\Phi^{-1}(0)}^{\Phi^{-1}(1-H_b(\Lambda(p_0,\tau)))} P(\gamma_X)\, d\gamma_X \\ &= \frac{m^m}{\bar{\gamma}_X^m \Gamma(m)} \exp\left(-\frac{m\Phi^{-1}(1 - H_b(\Lambda(p_0,\tau)))}{\bar{\gamma}_X}\right) \\ &\quad \times \sum_{t=0}^{m-1} a_{t,\bar{\gamma}_X} \left(\Phi^{-1}(1 - H_b(\Lambda(p_0,\tau)))\right)^{m-1-t}. \end{aligned} \tag{33}$$

where $a_{t,\bar{\gamma}} = \frac{(-1)^t t! \binom{m-1}{k}}{\left(-\frac{m}{\bar{\gamma}}\right)^{t+1}}$ according to [59, Eq. (2.321.2)].[4]

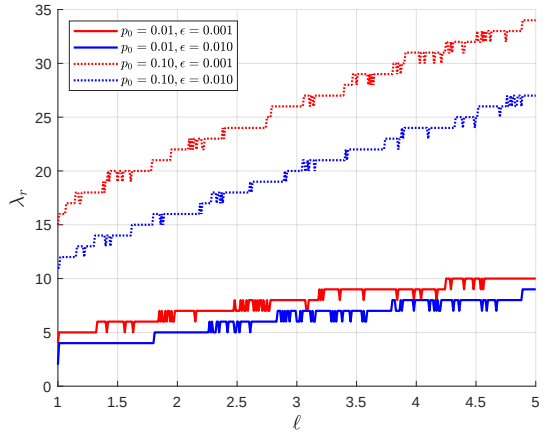
V. Numerical Evaluations

This section presents numerical results in three parts. First, we display the derived bounds for Type-I and Type-II error probabilities, using two examples based on binary linear block codes. Second, we illustrate the computation of the SNS and

[4]As observed in [45, Chapter 8, Q4], channels with memory generally exhibit higher capacity than memoryless channels when operating under the same average bit error rate (BER). Consequently, a time-varying fading channel can achieve greater capacity than a block-fading channel, which is effectively modeled as a static, memoryless AWGN channel within each block. Accordingly, the outage probability derived from the block-fading model serves as an upper bound on the outage probability.

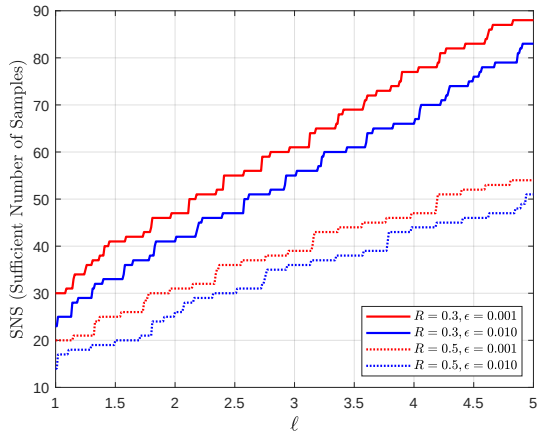


(a) Sufficient number of samples

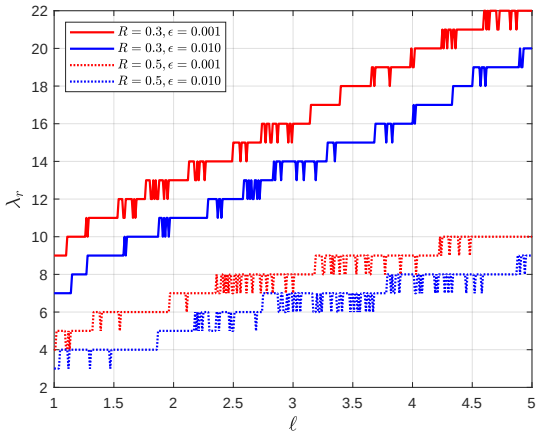


(b) Normalized decision threshold

Fig. 6: Sufficient number of samples and the corresponding random decision threshold λ_r for $p_0 = \{0.01, 0.1\}$ with $R_X = 0.5$.



(a) Sufficient number of samples



(b) Normalized decision threshold

Fig. 7: Sufficient number of samples and the corresponding random decision threshold λ_r for $R_X = \{0.3, 0.5\}$ with $p_0 = 0.01$.

the random decision threshold, as described in Section III. Third, we assess overall system performance in terms of outage probability under Nakagami-m fading, applying the numerical integration method introduced in Section IV.

A. Type-I and Type-II error probability bounds

The exact analytical values of Type-I and Type-II error probabilities, given by (15) and (16), along with the corresponding lower and upper bounds as derived in (17) and (18) are shown in Fig. 5. We consider a Bose–Chaudhuri–Hocquenghem (BCH) code $[31, 11]$ with $d_{\max} = 7$ and $\lambda_n = 15$, as well as a best known linear code (BKLC)[5] $[50, 25]$ with $d_{\max} = 10$ and $\lambda_n = 25$ as shown in Fig. 4(a).[6]

It is clear that a lower correlation, i.e., larger values of p_0, results in a higher Type-I error probability for a fixed threshold λ_n. This occurs because the NP test is more likely to exceed the threshold λ_n when p_0 is larger, causing the DC to more frequently select $\hat{\mathcal{H}} = \mathcal{H}_1$, thereby increasing the probability of Type-I error. Furthermore, as the number of samples n increases from 31 to 50, both the upper bound and the exact values of Type-I error probability decrease. However, the tightness of upper bounds does not vary significantly with n. This shows other parameters, such as λ_n and $\{N_i\}_{i \in [\![1, d_{\max}]\!]}$, also play a significant role in the case of short block lengths.

Fig. 4(b) depicts Type-II error probability with the corresponding lower and upper bounds in terms of $\lambda_n \in (0, \frac{n}{2})$ for block lengths $n = 31$ and 50 with a fixed crossover probability $p_0 = 0.1$. According to the NP test given in (11), as the threshold increases toward $\frac{n}{2}$, the acceptance region for deciding $\hat{\mathcal{H}} = \mathcal{H}_0$ expands. This causes the DC to be more likely to select $\mathcal{H}_0$, thereby increasing Type-II error probability. Additionally, as expected, the exact values of Type-II error probability, the lower and upper bounds decrease as the number of samples increases from $n = 31$ to 50 for the same threshold λ_n.

B. Sufficient number of samples

For a given positive integer ℓ, the computed values of the SNS and the random decision threshold λ_r, which satisfy (19) and (20) for a precision level $\delta = \{\delta_1, \delta_2\} = 10^{-\ell}$, are

[5]The best known linear block code (BKLC) is obtained using the Magma calculator [60].

[6]For the BCH code $[31, 11]$, we obtain $\{N_i\}_{i \in [\![0,7]\!]} = \{1, 31, 465, 4495, 31465, 1669911, 522009, 320199\}$, and for the BKLC $[50, 25]$, we have $\{N_i\}_{i \in [\![0,10]\!]} = \{1, 50, 1225, 19600, 230300, 1864982, 6906666, 11767026, 9435459, 3125556, 203565, 2\}$, as calculated by [60].

TABLE II: Magnitude of the representative metrics for a given number of samples with $R = 0.8$, $p_0 = 0.01$ and $\epsilon = 0.1$.

ϵ_n \ n	100	300	500	700	900	1100	1300	1500	1700	1900
α_n	0.2151	0.1543	0.0803	0.0668	0.0880	0.1059	0.0983	0.1177	0.1300	0.0967
$\bar{\alpha}_n$	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000
$UB_\beta - LB_\beta$	2.7639e-22	1.6470e-66	1.8648e-109	1.0386e-153	2.7652e-199	7.1097e-245	3.8760e-289	0	0	0
$\widetilde{\beta}_n - \widetilde{\beta}_n^*$	3.2700e-22	1.8234e-66	2.0138e-109	1.1085e-153	2.9308e-199	7.4971e-245	4.0692e-289	0	0	0

depicted in Figs. 6 and 7. Moreover, A heuristic brute-force search procedure is described in Algorithm 1 to determine the SNS and λ_r.[7] For given p_0, R, and ϵ, we first set $N = \lceil 2^{n(1-R)} \rceil$ to satisfy the sphere-covering bound as $N \geq 2^{n(1-R)}$ [55].[8] Following this, the values of N_i are generated as $N_i = \binom{n}{i}$ for $i \in [\![0, d_{\max}]\!]$, where they satisfy $N = \sum_{i=0}^{d_{\max}} N_i$.[9] Hence, $d_{\max}$ is the largest value of i required. Next, we compute λ_r using the randomized method in (22) to evaluate $\mathrm{LB}_\beta(n, \lambda_r)$ and $\mathrm{UB}_\beta(n, \lambda_r)$. Finally, the conditions (19) and (20) are verified. The SNS and λ_r are then returned as the feasible smallest n and the selected decision threshold, respectively.

In Fig. 6, the results are presented for $R = 0.5$, with $p_0 \in \{0.01, 0.1\}$ and $\epsilon \in \{0.001, 0.01\}$. As the precision level ℓ increases, i.e., as $\delta = 10^{-\ell}$ decreases from 10^{-1} to 10^{-5}, the SNS values exhibit a noticeable increase, and the selected values of threshold λ_r also increase accordingly. However, it can be seen that λ_r increases more slowly than the SNS values, leading to the ratio $\frac{\lambda_r}{n}$ decreasing as ℓ increases. Moreover, a more stringent Type-I constraint, e.g., $\epsilon = 0.001$, requires a larger SNS in order to satisfy $\alpha_n \leq \epsilon$. In contrast, stronger correlation between source and side information, i.e., smaller p_0, substantially reduces the required SNS; notably, in the highly correlated regime, e.g., $p_0 = 0.01$, an SNS of at most 50 is sufficient even for $\delta = 10^{-5}$, demonstrating rapid convergence to $\widetilde{\beta}_n^*$. The staircase-like patterns observed in both subfigures reflect the inherent discreteness of n and λ_r.

Fig. 7 analyzes a highly correlated source with $p_0 = 0.01$ under compression rates $R_X \in \{0.3, 0.5\}$ and Type-I error constraints $\epsilon \in \{0.001, 0.01\}$. As the precision parameter δ tightens, both the SNS and the threshold λ_r increase to meet the requirements of (19) and (20). A higher rate ($R_X = 0.5$) reduces the required SNS, indicating that lower Type-II error probabilities, closer to $\widetilde{\beta}_n^*$, can be achieved for the same Type-I constraint ϵ. This trend is consistent with the behavior observed in Fig. 6, where a stricter constraint $\epsilon = 0.001$ demands a larger SNS. Notably, in this strongly correlated regime ($p_0 = 0.01$), an SNS below 90 suffices even at $\delta = 10^{-5}$, illustrating rapid convergence to $\widetilde{\beta}_n^*$.

Algorithm 1: A heuristic method for determining SNS and λ_r

Require: $p_0 \in [0, 0.5)$, $R \in [0, 1]$, $\epsilon \in (0, 1)$, $\ell \in [\![1, 5]\!]$

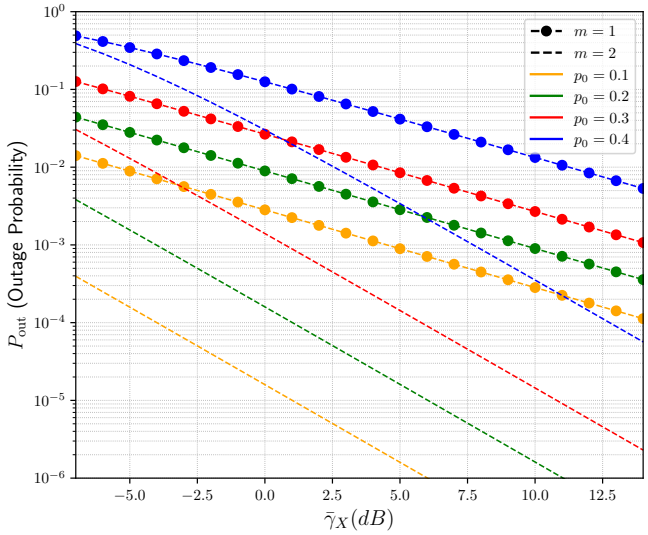
$\delta \leftarrow 10^{-\ell}$;
$n_{\text{start}} \leftarrow 1$;
$n_{\max} \leftarrow 500$;
for $n = n_{start}$ **to** $n_{\max}$ **do**
  $N \leftarrow \lceil 2^{n(1-R)} \rceil$;
  $i \leftarrow 0$, $S \leftarrow 0$;
  while $S < N$ **do**
    if $S + \binom{n}{i} \leq N$ **then**
      $N_i \leftarrow \binom{n}{i}$;
      $S \leftarrow S + N_i$;
      $i \leftarrow i + 1$;
    else
      $N_i \leftarrow N - S$;
      $S \leftarrow N$;
      $d_{\max} \leftarrow i$;
  for $\lambda_n = 0$ **to** $\lfloor n/2 \rfloor$ **do**
    UPDATE $F_n(\lambda_n)$ using (21);
    SELECT λ_r using (22);
    UPDATE $\mathrm{LB}_{\beta_n}(n, \lambda_r)$ using (17);
    UPDATE $\mathrm{UB}_{\beta_n}(n, \lambda_r)$ using (17);
    if $\mathrm{UB}_{\beta_n} - \mathrm{LB}_{\beta_n} \leq \delta$ **then**
      $\widetilde{\beta}_n \leftarrow \mathrm{UB}_{\beta_n}(n, \lambda_r)$;
      if $\widetilde{\beta}_n - \widetilde{\beta}_n^* \leq \delta$ **then**
        SNS $\leftarrow n$;
        return *SNS*, λ_r;

To determine the finite-blocklength regime in which the Type-II error bounds become meaningfully tight, the polynomial prefactor in $\mathrm{LB}_{\beta_n}(n, \lambda_r)$, namely, $\left(8\lambda_r(1 - \frac{\lambda_r}{n})\right)^{-1/2}$, must be negligible relative to the exponentially decaying $\mathrm{UB}_{\beta_n}(n, \lambda_r)$. This requires that the decay rate of the prefactor, on the order of $\mathcal{O}(1/\sqrt{n})$, be dominated by the exponential term. On the other hand, the search procedure must ensure that the selected values of λ_r and SNS satisfy the constraint $\alpha_n \leq \epsilon$, which directly affects the tightness of the Type-II bounds. In particular, smaller values of ϵ generally require larger SNS to meet this constraint. Consequently, the $\mathcal{O}(1/\sqrt{n})$ term provides a lower bound on the complexity of the approximation of $\widetilde{\beta}_n$.

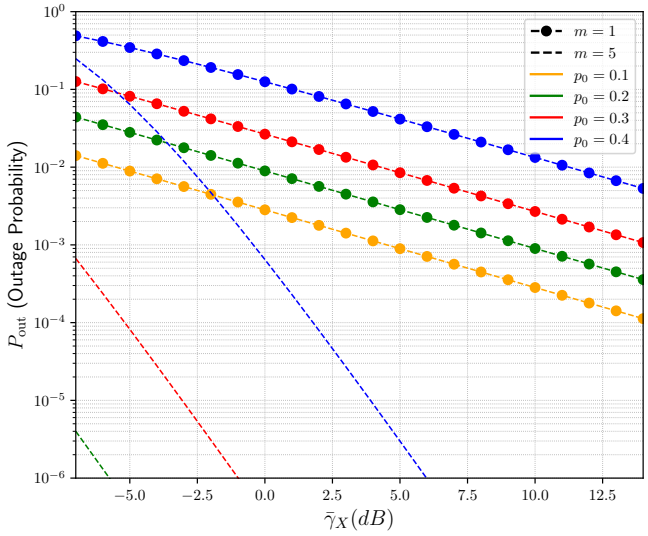
[7] Given the short-length nature of DHT [27], a brute-force search is utilized to identify the sufficient values of n.

[8] Since we treat $d_{\max}$ as the covering radius of a binary linear block code $[n, nR]$ [55, Theorem 2.1.11], reducing N decreases $d_{\max}$ which in turn improves the Receiver Operating Characteristic (ROC) that reflects the decision performance [28].

[9] Note that utilizing the minimum-distance quantizer in (10) implies the weight distribution satisfies $N_i = \binom{n}{i}$ for $i \in [\![0, d_{\max}]\!]$, truncated to the sum exactly to N.

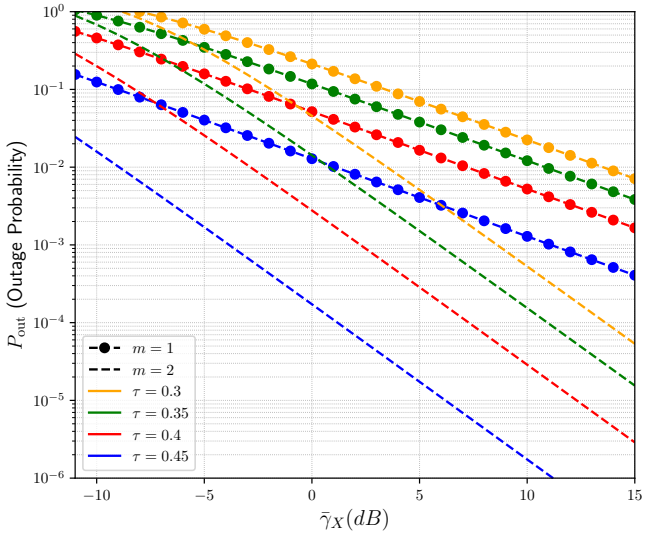


(a) $m = 1, 2$

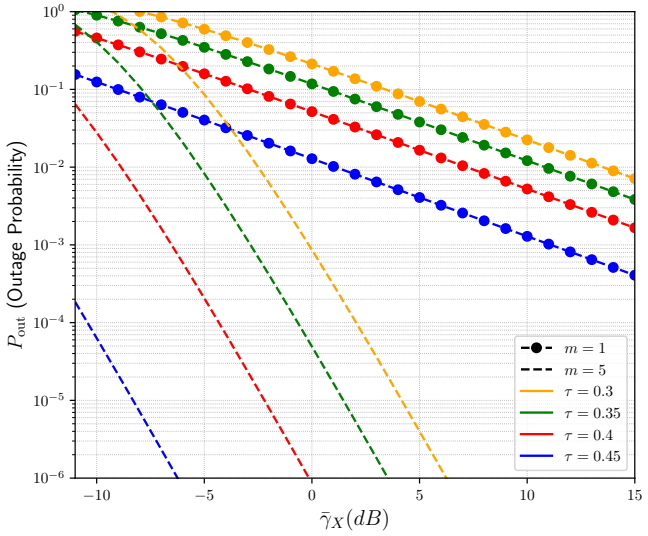


(b) $m = 1, 5$

Fig. 8: The impact of the crossover probability p_0 on the outage probability for a normalized decision threshold $\tau = 0.45$.



(a) $m = 1, 2$



(b) $m = 1, 5$

Fig. 9: The impact of the normalized decision threshold τ on the outage probability for a crossover probability $p_0 = 0.1$.

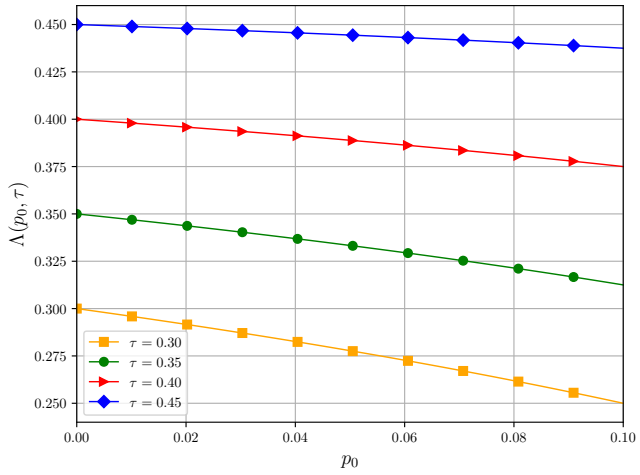
Table II represents key performance metrics for $R = 0.8$, $p_0 = 0.01$, and $\epsilon = 0.1$ across a broad range of sample sizes $n \in \{100, 300, \ldots, 1900\}$. The per-sample Type-I error probability α_n fluctuates around the constraint $\epsilon = 0.1$ due to the randomized selection of the decision threshold λ_r, whereas the average Type-I error $\bar{\alpha}_n$ matches the target value $\epsilon = 0.1$ exactly for each n. Both the interval length $\mathrm{UB}_{\beta_n}(n, \lambda_r) - \mathrm{LB}_{\beta_n}(n, \lambda_r)$ and the gap $\widetilde{\beta}_n - \widetilde{\beta}_n^*$ decrease rapidly with increasing n, approaching zero.

C. Outage Probability

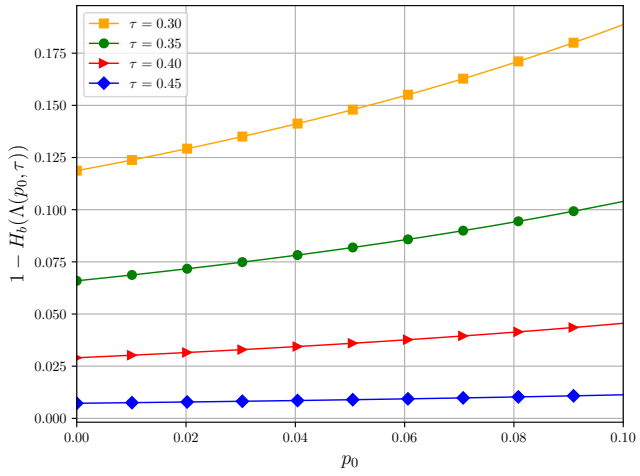
In this section, we compare the outage probability as a function of the average SNR $\bar{\gamma}_X$ for the SE-DC channel, considering the influence of the correlation level between between X and Y through the crossover probability p_0, the normalized decision threshold τ and the fading parameter m.

In Fig. 8, the decision threshold is fixed at $\tau = 0.45$, while the parameter p_0 increases from 0.1 to 0.4. It is obvious that the outage probability increases with higher levels of p_0. This behavior is anticipated because a higher rate R_X must be supported by the fading channel when the side information is less correlated, as described by (29). Fig. 8(a) presents results for $m = 1$, 2, while Fig. 8(b) illustrates the case for $m = 1$ and 5. By comparing the curves, it becomes clear that for the same values of p_0, a lower fading severity, which is indicated by a larger m, results in a lower outage probability.

Fig. 9 presents outage probability results for a scenario in which the crossover probability is at most $p_0 = 0.1$, while the decision threshold τ is varied within the range $(0.3, 0.45)$ at the DC. As in the previous demonstration, Fig. 9(a) displays the results for $m = 1$, 2, and Fig. 9(b) depicts the results for $m = 1$, 5. In these sub-figures, increasing τ leads to a decrease in the outage probability. Notably, there is a significant reduction in the outage probability is achieved for



(a) Distortion transfer function



(b) Minimum Required SE rate R_X

Fig. 10: The impact of τ on the source encoder rate R_X for $p_0 = [0, 0.1]$.

$\tau = 0.45$. Since τ increases in steps of 0.05, the value of $\Lambda(p_0, \tau)$ also monotonically increases for a fixed p_0, as shown in Fig. 10(a).

The minimum required SE rate, described in (29), is illustrated in terms of p_0 in Fig. 10(b). As p_0 increases, the required SE rate also rises, which must be supported by the fading channel; otherwise, an outage occurs. However, this behavior is less visible when $\tau = 0.45$ as shown by the blue line. This reflects as τ approaches 0.5, the minimum required rate tends zero, indicating there are almost no cases of fading variation resulting in an outage.

VI. Conclusions

This paper established a theoretical framework for analyzing DHT problem in a goal-oriented WSN consisting of a single sensor and a single DC. By establishing a direct connection between the decision threshold and the source encoding rate, we defined a rate-threshold region. Within this framework, we have derived the outage probability of the DHT system over a fading SE-to-DC wireless channel, assuming that perfect side information is available at the DC via an error-free link. By characterizing the rate–threshold region, we quantified system performance in the presence of Nakagami fading, considering fading severity, decision threshold, and the correlation between the source and side information as key parameters. It has been shown through numerical calculations that the outage probability decreases as the correlation reduces and as the decision threshold increases within its effective range. The findings provide constructive insights into the design of reliable decentralized decision-making mechanisms in WSNs operating under fading conditions. Future research directions include extending the proposed framework to multi-sensor systems, exploring a broader range of fading models, and developing adaptive decision thresholds for dynamic environments.

Acknowledgment

The authors sincerely thank Prof. Elsa Dupraz (IMT Atlantique University, France) and Prof. Daisuke Umehara (Kyoto Institute of Technology, Japan) for their insightful technical comments and valuable suggestions, which significantly enhanced the quality of this work.

Appendix

In this section, we provide a detailed proof of Prop. 1. According to (15), we have

$$\alpha_n = \frac{1}{N} \sum_{\lambda=\lambda_n+1}^{n} \sum_{i=0}^{d_{\max}} \sum_{u=0}^{\min(\lambda,i)} \Gamma_{\lambda,j,i} p_0^j q_0^{n-j} N_i, \tag{34}$$

where $j = i + \lambda - 2u$ and $0 \le u \le \min(i, \lambda) \le n$, $\Gamma_{\lambda,j,i} = \binom{i}{u}\binom{n-i}{\lambda-u}$ and $q_0 = 1 - p_0$. Note that N_i represents the number of sequences $\mathbf{x}^n$ with Hamming distance i that belong to the decision region of the all-zero codeword $\mathbf{u}_q^k = \mathbf{0}^k$. We define $N = \sum_{i=0}^{d_{\max}} N_i$, where $d_{\max}$ is the maximum Hamming distance for which the source words $\mathbf{x}^n$ are mapped to the all-zero codeword using a binary linear block code.

First, we provide an upper bound for α_n through mathematical manipulation. According to Vandermonde's identity [61], we observe that $\binom{i}{u}\binom{n-i}{\lambda-u} \le \binom{n}{\lambda}$. Consequently, (34) can be initially upper-bounded as follows

$$\le \frac{1}{N} \sum_{\lambda=\lambda_n+1}^{n} \sum_{i=0}^{d_{\max}} \sum_{u=0}^{\min(\lambda,i)} \binom{n}{\lambda} p_0^{(\lambda+i-2u)} q_0^{n-(\lambda+i-2u)} N_i$$

$$= \frac{1}{N} \sum_{\lambda=\lambda_n+1}^{n} \binom{n}{\lambda} p_0^{\lambda} q_0^{n-\lambda} \sum_{i=0}^{d_{\max}} \left(\frac{p_0}{q_0}\right)^i \sum_{u=0}^{\min(\lambda,i)} \left(\frac{q_0}{p_0}\right)^{2u} N_i$$

$$= \frac{1}{N} \sum_{\lambda=\lambda_n+1}^{n} \binom{n}{\lambda} p_0^{\lambda} q_0^{n-\lambda} \underbrace{\sum_{i=0}^{d_{\max}} \left(\frac{p_0}{q_0}\right)^i G_s N_i}_{B}, \tag{35}$$

here $G_s = \left(\frac{\kappa^{\min(\lambda,i)}-1}{\kappa-1}\right)$, which is derived from the formula for a geometric sum with ratio $\kappa = \left(\frac{q_0}{p_0}\right)^2$. On the other hand,

part B can be upper-bounded as follows

$$\begin{aligned} B &= \sum_{i=0}^{d_{\max}} (\frac{1}{\kappa})^{\frac{i}{2}} \Big(\frac{\kappa^{\min(\lambda,i)}-1}{\kappa-1}\Big) N_i \\ &\leq \sum_{i=0}^{d_{\max}} (\frac{1}{\kappa})^{\frac{i}{2}} \Big(\frac{\kappa^{i}-1}{\kappa-1}\Big) N_i \\ &\triangleq \Delta, \end{aligned} \tag{36}$$

where the last inequality holds because $\kappa \geq 1$, and also due to the property of the minimum function, $\min(\lambda, i) \leq i$. Therefore, by substituting (36) into (35), we have

$$\begin{aligned} \alpha_n &\leq \frac{\Delta}{N} \sum_{\lambda=\lambda_n+1}^{n} \binom{n}{\lambda} p_0^{\lambda} q_0^{n-\lambda} \\ &\leq \frac{\Delta}{N} 2^{-nD_b(\frac{\lambda_n}{n}+\frac{1}{n}||p_0)}, \end{aligned} \tag{37}$$

if $p_0 < \frac{\lambda_n+1}{n} < 1$, the last inequality follows from the upper bound on the binomial distribution given in [62, lemma 4.7.2].

Next, we investigate both the lower and upper bounds for Type-II error probability β_n.

Lemma 1 : With the blocklength n and an integer threshold $0 \leq \lambda_n \leq n$, Type-II error probability can be expressed as:

$$\beta_n = (\frac{1}{2})^n \sum_{\lambda=0}^{\lambda_n} \binom{n}{\lambda}. \tag{38}$$

Proof: According to (16), we define

$$W_{\beta_i} = (\frac{1}{2})^n \sum_{\lambda=0}^{\lambda_n} \sum_{u=0}^{\min(\lambda,i)} \binom{i}{u} \binom{n-i}{\lambda-u}. \tag{39}$$

Consider the summation term with respect to u in (39), we separate it into the following two cases:
i) if $\lambda < i$, by utilizing the Vandermonde's identity [61], we have

$$\sum_{u=0}^{\lambda} \binom{i}{u} \binom{n-i}{\lambda-u} = \binom{n}{\lambda}.$$

ii) if $\lambda > i$, by using the binomial formula, we have

$$\begin{aligned} (1+x)^n &= \sum_{\lambda=0}^{n} \binom{n}{\lambda} x^{\lambda}, \\ (1+x)^i (1+x)^{n-i} &= \sum_{u=0}^{i} \binom{i}{u} x^u \sum_{u'=0}^{n-i} \binom{n-i}{u'} x^{u'} \sum_{\lambda=0}^{n} \delta_{\lambda,u+u'} \\ &= \sum_{\lambda=0}^{n} x^{\lambda} \sum_{u=0}^{i} \binom{i}{u} \sum_{u'=0}^{n-i} \binom{n-i}{u'} \delta_{u',\lambda-u} \\ &= \sum_{\lambda=0}^{n} x^{\lambda} \sum_{u=0}^{i} \binom{i}{u} \binom{n-i}{\lambda-u} \Big|_{0\leq\lambda-u\leq n-i}, \end{aligned}$$

where $\delta_{(\cdot,\cdot)}$ denotes the Kronecker delta function. Then, the following equality is derived:

$$\sum_{\substack{u=0 \\ \lambda-(n-i)\leq u\leq\lambda}}^{i} \binom{i}{u} \binom{n-i}{\lambda-u} = \binom{n}{\lambda}, \tag{40}$$

given that $\lambda - (n-i) = \lambda - n < 0$. With $i = 0$ being the minimum possible value for the summation index, the summation in (40) is properly bounded. Therefore, we can express (16) as,

$$\begin{aligned} \beta_n &= \frac{1}{N} (\frac{1}{2})^n \sum_{i=0}^{d_{\max}} \sum_{\lambda=0}^{\lambda_n} \binom{n}{\lambda} N_i \\ &= (\frac{1}{2})^n \sum_{\lambda=0}^{\lambda_n} \binom{n}{\lambda}. \end{aligned}$$

According to Lemma 1 and [62, lemma 4.7.2], when $0 < \frac{\lambda_n}{n} < \frac{1}{2}$, lower and upper bounds for β_n can be established as follows

$$\Big(8\lambda_n\big(1-\frac{\lambda_n}{n}\big)\Big)^{-\frac{1}{2}} 2^{-nD_b(1-\frac{\lambda_n}{n}||\frac{1}{2})} \leq \beta_n \leq 2^{-nD_b(1-\frac{\lambda_n}{n}||\frac{1}{2})}.$$

■

REFERENCES

[1] R. Pohorecki and G. Hill, "Overconfidence in medical research: The role of *p*-value and hypothesis testing paradigm," *J. Cardiol. Clin. Pract.*, vol. 3, no. 2, pp. 1–3, 2020.

[2] A. Mpatziakas, S. Papadopoulos, A. Drosou, and D. Tzovaras, "A hypothesis testing tool for the comparison of different cyber-security mitigation strategies in IoT", *in Proc. IEEE International IoT, Electronics and Mechatronics Conference (IEMTRONICS)*, 2021, pp. 1-7.

[3] Q. H. Zhang and Y. Q. Ni, "A Bayesian hypothesis testing-based statistical decision philosophy for structural damage detection", *Structural Health Monitoring*, vol. 22, no. 4, pp. 2734-2754, 2022.

[4] J. F. Goldberg and S. M. Stahl, "*Hypothesis-testing and crafting patient-specific decision trees*," in Clinical Reasoning and Decision-Making in Psychiatry, Cambridge University Press, 2024.

[5] T. Berger, "Decentralized estimation and decision theory," *in IEEE 7th Spring Workshop on Inf. Theory*, Mt. Kisco, NY, Sept. 1979

[6] U. Ihsan, S. Yan, and R. Malaney, "Location verification for emerging wireless vehicular networks", *IEEE Internet of Things J.*, vol. 6, no. 6, pp. 10261-10272, Dec. 2019.

[7] J. F. Chamberland and V. V. Veeravalli, "Wireless sensors in distributed detection applications," *IEEE Signal Process. Mag.*, vol. 24, no. 3, pp. 16–25, May 2007.

[8] A. Gretton, K. Fukumizu, C. Teo, L. Song, B. Schölkopf, and A. Smola, "A kernel statistical test of independence," *in Advances in Neural Inf. Process. Systems*, vol. 20, 2007, pp. 1–8.

[9] P. Deheuvels, "A nonparametric test for independence," *Ann. Inst. Stat. Univ. Paris (Ann. ISUP)*, vol. 26, no. 2, pp. 29–50, 1981.

[10] B. Efron and V. Petrosian, "A simple test of independence for truncated data with applications to redshift surveys," *Astrophys. J.*, vol. 391, pp. 1–14, May 1992.

[11] A. C. Cameron and P. K. Trivedi, "Tests of independence in parametric models with applications and illustrations," *J. Bus. Econ. Stat.*, vol. 11, no. 1, pp. 29–43, Jan. 1993.

[12] Z. Zhang, Y. Qi, and X. Ma, "Asymptotic independence of correlation coefficients with application to testing hypothesis of independence," *Electron. J. Stat.*, vol. 5, pp. 466–484, 2011.

[13] M. E. Gonzalez, J. F. Silva, M. Videla, and M. E. Orchard, "Data-driven representations for testing independence: Modeling, analysis and connection with mutual information estimation," *IEEE Trans. Signal Process.*, vol. 69, pp. 3099–3113, 2021.

[14] S. Espinosa, J. F. Silva, and S. Céspedes, "A survey on error exponents in distributed hypothesis testing: Connections with information theory, interpretations, and applications," *Entropy*, vol. 26, no. 7, p. 596, Jul. 2024.

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

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

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

[18] Md. 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.

[19] 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.

[20] J. Körner and K. Marton, "How to encode the modulo-two sum of binary sources (Corresp.)," *IEEE Transactions on Information Theory*, vol. 25, no. 2, pp. 219–221, 1979.

[21] E. Haim and Y. Kochman, "Binary distributed hypothesis testing via Körner-Marton coding," *in IEEE Inf. Theory Workshop (ITW)*, Cambridge, UK, 2016, pp. 146-150.

[22] H. M. H. Shalaby and A. Papamarcou, "Multiterminal detection with zero-rate data compression," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 254–267, 1992.

[23] T. S. Han and K. Kobayashi, "Exponential-type error probabilities for multiterminal hypothesis testing," *IEEE Trans. Inf. Theory*, vol. 35, no. 1, pp. 2–14, 1989.

[24] S. Watanabe, "Neyman-Pearson test for zero-rate multiterminal hypothesis testing," in *Proc. 2017 IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 116–120, 2017.

[25] A. Zaidi, "Rate-exponent region for a class of distributed hypothesis testing against conditional independence problems," *IEEE Trans. Inf. Theory*, vol. 69, no. 2, pp. 703–718, 2023.

[26] I. S. Adamou, "Error exponent bounds and practical short-length coding schemes for distributed hypothesis testing," Networking and Internet Architecture [cs.NI]. Ecole nationale supérieure Mines-Télécom Atlantique, Paris, France, 2024.

[27] E. Dupraz, I. S. Adamou, R. Asvadi, and T. Matsumoto, "Practical short-length coding schemes for binary distributed hypothesis testing," in *2024 IEEE Int. Symp. Inf. Theory (ISIT)*, 2024, pp. 2915–2920.

[28] F. Khaledian, R. Asvadi, E. Dupraz, and T. Matsumoto, "Covering codes as near-optimal quantizers for distributed hypothesis testing against independence," in *2024 IEEE Inf. Theory Workshop (ITW)*, 2024, pp. 67–72.

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

[30] S. Salehkalaibar and M. Wigger, "Distributed hypothesis testing over multi-access channels," in *Proc. IEEE Inf. Theory Appl. Workshop (ITA)*, San Diego, CA, USA, 2018, pp. 1–5.

[31] S. Sreekumar and D. Gündüz, "Distributed hypothesis testing over discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 66, no. 4, pp. 2044–2066, Apr. 2020.

[32] S. Salehkalaibar and M. Wigger, "Distributed Hypothesis Testing Based on Unequal-Error Protection Codes," *in IEEE Trans. on Inf. Theory*, vol. 66, no. 7, pp. 4150-4182, Jul. 2020.

[33] Y. Xiang and Y.-H. Kim, "Interactive hypothesis testing with communication constraints," in *Proc. 50th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, pp. 1065–1072, 2012.

[34] G. Katz, P. Piantanida, and M. Debbah, "Collaborative distributed hypothesis testing with general hypotheses," in *Proc. 2016 IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 1705–1709, 2016, doi: 10.1109/ISIT.2016.7541590.

[35] S. Salehkalaibar and M. Wigger, "Distributed hypothesis testing over noisy broadcast channels," *Information*, vol. 12, no. 7, p. 268, 2021.

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

[37] P. Escamilla, M. Wigger, and A. Zaidi, "Distributed hypothesis testing: cooperation and concurrent detection," *IEEE Trans. Inf. Theory*, vol. 66, no. 12, pp. 7550–7564, Dec. 2020.

[38] M. Hamad, M. Wigger, and M. Sarkiss, "Two-hop network with multiple decision centers under expected-rate constraints," in *Proc. IEEE GLOBECOM*, 2021, pp. 1–6.

[39] M. Hamad, M. Wigger, and M. Sarkiss, "Multi-hop network with multiple decision centers under expected-rate constraints," *IEEE Trans. Inf. Theory*, vol. 69, no. 7, pp. 4255–4283, Jul. 2023.

[40] J. Liao, L. Sankar, V. Y. F. Tan, and F. du Pin Calmon, "Hypothesis testing under mutual information privacy constraints in the high privacy regime," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 4, pp. 1058–1071, 2018.

[41] S. Sreekumar, A. Cohen, and D. Gündüz, "Privacy-aware distributed hypothesis testing," *Entropy*, vol. 22, no. 6, Article no. 665, 2020.

[42] S. Sreekumar and D. Gündüz, "Testing against conditional independence under security constraints," in *Proc. 2018 IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 181–185, 2018.

[43] S. Faour, M. Hamad, M. Sarkiss, and M. Wigger, "Testing against independence with an eavesdropper," in *Proc. 2023 IEEE Inf. Theory Workshop (ITW)*, pp. 277–282, 2023.

[44] J. He et al., "A tutorial on lossy forwarding cooperative relaying," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 1, pp. 66–87, 1st Quart., 2019.

[45] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ, USA: Wiley, 2012.

[46] W. Lin, A. Zribi, R. Asvadi, E. Dupraz, L. Li and T. Matsumoto, "Bit-Flipping Helper-Assisted Lossy Communications: Performance Analyses Over Fading Multiple Access Channels," *in IEEE Sensors J.*, vol. 24, no. 14, pp. 23045-23061, Jul. 2024.

[47] S. Song, M. Cheng, J. He, X. Zhou, and T. Matsumoto, "Outage probability of one-source-with-one- helper sensor systems in block rayleigh fading multiple access channels," *IEEE Sensors J.*, vol. 21, no. 2, pp. 2140–2148, Jan. 2021.

[48] S. Espinosa, J. F. Silva, and P. Piantanida, "On the exponential approximation of type II error probability of distributed test of independence," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 7, no. 4, pp. 777–790, 2021.

[49] S. M. Moser, *Information Theory (Lecture Notes)*, 6th ed., Signal and Inf. Proc. Lab., ETH Zürich, Switzerland, and Inst. Commun. Eng., Nat. Chiao Tung Univ., Hsinchu, Taiwan, 2018. Available: http://moser-isi.ethz.ch/scripts.html

[50] T. Matsumoto, A. Zribi, R. Asvadi, E. Dupraz, and W. Lin, "Two-stage successive Wyner–Ziv lossy forward relaying for lossy communications: Rate-distortion and outage probability analyses," *IEEE Trans. Veh. Tech.*, vol. 73, no. 8, pp. 11394–11410, 2024.

[51] S. Salehkalaibar and M. Wigger, "Distributed hypothesis testing with variable-length coding," *IEEE J. Sel. Areas Inf. Theory*, vol. 1, no. 3, pp. 681-694, Nov. 2020.

[52] T. Richardson and R. Urbanke, *Modern coding theory*. Cambridge university press, 2008.

[53] R. Gray, "Vector quantization," *IEEE ASSP Mag.*, vol. 1, no. 2, pp. 4–29, April 1984.

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

[55] G. Cohen, I. Honkala, S. Litsyn, and A. Lobstein, *Covering Codes*. Elsevier, 1997.

[56] A. Goldsmith, *Wireless Communications*. Stanford University, USA: Cambridge University Press, 2005.

[57] M. K. Simon and M. S. Alouini, *Digital communications over fading channels*. Wiley, 2005.

[58] N. Weinberger, Y. Kochman, and M. Wigger, "Exponent trade-off for hypothesis testing over noisy channels," in Proc. IEEE Int. Symp. Inf. Theory, Paris, France, Jul. 2019, pp. 1852–1856.

[59] I. S. Gradshteyn and I. M. Ryzhik, *Table of integrals, series, and products*, 7th ed., A. Jeffrey and D. Zwillinger, Eds. Burlington, MA, USA: Academic, 2007.

[60] W. Bosma, J. Cannon, and C. Playoust, "The magma algebra system i: The user language," *J. of Symbolic Computation*, vol.24, no.3-4, pp.235–265, 1997.

[61] L. Comtet, *Advanced Combinatorics: The art of finite and infinite expansions*. Springer Science and Business Media, 2012.

[62] R. B. Ash, *Information Theory*. Courier Corporation, 2012.