Multi-Sensor Distributed Hypothesis Testing in the Low-Power Regime

Cécile Bouette, Michèle Wigger

Abstract

We characterize the Stein-exponent of a distributed hypothesis testing scenario where two sensors transmit information through a memoryless multiple access channel (MAC) subject to a sublinear input cost constraint with respect to the number of channel uses and where the decision center has access to an additional local observation. Our main theorem provides conditions on the channel and cost functions for which the Stein-exponent of this distributed setup is no larger than the Stein-exponent of the local test at the decision center. Under these conditions, communication from the sensors to the decision center is thus useless in terms of Stein-exponent. The conditions are satisfied for additive noise MACs with generalized Gaussian noise under a p-th moment constraint (including the Gaussian channel with second-moment constraint) and for the class of fully-connected (where all inputs can induce all outputs) discrete memoryless multiple-access channels (DMMACs) under arbitrary cost constraints. We further show that for DMMACs that are not fully-connected, the Stein-exponent is larger and coincides with that of a setup with zero-rate noiseless communication links from either both sensors or only one sensor, as studied in [1].

Index Terms

Hypothesis testing, sublinear input cost constraint, Stein's error exponent, multiple-access channels.

I. Introduction

Binary hypothesis testing refers to a problem that involves determining which of two joint distributions governs observed data. This a standard problem encountered in many sensor applications and as such also in sensor networks and the Internet of Things (IoT). A specificity of the IoT is that sensors have extremely stringent power budgets because their batteries are supposed to last for decades. Recent 6G standards tighten the requirement on the sensor's power consumptions even further, in particular under the framework of Ambient IoT [2]. Our goal in this paper is to study the performance of distributed binary hypothesis testing under stringent power budgets at the sensors. In particular, we will impose stringent power constraints on the signals that are transmitted by the sensors.

Formally, a sensor network consists of several sensors, all with local observations, and at least one decision center that is tasked to decide on one of two hypotheses $H \in \{0, 1\}$ based on the information that is communicated from the sensors and possibly also based on own local observations. The goal of the sensors and the decision center is to minimize the two

Cécile Bouette and Michèle Wigger are with LTCI, Télécom Paris, Institut Polytechnique de Paris, France (email: cecile.bouette@telecom-paris.fr; michele.wigger@telecom-paris.fr). This work was supported by the ERC under Grant Agreement 101125691.

types of error: the type-I error probability which refers to the probability that the decision center declares $\hat{H} = 1$ while the true hypothesis underlying the observations is $H = 0$; and the type-II error probability which refers to the probability that the decision center declares $\hat{H} = 0$ while the true hypothesis underlying the observations is $H = 1$. We are interested in asymmetric situations where type-II error probabilities are more harmful than type-I error probabilities because hypothesis $H = 0$ describes a normal situation while $H = 1$ describes an alert situation such as a tsunami or avalanche event. In such scenarios, type-I errors are often referred to false-alarm events and type-II errors as miss-detection events. To capture the asymmetry in the hypothesis test, the Stein-exponent [3] measures the largest possible decay rate to zero (in the number of observations n) of the type-II error probability under a fixed threshold $\epsilon \in [0, 1)$ on the type-I error probability.

The Stein-exponent is well-known for local tests where all the observations are locally available at the decision center [3], in which case it does not depend on the allowed type-I error threshold $\epsilon \in [0, 1)$. For most distributed scenarios where part of the observations are located at remote sensors and first need to be communicate to the decision center, the Stein-exponent is however still unknown. Notable exceptions are, for example, [1, 4, 5, 6, 7, 8, 9, 10], which make different assumptions on the communication from the sensors to the decision center.

A canonical line of work [4, 5, 8, 11, 12, 13, 14, 15, 16, 17] studied the Stein-exponent in a communication scenario where the single sensor can send Rn bits to the decision center over a noise- and error-free link. For a small class of source distributions the Stein-exponent has been determined, but it remains open in general and only upper and lower bounds are available. Extensions where also proposed for multi-sensors networks [8, 16, 17, 18] or under a variable-length coding framework [7, 19, 20].

A line of work that is more closely related to the present work, studied the Stein-exponent again in a noiseless-link setup, but under the constraint that the number of bits communicated from the sensor to the decision center grows only sublinearly in n. This setting is commonly known as *noiseless zero-rate communication*, and [1, 5] characterized the Stein-exponent for a broad class of source distributions. In particular, Han [5] considered the single-sensor setup where the sensor can only send a single bit to the decision center. The optimal strategy in this scenario is for the decision center to decide on $\hat{H} = 0$ if, and only if, both its own observation and the sensor's observation are *marginally typical* under the distribution corresponding to hypothesis $H = 0$. In this strategy, the sensor only sends the binary outcome of its typicality test to the decision center. Shalaby and Papamarcou [1] proved that this simple 1-bit communication strategy and the corresponding decision rule are optimal and achieve the Stein-exponent even in scenarios where the sensor can send a sublinear number of bits to the decision center. They further also extended the result to multi-sensor setups [1], and more recently, similar zero-rate results have been obtained in the quantum domain [21].

In this work we will relax the assumption of noise-free communication and instead consider general classes of memoryless multiple-access channels (MAC) from the sensors to the decision center. In the information-theoretic literature, noisy communication channels for hypothesis testing have been explored in [6, 22], presenting general lower bounds and exact characterizations of the Stein-exponent for special source distributions. In these works, channel input sequences have same length as the observations and no cost constraints are imposed.

A recent work [23] also considered distributed hypothesis testing over noisy communication links, but under the assumption that the input sequences either have to be much shorter than the observations or they are subject to stringent block-input cost constraints that grow only sublinearly in the number of observations. These assumptions are motivated by stringent resource constraints at the sensors, as mentioned in the first paragraph. More specifically, the work in [23] focuses on a single-sensor setup where communication is over a discrete-memoryless channel (DMC). It is shown that under the two mentioned input constraints (sublinear number of inputs or cost), the Stein-exponent only depends on whether the DMC is fully-connected, i.e., each input symbol induces each output symbol with positive probability, or not. For fully-connected DMCs, the Stein-exponent completely degrades to the Stein-exponent of the local test at the decision center rendering the sensor and its communication useless. For partially-connected DMCs, in contrast, the Stein-exponent is equal to the Stein-exponent of a setup where communication from the sensor to the decision center takes place over a zero-rate *noiseless* link as studied in [1]. It was thus shown that imposing stringent constraints on the input sequence strongly degrades the Stein-exponent over DMCs, and for certain channels even renders communication useless.

In this paper, we extend the results in [23] to scenarios with multiple sensors which can communicate to the decision center over a continuous or discrete memoryless MAC. We assume that the inputs from each of the two sensors are subject to stringent block-input cost constraints that grow only sublinearly in the number of observations n. For a large class of MACs, we show a complete degradation of the Stein-exponent in this setup to the Stein-exponent of a local test, rendering the communication and the sensors useless. As we further show, this is in particular the case for the class of continuous-valued MACs with p-th order generalized Gaussian noise subject to a p-th moment block-input constraint and for fully-connected DMMACs, i.e., DMMACs where each input pair induces each output pair with positive probability. For other connectivity patterns of the DMMAC, the Stein-exponent over DMMACs coincides with the Stein-exponent of a communication scenario where either one or both sensors can communicate to the decision center over zero-rate noiseless links as in [1]. These results for DMMACs are obtained under arbitrary block-input cost constraints that grow sublinearly in n.

To summarize, in this paper we show that for a multi-sensor distributed hypothesis testing setup the Stein-exponent can completely collapse to the local Stein-exponent when stringent sublinear cost-constraints are imposed on the channel input sequences. Exceptions are channels where some outputs can only be induced by certain input pairs, where the degradation can be milder and Stein-exponents as in noiseless zero-rate communication scenarios remain achievable. Here, again, the type of communication scenario (one or two noiseless links) to consider depends on the structure (connectivity) of the original MAC. Our results thus extend the previous results in [23] in two directions: multiple sensors and continuous channels. Both extensions cannot be obtained from the results in [23] but instead required new proof elements.

The rest of this paper is organized as follows. Section II introduces notation and Section III describes the general problem setup. Section IV presents our main results, which state following.

- Conditions for continuous and discrete memoryless MACs and associated cost-constraints are presented under which the Stein-exponent degrades to the local Stein-exponent at the decision center.
- The additive noise MAC with generalized Gaussian noise of parameter $p > 0$ under a p-th moment cost constraints satisfies above conditions, and thus its Stein-exponent degrades to the local Stein-exponent.
- The Stein-exponent for the general class of DMMACs under arbitrary cost constraints. Its Stein-exponent depends on the

connectivity of the DMMAC, ranging from the local Stein-exponent to the Stein-exponent of a zero-rate noiseless link scenario from either both or only a single sensor.

The subsequent Sections V and VI present the proofs of our main results, and Section VII presents concluding remarks.

II. Notation

Random variables are denoted by uppercase letters like U, while their realizations are denoted by lowercase letters like u. We abbreviate $(u_1,\ldots,u_n)$ by u^n and the subvector $(u_i,\ldots,u_j), i\leq j$, by u_i^j. Depending on the context, if a vector u_l is itself indexed by l, its subvector $(u_{l,i},\ldots,u_{l,j}), i\leq j$, is denoted by $u_{l,i}^j$. The set of all real numbers is denoted $\mathbb{R}$ and the set of nonnegative real numbers by $\mathbb{R}^+$. We use P_U to denote the law (also called distribution) of the random variable U and P_{U^n} that of the random vector U^n. The associated Lebesgue-Stieltjes measure [24, Section 3.11] of P_U is denoted by $\mathrm{d}P_U$. When it exists, the probability density function corresponding to the distribution P_U is denoted p_U. We denote the product of measures by $\otimes$ and the distribution of an independent and identically distributed sequence of random variables U^n by $P_U^{\otimes n}$. We abbreviate *independent and identically distributed* as *i.i.d.*. The set of all probability distributions over the set $\mathcal{U}$ is denoted by $\mathcal{M}(\mathcal{U})$. Given a sequence $u^n\in\mathcal{U}^n$, we denote its type [25, Ch. 11] by

$$\pi_{u^n}(a)\triangleq\frac{|\{i\in\{1,\ldots,n\}:u_i=a\}|}{n},\qquad a\in\mathcal{U}. \tag{1}$$

The set of all possible n-types is denoted by $\mathcal{P}_n(\mathcal{U})=\{\pi_{u^n}:u^n\in\mathcal{U}^n\}$. We denote the type class [25, Ch. 11] of π_{u^n} as $\mathcal{T}_n(\pi_{u^n})=\{\tilde{u}^n\in\mathcal{U}^n:\pi_{\tilde{u}^n}=\pi_{u^n}\}$. For $\mu>0$, the strongly-typical set [25, Ch. 10] for a given a random variable U is denoted by $\mathcal{T}_\mu(P_U)$ and is defined as the set of all sequences $u^n\in\mathcal{U}^n$ with type π_{u^n} satisfying $|\pi_{u^n}(a)-P_U(a)|\leq\mu$, $\forall a\in\mathcal{U}$, and $\pi_{u^n}(a)=0$ if $P_U(a)=0$. Likewise, we denote the joint type of the sequences $(u^n,v^n)\in\mathcal{U}^n\times\mathcal{V}^n$ by

$$\pi_{u^n,v^n}(a,b)\triangleq\frac{|\{i\in\{1,\ldots,n\}:u_i=a,v_i=b\}|}{n},\quad (a,b)\in\mathcal{U}\times\mathcal{V}. \tag{2}$$

Let P_1 and P_2 be two distributions over $\mathcal{U}$. Suppose that P_1 is absolutely continuous with respect to P_2 (denoted $P_1\ll P_2$) which means that for any $u\in\mathcal{U}$, $P_2(u)=0\implies P_1(u)=0$, then the Kullback-Leibler divergence between P_1 and P_2 is denoted

$$D(P_1\|P_2)=\int_{\mathcal{U}}\ln\left(\frac{\mathrm{d}P_1}{\mathrm{d}P_2}\right)\mathrm{d}P_1, \tag{3}$$

where $\frac{\mathrm{d}P_1}{\mathrm{d}P_2}$ is the Radon–Nikodym derivative [26, p. 128] of P_1 with respect to P_2. We denote the quasi p-norms by $\|a^n\|_p=\left(\sum_{i=1}^n|a_i|^p\right)^{1/p}$ for any $a^n\in\mathbb{R}^n$.

III. General Model

We consider the setup of Figure 1 where two sensors observe random sequences U_1^n and U_2^n, respectively, and a decision center observes the random sequence V^n. The observations take value in finite sets $\mathcal{U}_1^n,\mathcal{U}_2^n$, and $\mathcal{V}^n$, respectively, and their distribution depends on a binary hypothesis $H\in\{0,1\}$ in the sense that:

$$\text{under } H=0\colon\quad (U_1^n,U_2^n,V^n)\sim P_{U_1,U_2,V}^{\otimes n} \tag{4a}$$

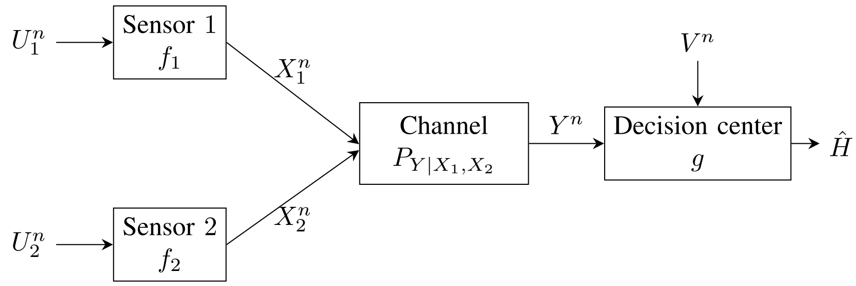

Fig. 1: Distributed hypothesis testing over a discrete memoryless channel.

$$\text{under } H = 1\colon \quad (U_1^n, U_2^n, V^n) \sim Q_{U_1,U_2,V}^{\otimes n}, \tag{4b}$$

where we assume that $P_{U_1,U_2,V} \ll Q_{U_1,U_2,V}$.

The two sensors communicate over a memoryless multiple-access channel (MAC) with a decision center, which then attempts to guess the hypothesis H based on its own local observations and on its observed channel outputs. More formally, each Sensor $\ell \in \{1, 2\}$ produces a sequence of channel inputs $X_\ell^n = f_\ell(U_\ell^n)$ using an encoding function of the form

$$\begin{aligned} f_\ell \colon \mathcal{U}_\ell^n &\to \mathcal{X}_\ell^n \\ u_\ell^n &\mapsto x_\ell^n = f_\ell(u_\ell^n), \quad \ell \in \{1, 2\}. \end{aligned} \tag{5}$$

The input sequences are subject to stringent cost constraints. That means, we are given per-symbol input cost functions

$$c_\ell \colon \mathcal{X}_\ell \to \mathbb{R}^+, \tag{6}$$

where we assume that there exists a unique zero-symbol $x_\ell \in \mathcal{X}_\ell$ with $c_\ell(x_\ell) = 0$. For simplicity, we call 0 the zero-symbol for both c_1 and c_2. We are also given stringent cost budgets $\Gamma_1(n) > 0$ and $\Gamma_2(n)$ that grow sublinearly in n:

$$\lim_{n\to\infty} \Gamma_\ell(n) = \infty \qquad \text{and} \qquad \lim_{n\to\infty} \frac{\Gamma_\ell(n)}{n} = 0, \quad \ell \in \{1, 2\}. \tag{7}$$

For each u_ℓ^n, the cost constraints impose that $f_\ell(u_\ell^n)$ lies in

$$\tilde{\mathcal{X}}_\ell \triangleq \left\{ x_\ell^n \in \mathcal{X}_\ell^n \colon \sum_{i=1}^n c_l(x_{\ell,i}) \leq \Gamma_\ell(n) \right\}, \quad \ell \in \{1, 2\}. \tag{8}$$

Given channel inputs x_1^n and x_2^n, the decision center observes a random output sequence Y^n with each Y_i distributed according to the conditional distribution $P_{Y|X_1,X_2}(\cdot|x_1, x_2)$.

The decision center applies a guessing function of the form

$$g \colon \mathcal{Y}^n \times \mathcal{V}^n \to \{0, 1\} \tag{9}$$

to produce a random guess $\hat{H} = g(Y^n, V^n)$. The decision's type-I (false positive) and type-II (false negative) error probabilities are respectively denoted

$$\alpha_n = \mathbb{P}\left[g(Y^n, V^n) = 1 \,\middle|\, H = 0\right] \tag{10a}$$

$$\beta_n = \mathbb{P}\left[g(Y^n, V^n) = 0 \,\middle|\, H = 1\right]. \tag{10b}$$

Definition 1 *Given $\epsilon \in [0,1)$, a type-II error exponent $\theta > 0$ is called $(\epsilon, \Gamma_1(\cdot), \Gamma_2(\cdot))$-achievable if there exists a sequence (in the blocklength n) of encoding and decision functions (f_1, f_2, g) satisfying the sublinear cost constraint* (8) *and*

$$\overline{\lim_{n\to\infty}}\, \alpha_n \leq \epsilon \tag{11a}$$

$$\underline{\lim_{n\to\infty}}\, -\frac{1}{n}\ln \beta_n \geq \theta. \tag{11b}$$

The supremum over all $(\epsilon, \Gamma_1(\cdot), \Gamma_2(\cdot))$-achievable type-II error exponents θ is denoted $\theta^*_{\text{sublin}}(\epsilon, \Gamma_1(\cdot), \Gamma_2(\cdot))$.

We shall see that $\theta^*_{\text{sublin}}(\epsilon, \Gamma_1(\cdot), \Gamma_2(\cdot))$ is the same for all functions $\Gamma_1(\cdot)$ and $\Gamma_2(\cdot)$ satisfying (7) and all $\epsilon \in [0,1)$. We therefore simply write θ^*_{sublin}.

The following proposition will be instrumental in the results of this paper. The proof can be found in Appendix A.

Proposition 1 *Allowing the decision center to take a randomized decision rule does not increase the Stein-exponent θ^*_{sublin}.*

IV. Results

We have the following general result on the type-II error exponent showing that for certain channels $P_{Y|X_1,X_2}$, communication from the sensors does not increase the Stein-exponent due to the stringent cost constraints (7).

Theorem 1 *Assume that the channel $P_{Y|X_1,X_2}$ is such that there exists a sequence of sets $\mathcal{D}_n \subseteq \mathcal{Y}^n$ satisfying*

$$\lim_{n\to\infty} \mathbb{P}[Y^n \notin \mathcal{D}_n] = 0 \tag{12}$$

and for all input sequences $x_1^n, \tilde{x}_1^n \in \tilde{\mathcal{X}}_1^n$ and $x_2^n, \tilde{x}_2^n \in \tilde{\mathcal{X}}_2^n$:

$$\overline{\lim_{n\to\infty}}\, -\frac{1}{n}\ln\left(\frac{\mathrm{d}P^{\otimes n}_{Y|X_1,X_2}(y^n|x_1^n, x_2^n)}{\mathrm{d}P^{\otimes n}_{Y|X_1,X_2}(y^n|\tilde{x}_1^n, \tilde{x}_2^n)}\right) \leq 0, \quad \forall y^n \in \mathcal{D}_n. \tag{13}$$

Then, for any $\epsilon \in [0,1)$ and $\Gamma_1(\cdot)$ and $\Gamma_2(\cdot)$ satisfying (7)*, the optimal type-II error exponent is*

$$\theta^*_{\text{sublin}} = D\left(P_V \| Q_V\right). \tag{14}$$

Proof: See Section V-A. □

Remark 1 *The optimal exponent θ^*_{sublin} of Theorem 1 coincides with the optimal exponent in a scenario where the decision center only observes the local observation V^n [27, Theorem 14.13]. Therefore, under the stringent cost constraints* (7)*, the sensors cannot improve the Stein-exponent.*

Remark 2 *The same Stein-exponent holds also in the related setup where the sensors communicate to the decision center only over $k(n)$ channel uses where $k(n)$ grows sublinearly in n, irrespective of whether a block-input power constraint or a per-symbol power constraint is imposed. This holds because achievability of the Stein-exponent does not need any communication and the converse is implied by the converse above, as the setup is weaker. In fact, in our original setup, the sensor can always choose to only transmit during the first $k(n)$ channel uses.*

A. Generalized Gaussian channels with a p-th moment constraint

Consider the MAC

$$Y_i = h_1 x_{1,i} + h_2 x_{2,i} + Z_i, \qquad i \in \{1, \ldots, n\}, \tag{15}$$

where h_1 and h_2 are given non-zero real channel coefficients, $(Z_i)_{i\geq 1}$ is an i.i.d. sequence independent of the inputs x_1^n and x_2^n, and each Z_i follows a generalized Gaussian distribution [28, 29] with parameters $p, \sigma > 0$, i..e, of probability density function

$$p_Z(z) = \frac{c_p}{\sigma} e^{-\frac{|z|^p}{2\sigma^p}}, \qquad z \in \mathbb{R}, \tag{16}$$

where

$$c_p \triangleq \frac{p}{2^{\frac{p+1}{p}} \Gamma\left(\frac{1}{p}\right)} \tag{17}$$

and Γ denotes the gamma function [30]. A p-th moment cost constraint is imposed on the input sequences:

$$\|x_\ell^n\|_p^p \leq \Gamma_\ell(n), \quad \ell \in \{1, 2\}, \tag{18}$$

i.e., $c_\ell(x) = |x|^p, \ell = 1, 2$.

For $p = 2$, the noise is Gaussian and the cost constraint is a standard average block-power constraint.

Corollary 1 *For the above generalized Gaussian setup, for any $p > 0$, $\epsilon \in [0, 1)$, and p-th moment cost constraints $\Gamma_1(\cdot)$ and $\Gamma_2(\cdot)$ satisfying* (7)*:*

$$\theta^*_{\text{sublin}} = D\left(P_V \| Q_V\right). \tag{19}$$

Proof: See Section V-B. □

B. Discrete memoryless channels with arbitrary cost constraints

Communication takes place over a discrete memoryless MAC (DMMAC). Accordingly, input and output sets $\mathcal{X}_1, \mathcal{X}_2$, and $\mathcal{Y}$ are finite and the channel law is described by a probability mass function (pmf) $P_{Y|X_1,X_2}$. As we shall see, θ^*_{sublin} depends on the topology (connectivity) of the DMMAC. We therefore define the following classes of channels.

- The set $\mathcal{C}_{\text{full}}$ contains all DMMACs $P_{Y|X_1,X_2}$ for which

$$P_{Y|X_1,X_2}(y|x_1,x_2) > 0, \qquad \forall y, x_1, x_2 \in \mathcal{Y} \times \mathcal{X}_1 \times \mathcal{X}_2. \tag{20}$$

- The set $\mathcal{C}_{\text{sparse}}$ contains all DMMACs $P_{Y|X_1,X_2}$ where there exists (not necessarily distinct) $x_1, x_1', \tilde{x}_1 \in \mathcal{X}_1$, $x_2, x_2', \tilde{x}_2 \in \mathcal{X}_2$ and $y, \tilde{y} \in \mathcal{Y}$ so that:

$$P_{Y|X_1,X_2}(y|x_1,\tilde{x}_2) = 0 \tag{21a}$$

$$P_{Y|X_1,X_2}(y|x_1',\tilde{x}_2) > 0 \tag{21b}$$

$$P_{Y|X_1,X_2}(\tilde{y}|\tilde{x}_1,x_2) = 0 \tag{21c}$$

$$P_{Y|X_1,X_2}(\tilde{y}|\tilde{x}_1,x_2') > 0. \tag{21d}$$

- The set $\mathcal{C}_{\text{sparse-full}}$ contains all DMMACs $P_{Y|X_1,X_2}$ for which there exists $x_1, x_1' \in \mathcal{X}_1$, $\tilde{x}_2 \in \mathcal{X}_2$ and $y^* \in \mathcal{Y}$ so that

$$P_{Y|X_1,X_2}(y^*|x_1,\tilde{x}_2) = 0 \tag{22a}$$

$$P_{Y|X_1,X_2}(y^*|x_1',\tilde{x}_2) > 0, \tag{22b}$$

and for any pair $(x_1, y) \in \mathcal{X}_1 \times \mathcal{Y}$ we either have

$$P_{Y|X_1,X_2}(y|x_1,x_2) = 0, \quad \forall x_2 \in \mathcal{X}_2 \tag{23a}$$

or

$$P_{Y|X_1,X_2}(y|x_1,x_2) > 0, \quad \forall x_2 \in \mathcal{X}_2. \tag{23b}$$

- Finally, the set $\mathcal{C}_{\text{full-sparse}}$ contains all DMMACs $P_{Y|X_1,X_2}$ for which there exists $\tilde{x}_1 \in \mathcal{X}_1$, $x_2, x_2' \in \mathcal{X}_2$ and $y^* \in \mathcal{Y}$ so that

$$P_{Y|X_1,X_2}(y^*|\tilde{x}_1,x_2) = 0 \tag{24a}$$

$$P_{Y|X_1,X_2}(y^*|\tilde{x}_1,x_2') > 0, \tag{24b}$$

and for any pair $(x_2, y) \in \mathcal{X}_2 \times \mathcal{Y}$ we either have

$$P_{Y|X_1,X_2}(y|x_1,x_2) = 0, \quad \forall x_1 \in \mathcal{X}_1 \tag{25a}$$

or

$$P_{Y|X_1,X_2}(y|x_1,x_2) > 0, \quad \forall x_1 \in \mathcal{X}_1. \tag{25b}$$

To see that the four sets $\mathcal{C}_{\text{full}}, \mathcal{C}_{\text{sparse}}, \mathcal{C}_{\text{sparse-full}}, \mathcal{C}_{\text{full-sparse}}$ are a partition of all channel laws, notice that the event

$$\{\exists x_1, \tilde{x}_1, y^* \colon \quad P_{Y|X_1,X_2}(y^*|x_1,\tilde{x}_2) = 0 \qquad \text{and} \qquad P_{Y|X_1,X_2}(y^*|x_1',\tilde{x}_2) > 0\} \tag{26}$$

is the complementary event of

$$\{\forall x_1 \colon\; P_{Y|X_1,X_2}(y^*|x_1,\tilde{x}_2) = 0\} \quad \cup \quad \{\forall x_1 \colon\; P_{Y|X_1,X_2}(y^*|x_1,\tilde{x}_2) > 0\}. \tag{27}$$

Example 1 *Consider the DMMAC*

$$Y = S_1 \cdot x_1 + S_2 \cdot x_2 + Z, \tag{28}$$

where multiplications and additions are the standard operations in $\mathbb{R}$ and Z is a discrete noise over any discrete and finite set $\mathcal{Z} \subset \mathbb{R}$. Inputs x_1 and x_2 as well as the auxiliary variables S_1 and S_2 take value in $\{-1, 1\}$. Assuming that Z is not deterministic, depending on the law of the "states" S_1 and S_2, the DMC belongs to one of the four classes above. In fact,

- *For S_1 and S_2 both deterministic, the DMMAC belongs to $\mathcal{C}_{\text{sparse}}$.*
- *For S_1 and S_2 both non-deterministic, the DMMAC belongs to $\mathcal{C}_{\text{full}}$.*
- *For S_1 deterministic and S_2 non-deterministic, the DMMAC belongs to $\mathcal{C}_{\text{sparse-full}}$.*
- *For S_1 non-deterministic and S_2 deterministic, the DMMAC belongs to $\mathcal{C}_{\text{full-sparse}}$.*

Theorem 2 1) *If $P_{Y|X_1,X_2} \in \mathcal{C}_{\text{full}}$, then*

$$\theta^*_{\text{sublin}} = D\left(P_V \| Q_V\right). \tag{29}$$

2) *If $P_{Y|X_1,X_2} \in \mathcal{C}_{\text{sparse}}$, then*

$$\theta^*_{\text{sublin}} = \min D\left(\tilde{P}_{U_1,U_2,V} \| Q_{U_1,U_2,V}\right), \tag{30}$$

where the minimum is taken over all distributions $\tilde{P}_{U_1,U_2,V}$ satisfying

$$\tilde{P}_{U_1} = P_{U_1}, \quad \tilde{P}_{U_2} = P_{U_2}, \quad \tilde{P}_V = P_V. \tag{31}$$

3) *If $P_{Y|X_1,X_2} \in \mathcal{C}_{\text{sparse-full}}$, then*

$$\theta^*_{\text{sublin}} = \min D\left(\tilde{P}_{V,U_1} \| Q_{V,U_1}\right), \tag{32}$$

where the minimum is taken over all distributions $\tilde{P}_{V,U_1}$ such that

$$\tilde{P}_{U_1} = P_{U_1}, \quad \tilde{P}_V = P_V. \tag{33}$$

4) *If* $P_{Y|X_1,X_2} \in \mathcal{C}_{\text{full-sparse}}$*, then*

$$\theta^*_{\text{sublin}} = \min D\left(\tilde{P}_{V,U_2} \| Q_{V,U_2}\right), \tag{34}$$

where the minimum is taken over all distributions $\tilde{P}_{V,U_2}$ *such that*

$$\tilde{P}_{U_2} = P_{U_2}, \quad \tilde{P}_V = P_V. \tag{35}$$

Proof: See Section VI. □

Remark 3 *In case 1) the optimal exponent is the same as in a scenario without any sensor or without communication from the sensors to the decision center. In case 2), the optimal exponent is the same as when both sensors can communicate to the decision center over independent zero-rate noiseless links. In case 3), the optimal exponent is the same as in a scenario without Sensor* 2 *and where Sensor 1 communicates to the decision center over a zero-rate noiseless link, see [1, Theorem 1].*

Remark 4 *Our results in Theorem 2 remain valid when the DMMAC can only be used for a sublinear (in* n*) number of times* $k(n)$*, irrespective of whether an input-cost constraint is imposed or not. The converse results follow in a straight-forward way because this new setup is weaker (in our original setup the sensors can always choose to transmit only during the first* $k(n)$ *channel uses). Inspecting the achievability proofs of Theorem 2, we see that communication from each sensor effectively only takes place over a sublinear number of channel uses at the beginning, while during the rest of the communication both sensors send the all-zero sequence. These latter channel inputs can thus be omitted without any loss of information at the decision center. The proposed Stein-exponents in Theorem 2 can thus also be achieved in our new setup where communication is only over* $k(n)$ *channel uses.*

V. Proofs of Theorem 1 and Corollary 1

A. Proof of Theorem 1

Achievability follows directly from Stein's lemma where the decision center can ignore the channel outputs Y^n. The converse is proved by relating the type-I and type-II error probabilities of our distributed hypothesis testing problem to the error probabilities of a randomized local hypothesis testing setup. We start our proof by fixing any sequence of encoding and decision functions $\{f_1, f_2, g\}_{n=1}^{\infty}$ with type-I error satisfying

$$\overline{\lim_{n\to\infty}} \alpha_n \leq \epsilon. \tag{36}$$

For each observation $v^n \in \mathcal{V}^n$, define the acceptance regions

$$\mathcal{A}(v^n) \triangleq \{y^n \in \mathbb{R}^n \colon g(v^n, y^n) = 0\}. \tag{37}$$

The chosen functions f_1, f_2, g imply a joint distribution on $U^n, V^n, X_1^n, X_2^n, Y^n$ under both hypotheses $H = 0$ and $H = 1$. We are particularly interested in the induced conditional probability distribution

$$P_{\tilde{Y}^n|V^n}(y^n|v^n) \triangleq \mathbb{P}[Y^n = y^n | V^n = v^n, H = 0] \tag{38}$$

and introduce the new binary hypothesis testing setup depicted in Figure 2 where under $H=0$ the decision center observes V^n i.i.d. $\sim P_V$ and under $H=1$ it observes V^n i.i.d. $\sim Q_V$. Moreover, under both hypotheses it has access to an additional local randomness $\tilde{Y}^n$ that is obtained from V^n based on the non-i.i.d. conditional distribution $P_{\tilde{Y}^n|V^n}$, irrespectively of the hypothesis H. The decision function in this auxiliary setup is thus of the form $\tilde{g}\colon \mathcal{V}^n \times \mathcal{Y}^n \to \{0,1\}$ and we denote its type-I and type-II error probabilities by $\tilde{\alpha}_n$ and $\tilde{\beta}_n$:

$$\tilde{\alpha}_n \triangleq \mathbb{P}[\tilde{g}(V^n,\tilde{Y}^n)=1|H=0] \tag{39a}$$

$$\tilde{\beta}_n \triangleq \mathbb{P}[\tilde{g}(V^n,\tilde{Y}^n)=0|H=1]. \tag{39b}$$

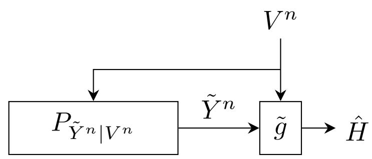

Fig. 2: Randomized local hypothesis test.

Choosing the decision rule

$$\tilde{g}(v^n,y^n) = g(v^n,y^n)\cdot \mathbb{1}\{y^n \in \mathcal{D}_{\delta,n}\} \tag{40}$$

for the setup of Figure 2, we have:

$$\begin{aligned}
1-\alpha_n &= \mathbb{P}\left[g(V^n,Y^n)=0 \;\middle|\; H=0\right] && (41)\\
&= \mathbb{P}\left[g(V^n,Y^n)=0, Y^n\in\mathcal{D}_n \;\middle|\; H=0\right] + \mathbb{P}\left[g(V^n,Y^n)=0, Y^n\notin\mathcal{D}_n \;\middle|\; H=0\right] && (42)\\
&\le \mathbb{P}\left[g(V^n,Y^n)=0, Y^n\in\mathcal{D}_n \;\middle|\; H=0\right] + \mathbb{P}\left[Y^n\notin\mathcal{D}_n \;\middle|\; H=0\right] && (43)\\
&= \sum_{v^n\in\mathcal{V}^n}\left(\mathbb{P}\left[g(V^n,Y^n)=0, Y^n\in\mathcal{D}_n \;\middle|\; V^n=v^n, H=0\right] P_V^{\otimes n}(v^n)\right) + \mathbb{P}\left[Y^n\notin\mathcal{D}_n \;\middle|\; H=0\right] && (44)\\
&= \sum_{v^n\in\mathcal{V}^n}\left(\mathbb{P}\left[g(V^n,\tilde{Y}^n)=0, \tilde{Y}^n\in\mathcal{D}_n \;\middle|\; V^n=v^n, H=0\right] P_V^{\otimes n}(v^n)\right) + \mathbb{P}\left[Y^n\notin\mathcal{D}_n \;\middle|\; H=0\right] && (45)\\
&= \mathbb{P}\left[\tilde{g}(V^n,\tilde{Y}^n)=0 \;\middle|\; H=0\right] + \mathbb{P}\left[Y^n\notin\mathcal{D}_n \;\middle|\; H=0\right] && (46)\\
&= 1-\tilde{\alpha}_n + \mathbb{P}\left[Y^n\notin\mathcal{D}_n \;\middle|\; H=0\right], && (47)
\end{aligned}$$

where (44) holds because we defined $\tilde{Y}^n$ to have the same conditional distribution given V^n as Y^n under hypothesis $H=0$ and (45) holds by the definition of the $\tilde{g}$ function in (40). Combining above inequality with (12), we deduce that

$$\varlimsup_{n\to\infty} \tilde{\alpha}_n \le \varlimsup_{n\to\infty} \alpha_n \le \epsilon. \tag{48}$$

Define

$$\zeta \triangleq \inf_{\substack{y^n\in\mathcal{D}_n,\\ x_1^n,\tilde{x}_1^n\in\tilde{\mathcal{X}}_1^n,\\ x_2^n,\tilde{x}_2^n\in\tilde{\mathcal{X}}_2^n}} \frac{\mathrm{d}P^{\otimes n}_{Y|X_1,X_2}(y^n|x_1^n,x_2^n)}{\mathrm{d}P^{\otimes n}_{Y|X_1,X_2}(y^n|\tilde{x}_1^n,\tilde{x}_2^n)}. \tag{49}$$

We can now bound β_n in terms of $\tilde{\beta}_n$:

$$\beta_n = \sum_{v^n \in \mathcal{V}^n} \mathbb{P}[Y^n \in \mathcal{A}(v^n), V^n = v^n | H = 1] \tag{50}$$

$$= \sum_{v^n \in \mathcal{V}^n} \left(Q_V^{\otimes n}(v^n) \int_{\mathcal{A}(v^n)} \mathrm{d}P_{Y^n|V^n,H}(y^n|v^n, 1) \right) \tag{51}$$

$$\geq \sum_{v^n \in \mathcal{V}^n} \left(Q_V^{\otimes n}(v^n) \int_{\mathcal{A}(v^n)\cap\mathcal{D}_n} \sum_{(x_1^n, x_2^n)\in\tilde{\mathcal{X}}_1^n\times\tilde{\mathcal{X}}_2^n} \left(P_{X_1^n,X_2^n|V^n,H}(x_1^n, x_2^n|v^n, 1) \cdot \mathrm{d}P_{Y|X_1,X_2}^{\otimes n}(y^n|x_1^n, x_2^n) \right) \right) \tag{52}$$

$$\geq \zeta \sum_{v^n \in \mathcal{V}^n} \left(Q_V^{\otimes n}(v^n) \int_{\mathcal{A}(v^n)\cap\mathcal{D}_n} \sum_{(x_1^n, x_2^n)\in\tilde{\mathcal{X}}_1^n\times\tilde{\mathcal{X}}_2^n} \left(P_{X_1^n,X_2^n|V^n,H}(x_1^n, x_2^n|v^n, 0) \cdot \mathrm{d}P_{Y|X_1,X_2}^{\otimes n}(y^n|x_1^n, x_2^n) \right) \right) \tag{53}$$

$$= \zeta \sum_{v^n \in \mathcal{V}^n} \left(Q_V^{\otimes n}(v^n) \int_{\mathcal{A}(v^n)\cap\mathcal{D}_n} \mathrm{d}P_{\tilde{Y}^n|V^n}(y^n|v^n) \right) \tag{54}$$

$$= \zeta \cdot \tilde{\beta}_n, \tag{55}$$

where (52) holds by restricting the integral and by the total law of probability; (53) holds by the definition of ζ in (49) and because for any bounded function f[1] the expectations of f with respect to any two measures μ and ν satisfy $\mathbb{E}_\nu[f]/\mathbb{E}_\mu[f] \geq \frac{f_{\min}}{f_{\max}}$ where $f_{\min}$ and $f_{\max}$ denote the infimum and supremum of f; (54) holds by the definition of $\tilde{Y}^n$; (55) holds by the definition of $\tilde{\beta}_n$.

Notice now that we can specialize Proposition 1 to a setup with a useless MAC from the two sensors to the decision center, in which case the decision center can base its decision only on the local observation V^n and the local randomness. Applying this proposition to the setup of Figure 2 where the local randomness is $\tilde{Y}^n$, we conclude that $\frac{1}{n}\ln\tilde{\beta}_n$ is asymptotically upper bounded by the Stein-exponent of a non-random test based on V^n only, i.e.,

$$\overline{\lim_{n\to\infty}} -\frac{1}{n}\ln\tilde{\beta}_n \leq D(P_V \| Q_V). \tag{56}$$

Plugging (56) and Assumption (13) into (55), we obtain the desired converse result.

B. Proof of Corollary 1

We only prove the converse since achievability is obvious. To this end, we first notice the following.

- For $p \in (0, 1]$ and for all $a, b \in \mathbb{R}$,

$$|a + b|^p \leq (|a| + |b|)^p \leq |a|^p + |b|^p, \tag{57}$$

where the second inequality is proved in [31, Eq (2.12.2)]. Above inequalities also imply:

$$||a|^p - |b|^p| \leq |a - b|^p. \tag{58}$$

[1]Notice that by Assumption (13), for any $y^n \in \mathcal{D}_n$, the function $(x_1^n, x_2^n) \mapsto \mathrm{d}P_{Y|X_1,X_2}^{\otimes n}(y^n|x_1^n, x_2^n)$ is bounded with finite non-zero infimum and supremum.

- For $p \in (1, \infty)$ and for all $a, b \in \mathbb{R}$, it holds that:

$$|a+b|^p \leq (|a|+|b|)^p \leq 2^{p-1}(|a|^p + |b|^p), \tag{59}$$

where the first inequality holds by the triangleq inequality and the second by the convexity of the $t \mapsto |t|^p$ function.

In particular, we have for any $p > 0$:

$$|a+b|^p \leq 2^p(|a|^p + |b|^p). \tag{60}$$

For ease of notation, we define

$$b_i \triangleq h_1 x_{1,i} + h_2 x_{2,i}, \quad i \in \{1, \ldots, n\}, \tag{61}$$

and notice that (60) and the input power constraints (18) imply that

$$\|b^n\|_p^p \leq 2^p(h_1^p \Gamma_1(n) + h_2^p \Gamma_2(n)), \tag{62}$$

for any $x_1^n \in \tilde{\mathcal{X}}_1$ and $x_2^n \in \tilde{\mathcal{X}}_2$.

Case $p \in (0, 1]$: The converse follows directly from Theorem 1 by choosing $\mathcal{D}_n = \mathcal{Y}^n$. In fact, for any input sequences $x_1^n, \tilde{x}_1^n \in \tilde{\mathcal{X}}_1$ and $x_2^n, \tilde{x}_2^n \in \tilde{\mathcal{X}}_2$, define $b^n = h_1 x_1^n + h_2 x_2^n$ and $\tilde{b}^n = h_1 \tilde{x}_1^n + h_2 \tilde{x}_2^n$, and notice:

$$\frac{p_{Y|X_1,X_2}^{\otimes n}(y^n|x_1^n, x_2^n)}{p_{Y|X_1,X_2}^{\otimes n}(y^n|\tilde{x}_1^n, \tilde{x}_2^n)} = \exp\left(-\frac{\|y^n - b^n\|_p^p - \|y^n - \tilde{b}^n\|_p^p}{2\sigma^p}\right) \tag{63}$$

$$\geq \exp\left(-\frac{\|b^n - \tilde{b}^n\|_p^p}{2\sigma^p}\right) \tag{64}$$

$$\geq \exp\left(-\frac{\|b^n\|_p^p + \|\tilde{b}^n\|_p^p}{2\sigma^p}\right) \tag{65}$$

$$\geq \exp\left(-\frac{2^p(h_1^p \Gamma_1(n) + h_2^p \Gamma_2(n))}{\sigma^p}\right) \tag{66}$$

for any sequence $y^n \in \mathbb{R}^n$. Here, (64) holds by (58); (65) holds by (57); and (66) holds by (62).

The proof is concluded by noting that (66) implies (13) because p, h_1, h_2, σ are fixed and by our assumption (7).

Case $p > 1$: Follows from Theorem 1 by choosing

$$\mathcal{D}_n = \left\{ y^n \in \mathbb{R}^n : \|y^n\|_p^p \leq \nu \right\}, \tag{67}$$

where

$$\nu \triangleq 2^{2p-2}\left(h_1^p \Gamma_1(n) + h_2^p \Gamma_2(n)\right) + 2^{p-1}\left(n\frac{2\sigma^p}{p} + \delta n\right) \tag{68}$$

for an arbitrary small number $\delta > 0$.

We check that the proposed set $\mathcal{D}_n$ satisfies the two Conditions (12) and (13) in the theorem. For any $x_1^n, \tilde{x}_1^n \in \tilde{\mathcal{X}}_1^n$ and

$x_2^n, \tilde{x}_2^n \in \tilde{\mathcal{X}}_2^n$, define b^n and $\tilde{b}^n$ as above and notice:

$$\frac{p^{\otimes n}_{Y|X_1,X_2}(y^n|x_1^n, x_2^n)}{p^{\otimes n}_{Y|X_1,X_2}(y^n|\tilde{x}_1^n, \tilde{x}_2^n)} = \exp\left(-\frac{\|y^n - b^n\|_p^p - \|y^n - \tilde{b}^n\|_p^p}{2\sigma^p}\right) \tag{69}$$

$$\geq \exp\left(-\sum_{i=1}^n \frac{p|b_i - \tilde{b}_i| \sup_{t\in(y_i - b_i, y_i - \tilde{b}_i)} |t|^{p-1}}{2\sigma^p}\right) \tag{70}$$

$$\geq \exp\left(-\sum_{i=1}^n \frac{p\left(|b_i| + |\tilde{b}_i|\right) \cdot \max\left((|y_i| + |b_i|)^{p-1}, (|y_i| + |\tilde{b}_i|)^{p-1}\right)}{2\sigma^p}\right) \tag{71}$$

$$\geq \exp\left(-\sum_{i=1}^n \frac{2^{p-1}p\left(|b_i| + |\tilde{b}_i|\right) \cdot \max\left((|y_i|^{p-1} + |b_i|^{p-1}, |y_i|^{p-1} + |\tilde{b}_i|^{p-1}\right)}{2\sigma^p}\right) \tag{72}$$

$$\geq \exp\left(-\frac{2^{p-1}p}{2\sigma^p}\sum_{i=1}^n \left(\left(|b_i| + |\tilde{b}_i|\right)\left(|y_i|^{p-1} + |b_i|^{p-1} + |\tilde{b}_i|^{p-1}\right)\right)\right) \tag{73}$$

$$\geq \exp\Bigg(-\frac{2^{p-2}p}{\sigma^p}\Big(\|b^n\|_p^p + \|\tilde{b}^n\|_p^p + \left(\|b^n\|_p + \|\tilde{b}^n\|_p\right)\|y^n\|_p^{p-1} + \|b^n\|_p \|\tilde{b}^n\|_p^{p-1} + \|\tilde{b}^n\|_p \|b^n\|_p^{p-1}\Big)\Bigg) \tag{74}$$

$$\geq \exp\left(-\frac{2^{p-2}p}{\sigma^p}\left(4\cdot 2^p(h_1^p\Gamma_1(n) + h_2^p\Gamma_2(n)) + 2\left(h_1^p\Gamma_1(n) + h_2^p\Gamma_2(n)\right)^{\frac{1}{p}} \nu^{\frac{p-1}{p}}\right)\right) \tag{75}$$

where (70) holds by the mean value theorem and because the derivative of $|t|^p$ is $p|t|^{p-1}\cdot \mathrm{sign}(t)$; (71) holds because the supremum is achieved at the borders of the interval; (72) holds by (60); (74) holds by factoring out and applying Hölder's inequality with the parameters p and $\frac{p}{p-1}$ to the cross terms; and finally (75) holds by (62) and the definition of $\mathcal{D}_n$ in (67).

Recalling the choice of ν in (68) and the stringent power constraints (7), we can conclude from (75) that Condition (13) in the theorem is satisfied.

To verify that remaining Condition (12), define $B^n = h_1X_1^n + h_2X_2^n$ and notice the following:

$$\|Y^n\|_p^p \leq 2^{p-1}(\|B^n\|_p^p + \|Z^n\|_p^p) \tag{76}$$

$$\leq 2^{2p-2}\left(h_1^p\Gamma_1(n) + h_2^p\Gamma_2(n)\right) + 2^{p-1}\|Z^n\|_p^p, \tag{77}$$

where (76) holds by (59) and (77) by the power constraint (18) and again (59). We thus have

$$\mathbb{P}\left[Y^n \notin \mathcal{D}_{\delta,n} \mid H=0\right] = \mathbb{P}\left[\|Y^n\|_p^p \geq \nu \;\middle|\; H=0\right] \tag{78}$$

$$\leq \mathbb{P}\left[\frac{1}{n}\|Z^n\|_p^p \geq \left(\frac{2\sigma^p}{p} + \delta n\right)\right] \tag{79}$$

$$= \mathbb{P}\left[\frac{1}{n}\sum_{i=1}^n(|Z_i|^p - \mathbb{E}[|Z_i|^p]) \geq \delta\right] \tag{80}$$

where (78) holds by (68) and (77); (80) holds because $\mathbb{E}[|Z|^p] = 2\sigma^p/p$, see [29, Eq (3)].

By the weak law of large numbers, the right-hand side of (80) vanishes with n, which establishes (12).

VI. Proof of Theorem 2

Define

$$\gamma \triangleq \min_{y,x_1,x_2,\tilde{x}_1,\tilde{x}_2} \frac{P_{Y|X_1,X_2}(y|x_1,x_2)}{P_{Y|X_1,X_2}(y|\tilde{x}_1,\tilde{x}_2)} \tag{81}$$

and the minimum costs

$$c_{\min,\ell} \triangleq \min_{x_\ell \in \mathcal{X}_\ell \backslash \{0\}} c_\ell(x_\ell), \quad \ell \in \{1,2\}. \tag{82}$$

Notice that the maximum Hamming weight, i.e., the number of symbols different from 0 for any input sequence $x_\ell^n \in \tilde{\mathcal{X}}_\ell^n$ is upper bounded by

$$k_{\max,\ell} \triangleq \frac{\Gamma_\ell(n)}{c_{\min,\ell}}, \tag{83}$$

and thus the number of positions on which any two pairs of input sequences (x_1^n, x_2^n) and $(\tilde{x}_1^n, \tilde{x}_2^n)$ differ is at most

$$\tau_{\max} \triangleq 2k_{\max,1} + 2k_{\max,2}. \tag{84}$$

Set also

$$k \triangleq \frac{1}{2} \min\{k_{\max,1}, k_{\max,2}\} \tag{85}$$

and notice that for sufficiently large blocklengths n we have $2k < n$ because $\Gamma_1(n)$ and $\Gamma_2(n)$ grow sublinearly in n. Notice finally that, as shown for example in [23, Equations (50)–(52)], the sets $\tilde{\mathcal{X}}_\ell$ grow sublinearly in n:

$$\lim_{n\to\infty} \frac{1}{n} \ln |\tilde{\mathcal{X}}_\ell| = 0. \tag{86}$$

A. Proof of 1)

Only the converse requires proof. To this end, for all pairs $x_1^n, \tilde{x}_1^n \in \tilde{\mathcal{X}}_1^n$ and $x_2^n, \tilde{x}_2^n \in \tilde{\mathcal{X}}_2^n$ and $y^n \in \mathcal{Y}^n$:

$$\frac{P_{Y|X_1,X_2}^{\otimes n}(y^n|x_1^n,x_2^n)}{P_{Y|X_1,X_2}^{\otimes n}(y^n|x_1^n,x_2^n)} \in \left[\gamma^{-\tau_{\max}}, \gamma^{\tau_{\max}}\right], \tag{87}$$

and thus

$$\lim_{n\to\infty} \frac{1}{n} \ln \frac{P_{Y|X_1,X_2}^{\otimes n}(y^n|x_1^n,x_2^n)}{P_{Y|X_1,X_2}^{\otimes n}(y^n|\tilde{x}_1^n,\tilde{x}_2^n)} = 0. \tag{88}$$

The converse is thus obtained directly from Theorem 1 by choosing $\mathcal{D}_n = \mathcal{Y}^n$.

B. Proof of 2)

Converse: The converse is directly obtained from [1] by giving the decision center direct access to inputs x_1^n and x_2^n. In this enhanced scenario the outputs Y^n become useless and we fall back to a scenario that is equivalent to the zero-rate scenario studied in [1] because the number of different input sequences $|\tilde{\mathcal{X}}_1|$ and $|\tilde{\mathcal{X}}_2|$ is sublinear in n, see (86).

Achievability: The achievability is proved based on the following scheme. Fix $\mu > 0$ and pick a set of symbols $x_1, x_1', \tilde{x}_1 \in \mathcal{X}_1$, $x_2, x_2', \tilde{x}_2 \in \mathcal{X}_2$ and $y, \tilde{y} \in \mathcal{Y}$ satisfying the conditions (21).

Sensor 1*:*

- During the first k channel uses, it sends the symbol x_1' if $u_1^n \in T_\mu(P_{U_1})$; otherwise it sends the symbol x_1.
- During the next k channel uses, it sends the symbol $\tilde{x}_1$.
- During the remaining channel uses, it sends the 0 symbol.

Sensor 2*:*

- During the first k channel uses, it sends the symbol $\tilde{x}_2$.
- During the next k channel uses, it sends the symbol x_2' if $u_2^n \in T_\mu(P_{U_2})$; otherwise it sends the symbol x_2.
- During the remaining channel uses, it sends the 0 symbol.

Decision center: It declares $\hat{H} = 0$ if the following three conditions are simultaneously satisfied:

- the symbol y occurs during the first k channel outputs $Y_1, \ldots, Y_k$,
- the symbol $\tilde{y}$ occurs during channel outputs $Y_{k+1}, \ldots, Y_{2k}$,
- $v^n \in T_\mu(P_V)$;

otherwise it declares $\hat{H} = 1$.

Analysis: We start by showing that the type-I error probability of this proposed scheme vanishes when $n \to \infty$. To this end, notice that

$$
\begin{aligned}
1 - \alpha_n &= \mathbb{P}\left[\hat{H} = 0 \;\middle|\; H = 0\right] && (89)\\
&= \mathbb{P}\left[y \in Y^k,\ \tilde{y} \in Y_{k+1}^{2k}, V^n \in \mathcal{T}_\mu(P_V) \;\middle|\; H = 0\right] && (90)\\
&\geq \mathbb{P}\left[y \in Y^k,\ \tilde{y} \in Y_{k+1}^{2k},\ V^n \in \mathcal{T}_\mu(P_V),\ U_1^n \in \mathcal{T}_\mu(P_{U_1}),\ U_2^n \in \mathcal{T}_\mu(P_{U_2}) \;\middle|\; H = 0\right] && (91)\\
&= \mathbb{P}\left[V^n \in \mathcal{T}_\mu(P_V),\ U_1^n \in \mathcal{T}_\mu(P_{U_1}),\ U_2^n \in \mathcal{T}_\mu(P_{U_2}) \;\middle|\; H = 0\right]\\
&\quad \cdot \mathbb{P}\left[y \in Y^k, \tilde{y} \in Y_{k+1}^{2k} \;\middle|\; U_1^n \in \mathcal{T}_\mu(P_{U_1}),\ U_2^n \in \mathcal{T}_\mu(P_{U_2}), H = 0\right] && (92)\\
&= \mathbb{P}\left[V^n \in \mathcal{T}_\mu(P_V), U_1^n \in \mathcal{T}_\mu(P_{U_1}),\ U_2^n \in \mathcal{T}_\mu(P_{U_2}) \;\middle|\; H = 0\right]\\
&\quad \cdot \mathbb{P}\left[y \in Y^k,\ \tilde{y} \in Y_{k+1}^{2k} \;\middle|\; X_1^{2k} = ({x_1'}^k, \tilde{x}_1^k), X_2^{2k} = (\tilde{x}_2^k, {x_2'}^k)\right] && (93)\\
&= \mathbb{P}\left[V^n \in \mathcal{T}_\mu(P_V), U_1^n \in \mathcal{T}_\mu(P_{U_1}), U_2^n \in \mathcal{T}_\mu(P_{U_2}) \;\middle|\; H = 0\right]\\
&\quad \cdot \mathbb{P}\left[y \in Y^k \;\middle|\; X_1^k = {x_1'}^k, X_2^k = \tilde{x}_2^k\right]\\
&\quad \cdot \mathbb{P}\left[\tilde{y} \in Y_{k+1}^{2k} \;\middle|\; X_{1,k+1}^{2k} = \tilde{x}_1^k, X_{2,k+1}^{2k} = {x_2'}^k\right] && (94)\\
&= \mathbb{P}\left[V^n \in \mathcal{T}_\mu(P_V), U_1^n \in \mathcal{T}_\mu(P_{U_1}), U_2^n \in \mathcal{T}_\mu(P_{U_2}) \;\middle|\; H = 0\right]\\
&\quad \cdot \left(1 - (1 - p_{Y|X_1,X_2}(y|x_1', \tilde{x}_2))^k\right) \cdot \left(1 - (1 - p_{Y|X_1,X_2}(\tilde{y}|\tilde{x}_1, x_2'))^k\right) && (95)
\end{aligned}
$$

where (93) follows by the design of the coding scheme and because the channel transition law does not depend on the hypothesis;

(94) follows because the channel is memoryless. Notice further that by the weak law of large numbers, irrespectively of μ,

$$\lim_{n\to\infty} \mathbb{P}\left[V^n \in \mathcal{T}_\mu(P_V), U_1^n \in \mathcal{T}_\mu(P_{U_1}), U_2^n \in \mathcal{T}_\mu(P_{U_2}) \;\middle|\; H=0\right] = 1. \tag{96}$$

Since $1-p_{Y|X_2,X_1}(y|x_1', \tilde{x}_2)$ and $1-p_{Y|X_1,X_2}(\tilde{y}|\tilde{x}_1, x_2')$ lie in the half-open interval $(0,1]$, we can thus conclude by (95) that $\lim \alpha_n = 0$ as $n \to \infty$. Next, we proceed to upper-bound the type-II error probability. To this end, for each n, we introduce the set $\mathcal{B}_{n,\mu}(P_{U_1}, P_{U_2}, P_V)$ of n-types such that

$$\mathcal{B}_{n,\mu}(P_{U_1}, P_{U_2}, P_V) = \left\{\pi_{u_1^n,u_2^n,v^n} \in \mathcal{P}_n(\mathcal{U}_1 \times \mathcal{U}_2 \times \mathcal{V}) : u_1^n \in \mathcal{T}_\mu\left(P_{U_1}\right), u_2^n \in \mathcal{T}_\mu\left(P_{U_2}\right), v^n \in \mathcal{T}_\mu\left(P_V\right)\right\}, \tag{97}$$

and notice

$$\begin{aligned}
\beta_n &= \mathbb{P}\left[\hat{H}=0|H=1\right] && (98)\\
&= \mathbb{P}\left[y \in Y^k, \tilde{y} \in Y_{k+1}^{2k}, V^n \in \mathcal{T}_\mu(P_V), U_1^n \in \mathcal{T}_\mu(P_{U_1}), U_2^n \in \mathcal{T}_\mu(P_{U_2}) \;\middle|\; H=1\right] && (99)\\
&\leq \mathbb{P}\left[V^n \in \mathcal{T}_\mu(P_V), U_1^n \in \mathcal{T}_\mu(P_{U_1}), U_2^n \in \mathcal{T}_\mu(P_{U_2}) \;\middle|\; H=1\right] && (100)\\
&= \sum_{\pi_{u_1^n,u_2^n,v^n} \in \mathcal{B}_{n,\mu}(P_{U_1},P_{U_2},P_V)} \mathbb{P}\left[(V^n, U_1^n, U_2^n) \in \mathcal{T}_n(\pi_{u_1^n,u_2^n,v^n}) \;\middle|\; H=1\right] && (101)\\
&\leq \sum_{\pi_{u_1^n,u_2^n,v^n} \in \mathcal{B}_{n,\mu}(P_{U_1},P_{U_2},P_V)} 2^{-nD\left(\pi_{u_1^n,u_2^n,v^n} \| Q_{U_1,U_2,V}\right)} && (102)\\
&\leq (n+1)^{|\mathcal{U}_1||\mathcal{U}_2||\mathcal{V}|} 2^{-n \min D\left(\pi_{u_1^n,u_2^n,v^n} \| Q_{U_1,U_2,V}\right)} && (103)
\end{aligned}$$

where the minimum is taken over all types $\pi_{u_1^n,u_2^n,v^n} \in \mathcal{B}_{n,\mu}(P_{U_1}, P_{U_2}, P_V)$. Here, (102) holds by [25, Theorem 11.1.4]; and (103) by [25, Theorem 11.1.1].

Dropping the restriction on $\pi_{u_1^n,u_2^n,v^n}$ being an n-type and defining the μ-marginal neighborhood of $P_{U_1,U_2,V}$ as

$$\begin{aligned}
\mathcal{B}_\mu(P_{U_1}, P_{U_2}, P_V) = \big\{\tilde{P}_{U_1,U_2,V} \in \mathcal{M}(\mathcal{U}_1 \times \mathcal{U}_2 \times \mathcal{V}) \quad \text{s.t.} \quad & \forall u_1, u_2, v \in \mathcal{U}_1 \times \mathcal{U}_2 \times \mathcal{V}: \\
& |\tilde{P}_{U_1}(u_1) - P_{U_1}(u_1)| \leq \mu, \\
& |\tilde{P}_{U_2}(u_2) - P_{U_2}(u_2)| \leq \mu, \\
& |\tilde{P}_V(v) - P_V(v)| \leq \mu\big\},
\end{aligned} \tag{104}$$

we conclude from (103) that

$$\beta_n \leq (n+1)^{|\mathcal{U}_1||\mathcal{U}_2||\mathcal{V}|} 2^{-n \min D\left(\tilde{P}_{U_1,U_2,V} \| Q_{U_1,U_2,V}\right)} \tag{105}$$

where now the minimum is taken over all distributions $\tilde{P}_{U_1,U_2,V} \in \mathcal{B}_\mu(P_{U_1}, P_{U_2}, P_V)$. Taking first the limit $n \to \infty$ and then $\mu \to 0$ we finally conclude that

$$\varliminf_{n\to\infty} -\frac{1}{n} \ln \beta_n \geq \min D\left(\tilde{P}_{U_1,U_2,V} \| Q_{U_1,U_2,V}\right), \tag{106}$$

where the minimum is now over all distributions $\tilde{P}_{U_1,U_2,V} \in \mathcal{M}(\mathcal{U}_1 \times \mathcal{U}_2 \times \mathcal{V})$ with marginals satisfying $\tilde{P}_{U_1} = P_{U_1}$, $\tilde{P}_{U_2} = P_{U_2}$

and $\tilde{P}_V = P_V$.

C. Proof of 3)

a) Achievability: Pick $x_1, x_1', \tilde{x}_2, y^*$ satisfying Condition (22a).

Sensor 1: During the first k channel uses, it sends the symbol x_1' if $U_1^n \in T_\mu(P_{U_1})$; otherwise it sends the symbol x_1. Subsequently, it sends the 0 symbol.

Sensor 2: During the first k channel uses, it sends the symbol $\tilde{x}_2$. Subsequently, it sends the 0 symbol.

Decision center: It declares $\hat{H} = 0$ if the following two conditions are simultaneously satisfied:

- the symbol y^* occurs in the output sequence Y^k ,
- $v^n \in T_\mu(P_V)$;

otherwise, it declares $\hat{H} = 1$.

In the following we analyze the type-I and type-II error probabilities of this scheme. First, we show that the type-I error probability of this proposed scheme vanishes using the same proof steps as in (95):

$$1 - \alpha_n \geq \mathbb{P}\left[V^n \in \mathcal{T}_\mu(P_V), U_1^n \in \mathcal{T}_\mu(P_{U_1}) \;\middle|\; H = 0\right] \cdot \left(1 - (1 - p_{Y|X_1,X_2}(y^*|x_1', \tilde{x}_2))^k\right). \tag{107}$$

By the weak law of large numbers, irrespectively of μ,

$$\lim_{n\to\infty} \mathbb{P}\left[V^n \in \mathcal{T}_\mu(P_V), U_1^n \in \mathcal{T}_\mu(P_{U_1}) \;\middle|\; H = 0\right] = 1. \tag{108}$$

Since $1 - p_{Y|X_1,X_2}(y^*|x_1', \tilde{x}_2)$ lies in the half-open interval $(0, 1]$, by (107), we conclude that $1 - \alpha_n$ tends to 1 as $n \to \infty$ and consequently $\lim \alpha_n = 0$, when $n \to \infty$. Next, we analyze the type-II error probability, using the same proof steps as in (103):

$$\beta_n \leq (n+1)^{|\mathcal{U}_1||\mathcal{V}|} 2^{-n \min D\left(\pi_{u_1^n, v^n} \| Q_{U_1,V}\right)} \tag{109}$$

where such that the minimum is over all n-types in the set

$$\mathcal{B}_{n,\mu}(P_{U_1}, P_V) = \left\{\pi_{u_1^n, v^n} \in \mathcal{P}_n(\mathcal{U}_1 \times \mathcal{V}) : u_1^n \in \mathcal{T}_\mu\left(P_{U_1}\right), v^n \in \mathcal{T}_\mu\left(P_V\right)\right\}. \tag{110}$$

Next, we upper-bound (109), taking the minimum over the μ-marginal neighborhood of $P_{U_1,V}$

$$\begin{aligned} \mathcal{B}_\mu(P_{U_1}, P_V) = \big\{\tilde{P}_{U_1,V} \in \mathcal{M}(\mathcal{U}_1 \times \mathcal{V}) \quad \text{s.t.} \quad & \forall u_1, v \in \mathcal{U}_1 \times \mathcal{V}\colon \\ & |\tilde{P}_{U_1}(u_1) - P_{U_1}(u_1)| \leq \mu, \\ & |\tilde{P}_V(v) - P_V(v)| \leq \mu\big\}. \end{aligned} \tag{111}$$

Finally, we let $n \to \infty$ and $\mu \to 0$ and we conclude that

$$\varliminf_{n\to\infty} -\frac{1}{n}\ln\beta_n \geq \min D\left(\tilde{P}_{V,U_1}\|Q_{V,U_1}\right), \tag{112}$$

where the minimum is over all distributions $\tilde{P}_{V,U_1}$ with marginals P_V and P_{U_1}.

b) Converse: We first introduce some useful notation. Define

$$\gamma_1 \triangleq \min_{\substack{y,x_1,x_2,\tilde{x}_1,\tilde{x}_2:\\ P_{Y|X_1,X_2}(y|x_1,x_2)>0}} \frac{P_{Y|X_1,X_2}(y|x_1,x_2)}{P_{Y|X_1,X_2}(y|\tilde{x}_1,\tilde{x}_2)} > 0 \tag{113}$$

We consider an enhanced setup where the detector has access not only to the side information V^n and the channel output Y^n, but also directly to the channel input X_1^n. The Stein-exponent of this enhanced setup (depicted in Figure 3) is an upper bound to the Stein-exponent of our original setup.

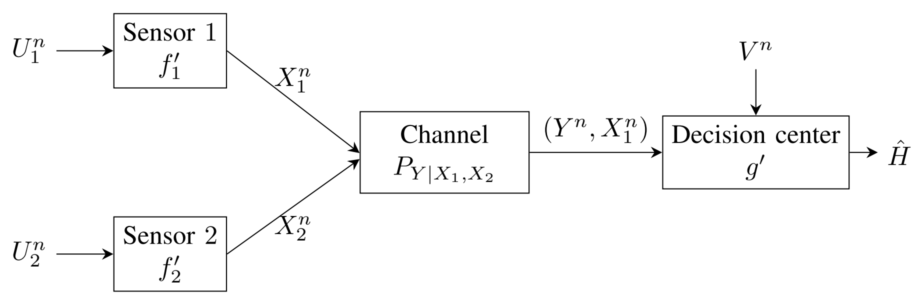

Fig. 3: Enhanced distributed hypothesis testing setup where the output consists of the pair (Y, X_1).

Encoding functions in the enhanced setup are the same as in our original setup and denoted f_1' and f_2', the decision function is denoted g', and the corresponding type-I and type-II error probabilities α_n' and β_n'.

Fix $\epsilon \in [0,1)$ and consider any sequence of encoding and decision functions (f_1', f_2', g') such that $\overline{\lim}\,\alpha_n' \leq \epsilon < 1$, $n \to \infty$. For the chosen decision function g' and a fixed blocklength n, define for each observation $v^n \in \mathcal{V}^n$:

$$\mathcal{A}(v^n) \triangleq \{(x_1^n, y^n) \in \mathcal{X}_1^n \times \mathcal{Y}^n \colon g(v^n, x_1^n, y^n) = 0\} \tag{114}$$

and

$$\mathcal{A}'(v^n) \triangleq \{(x_1^n, y^n) \in \mathcal{X}_1^n \times \mathcal{Y}^n \colon g(v^n, x_1^n, y^n) = 0,\;\; P_{Y|X_1,X_2}(y|x_1,x_2) > 0 \;\; \forall x_2 \in \mathcal{X}_2\}^2 \tag{115}$$

Notice that the two regions $\mathcal{A}(v^n)$ and $\mathcal{A}'(v^n)$ have same probability to occurr under both $H=0$ and $H=1$. Define further the conditional pmf

$$P_{\tilde{Y}^n|X_1^n,V^n}(y^n|x_1^n, v^n) \triangleq \mathbb{P}[Y^n = y^n | X_1^n = x_1, V^n = v^n, H = 0], \tag{116}$$

and introduce the random binary hypothesis testing setup where the decision center observes (X_1^n, V^n) which has the same distribution as in our original setup, and has access to the local randomness $\tilde{Y}^n \sim P_{\tilde{Y}^n|X_1^n,V^n}$, irrespectively of the hypothesis $H \in \{0,1\}$. The randomized test is depicted in Figure 4.

[2] By our assumption, for given (y, x_1) either $P_{Y|X_1,X_2}(y|x_1,x_2) > 0$ for all x_2 or for no x_2.

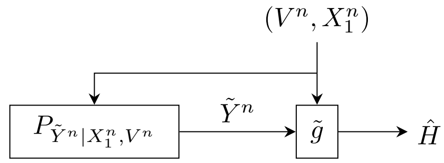

Fig. 4: Randomized hypothesis test.

We consider the same decision function g' for this auxiliary setup as in enhanced setup, simply applied to $\tilde{Y}^n$ instead of Y^n. The type-I and type-II error probabilities of this test are:

$$\tilde{\alpha}_n \triangleq \mathbb{P}[g'(X_1^n, V^n, \tilde{Y}^n) = 1 | H = 0] \tag{117a}$$

$$\tilde{\beta}_n \triangleq \mathbb{P}[g'(X_1^n, V^n, \tilde{Y}^n) = 0 | H = 1]. \tag{117b}$$

Notice that the local randomness $\tilde{Y}^n$ follows the same joint distribution with (X_1^n, V^n) as Y^n under $H = 0$. Therefore, $\tilde{\alpha}_n = \alpha_n'$ and $\overline{\lim}_{n\to\infty} \tilde{\alpha}_n \leq \epsilon$. Moreover, for the type-II error probabilities of the two tests, we obtain similarly as in the proof of Theorem 1:

$$\beta_n' = \sum_{v^n \in \mathcal{V}^n} \mathbb{P}[(X_1^n, Y^n) \in \mathcal{A}'(v^n), V^n = v^n | H = 1] \tag{118}$$

$$= \sum_{v^n \in \mathcal{V}^n} \left(Q_V^{\otimes n}(v^n) \sum_{(x_1^n, y^n) \in \mathcal{A}'(v^n)} \sum_{x_2^n} \left(P_{X_1^n, X_2^n | V^n, H}(x_1^n, x_2^n | v^n, 1) P_{Y|X_1,X_2}^{\otimes n}(y^n | x_1^n, x_2^n) \right) \right) \tag{119}$$

$$= \sum_{v^n \in \mathcal{V}^n} \left(Q_V^{\otimes n}(v^n) \sum_{(x_1^n, y^n) \in \mathcal{A}'(v^n)} P_{X_1^n|V^n,H}(x_1^n, v^n, 1) \sum_{x_2^n} \left(P_{X_2^n|X_1^n,V^n,H}(x_1^n, x_2^n | v^n, 1) P_{Y|X_1,X_2}^{\otimes n}(y^n | x_1^n, x_2^n) \right) \right) \tag{120}$$

$$\geq \gamma_1^{\tau_{\max}} \sum_{v^n \in \mathcal{V}^n} \left(Q_V^{\otimes n}(v^n) \sum_{(x_1^n, y^n) \in \mathcal{A}'(v^n)} P_{X_1^n|V^n,H}(x_1^n, v^n, 1) \sum_{x_2^n} \left(P_{X_2^n|X_1^n,V^n,H}(x_1^n, x_2^n | v^n, 0) P_{Y|X_1,X_2}^{\otimes n}(y^n | x_1^n, x_2^n) \right) \right) \tag{121}$$

$$= \gamma_1^{\tau_{\max}} \sum_{v^n \in \mathcal{V}^n} \left(Q_V^{\otimes n}(v^n) \sum_{(x_1^n, y^n) \in \mathcal{A}'(v^n)} P_{X_1^n|V^n,H}(x_1^n, v^n, 1) P_{Y^n|X_1^n,V^n,H}(x_1^n, x_2^n | v^n, 0) \right) \tag{122}$$

$$= \gamma_1^{\tau_{\max}} \sum_{v^n \in \mathcal{V}^n} \left(Q_V^{\otimes n}(v^n) \sum_{(x_1^n, y^n) \in \mathcal{A}'(v^n)} P_{X_1^n|V^n,H}(x_1^n, v^n, 1) P_{\tilde{Y}^n|X_1^n,V^n}(x_1^n, x_2^n | v^n) \right) \tag{123}$$

$$= \gamma_1^{\tau_{\max}} \cdot \tilde{\beta}_n, \tag{124}$$

where $\tau_{\max}$ is defined in (84) and grows sublinearly in n. Here, (121) holds because the two pairs (x_1^n, x_2^n) and $(\tilde{x}_1^n, \tilde{x}_2^n)$ differ in at most $\tau_{\max}$ positions and by the definition of γ_1 in (113); (123) holds by the definition of $\tilde{Y}^n$, see (116).

Since γ_1 is a constant and $\tau_{\max}$ grows sublinearly in n, see (83), (84), and (7), we conclude from (124) that

$$\overline{\lim_{n\to\infty}} -\frac{1}{n} \ln \beta_n' \leq \overline{\lim_{n\to\infty}} -\frac{1}{n} \ln \tilde{\beta}_n \tag{125}$$

and the Stein-exponent of our original setup cannot be larger than the Stein-exponent of the auxiliary setup where the local randomness $\tilde{Y}^n$ replaces the observation Y^n. Consider now the special case of Proposition 1 where the MAC is the channel

$Y = X_1$. The randomized hypothesis test in our enhanced setup is of this form and we can thus deduce that the local randomness $\tilde{Y}^n$ does not increase the Stein-exponent of our enhanced setup. Without local randomness, the enhanced setup is however equivalent to a single-sensor setup with a noise-less link from Sensor 1 to the decision center, and since the number of input sequences $|\tilde{\mathcal{X}}_1|$ is sublinear in n, see (86), the Stein-exponent is upper bounded by the exponent in [1, Theorem 1], i.e.,

$$\varlimsup_{n\to\infty} -\frac{1}{n}\ln \beta'_n \leq \min D\left(\tilde{P}_{U_1,V}\|Q_{U_1,V}\right), \tag{126}$$

where the minimization is over all probability mass functions $\tilde{P}_{U_1,V}$ with marginals P_{U_1} and P_V.

Combining (126) with (125) concludes the proof.

D. Proof of 4)

The proof follows the same steps as the one for 3) by symmetry.

VII. Summary and Discussion

We characterized the Stein-exponent for two-sensors distributed detection over noisy memoryless channels with stringent input cost constraints that grow sublinearly in the blocklength n. For a large class of MACS, like Gaussian MACs and fully-connected DMMACs, the sublinear cost constraints render communication from the two sensors to the decision center useless in terms of Stein's exponent. In these setups, the Stein-exponent coincides with the exponent in the non-distributed case where the decision center has to take its decision solely based on its own observations. In contrast, for the class of partially-connected DMMACs where certain outputs can be induced only by a subset of the inputs from each user, the Stein-exponent coincides with the exponent in a scenario where communication from both sensors takes place over noiseless links of zero-rate, a scenario solved in [1]. For the case where the partial-connectivity only holds from the first sensor but not the second, the Stein-exponent of our setup coincides with the Stein-exponent in a setup without the second sensor and noise-free zero-rate communication from the first sensor.

While this manuscript focuses on two-sensor setups, our proofs and results readily extend to scenarios with an arbitrary number of sensors.

Comparing our results to the Stein-exponent of distributed hypothesis testing over DMMACs without cost constraints studied in [22], we observe that the stringent resource constraint severely degrades the Stein-exponent. In particular, without sublinear cost constraints, the Stein-exponent does not degrade to the exponent of the local setup, but the information from the sensor is useful even when communicated over a noisy channel.

While in this paper we solved the problem for general DMMACs and general cost constraints, we only considered the class of generalized Gaussian channels with moment constraints. It will be interesting to extend our study to more general classes of continuous-valued channels.

Appendix A
Proof of Proposition 1

Before proving the proposition, we make the considered setup more precise. For each blocklength n, let $\boldsymbol{S}_n$ denote the local randomness at the decision center that under both hypotheses follows the same distribution

$$P_{\boldsymbol{S}_n|V^n}(\boldsymbol{s}|v^n). \tag{127}$$

The decision center thus can now base its decision on the triple $V^n, Y^n, \boldsymbol{S}_n$, where V^n, Y^n are described as in our original setup and $\boldsymbol{S}_n$ is the newly introduced local randomness. The decision rule thus is of the form $\tilde{g}(V^n, Y^n, \boldsymbol{S}_n)$ and $\tilde{\alpha}_n$ and $\tilde{\beta}_n$ denote the type-I and type-II error probabilities of the randomized decision rule:

$$\tilde{\alpha}_n = \mathbb{P}\left[\tilde{g}(V^n, Y^n, \boldsymbol{S}_n) = 1 \;\middle|\; H = 0\right] \tag{128}$$

$$\tilde{\beta}_n = \mathbb{P}\left[\tilde{g}(V^n, Y^n, \boldsymbol{S}_n) = 0 \;\middle|\; H = 1\right]. \tag{129}$$

To prove the remark, we show that for any choice of the sequence of randomized decision function $\tilde{g}$ there exists a sequence of deterministic tests g (without local randomness $\boldsymbol{S}_n$) that achieves same asymptotic error probabilities.

To this end, choose a sequence γ_n satisfying

$$\lim_{n\to\infty} \gamma_n = 0 \tag{130a}$$

$$\lim_{n\to\infty} \frac{1}{n} \ln \gamma_n = 0. \tag{130b}$$

For each blocklength n, define a new deterministic decision rule

$$g(v^n, y^n) = \mathbb{1}\left\{\Pr\left[\tilde{g}\left(v^n, y^n, \boldsymbol{S}_n\right) = 0\;\right] \le \gamma_n\right\} \tag{131}$$

and the associated acceptance region

$$\mathcal{G}_{\gamma_n} \triangleq \left\{(v^n, y^n) \colon\; \Pr\left[\tilde{g}\left(v^n, y^n, \boldsymbol{S}_n\right) = 0\right] > \gamma_n\right\}. \tag{132}$$

Using the definition in (132), we can relate the error probabilities of the two tests as follows:

$$1 - \tilde{\alpha}_n = \sum_{(v^n, y^n)} \left(\Pr[V^n = v^n, Y^n = y^n | H = 0] \Pr\left[\tilde{g}(v^n, Y^n, \boldsymbol{S}_n) = 0 \;\middle|\; V^n = v^n, Y^n = y^n\right]\right) \tag{133}$$

$$\le \sum_{(v^n, y^n) \in \mathcal{G}_{\gamma_n}} \Pr[V^n = v^n, Y^n = y^n | H = 0]$$
$$+ \sum_{(v^n, y^n) \notin \mathcal{G}_{\gamma_n}} \left(\Pr[V^n = v^n, Y^n = y^n | H = 0] \underbrace{\Pr\left[\tilde{g}(v^n, y^n, \boldsymbol{S}_n) = 0 \;\middle|\; V^n = v^n, Y^n = y^n\right]}_{\le \gamma_n}\right) \tag{134}$$

$$\le 1 - \alpha_n + \gamma_n \tag{135}$$

and as

$$\tilde{\beta}_n = \sum_{(v^n,y^n)} \Pr[V^n = v^n, Y^n = y^n | H = 1] \Pr\left[\tilde{g}(v^n, y^n, \boldsymbol{S}_n) = 0 \,\middle|\, V^n = v^n, Y^n = y^n\right] \tag{136}$$

$$\geq \sum_{(v^n,y^n)\in\mathcal{G}_{\gamma_n}} \Big(\Pr[V^n = v^n, Y^n = y^n | H = 1] \underbrace{\Pr\left[\tilde{g}(v^n, y^n, \boldsymbol{S}_n) = 0 \,\middle|\, V^n = v^n, Y^n = y^n\right]}_{>\gamma_n} \Big) \tag{137}$$

$$\geq \beta_n \cdot \gamma_n. \tag{138}$$

Combining (135) and (138) with (130), we conclude

$$\varlimsup_{n\to\infty} \alpha_n \leq \varlimsup_{n\to\infty} \tilde{\alpha}_n \tag{139}$$

$$\varlimsup_{n\to\infty} -\frac{1}{n} \ln \beta_n \geq \varlimsup_{n\to\infty} -\frac{1}{n} \ln \tilde{\beta}_n. \tag{140}$$

This establishes that the Stein-exponent without randomized decision needs to be at least as large as the Stein-exponent with randomized decisions.

References

[1] 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.

[2] K. Zheng, R. Xu, J. Mei, H. Yang, L. Lei, and X. Wang, "Ambient IoT toward 6G: Standardization, potentials, and challenges," *IEEE Access*, vol. 12, pp. 146 668–146 677, October 2024.

[3] C. Stein, "A bound for the error in the normal approximation to the distribution of a sum of dependent random variables," in *Proc. of the sixth Berkeley symposium on mathematical statistics and probability*, vol. 6, California, USA, 1972.

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

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

[6] S. Sreekumar and D. Gündüz, "Distributed hypothesis testing over discrete memoryless channels," *IEEE Transactions on Information Theory*, vol. 66, no. 4, pp. 2044–2066, April 2020.

[7] S. Salehkalaibar and M. Wigger, "Distributed hypothesis testing with variable-length coding," *IEEE Journal on Selected Areas in Information Theory*, vol. 1, no. 3, pp. 681–694, November 2020.

[8] 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, October 2012.

[9] S. Sreekumar and D. Gündüz, "Hypothesis testing over a noisy channel," in *Proc. of IEEE International Symposium on Information Theory*, Paris, France, July 7–12 2019.

[10] 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, December 2020.

[11] N. Weinberger, Y. Kochman, and M. Wigger, "Exponent trade-off for hypothesis testing over noisy channels," in *Proc. of IEEE International Symposium on Information Theory*, Paris, France, July 7–12 2019.

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

[13] H. Shimokawa, T. Han, and S. I. Amari, "Error bound for hypothesis testing with data compression," in *Proc. of IEEE International Symposium on Information Theory*, Trondheim, Norway, June 27 – July 1 1994, p. 114.

[14] Z. Wen and A. Gohari, "A new upper bound for distributed hypothesis testing using the auxiliary receiver approach," 2025. [Online]. Available: https://arxiv.org/abs/2409.14148

[15] S. Watanabe, "On sub-optimality of random binning for distributed hypothesis testing," in *Proc. of IEEE International Symposium on Information Theory*, Espoo, Finland, 26 June – July 1 2022.

[16] W. Zhao and L. Lai, "Distributed testing with cascaded encoders," *IEEE Transactions on Information Theory*, vol. 64, no. 11, pp. 7339–7348, November 2018.

[17] E. Haim and Y. Kochman, "Binary distributed hypothesis testing via Körner-Marton coding," in *Proc. of IEEE Information Theory Workshop*, Cambridge, UK, September 11–14 2016.

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

[19] M. Hamad, M. Wigger, and M. Sarkiss, "Multi-hop network with multiple decision centers under expected-rate constraints," *IEEE Transactions on Information Theory*, vol. 69, no. 7, pp. 4255–4283, July 2023.

[20] ——, "Cooperative multi-sensor detection under variable-length coding," in *Proc. of IEEE Information Theory Workshop*, Riva del Garda, Italy, 11-15 April 2021.

[21] S. Sreekumar, C. Hirche, H.-C. Cheng, and M. Berta, "Distributed quantum hypothesis testing under zero-rate communication constraints," 2025. [Online]. Available: https://arxiv.org/abs/2410.08937

[22] S. Salehkalaibar and M. Wigger, "Distributed hypothesis testing based on unequal-error protection codes," *IEEE Transactions on Information Theory*, vol. 66, no. 7, pp. 4150–41 820, July 2020.

[23] A. Bounhar, M. Sarkiss, and M. Wigger, "A dichotomy for distributed detection with limited communication," in *Proc. of IEEE Information Theory Workshop*, Sidney, Australia, September 29–October 3 2025.

[24] D. Williams, *Probability with martingales*. Cambridge University Press, 1991.

[25] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. John Wiley & Sons Inc., 2006.

[26] P. R. Halmos, *Measure theory*. Springer, 2013, vol. 18.

[27] Y. Polyanskiy and Y. Wu, *Information Theory: From Coding to Learning*. Cambridge University Press, 2025.

[28] S. Nadarajah, "A generalized normal distribution," *Journal of Applied statistics*, vol. 32, no. 7, pp. 685–694, September 2005.

[29] A. Dytso, R. Bustin, H. V. Poor, and S. Shamai Shitz, "On additive channels with generalized Gaussian noise," in *Proc. of IEEE International Symposium on Information Theory*, Aachen, Germany, June 25–30 2017.

[30] P. J. Davis, "Leonhard Euler's integral: A historical profile of the gamma function," *The American Mathematical Monthly*, vol. 66, no. 10, pp. 849–869, 1959.

[31] G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*. Cambridge university press, 1952.