Distributed Hypothesis Testing Under A Covertness Constraint

Ismaila Salihou Adamou, Michèle Wigger
LTCI, Télécom Paris, Institut Polytechnique de Paris, 91120 Palaiseau, France
{ismaila.salihou, michele.wigger}@telecom-paris.fr

***Abstract*—We study distributed hypothesis testing under a covertness constraint in the non-alert situation, which requires that under the null-hypothesis an external warden be unable to detect whether communication between the sensor and the decision center is taking place. We characterize the achievable Stein exponent of this setup when the channel from the sensor to the decision center is a partially-connected discrete memoryless channel (DMC), i.e., when certain output symbols can only be induced by some of the inputs. The Stein-exponent in this case, does not depend on the specific transition law of the DMC and equals Shalaby and Papamarcou's exponent without a warden but where the sensor can send k noise-free bits to the decision center, for k a function that is sublinear in the observation length n. For fully-connected DMCs, we propose an achievable Stein-exponent and show that it can improve over the local exponent at the decision center. All our coding schemes do not require that the sensor and decision center share a common secret key, as commonly assumed in covert communication. Moreover, in our schemes the divergence covertness constraint vanishes (almost) exponentially fast in the observation length n, again, an atypical behavior for covert communication.**

***Index Terms*—Distributed hypothesis testing, covert communication, error exponents.**

I. Introduction

In distributed hypothesis testing, two distant terminals, a remote sensor and a decision center, observe correlated sources, see Figure 1 (without the external warden). The sources' underlying joint distribution depends on one of two hypotheses $\mathcal{H} = 0$ or $\mathcal{H} = 1$ and the goal of the decision center is to guess this underlying hypothesis based on its own observations and communication from the sensor. The performance of the decision center is characterized by two error probabilities [1]: the Type-I error probability of declaring $\hat{\mathcal{H}} = 1$ while $\mathcal{H} = 0$, denoted α_n, and the Type-II error probability of declaring $\hat{\mathcal{H}} = 0$ while $\mathcal{H} = 1$, denoted β_n.

An important line of work in information-theory aims at characterizing the largest exponential decay rate of the Type-II error probability β_n under the constraint that the Type-I error probability α_n lies below a given threshold ϵ [2]–[8]. This largest exponent is commonly referred to as *Stein-exponent*. It was mostly studied in a centralized setup or when communication from the sensor to the decision center is over a noiseless link. Even in this latter case, the Stein-exponent remains generally an open problem, with the notable exception of certain source distributions [3], [9] or when the number of transmitted bits scales only sublinearly in the observation length [10].

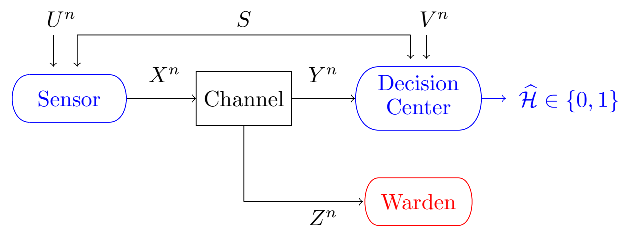

Fig. 1. Distributed hypothesis testing with an external warden.

In a recent line of work [11], [12], we extended the results in [10] to a setup where the sensor communicates to the decision center using a discrete memoryless channels (DMC) a number of times that is sublinear in the observation length. Such a scenario arises naturally in practical scenarios where sensors nowadays are often highly energy-limited. Our results [11] showed a dichotomy of the Stein-exponent with respect to the connectivity of the DMC. Whenever the DMC is fully-connected, i.e., each input symbol can induce each output symbol, then the exponent is no better as the local exponent at the decision center. Communication from the sensor thus becomes useless in terms of Stein-exponent. In contrast, when the DMC is only partially-connected, i.e., some outputs can only be induced by some inputs, then the same exponent is achievable as over a noiseless link [10]. We showed that the same result also remains valid when the sensor can use the DMC n (the observation length) times, but a stringent sublinear cost-constraint is imposed.

In this work, the sensor again communicates to the decision center over n uses of a DMC. However, under $\mathcal{H} = 0$, communication needs to be such that an external warden cannot detect the mere fact that communication is happening. Depending on the practical situation, it might be important that communication be covert under both hypotheses, or only in the case where there is no alert situation, i.e., under $\mathcal{H} = 0$. In this paper, we consider the latter scenario. Implicitly, such a covertness constraint imposes that under $\mathcal{H} = 0$, the sensor employs most of the time a dedicated zero-symbol that also represents the absence of communication. It can be shown that the covertness constraint is weaker than imposing that the channel can only be used a sublinear number of times but stronger than the expected cost constraint. However, as is common for covert communication, here we allow the sensor and the decision center to share a uniform secret key that is unknown to the warden, see Figure 1.

In this article we show that for some partially-connected DMCs, the Stein-exponent is the same as for the noiseless link with sublinear communication setup in [10], irrespective of the size of the shared secret key. Moreover, it can be achieved without using the secret key at all, and the classical divergence covertness measure [13]–[15] can be made to vanish arbitrarily close to exponentially fast in the observation length n. These results are in contrast to classical covert communication where the shared secret key helps in improving performance and the covertness constraint vanishes more slowly [14].

For fully-connected DMCs we show that the Stein-exponent can improve over the local exponent at the decision center. Our coding and testing scheme again does not use the shared secret key and achieves divergence covertness metrics that vanish exponentially fast in the blocklength n. All our results show that the Stein-exponent does not depend on the imposed Type-I error thresholds ϵ.

Our setup has previously been considered in [16], where it was shown that the noiseless link exponent provides an upper bound on the Stein-exponent under a covertness constraint. The proof however only holds when the key length grows logarithmically in n. In this paper we extend this upper bound for arbitrary key lengths. Moreover, the work in [16] did not establish any achievability result improving over the local exponent.

Notation: We mostly follow standard notation. Random variables are denoted by uppercase letters (e.g., X), while their realizations are denoted by lowercase letters (e.g., x). We abbreviate $(x_1, \ldots, x_n)$ by x^n and $(x_{t+1}, \ldots, x_n)$ by x_{t+1}^n. The Hamming weight and Hamming distance are denoted by $w_{\mathrm{H}}(\cdot)$ and $d_{\mathrm{H}}(\cdot, \cdot)$, respectively. We further abbreviate *independent and identically distributed* as *i.i.d.* and *probability mass function* as *pmf*.

We denote by $\pi_{x^n y^n}$ the joint type of the sequences (x^n, y^n), defined as

$$\pi_{x^n y^n}(a,b) \triangleq \frac{n_{x^n,y^n}(a,b)}{n}, \tag{1}$$

where $n_{x^n,y^n}(a,b)$ is the number of occurrences of the pair (a,b) in (x^n, y^n). We use $\mathcal{T}_{\mu}^{(n)}(P_{XY})$ to denote the jointly strongly typical set as defined in [17, Definition 2.9].

II. Problem Setup

Consider the hypothesis testing problem in Figure 1, where the sensor observes a sequence U^n, and a secret key S and communicates to a decision center, which also knows the secret key S in addition to its local observations V^n. Under the null hypothesis

$$\mathcal{H}_0 : (U^n, V^n) \text{ i.i.d. } P_{UV} \tag{2}$$

whereas under the alternative hypothesis

$$\mathcal{H}_1 : (U^n, V^n) \text{ i.i.d. } Q_{UV}. \tag{3}$$

Similarly to [10] we assume that $P_{UV}(u,v) = 0$ whenever $Q_{UV}(u,v) = 0$.

The sensor sends an input sequence

$$X^n = f^{(n)}(U^n, S) \in \mathcal{X}^n \tag{4}$$

over the channel, where $f^{(n)}$ is a chosen encoding function on appropriate domains. The decision center observes the corresponding outputs Y^n of the discrete memoryless channel (DMC) $\Gamma_{YZ|X}$ and the warden the outputs Z^n.

It is assumed that the input alphabet $\mathcal{X}$ contains the 0 symbol. In fact, the covertness constraint imposes that under $\mathcal{H} = 0$ the pmf of the warden's output distribution

$$P_{Z^n|\mathcal{H}=0} = \frac{1}{\mathcal{K}} \sum_{s\in\mathcal{K}} \sum_{u^n\in\mathcal{U}^n} P_U^{\otimes n}(u^n) \Gamma_{Z|X}^{\otimes n}(\cdot|f^{(n)}(u^n, s)), \tag{5}$$

be close to the output distribution $\Gamma_{Z|X=0}^{\otimes n}$ induced by the all-zero input sequence. Specifically, covertness is measured by the Kullback-Leibler divergence:

$$d_n := D(P_{Z^n|\mathcal{H}=0} || \Gamma_{Z|X=0}^{\otimes n}) \tag{6}$$

and is required to stay below a given threshold ϵ_n for sufficiently large blocklengths n. Following standard convention in covert communication, we require that the DMC $\Gamma_{YZ|X}$ satisfies the following conditions:

$$\sum_{x\in\mathcal{X}\backslash\{0\}} \psi(x)\Gamma_{Z|X}(\cdot|x) \neq \Gamma_{Z|X}(\cdot|0), \quad \forall\, \psi(\cdot), \tag{7a}$$

$$\text{Supp}\left(\Gamma_{Z|X}(\cdot|x)\right) \subseteq \text{Supp}\left(\Gamma_{Z|X}(\cdot|0)\right), \quad \forall x \in \mathcal{X}, \tag{7b}$$

$$\text{Supp}\left(\Gamma_{Y|X}(\cdot|x)\right) \subseteq \text{Supp}\left(\Gamma_{Y|X}(\cdot|0)\right), \quad \forall x \in \mathcal{X} \tag{7c}$$

where in the above, $\psi(\cdot)$ indicates a pmf over $\mathcal{X}\backslash\{0\}$.

Based on the received sequence Y^n and its observations V^n, the decision center produces a guess of the hypothesis:

$$\hat{\mathcal{H}} = g^{(n)}(V^n, Y^n) \in \{0, 1\}. \tag{8}$$

The goal is to design encoding and decision functions such that the Type-I error probability

$$\alpha_n \triangleq \Pr\left[\hat{\mathcal{H}} = 1 | \mathcal{H} = 0\right] \tag{9}$$

stays below given threshold $\epsilon \geq 0$ and the Type-II error probability

$$\beta_n \triangleq \Pr\left[\hat{\mathcal{H}} = 0 | \mathcal{H} = 1\right] \tag{10}$$

decays to 0 with largest possible exponential decay, and under hypothesis $\mathcal{H}_0$:

$$\lim_{n\to\infty} d_n = 0. \tag{11}$$

Definition 1. *Given $\epsilon \in [0, 1)$, a miss-detection error exponent $\theta > 0$ is called ϵ-achievable* under a covertness constraint *if there exists a sequence of encoding and decision functions $\{(f^{(n)}, g^{(n)})\}_{n=1}^{\infty}$ satisfying* (11) *and*

$$\overline{\lim_{n\to\infty}}\, \alpha_n \leq \epsilon \tag{12a}$$

$$\underline{\lim_{n\to\infty}}\, -\frac{1}{n} \log \beta_n \geq \theta. \tag{12b}$$

The supremum over all ϵ-achievable miss-detection error exponents θ is denoted $\theta_{\text{covert}}^{\star}(\epsilon)$ and called covert Stein-exponent*.*

The achievable error exponent depends on the transition law $\Gamma_{YZ|X}$ and on the source pmfs P_{UV} and Q_{UV}.

III. Results

Define the following three exponents:

$$E_1 \triangleq \min_{\substack{\pi_{UV}:\\ \pi_U=P_U\\ \pi_V=P_V}} D(\pi_{UV}\|Q_{UV}) \tag{13a}$$

$$E_2(x_1) \triangleq \min_{\substack{\pi_{UV}:\\ \pi_V=P_V,\ \pi_U\notin\bar{\mathcal{P}}_U(x_1)}} D\big(\pi_{UV}\,\|\,Q_{UV}\big) \tag{13b}$$

$$E_3(x_1) \triangleq \min_{\substack{\pi_{UV}:\\ \pi_V=P_V,\ \pi_U\in\bar{\mathcal{P}}_U(x_1)}} D\big(\pi_{UV}\,\|\,Q_{UV}\big) + D(\Gamma_{Y|X=0}\|\Gamma_{Y|X=x_1}), \tag{13c}$$

where $\forall\ x_1 \in \mathcal{X}\setminus\{0\}$, $\bar{\mathcal{P}}_U(x_1)$ is defined in (14) on top of the next page. Notice that $P_U \notin \bar{\mathcal{P}}_U(x_1)$ for all $x_1 \in \mathcal{X}\setminus\{0\}$.

Theorem 1. *Fix $\epsilon \in [0,1)$.*

1) *If the DMC is such that there exists an input $\hat{x} \in \mathcal{X}\setminus\{0\}$ and an output $y^* \in \mathcal{Y}$ satisfying the two conditions:*

$$\Gamma_{Y|X}(y^*|0) > 0 \tag{15a}$$

$$\Gamma_{Y|X}(y^*|\hat{x}) = 0, \tag{15b}$$

then the miss-detection error probability is given by:

$$\theta^*_{\text{covert}}(\epsilon) = E_1. \tag{16}$$

2) *Otherwise*

$$\theta^*_{\text{covert}}(\epsilon) \geq \max_{x_1\in\mathcal{X}\setminus\{0\}} \min\big\{E_2(x_1), E_3(x_1)\big\}. \tag{17}$$

Our achievability results do not require that the sensor and decision center use the shared secret key S.

The covertness constraint d_n can be made to vanish exponentially fast in n for the result in (17) *and arbitrary close to exponentially fast in n for the achievability result in* (16).

Remark 1. *Exponent E_1 is also the Stein-exponent in a distributed setup without warden where the sensor can send a sublinear (in n) number of noise-free bits to the decision center [10]. For partially-connected DMCs satisfying* (15)*, the covertness constraint* (11) *thus has the same impact as limiting communication to a sublinear (in n) number of noise-free bits.*

Remark 2 (Improvement over the Local Test)**.** *Consider the pmf*

$$T_U(u) := \sum_v P_V(v)\, Q_{U|V}(u\mid v), \quad u\in\mathcal{U}. \tag{18}$$

We observe that if $T_U \notin \bar{\mathcal{P}}_U(x_1)$ for all x_1, then $E_2(x_1) = D(P_V\|Q_V)$, which coincides with the local exponent at the decision center without communication from the sensor.

In contrast, we have the strict inequality

$$\theta^*_{\text{covert}}(\epsilon) > D(P_V\|Q_V) \tag{19}$$

for partially-connected DMCs whenever $T_U \neq P_U$ and for fully-connected DMCs whenever $T_U \in \bar{\mathcal{P}}_U(x_1)$ for some x_1.

Proof: We simply prove (19). To this end, start by noticing that for any π_{UV} with marginal $\pi_V = P_V$:

$$D\big(\pi_{UV}\big\|Q_{UV}\big) = D\big(P_V\big\|Q_V\big) + \sum_v P_V(v)\, D\big(\pi_{U|V}(\cdot|v)\big\|Q_{U|V}(\cdot|v)\big). \tag{20}$$

The second summand is strictly positive, except when $\pi_{U|V}(u|v) = Q_{U|V}(u|v)$ for all u and all v with $P_V(v) > 0$. Thus, to deduce that $E_1 > D(P_V\|Q_V)$ or $E_2(x_1) > D(P_V\|Q_V)$ it suffices to prove that $\pi_{U|V} = Q_{U|V}$ is not a valid choice in the minimizations. Inspecting the two minimizations, we see that $\pi_{U|V} = Q_{U|V}$ is not a permissible choice in the minimization of E_1 when $T_U \neq P_U$ and of $E_2(x_1)$ when $T_U \in \bar{\mathcal{P}}_U(x_1)$ for some x_1.

Notice also that $E_3(x_1) > D(P_V\|Q_V)$ by Assumptions (7). ■

We now provide an example for which $\theta^*_{\text{covert}}(\epsilon) > D(P_V\|Q_V)$ even for fully-connected DMCs.

Example 1. *Let P_U be Bern(0.2) and Q_U be Bern(0.7). We assume degenerate observations V =const. at the receiver under both hypotheses.*

Consider a BSC(0.4) for both $\Gamma_{Z|X}$ and $\Gamma_{Y|X}$.

We start by characterizing the set $\bar{\mathcal{P}}_U(1)$. To this end, parametrize $\pi_z(0) = 1-q$ and $\pi_z(1) = q$ and write

$$\begin{aligned} &D(\pi_z\,\|\,\Gamma_{Z|X}(\cdot\mid 0)) - \tfrac{3}{2}D(\pi_z\,\|\,\Gamma_{Z|X}(\cdot\mid \hat{x}))\\ &\quad= (1-q)\log\frac{1-q}{0.6} + q\log\frac{q}{0.4}\\ &\quad\quad -\frac{3}{2}\Big[(1-q)\log\frac{1-q}{0.4} + q\log\frac{q}{0.6}\Big] \qquad (21)\\ &\quad=\frac{1}{2}H_{\text{q}}(q) + \left(\frac{3}{2}-\frac{5}{2}q\right)\log 0.4 + \left(-1+\frac{5}{2}q\right)\log 0.6. \qquad (22)\end{aligned}$$

Above function is strictly concave in q and thus the maximum is obtained by setting the derivative to 0:

$$q^* = \frac{0.6^5}{0.4^5+0.6^5} = 0.884. \tag{23}$$

Plugging back this value into (22) *and simplifying, we obtain:*

$$\begin{aligned}&\max_{\pi_z} D(\pi_z\,\|\,\Gamma_{Z|X}(\cdot\mid 0)) - \tfrac{3}{2}D(\pi_z\,\|\,\Gamma_{Z|X}(\cdot\mid \hat{x}))\\ &\quad= \frac{1}{2}\log\frac{0.4^5+0.6^5}{0.4^2 0.6^2} = 0.306 \qquad (24)\end{aligned}$$

and thus

$$\bar{\mathcal{P}}_U(1) = \left\{\pi_U\colon D(\pi_U\|P_U) \geq \frac{1}{2}\log\frac{0.4^5+0.6^5}{0.4^2 0.6^2}\right\}. \tag{25}$$

Parameterizing the binary type

$$\pi_U(1) = m, \quad \pi_U(0) = 1-m, \tag{26}$$

for $m \in [0,1]$, numerical evaluation shows that $\bar{\mathcal{P}}_U(1)$ contains all types π_U parametrized by $m \geq 0.45$, i.e.,

$$\bar{\mathcal{P}}_U(1) = \{\pi_U\colon \pi_U(1) \geq 0.45\}. \tag{27}$$

We are now ready to evaluate the exponents E_1, E_2, E_3:

$$E_1 = D(P_U\|Q_U) = 0.7706 \tag{28}$$

$$E_2(1) = \min_{\pi_U\notin\bar{\mathcal{P}}_U(1)} D(\pi_U\|Q_U) = 0.2095 \tag{29}$$

$$E_3(1) = D(\Gamma_{Y|X}(\cdot|0)\|\Gamma_{Y|X}(\cdot|1)) = 0.1170. \tag{30}$$

$$\bar{\mathcal{P}}_U(x_1) := \left\{ \pi_U \in \mathcal{P}(\mathcal{U}) : D\left(\pi_U \,\|\, P_U\right) \geq \max_{\pi_z \in \mathcal{P}(\mathcal{Z})} \left[D\left(\pi_z \,\|\, \Gamma_{Z|X}(\cdot|0)\right) - \tfrac{3}{2} D\left(\pi_z \,\|\, \Gamma_{Z|X}(\cdot|x_1)\right) \right] \quad \right\}. \tag{14}$$

IV. Proof of achievability results

We start with the following lemma.

Lemma 1. *Assume that $X^n = 0^n$ with probability $1-\delta_n$ and $X^n = x_1^n$ with probability $\delta_n > 0$, then the warden's divergence satisfies*

$$D\left(P_{Z^n} \,\|\, \Gamma_{Z|X}^{\otimes n}(\cdot|0^n)\right) \leq \delta_n^2 \chi^2\big(\Gamma_{Z|X}^{\otimes n}(\cdot|x_1^n) \,\|\, \Gamma_{Z|X}^{\otimes n}(\cdot|0^n)\big), \tag{31}$$

for $\chi^2(\cdot\|\cdot)$ the chi-squared distance between two distributions:

$$\chi^2(P\|Q) := \sum_{z\in\mathcal{Z}} \frac{\big(P(z)-Q(z)\big)^2}{Q(z)}.$$

Moreover,

$$\begin{aligned}&\chi^2\big(\Gamma_{Z|X}^{\otimes n}(\cdot|x_1^n) \,\|\, \Gamma_{Z|X}^{\otimes n}(\cdot|0^n)\big)\\ &\leq (n+1)^{|\mathcal{Z}|} \max_{\pi\in\mathcal{P}_n(\mathcal{Z})} e^{-nD\left(\pi \| \Gamma_{Z|X}(\cdot|x_1)\right)}\\ &\quad\cdot\left(e^{-n\left(D(\pi\|\Gamma_{Z|X}(\cdot|x_1)) - D(\pi\|\Gamma_{Z|X}(\cdot|0))\right)} - 1\right)^2. \end{aligned}\tag{32}$$

Proof: See [18]. ■

With this result at hand, we now prove the achievability results. Notice that in our proposed schemes, the sensor and decision center do not use their shared secret key S.

A. Achievability of (16)

Fix a small number $\mu > 0$ and let x_1, y^* be as in the theorem.

Choose further a function $k(\cdot)$ satisfying

$$\lim_{n\to\infty} k(n) = \infty \qquad \text{and} \qquad \lim_{n\to\infty} \frac{k(n)}{n} = 0. \tag{33a}$$

We propose a scheme where the sensor uses only the first $k(n)$ channel uses, and sends 0 during the rest of the time.

For ease of notation we will also write k instead of $k(n)$. The scheme works as follows

Sensor: If $U^n \in \mathcal{T}_\mu^{(n)}(P_U)$, the sensor sends $X^k = 0^k$. Otherwise, it sends $X^k = x_1^k$.

Decision Center: If at least one of the channel outputs is y^* and if $V^n \in \mathcal{T}_\mu^{(n)}(P_V)$, then it declares $\hat{\mathcal{H}} = 0$. Otherwise, it declares $\hat{\mathcal{H}} = 1$.

As shown in [16], the Type-I error probability α_n tends to 0 as $n\to\infty$ and $\mu\to 0$, while the Type-II error probability β_n exceeds E_1. It remains to verify that the scheme satisfies the covertness constraint (11).

Analysis of the covertness constraint: Apply Lemma 1 but only to the first k channel uses because the remaining channel uses do not contribute to the divergence. This allows to obtain:

$$\begin{aligned}&D\left(P_{Z^n|\mathcal{H}_0} \,\|\, \Gamma_{Z|X}^{\otimes n}(\cdot|0^n)\right)\\ &= D\left(P_{Z^k|\mathcal{H}_0} \,\|\, \Gamma_{Z|X}^{\otimes k}(\cdot|0^k)\right) &(34)\\ &\leq \delta_n^2\, \chi^2\big(\Gamma_{Z|X}^{\otimes k}(\cdot|x_1^k) \,\|\, \Gamma_{Z|X}^{\otimes k}(\cdot|0^k)\big) &(35)\\ &\overset{(a)}{\leq} (k+1)^{|\mathcal{Z}|} 2|\mathcal{U}| e^{-2n\mu^2} \max_{\pi\in\mathcal{P}_k(\mathcal{Z})} e^{-kD(\pi\|\Gamma_{Z|X}(\cdot|0))}\\ &\quad\cdot\left(e^{-k\left(D(\pi\|\Gamma_{Z|X}(\cdot|x_1)) - D(\pi\|\Gamma_{Z|X}(\cdot|0))\right)} - 1\right)^2, &(36)\end{aligned}$$

where (a) holds by the second part of Lemma 1 and [17, Remark to Lemma 2.12].

Since $k/n \to 0$ as $n\to\infty$ and $\mu > 0$, we can conclude that the right-hand side of (36) tends to 0 almost exponentially fast in n, and the covertness constraint (11) is satisfied.

Notice that our proposed scheme does not require using the shared secret key S.

B. Achievability result in (17)

Pick an input symbol $x_1 \in \mathcal{X}\setminus\{0\}$ and fix a small $\mu > 0$.

We consider the following scheme:

Sensor: If $\pi_U \in \bar{\mathcal{P}}_U(x_1)$, it transmits $X^n = x_1^n$. Otherwise, it transmits $X^n = 0^n$.

Decision Center: If $V^n \in \mathcal{T}_\mu^{(n)}(P_V)$ and $Y^n \in \mathcal{T}_\mu^{(n)}(\Gamma_{Y|X=0})$, it declares $\hat{\mathcal{H}} = 0$. Otherwise, $\hat{\mathcal{H}} = 1$.

We analyze above scheme.

Analysis of covertness constraint: We again employ Lemma 1, but now to the entire blocklength, and we again use the bound

$$\delta_n := \Pr[X^n = x_1^n] \leq (n+1)^{|\mathcal{U}|} \max_{\pi_U \in \bar{\mathcal{P}}_U(x_1)} e^{-nD(\pi_U\|P_U)}. \tag{37}$$

Similarly to above, we obtain inequalities (38)-(41) on the top of the next page.

As we argue next, all three exponential terms in (41) vanish as $n\to\infty$ because the terms multiplying $-n$ are positive. In fact,

$$D(\pi\|\Gamma_{Z|X}(\cdot|x_1)) + 2D(\pi_U\|P_U) > 0 \tag{42}$$

because $P_U \notin \bar{\mathcal{P}}_U(x_1)$ and thus the second divergence is strictly positive for all π_U and π. Similarly, by the definition of the set $\bar{\mathcal{P}}_U(x_1)$:

$$3D(\pi\|\Gamma_{Z|X}(\cdot|x_1)) - 2D(\pi\|\Gamma_{Z|X}(\cdot|0)) + 2D(\pi_U\|P_U)\big) > 0. \tag{43}$$

Finally, to see that also

$$2D(\pi\|\Gamma_{Z|X}(\cdot|x_1)) - D(\pi\|\Gamma_{Z|X}(\cdot|0)) + 2D(\pi_U\|P_U) > 0, \tag{44}$$

distinguish the cases where $D(\pi\|\Gamma_{Z|X}(\cdot|x_1)) > D(\pi\|\Gamma_{Z|X}(\cdot|0))$ or not. In the former case, positivity of (44) is obvious. In the latter case, positivity of (43) implies also positivity of (44), thus concluding that the divergence d_n tends to 0 exponentially fast in the blocklength n.

$$
\begin{aligned}
&D\left(P_{Z^n|\mathcal{H}=0} \,\middle\|\, \Gamma_{Z|X}^{\otimes n}(\cdot|0^n)\right)\\
&\leq \delta^2\, \chi^2\Big(\Gamma_{Z|X}^{\otimes n}(\cdot|x_1^n) \,\big\|\, \Gamma_{Z|X}^{\otimes n}(\cdot|0^n)\Big) && (38)\\
&\leq (n+1)^{2|\mathcal{U}|} \max_{\pi_U \in \bar{\mathcal{P}}_U(x_1)} e^{-2nD(\pi_U\|P_U)} \cdot (n+1)^{|\mathcal{Z}|} \max_{\pi\in\mathcal{P}_n(\mathcal{Z})} \Bigg\{ e^{-nD(\pi\|\Gamma_{Z|X}(\cdot|x_1))}\\
&\qquad \left(e^{-n\big(D(\pi\|\Gamma_{Z|X}(\cdot|x_1))-D(\pi\|\Gamma_{Z|X}(\cdot|0))\big)} - 1\right)^2 \Bigg\} && (39)\\
&= (n+1)^{2|\mathcal{U}|} \max_{\pi_U \in \bar{\mathcal{P}}_U(x_1)} (n+1)^{|\mathcal{Z}|} \max_{\pi\in\mathcal{P}_n(\mathcal{Z})} \Bigg\{ e^{-n\big(3D(\pi\|\Gamma_{Z|X}(\cdot|x_1))-2D(\pi\|\Gamma_{Z|X}(\cdot|0))+2D(\pi_U\|Q_U)\big)}\\
&\qquad -2e^{-n\big(2D(\pi\|\Gamma_{Z|X}(\cdot|x_1))-D(\pi\|\Gamma_{Z|X}(\cdot|0))+2D(\pi_U\|P_U)\big)} + e^{-n\big(D\big(\pi\|\Gamma_{Z|X}(\cdot|x_1)+2nD(\pi_U\|Q_U)\big)\big)} \Bigg\} && (40)\\
&\leq (n+1)^{2|\mathcal{U}|}\, (n+1)^{|\mathcal{Z}|} \max_{\pi_U \in \bar{\mathcal{P}}_U(x_1)} \max_{\pi\in\mathcal{P}_n(\mathcal{Z})} \Bigg\{ e^{-n\big(3D(\pi\|\Gamma_{Z|X}(\cdot|x_1))-2D(\pi\|\Gamma_{Z|X}(\cdot|0))+2D(\pi_U\|P_U)\big)}\\
&\qquad -2e^{-n\big(2D(\pi\|\Gamma_{Z|X}(\cdot|x_1))-D(\pi\|\Gamma_{Z|X}(\cdot|0))+2D(\pi_U\|P_U)\big)} + e^{-n\big(D(\pi\|\Gamma_{Z|X}(\cdot|x_1))+2D(\pi_U\|P_U)\big)} \Bigg\}. && (41)
\end{aligned}
$$

The Type-I and Type-II error probabilities of the proposed scheme can be bounded using standard tools, for instance following arguments similar to those in [12]. Detailed derivations are provided in [18].

V. Proof of Converse Result in (16)

We only provide a sketch of the proof. Details are given in [18]. Similarly, to [16] we prove a converse for the stronger setup where the decision center directly observes the DMC inputs X^n instead of the outputs Y^n. For this stronger setup, we follow the same arguments as in [16, Equations (20)–(39)] to show that for any $\eta \in (0, 1-\epsilon)$ and large n:

$$
\Pr\left[\hat{\mathcal{H}} = 0, X^n \in \tilde{\mathcal{X}}^n | \mathcal{H} = 0\right] \geq 1 - \epsilon - \eta \qquad (45)
$$

$$
\lim_{n\to\infty} \frac{1}{n} \log \left|\tilde{\mathcal{X}}^n\right| = 0, \qquad (46)
$$

where $\tilde{\mathcal{X}}^n$ is an appropriately defined set of low-weight input sequences, see [16, Equation (31)]. In particular, there must be a special sequence $\bar{x}^n \in \tilde{\mathcal{X}}^n$ so that:

$$
\Pr\left[\hat{\mathcal{H}} = 0, X^n = \bar{x}^n \middle| \mathcal{H} = 0\right] \geq \frac{1-\epsilon-\eta}{|\tilde{\mathcal{X}}^n|}. \qquad (47)
$$

For an arbitrary constant $c > 1$, define:

$$
\bar{\mathcal{S}}_{\bar{x}^n} := \Bigg\{ \bar{s} \in \mathcal{K} \text{ s.t :} \quad \Pr\left[\hat{\mathcal{H}} = 0, X^n = \bar{x}^n \middle| \mathcal{H} = 0, S = \bar{s}\right] \geq \frac{1-\epsilon-\eta}{c|\tilde{\mathcal{X}}^n|} \Bigg\}.
$$

It can then be shown that there exists a vanishing sequence η_n so that for any $\bar{s} \in \bar{\mathcal{S}}_{\bar{x}^n}$ (see Lemma 3 in the long version of this article [18]):

$$
\Pr[\hat{\mathcal{H}} = 0, X^n = \bar{x}^n | \mathcal{H} = 1, S = \bar{s}] \geq 2^{-n(\min_{\pi_{UV}} D(\pi_{UV}\|Q_{UV})+\eta_n)} \qquad (48)
$$

where the minimum is over types π_{UV} with marginals P_U and P_V.

As a consequence:

$$
\begin{aligned}
\beta_n \geq \sum_{s\in\bar{\mathcal{S}}_{\bar{x}^n}} &\Pr[S = s]\\
&\cdot \Pr\left[\hat{\mathcal{H}} = 0, X^n = \bar{x}^n | \mathcal{H} = 1, S = s\right] && (49)\\
\geq \Pr[S \in \bar{\mathcal{S}}_{\bar{x}^n}]& \left(2^{-n(\min_{\pi_{UV}} D(\pi_{UV}\|Q_{UV})+\eta_n)}\right). && (50)
\end{aligned}
$$

Proving $\Pr\left[S \in \bar{\mathcal{S}}_{\bar{x}^n}\right] \geq \frac{1-\epsilon-\eta}{|\tilde{\mathcal{X}}^n|}(1-\frac{1}{c})$ concludes the proof.

VI. Conclusion

We studied distributed hypothesis testing under a covertness constraint in the non-alert setting, where an external warden must be unable to detect whether communication is taking place under the null hypothesis. For some partially-connected DMCs, we showed that the Stein exponent coincides with that of the noiseless sublinear-rate communication model of Shalaby and Papamarcou, and is independent of the specific transition law. For the remaining DMCs, we proposed an achievable Stein-exponent and showed that covert communication can strictly improve over the local exponent at the decision center. All proposed schemes operate without any shared secret key and satisfy a strong divergence-based covertness constraint. An interesting direction for future work is to establish a converse for fully connected DMCs case and to study covertness constraints imposed under both hypotheses.

Acknowledgments

The work of the authors has been supported by the ERC under grant agreement 101125691.

References

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

[2] T. S. Han, "Hypothesis testing with multiterminal data compression," *IEEE Transactions on Information Theory*, vol. 33, no. 6, pp. 759–772, 1987.

[3] R. Ahlswede and I. Csiszár, "Hypothesis testing with communication constraints," *IEEE Transactions on Information Theory*, vol. 32, no. 4, pp. 533–542, 1986.

[4] H. Shimokawa, T. S. Han, and S. Amari, "Error bound of hypothesis testing with data compression," in *IEEE International Symposium on Information Theory (ISIT)*, p. 114, 1994.

[5] S. Watanabe, "Neyman-pearson test for zero-rate multiterminal hypothesis testing," *IEEE Transactions on Information Theory*, vol. 64, no. 7, pp. 4923–4939, 2018.

[6] Y. Kochman and L. Wang, "Improved random-binning exponent for distributed hypothesis testing," *IEEE Transactions on Information Theory*, 2025.

[7] N. Weinberger, Y. Kochman, and M. Wigger, "Exponent trade-off for hypothesis testing over noisy channels," in *2019 IEEE International Symposium on Information Theory (ISIT)*, pp. 1852–1856, 2019.

[8] S. Salehkalaibar, M. Wigger, and L. Wang, "Hypothesis testing over the two-hop relay network," *IEEE Transactions on Information Theory*, vol. 65, pp. 4411–4433, Jul. 2019.

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

[10] H. Shalaby and A. Papamarcou, "Multiterminal detection with zero-rate data compression," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 254–267, 1992.

[11] A. Bounhar, M. Sarkiss, and M. Wigger, "A dichotomy for distributed detection with limited communication," in *2025 IEEE Information Theory Workshop (ITW)*, 2025.

[12] A. Bounhar, M. Sarkiss, and M. Wigger, "Distributed detection under stringent resource constraints," *arXiv preprint arXiv:2601.07989*, 2026.

[13] B. A. Bash, D. Goeckel, and D. Towsley, "Limits of reliable communication with low probability of detection on awgn channels," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 9, pp. 1921–1930, 2013.

[14] M. R. Bloch, "Covert communication over noisy channels: A resolvability perspective," *IEEE Transactions on Information Theory*, vol. 62, no. 5, pp. 2334–2354, 2016.

[15] L. Wang, G. W. Wornell, and L. Zheng, "Fundamental limits of communication with low probability of detection," *IEEE Transactions on Information Theory*, vol. 62, no. 6, pp. 3493–3503, 2016.

[16] A. Bounhar, M. Sarkiss, and M. Wigger, "Covert distributed detection over discrete memoryless channels," in *2024 IEEE International Symposium on Information Theory (ISIT)*, pp. 172–177, IEEE, 2024.

[17] I. Csiszár and J. Körner, *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press, 2011.

[18] I. S. Adamou and M. Wigger, "Distributed hypothesis testing under a covertness constraint," *arXiv preprint arXiv:2601.09837*, 2026.