A Dichotomy for Distributed Detection With Limited Communication

Abdelaziz Bounhar[⋆], Mireille Sarkiss[§], and Michèle Wigger[†]
⋆MBZUAI, Paris, France. Email: abdelaziz.bounhar@mbzuai.ac.ae
[§]SAMOVAR, Télécom SudParis, Institut Polytechnique de Paris, 91120 Palaiseau, France
Email: mireille.sarkiss@telecom-sudparis.eu
[†]LTCI, Télécom Paris, Institut Polytechnique de Paris, 91120 Palaiseau, France, Email: michele.wigger@telecom-paris.fr

***Abstract*—This paper identifies the Stein exponent of two distributed detection (binary hypothesis testing) setups with limited communication over a discrete memoryless channel (DMC). In the first setup, the DMC can only be used $k(n)$ times, where $k(n)$ grows sublinearly in the length of the observations n. In the second setup, the DMC can be used n times, however a block-input cost constraint C_n is imposed and C_n grows sublinearly in n. The optimal Stein exponent coincides for both setups and depends on whether the DMC is partially-connected, i.e., one of the output symbols can only be induced by a strict subset of the input symbols, or fully-connected. For *partially-connected* DMCs, the optimal Stein exponent of our setups coincides with the optimal Stein exponent (identified by Han and by Shalaby and Papamarcou) for the scenario where the sensor can communicate a sublinear (in n) number of bits to the decision center and communication is over a noiseless link. In contrast, for *fully-connected* DMCs the optimal Stein exponent collapses and is given by the optimal Stein exponent of the local test at the decision center. In this case, the sensor and the DMC do not help in improving the Stein exponent. Our results hold for general independent and identically distributed sources.**

***Index Terms*—Hypothesis testing, Stein exponents, DMC, sublinear cost constraint.**

I. Introduction

Distributed binary hypothesis testing has been extensively studied in information theory, with a key focus on the setup involving a single sensor and a single decision center. In this framework, both terminals observe correlated sources, whose joint distribution depends on an underlying binary hypothesis. The sensor transmits information to the decision center over either a perfect communication link or a noisy channel. Based on the received symbols and its local observations, the decision center then makes a guess of the underlying hypothesis.

For certain classes of source distributions, Ahlswede and Csiszàr [1], and later Rahman and Wagner [2], established the optimal Stein exponent when communication occurs over a noise-free but rate-limited channel. The Stein exponent quantifies the best achievable exponential decay rate of the probability of error under the alternative hypothesis, given a constraint on the probability of error under the null hypothesis. This type of asymmetric constraint is particularly relevant in alert systems, where keeping the false alarm rate below a certain threshold is sufficient, but minimizing the missed detection probability is critical. Despite these early breakthroughs, the optimal Stein exponent remains unknown for general source distributions in this noise-less link setup. A prominent line of works has established interesting lower bounds [3]–[8]. Similar results have also been obtained when communication takes place over a discrete memoryless channel (DMC) [9], [10] or when security constraints are imposed [11]–[16].

A somehow separate line of work [4], [17], [18] considered the distributed hypothesis testing problem where the sensor can send only a sublinear (in the observation blocklength) number of bits over a noise-free link to the decision center, which we call zero-rate. In this case, the sensor's optimal strategy [4], [17] is to send a single bit indicating whether its observed source sequence is typical according to the distribution under the null hypothesis. The decision center then declares this null hypothesis if also its own observation is typical according to the distribution under the null hypothesis, and it declares the alternative hypothesis in all other cases.

Our contributions: In this work, we reexamine the results in [4], but now in the realm of noisy communication channels. More precisely, we assume that communication takes place over a DMC and while the sensor and the decision centers observe sequences of given blocklength n, the DMC can only be used $k(n)$ times, for $k(n)$ a sequence that grows only sublinearly in n. Our results reveal a dichotomy of the Stein exponent for this setup. When the DMC has an output symbol that can only be induced by some of the inputs but not all of them, then it is possible to achieve the same Stein exponent as reported in [4] for noiseless channels. In contrast, when the channel is fully-connected, the optimal Stein exponent collapses. In this case, it is reduced to the optimal exponent of a local test at the decision center only; the sensor and the DMC thus become completely useless in this case, in the sense that they do not improve the Stein exponent. Notice that our results hold for generally correlated but independent and identically distributed (i.i.d.) observations. Results for a finite number of channel uses are also discussed.

In the second part of the paper, we extend our results to a setup where the DMC can be used n times but a stringent cost constraint is imposed on the channel inputs. Specifically, we assume that the DMC contains a designate zero cost symbol, while all other symbols have positive costs and that the sum of all input costs over a blocklength n is bounded by a block

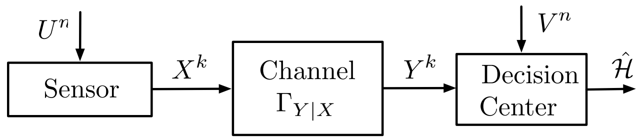

Fig. 1: System Setup

constraint C_n which grows sublinearly in n. As our results show, the optimal Stein exponent for this relaxed setup is the same as for our original setup where the DMC could only be used a sublinear number of times.

Notation: We mostly follow standard notation. In particular, random variables are denoted by upper case letters (e.g., X), while their realizations are denoted by lowercase (e.g. x). We abbreviate $(x_1, \ldots, x_n)$ by x^n. To indicate the Hamming weight, we use $w_{\mathrm{H}}(\cdot)$. We further abbreviate *independent and identically distributed* as *i.i.d.* and *probability mass function* as *pmf.* We denote by $\pi_{u^n v^n}$ the joint type of sequences (u^n, v^n):

$$\pi_{u^n v^n}(a,b) \triangleq \frac{n_{u^n v^n}(a,b)}{n}, \quad (a,b) \in \mathcal{U} \times \mathcal{V}, \tag{1}$$

and $\pi_{u^n}(a)$ the type of u^n. The set of all types of n-length sequences over $\mathcal{U}$ is denoted $\mathcal{P}_n(\mathcal{U})$. The jointly strongly-typical set of all u^n such that $\pi_{u^n}(a) = 0$ if $P_U(a) = 0$ and $|\pi_{u^n}(a) - P_U(a)| \leq \mu$ otherwise, is denoted $\mathcal{T}_\mu^{(n)}(P_U)$.

II. A Sublinear Number of Channel Uses

A. Problem Setup

Consider the distributed hypothesis testing problem in Figure 1 where for a given blocklength n, a sensor observes a sequence U^n and communicates to a decision center with local observations V^n. The distribution of the observations (U^n, V^n) depends on a binary hypothesis $\mathcal{H} \in \{0,1\}$:

$$\text{if } \mathcal{H} = 0: \quad (U^n, V^n) \text{ i.i.d. } \sim P_{UV}; \tag{2a}$$

$$\text{if } \mathcal{H} = 1: \quad (U^n, V^n) \text{ i.i.d. } \sim Q_{UV}, \tag{2b}$$

for given pmfs P_{UV} and Q_{UV} over the product alphabet $\mathcal{U} \times \mathcal{V}$, where we assume that $Q_{UV}(u,v) > 0$ for all $(u,v) \in \mathcal{U} \times \mathcal{V}$. Let P_U and P_V denote the marginal pmfs of P_{UV}. Similarly to the results in [4], we assume that the support of P_{UV} is included in the support of Q_{UV}.

Communication from sensor to decision center takes place over $k(n)$ uses of a discrete memoryless channel DMC with finite input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$ and transition law $\Gamma_{Y|X}$. The number of channel uses grows sublinearly in n:

$$\lim_{n \to \infty} k(n) = \infty \tag{3a}$$

$$\lim_{n \to \infty} \frac{k(n)}{n} = 0. \tag{3b}$$

For ease of notation we will also write k instead of $k(n)$.

The encoder and decoder are thus formalized by two functions $f^{(n)}$ and $g^{(n)}$ on appropriate domains, where $f^{(n)}$ describes how observations U^n are mapped to channel inputs:

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

and $g^{(n)}$ describes how channel outputs Y^k and observations V^n are used to generate the decision $\hat{\mathcal{H}}$:

$$\hat{\mathcal{H}} = g^{(n)}(V^n, Y^k) \in \{0,1\}. \tag{5}$$

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

$$\overline{\lim_{n \to \infty}} \Pr\left[\hat{\mathcal{H}} = 1 | \mathcal{H} = 0\right] \leq \epsilon \tag{6a}$$

$$\underline{\lim_{n \to \infty}} -\frac{1}{n} \log \Pr\left[\hat{\mathcal{H}} = 0 | \mathcal{H} = 1\right] \geq \theta. \tag{6b}$$

The supremum over all ϵ-achievable miss-detection error exponents θ is denoted $\theta^\star_{\text{sublin}}(\epsilon)$.

B. Results

The following theorem determines the largest miss-detection error exponent $\theta_{\text{sublin}}(\epsilon)$, which depends on the source distributions P_{UV} and Q_{UV} as well as on the DMC transition law $\Gamma_{Y|X}$, however not on $\epsilon \in [0,1)$. In particular, the theorem illustrates a dichotomy of this largest exponent with respect to the transition law $\Gamma_{Y|X}$, depending on whether the DMC allows to transmit a symbol with zero error probability. (I.e., there is no detection error when this symbol transmitted, but there can be an error when a different symbol is transmitted.)

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

1) If the DMC is such that there exist two inputs $x_0, x_1 \in \mathcal{X}$ and an output $y^* \in \mathcal{Y}$ satisfying the two conditions:

$$\Gamma_{Y|X}(y^*|x_0) > 0 \tag{7a}$$

$$\Gamma_{Y|X}(y^*|x_1) = 0, \tag{7b}$$

the largest miss-detection error probability is given by:

$$\theta^\star_{\text{sublin}}(\epsilon) = \min_{\substack{\pi_{UV}: \\ \pi_U = P_U \\ \pi_V = P_V}} D(\pi_{UV} \| Q_{UV}). \tag{8}$$

2) Otherwise, it is given by

$$\theta^\star_{\text{sublin}}(\epsilon) = D(P_V \| Q_V). \tag{9}$$

Notice that the largest miss-detection error exponent for case 1) coincides with the largest exponent that is achievable when communication takes place over a noise-free channel that can be used $k(n)$ times, while the largest exponent for case 2) is obtained by a simple test at the decision center without any communication. In this sense, the channels satisfying condition (7) are equally-good for distributed detection as noiseless links in the regime where the number of channel uses $k(n)$ is sublinear in n, while all other channels are completely useless.

Remark 1 (Finite Values of k): Above theorem holds under the assumption that $k \to \infty$. Most of the results however remain valid also for fixed and finite $k \geq 1$. Specifically, the theorem remains valid for all DMCs violating Condition (7) for all triples (x_0, x_1, y^*). For DMCs satisfying (7) for at least one triple (x_0, x_1, y^*), the converse proof trivially remains valid. By inspecting the proof in Subsection II-C, we see that achievability continues to hold for all allowed type-I error probabilities $\epsilon \geq (1-\Gamma_{Y|X}(y^*|x_0))^k$.

Remark 2: Result (9) holds also when the support of P_{UV} is not included in the support of Q_{UV}.

C. Proof of Theorem 1

Case 1): The converse follows from the result in [4], which proves that the largest exponent over a noise-less link cannot exceed the exponent on the right-hand side of (8). To prove achievability, consider the following scheme. Fix a small number $\mu > 0$ and let x_0, x_1, y^* be as in the theorem. Define

$$\gamma_{x_0} \triangleq \Gamma_{Y|X}(y^*|x_0), \tag{10}$$

which by our assumptions is strictly positive.

Sensor: If $U^n \in \mathcal{T}_\mu(P_U)$, send $X^k = x_0^k$. Else, send $X^k = x_1^k$.

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

Denote by α_n and β_n the probability of false alarm and the probability of missed detection, respectively.

Analysis of α_n:

$$\begin{aligned}
&1-\alpha_n && (11)\\
&= \Pr\left[\hat{\mathcal{H}}=0|\mathcal{H}=0\right] && (12)\\
&= \Pr\left[\exists i\in\{1,\dots,k\}\colon Y_i=y^* \text{ and } V^n\in\mathcal{T}_\mu(P_V)|\mathcal{H}=0\right] && (13)\\
&\overset{(a)}{=} \Pr\left[\exists i\in\{1,\dots,k\}\colon Y_i=y^* \text{ and } V^n\in\mathcal{T}_\mu(P_V) \text{ and } X^k=x_0^k\middle|\mathcal{H}=0\right] && (14)\\
&\overset{(b)}{=} \Pr\left[\exists i\in\{1,\dots,k\}\colon Y_i=y^* \text{ and } V^n\in\mathcal{T}_\mu(P_V) \text{ and } X^k=x_0^k, U^n\in\mathcal{T}_\mu(P_U)\middle|\mathcal{H}=0\right] && (15)\\
&= \Pr\left[V^n\in\mathcal{T}_\mu(P_V) \text{ and } U^n\in\mathcal{T}_\mu(P_U)|\mathcal{H}=0\right] \\
&\quad\cdot \Pr\left[X^k=x_0^k|U^n\in\mathcal{T}_\mu(P_U)\right] \\
&\quad\cdot \Pr\left[\exists i\in\{1,\dots,k\}\colon Y_i=y^*|X^k=x_0^k,\mathcal{H}=0\right] && (16)\\
&= \Pr\left[V^n\in\mathcal{T}_\mu(P_V) \text{ and } U^n\in\mathcal{T}_\mu(P_U)|\mathcal{H}=0\right] \\
&\quad\cdot(1-(1-\gamma_{x_0})^k), && (17)
\end{aligned}$$

where (a) holds because the output symbol y^* can occur from input x_0 but not from input x_1 and (b) holds because input $X^k = x_0^k$ is sent only if $U^n \in \mathcal{T}_\mu(P_U)$.

Since γ_{x_0} lies in the half-open interval $(0, 1]$, we have $(1-\gamma_{x_0})^{k(n)}$ that tends to 0 as $n \to \infty$. Moreover, by the weak law of large numbers, irrespective of $\mu > 0$:

$$\lim_{n\to\infty} \Pr\left[V^n\in\mathcal{T}_\mu(P_V) \text{ and } U^n\in\mathcal{T}_\mu(P_U)\middle|\mathcal{H}=0\right] = 1 \tag{18}$$

Plugging these limits into (17), we can conclude that the type-I error probability of our scheme vanishes:

$$\lim_{n\to\infty} \alpha_n = 0. \tag{19}$$

Analysis of β_n and θ: Similarly to above, we have:

$$\begin{aligned}
\beta_n &= \Pr\left[\hat{\mathcal{H}}=0|\mathcal{H}=1\right] && (20)\\
&= \Pr\left[\exists i\in\{1,\dots,k\}\colon Y_i=y^* \text{ and } V^n\in\mathcal{T}_\mu(P_V)\ X^k=x_0^k \text{ and } U^n\in\mathcal{T}_\mu(P_U)|\mathcal{H}=1\right] && (21)\\
&= \Pr\left[V^n\in\mathcal{T}_\mu(P_V) \text{ and } U^n\in\mathcal{T}_\mu(P_U)|\mathcal{H}=1\right] \\
&\quad\cdot \Pr\left[X^k=x_0^k|U^n\in\mathcal{T}_\mu(P_U)\right] \\
&\quad\cdot \Pr\left[\exists i\in\{1,\dots,k\}\colon Y_i=y^*|X^k=x_0^k\right] && (22)\\
&\leq \Pr\left[V^n\in\mathcal{T}_\mu(P_V) \text{ and } U^n\in\mathcal{T}_\mu(P_U)|\mathcal{H}=1\right] && (23)\\
&\leq (n+1)^{|\mathcal{U}||\mathcal{V}|} 2^{-n\min D(\pi_{UV}\|Q_{UV})} && (24)
\end{aligned}$$

where the minimum is over types $\pi_{UV} \in \mathcal{P}_n(\mathcal{U}\times\mathcal{V})$ whose marginals satisfy $|\pi_U(u)-P_U(u)| \leq \mu$ and $|\pi_V(u)-P_V(u)| \leq \mu$.

We can conclude that

$$\lim_{n\to\infty} -\frac{1}{n}\log\beta_n = \min D(\pi_{UV}\|Q_{UV}), \tag{25}$$

where the minimum is now over all pmfs $\pi_{UV} \in \mathcal{P}(\mathcal{U}\times\mathcal{V})$ with marginals satisfying $|\pi_U(u) - P_U(u)| \leq \mu$ and $|\pi_V(u) - P_V(u)| \leq \mu$. Picking $\mu > 0$ sufficiently small, all type-II error exponents smaller than the right-hand side of (8) can be shown to be achievable.

Case 2): We now turn to case 2). Achievability of $\theta^\star_{\text{sublin}}(\epsilon)$ is trivial, because it is achieved by a local test at the decision center, without any communication.

To prove the converse, we fix a sequence of encoding and decision functions $\{f^{(n)}, g^{(n)}\}_{n=1}^\infty$ such that $\overline{\lim}_{n\to\infty} \alpha_n \leq \epsilon$. To analyze the type-II error probability of such a scheme, we introduce the notions of acceptance regions:

$$\mathcal{A}_V(y^k) \triangleq \{v^n\in\mathcal{V}^n\colon g^{(n)}(v^n, y^k) = 0\}, \quad y^k\in\mathcal{Y}^k, \tag{26}$$

and

$$\mathcal{A}_V \triangleq \bigcup_{y^k\in\mathcal{Y}^k} \mathcal{A}_V(y^k). \tag{27}$$

We can then write the miss-detection error probability as:

$$\begin{aligned}
\beta_n &= \Pr\left[\hat{\mathcal{H}}=0|\mathcal{H}=1\right] && (28)\\
&= \sum_{y^k} \Pr\left[Y^k=y^k, V^n\in\mathcal{A}_V(y^k)|\mathcal{H}=1\right] && (29)\\
&= \sum_{y^k}\sum_{v^n\in\mathcal{A}_V(y^k)} \sum_{u^n} \Pr\left[Y^k=y^k, V^n=v^n, U^n=u^n|\mathcal{H}=1\right] && (30)\\
&= \sum_{y^k}\sum_{v^n\in\mathcal{A}_V(y^k)}\sum_{u^n} \Pr\left[V^n=v^n, U^n=u^n|\mathcal{H}=1\right] \\
&\quad\cdot \Pr\left[Y^k=y^k|U^n=u^n\right]. && (31)
\end{aligned}$$

We continue to bound the second probability, by noticing that

$$\Pr\left[Y^k=y^k|U^n=u^n\right] \geq \gamma_{\min}^k, \tag{32}$$

where we define

$$\gamma_{\min} \triangleq \min_{x,y} \Gamma_{Y|X}(y|x), \tag{33}$$

which is strictly positive by assumption. Thus:

$$\beta_n \geq \gamma_{\min}^k \sum_{y^k}\sum_{v^n\in\mathcal{A}_V(y^k)}\sum_{u^n} \Pr\left[V^n=v^n, U^n=u^n|\mathcal{H}=1\right] \tag{34}$$

$$= \gamma_{\min}^k \sum_{y^k} \sum_{v^n \in \mathcal{A}_V(y^k)} \Pr\left[V^n = v^n | \mathcal{H}=1\right] \quad (35)$$

$$\geq \gamma_{\min}^k \cdot \Pr[V^n \in \mathcal{A}_V | \mathcal{H}=1]. \quad (36)$$

Since k grows sublinearly in n, the type-II error probability of the chosen encoding and decision functions is bounded by

$$\overline{\lim_{n\to\infty}} -\frac{1}{n}\log \beta_n \leq \overline{\lim_{n\to\infty}} -\frac{1}{n}\log \Pr[V^n \in \mathcal{A}_V | \mathcal{H}=1], \quad (37)$$

and thus is bounded by the type-II error exponent of a local test at the decision center with acceptance region $\mathcal{A}_V$.

Notice next that under $\mathcal{H}=0$, the V^n sequence falls in $\mathcal{A}_V$ with probability at least

$$\Pr[V^n \in \mathcal{A}_V | \mathcal{H}=0] \geq \Pr[V^n \in \mathcal{A}_V(Y^k) | \mathcal{H}=0] = 1-\alpha_n, \quad (38)$$

and thus, by assumption, the type-I error probability of the local test on V^n with acceptance region $\mathcal{A}_V$ satisfies

$$\overline{\lim_{n\to\infty}} \Pr[V^n \notin \mathcal{A}_V | \mathcal{H}=0] \leq \epsilon < 1. \quad (39)$$

We can now invoke the standard Stein lemma, which states that the type-II error probability of any local test on V^n with type-I error probability bounded away from 1 satisfies

$$\overline{\lim_{n\to\infty}} -\frac{1}{n}\log \Pr[V^n \in \mathcal{A}_V | \mathcal{H}=1] \leq D(P_V \| Q_V), \quad (40)$$

which concludes the proof.

III. Sublinear Resources

We reconsider almost the same setup, but now communication takes place over n channel uses and a stringent, sublinear block-power constraint is imposed. Encoder and decoder are formalized by functions $f^{(n)}$ and $g^{(n)}$:

$$X^n = f^{(n)}(U^n) \in \mathcal{X}^n \quad (41)$$

$$\hat{\mathcal{H}} = g^{(n)}(V^n, Y^n) \in \{0,1\}. \quad (42)$$

The encoding function is required to produce inputs satisfying a stringent resource constraint, described by a bounded and nonnegative cost function $c(\cdot)\colon \mathcal{X} \to \mathbb{R}_0^+$, for which we assume that there exists a unique zero-symbol. We assume:

$$0 \in \mathcal{X} \quad (43)$$

$$c(x) = 0 \quad \text{if, and only if, } x = 0. \quad (44)$$

The stringent resource constraint is described by condition

$$\sum_{i=1}^n c(X_i) \leq C_n, \text{ with prob. } 1, \quad (45)$$

for a given sequence $\{C_n\}$ that grows sublinearly in n:

$$\lim_{n\to\infty} C_n = \infty \quad (46a)$$

$$\lim_{n\to\infty} \frac{C_n}{n} = 0. \quad (46b)$$

Definition 2: Given $\epsilon \in [0,1)$, a miss-detection error exponent $\theta > 0$ is called ϵ-achievable *under stringent resource constraints* $\{C_n\}$ if there exists a sequence of encoding and decision functions $\{(f^{(n)}, g^{(n)})\}_{n=1}^\infty$ satisfying (45) and (6). The supremum over all miss-detection error exponents θ that are ϵ-achievable *under stringent resource constraints* $\{C_n\}$ is denoted $\theta^\star_{\text{str-cost}}(\epsilon, \{C_n\})$.

Theorem 2: For any DMC $\Gamma_{Y|X}$, and sequence of cost-constraints $\{C_n\}$ satisfying (46), we have

$$\theta^\star_{\text{str-cost}}(\epsilon, \{C_n\}) = \theta^\star_{\text{sublin}}(\epsilon), \quad \forall \epsilon \in [0,1). \quad (47)$$

Proof: Let x_0, x_1, y^* be as in (7) and choose a sequence of increasing blocklengths $k(n)$ satisfying

$$\max\{c(x_0), c(x_1)\} \cdot k(n) \leq C_n \quad (48)$$

Notice that $k(n)$ grows sublinearly in n by (46). Achievability of (8) can thus be proved by employing the achievability proof of (8) to the first $k(n)$ channel uses of the DMC.

Achievability of (9) is trivial as it requires no communication and is thus not affected by the resource constraint.

To prove the two converses to (8) and (9), define

$$k'(n) \triangleq \frac{C_n}{c_{\min}}, \quad (49)$$

where $c_{\min} \triangleq \min_{x\neq 0} c(x) > 0$. Notice that $k'(n)$ bounds the number of non-zero symbols in each possible input sequence x^n. The channel input sequence X^n thus lies with probability 1 in the set of all low-weight inputs

$$\tilde{\mathcal{X}}^n := \{x^n \in \mathcal{X}^n \colon w_{\text{H}}(x^n) \leq k'(n)\}. \quad (50)$$

We shall show that the cardinality of $\tilde{\mathcal{X}}^n$ grows sublinearly in n, which immediately establishes the converse result to (8) because the type-II error exponent in this setup cannot exceed the type-II error exponent in the setup of [4], where the sensor noiselessly sends a sublinear number of bits to the detector.

To see that the cardinality of $\tilde{\mathcal{X}}^n$ grows sublinearly in n, notice that this set can be described as the union over all type-classes (i.e., sets of sequences with same type) for types that assign frequency larger or equal to $1-\frac{k'(n)}{n}$ to the 0 symbol. Since the type-class for type $\boldsymbol{\pi}$ is of size at most $2^{nH(\boldsymbol{\pi})}$ and the number of type-classes is bounded by $(n+1)^{|\mathcal{X}|}$, we have:

$$\left|\tilde{\mathcal{X}}^n\right| \leq (n+1)^{|\mathcal{X}|} 2^{n \max_{\boldsymbol{\pi}} H(\boldsymbol{\pi})}, \quad (51)$$

where the maximum is over types satisfying $\boldsymbol{\pi}(0) \geq 1-\frac{k'(n)}{n}$. Since $k'(n)$ grows sublinearly in n, by continuity of entropy:

$$\lim_{n\to\infty} \frac{1}{n}\log\left|\tilde{\mathcal{X}}^n\right| \leq \lim_{n\to\infty}\left[\frac{|\mathcal{X}|}{n}\log(n+1) + \max_{\substack{\boldsymbol{\pi}\colon \\ \boldsymbol{\pi}(0)\geq 1-\frac{k'(n)}{n}}} H(\boldsymbol{\pi})\right] = 0. \quad (52)$$

To prove the converse to (9), we notice that for any two input sequences $x_1^n, x_2^n \in \tilde{\mathcal{X}}^n$ and output sequence $y^n \in \mathcal{Y}^n$:

$$\gamma_Q^{2k'(n)} \leq \frac{\Pr\left[Y^n = y^n | X^n = x_1^n\right]}{\Pr\left[Y^n = y^n | X^n = x_2^n\right]} \leq \gamma_Q^{-2k'(n)}, \quad (53)$$

where

$$\gamma_Q \triangleq \min_{\substack{x_1,x_2,y:\\ x_1\neq x_2}} \frac{\Gamma_{Y|X}(y|x_1)}{\Gamma_{Y|X}(y|x_2)} \in (0,1]. \tag{54}$$

Trivially, this implies also the bounds:

$$\gamma_Q^{2k'(n)} \leq \frac{\Pr\left[Y^n=y^n|U^n=u_1^n\right]}{\Pr\left[Y^n=y^n|U^n=u_2^n\right]} \leq \gamma_Q^{-2k'(n)}, \tag{55}$$

for any input and output sequences $u_1^n, u_2^n \in \mathcal{U}^n$ and $y^n \in \mathcal{Y}^n$.

Noting that the conditional law of Y^n given U^n is the same under both hypotheses and also that the Markov chain $Y^n \to X^n = f^{(n)}(U^n) \to V^n$ holds under both hypotheses, we can conclude that:

$$\begin{aligned}
&\Pr\left[Y^n=y^n|U^n=u^n,\mathcal{H}=1\right]\\
&= \Pr\left[Y^n=y^n|U^n=u^n,\mathcal{H}=0\right] && (56)\\
&= \sum_{\tilde{u}^n} \Pr[U^n=\tilde{u}^n|V^n=v^n,\mathcal{H}=0]\\
&\qquad \cdot \Pr\left[Y^n=y^n|U^n=u^n,\mathcal{H}=0\right] && (57)\\
&\overset{(a)}{\geq} \sum_{\tilde{u}^n} \Pr[U^n=\tilde{u}^n|V^n=v^n,\mathcal{H}=0]\\
&\qquad \cdot \Pr\left[Y^n=y^n|U^n=\tilde{u}^n,\mathcal{H}=0\right]\cdot \gamma_Q^{2k'(n)} && (58)\\
&\overset{(b)}{=} \sum_{\tilde{u}^n} \Pr[U^n=\tilde{u}^n|V^n=v^n,\mathcal{H}=0]\\
&\qquad \cdot \Pr\left[Y^n=y^n|U^n=\tilde{u}^n,V^n=v^n,\mathcal{H}=0\right]\cdot \gamma_Q^{2k'(n)} && (59)\\
&= \Pr\left[Y^n=y^n|V^n=v^n,\mathcal{H}=0\right]\gamma_Q^{2k'(n)}, && (60)
\end{aligned}$$

where Inequality (a) holds by (55) and Equality (b) because $V^n \to U^n \to Y^n$ forms a Markov chain.

Define the acceptance regions

$$\mathcal{A}_V(y^n) \triangleq \{v^n \in \mathcal{V}^n \colon g^{(n)}(v^n,y^n)=0\}, \quad y^n \in \mathcal{Y}^n. \tag{61}$$

Similarly to the converse proof to (9)—but where y^k needs to be replaced by y^n—we have:

$$\begin{aligned}
\beta_n &= \sum_{y^n}\sum_{v^n\in\mathcal{A}_V(y^n)}\sum_{u^n} \Pr\left[V^n=v^n,U^n=u^n|\mathcal{H}=1\right]\\
&\qquad \cdot \Pr\left[Y^n=y^n|U^n=u^n,\mathcal{H}=1\right] && (62)\\
&\overset{(a)}{\geq} \sum_{y^n}\sum_{v^n\in\mathcal{A}_V(y^n)}\sum_{u^n} \Pr\left[V^n=v^n,U^n=u^n|\mathcal{H}=1\right]\\
&\qquad \cdot \Pr\left[Y^n=y^n|V^n=v^n,\mathcal{H}=0\right]\gamma_Q^{2k'(n)} && (63)\\
&\geq \sum_{y^n}\sum_{v^n\in\mathcal{A}_V(y^n)} \Pr\left[V^n=v^n|\mathcal{H}=1\right]\\
&\qquad \cdot \Pr\left[Y^n=y^n|V^n=v^n,\mathcal{H}=0\right]\gamma_Q^{2k'(n)}. && (64)
\end{aligned}$$

where Inequality (a) holds by (60).

By assumption, $\gamma_Q^{2k'(n)}$ is subexponential in n, and thus

$$\begin{aligned}
&\varlimsup_{n\to\infty} -\frac{1}{n}\log\beta_n\\
&\leq \varlimsup_{n\to\infty} -\frac{1}{n}\log\sum_{y^n}\sum_{v^n\in\mathcal{A}_V(y^n)} \Pr\left[V^n=v^n|\mathcal{H}=1\right]\\
&\qquad \cdot \Pr\left[Y^n=y^n|V^n=v^n,\mathcal{H}=0\right]. && (65)
\end{aligned}$$

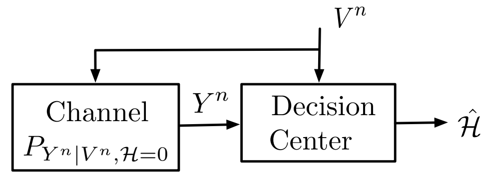

Fig. 2: Derived binary hypothesis test with channel $P_{Y^n|V^n,\mathcal{H}=0}$ used under both hypotheses.

The right-hand side of above equation corresponds to the miss-detection error exponent of a local detection problem of the form in Figure 2, where the decision center observes V^n, which is i.i.d. P_V or Q_V depending on the two hypotheses, and the outcome $\tilde{Y}^n$, which is generated from V^n according to the same law $\Pr\left[Y^n=y^n|V^n=v^n,\mathcal{H}=0\right]$, irrespective of the hypothesis. So, we encounter a randomized binary hypothesis test with observation V^n and acceptance region $\mathcal{A}_V(Y^n)$. We notice that the probability of correct detection under $\mathcal{H}=0$ of the randomized test satisfies

$$\begin{aligned}
&\sum_{y^n}\sum_{v^n\in\mathcal{A}_V(y^n)} \Pr\left[V^n=v^n|\mathcal{H}=0\right]\\
&\qquad \cdot \Pr\left[Y^n=y^n|V^n=v^n,\mathcal{H}=0\right] \geq 1-\epsilon. && (66)
\end{aligned}$$

By standard Stein exponent arguments, the limit on the right-hand side of (65) is thus upper bounded by $D(P_V\|Q_V)$. ■

IV. Conclusion and Outlook

We derived the optimal Stein exponent for two distributed hypothesis testing problems where communication resources are scarce in the sense that either the channel can be used only a much smaller number of times than the length of the observations or in the sense that the transmitted signal is subject to a stringent block-power constraint that grows sublinearly in the observation length. Our results revealed a dichotomy with respect to the DMC $\Gamma_{Y|X}$ over which communication takes place. If the DMC is fully-connected then the optimal Stein exponent is no better than the exponent of a local test at the decision center. The sensor and the communication over the DMC thus do not lead to a larger exponent in this case. If in contrast the DMC is only partially-connected, then the optimal Stein exponent coincides with the exponent that is achievable when the DMC is noise-free.

An interesting research direction is to consider in the future concerns more relaxed resource constraints such as expected resource constraints. In this case we expect larger exponents to be achievable when the DMC is fully-connected. Proving matching converses however seems to require additional tools. Extensions to multiple sensors are also of practical interest. It might be possible to obtain them with the presented tools.

Acknowledgments

The authors acknowledge funding from the ERC under Grant Agreement 101125691 and from the PEPR "Networks of the Future" Initiative.

References

[1] R. Ahlswede and I. Csiszár, "Hypothesis testing with communication constraints," *IEEE Trans. Inf. Theory*, vol. 32, pp. 533–542, Jul. 1986.

[2] M. S. Rahman and A. B. Wagner, "On the optimality of binning for distributed hypothesis testing," *IEEE Trans. Inf. Theory*, vol. 58, pp. 6282–6303, Oct. 2012.

[3] T. S. Han, "Hypothesis testing with multiterminal data compression," *IEEE Trans. Inf. Theory*, vol. 33, pp. 759–772, Nov. 1987.

[4] 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.

[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] E. Haim and Y. Kochman, "Binary distributed hypothesis testing via korner-marton coding," in *Proc. IEEE Info. Theory Work. (ITW)*, 2016.

[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] Y. Kochman and L. Wang, "Improved random-binning exponent for distributed hypothesis testing," 2023.

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

[10] S. Salehkalaibar and M. Wigger, "Distributed hypothesis testing based on unequal-error protection codes," *IEEE Trans. Inf. Theory*, vol. 66, pp. 4150–41820, Jul. 2020.

[11] M. Mhanna and P. Piantanida, "On secure distributed hypothesis testing," in *2015 IEEE International Symposium on Information Theory (ISIT)*, pp. 1605–1609, 2015.

[12] 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 Transactions on Information Forensics and Security*, vol. 13, no. 4, pp. 1058–1071, 2018.

[13] S. B. Amor, A. Gilani, S. Salehkalaibar, and V. Y. F. Tan, "Distributed hypothesis testing with privacy constraints," in *2018 International Symposium on Information Theory and Its Applications (ISITA)*, pp. 742–746, 2018.

[14] S. Sreekumar and D. Gündüz, "Testing against conditional independence under security constraints," in *2018 IEEE International Symposium on Information Theory (ISIT)*, pp. 181–185, 2018.

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

[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, 2024.

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

[18] P. Escamilla, M. Wigger, and A. Zaidi, "Distributed hypothesis testing: cooperation and concurrent detection," *IEEE Transactions on Information Theory*, vol. 66, no. 12, pp. 7550–7564, 2020.