Distributed Detection over Noisy Channels under Stringent Resource Constraints

Abdelaziz Bounhar, Mireille Sarkiss, Michèle Wigger

***Abstract*—This paper identifies the Stein-exponent of distributed detection when the sensor communicates to the decision center over a discrete memoryless channel (DMC) subject to one of three stringent communication constraints: 1) The number of channel uses of the DMC grows sublinearly in the number of source observations n; 2) The number of channel uses is n but a block-input cost constraint C_n is imposed almost surely and C_n grows sublinearly in n; 3) The block-input constraint is imposed only on expectation. We show a dichotomy in the Stein-exponent in all three scenarios depending on the structure of the DMC's transition law. When the DMC is partially-connected (i.e., some outputs cannot be induced by certain inputs) in all three scenarios the Stein-exponent matches the exponent identified by Han and by Shalaby and Papamarcou for the scenario where communication is of zero-rate but over a noiseless link, while otherwise the exponent is smaller. Specifically, in scenarios 1) and 2), for fully-connected DMCs the Stein-exponent collapses to that of a local test at the decision center, rendering the remote sensor and communication useless. In scenario 3), the sensor remains beneficial even for fully-connected DMCs, and the Stein-exponent is larger than the local exponent at the decision center. Here, the exact value of the Stein-exponent depends on whether the expectation constraint is imposed only under the null hypothesis or under both hypotheses.**

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

I. Introduction

Consider a distributed binary hypothesis testing problem where a sensor and a decision center observe correlated sources and the sensor can transmit information to the decision center over a channel. The joint distribution of the observations at the sensor and the decision center depends on a binary hypothesis $\mathcal{H} \in \{0, 1\}$ that should be guessed by the decision center based on its own observed samples and the outputs of the channel. The goal is to reduce the error probabilities under the two hypotheses. In particular, we wish to keep the type-I error probability (deciding $\hat{\mathcal{H}} = 1$ if $\mathcal{H} = 0$) below a given threshold $\epsilon \in [0, 1)$ while driving the type-II error probability (deciding $\hat{\mathcal{H}} = 0$ if $\mathcal{H} = 1$) to 0 exponentially fast in the number of observation samples n. 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. The largest exponential decay of the type-II error probability in such an asymmetric setup is known as the Stein-exponent.

This canonical distributed hypothesis testing problem has been studied in various previous works under different assumptions on the communication channel. An important line of work started in [1] and continued in [2]–[7]; it studied the Stein-exponent in a communication scenario where the sensor can send Rn bits to the decision center over an error-free link. Achievability results were reported, but the exact exponent is known only for special classes of source distributions. Extensions to network scenarios with multiple sensors and/or decision centers have been reported since, see [4], [8], [9], as well as extensions to include security constraints [10]–[15], or to allow for variable-length coding [16], [17].

A somehow separate line of work [2], [18]–[20] derived the Stein-exponent of binary distributed hypothesis testing in a scenario where the number of communication bits from the sensor to the decision center is either one [2] or sublinear in the number of observations n [18]. Such a scenario is commonly referred to as zero-rate communication. The optimal strategy in both these scenarios is that the decision center decides on $\hat{\mathcal{H}} = 0$ if, and only if, its own local observation and the sensor's observations are each *marginal*-typical according to the distributions under hypothesis $\mathcal{H} = 0$. Implementing this strategy requires only that the sensor sends a single bit, namely the binary outcome of its local typicality test to the decision center. These zero-rate results have recently also been extended to the quantum case [21].

While above described results consider error-free communication channels (except for [15]), in this work, we assume that communication from the sensor to the decision center takes place over a discrete memoryless channel (DMC). Distributed hypothesis testing over a DMC has already been considered in [22], [23] under the assumption that the DMC can be used κn times, i.e., the number of channel uses is linear in the observation length n. The authors in [22] determined the Stein-exponent for the class of source distributions termed "testing against conditional independence" and showed that it coincides with the exponent of a noise-free link of same capacity. The fact that the DMC is noisy thus does not decrease the Stein-exponent in these setups. For other source distributions however this is not the case and the noise in the DMC decreases the Stein-exponent and mechanisms such as "unequal error protection" need to be employed to partially mitigate the noise [22], [23]. Similar results were also reported for the MAC [23] and the BC [24].

Contributions: In this work, we consider distributed hypothesis testing over a DMC and we establish the Stein-exponents for three communication scenarios: 1) The number of channel uses of the DMC grows sublinearly in the number of source observations n. 2) The number of channel uses is n but a block-input cost constraint C_n is imposed almost surely and C_n is sublinear in n; 3) The stringent block-input constraint C_n is imposed only on expectation. Our results reveal a dichotomy

This work was presented in part at the 2025 IEEE Information Theory Workshop. A. Bounhar was with MBZUAI, Paris, France. He is now with Mistral AI. Email: bounhar.abdelaziz@gmail.com. Mireille Sarkiss is with SAMOVAR, Télécom SudParis, Institut Polytechnique de Paris, Palaiseau, France. Email: mireille.sarkiss@telecom-sudparis.eu. Michèle Wigger was with LTCI, Télécom Paris, Institut Polytechnique de Paris, Palaiseau, France. She is now with Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des Signaux et Systèmes. Email: michele.wigger@centralesupelec.fr.

of the Stein-exponent in these setups depending on whether the DMC is fully- or partially-connected. When the DMC is partially-connected, it is possible to achieve the same Stein-exponent as reported in [18] for noiseless channels that can be used a sublinear (in the observation length n) number of times. In contrast, when the channel is fully-connected, the optimal Stein-exponent collapses. While in cases 1) and 2) the exponent collapses to the local exponent at the decision center and the sensor is rendered useless, in case 3) the collapse is only partial and the sensor still helps improving the Stein-exponent. In this case 3), the Stein-exponent moreover not only depends on the source distributions but also on the largest error exponent that can be attained over the DMC when transmitting a binary symbol. Moreover, it depends on whether the expected cost constraint is imposed under both hypotheses or only under the null hypothesis. The single-hypothesis constraint is motivated by highly critical alarm systems where in case of alarm situation the consumed power is of no interest.

For the partially-connected case, the converse proof is obtained by proving that the logarithm of the number of different input sequences that can be transmitted is sublinear in n, and by noting that any DMC can be simulated on a noiseless link, implying that the Stein-exponent cannot exceed the exponent in [18]. For the fully-connected case, the converse proof for Scenarios 1) and 2), is obtained by means of simple bounding steps. For Scenario 3, it is obtained through different carefully chosen change-of-measure arguments as well as an application of the the blowing-up lemma [25].

The achievability result for partially-connected DMCs (in all three scenarios) is obtained by adapting the noiseless zero-rate scheme in [18]. In this scheme, the decision center declares $\hat{\mathcal{H}} = 0$ only when the marginal typicality tests for the distribution under $\mathcal{H} = 0$ are satisfied both at the sensor and at the decision center. Over a noisy but partially-connected DMC, this decision strategy can be implemented by having the sensor send k' (a number that is sublinear in n) times symbol x' if its own marginal typicality check is successful and k' times a different symbol x' otherwise, where x and x' are chosen so that a certain output y^* is never reached by input x but by input x'. On the remaining time the sensor sends a zero-cost symbol. The decision center then only declares $\hat{\mathcal{H}} = 0$ if it observes at least one output y^* and its local typicality check is satisfied. This way, the type-II error is not increased compared to the noiseless link scenario. Moreover, for increasing $k \to \infty$ the type-I error vanishes as for noise-free communication. For finite number of channel uses k the strategy remains applicable if the tolerated type-I error probability is sufficiently large.

For fully-connected DMCs, in Scenario 3 the sensor can still send codewords with large costs, albeit this can happen only with very small probabilities. We can therefore employ the previously described scheme with the marginal typicality checks where the sensor sends the zero-cost symbol over the entire blocklength n in case its own marginal typicality test succeeds, and otherwise it sends any other symbol x. Since the sensor's marginal typicality test fails with exponentially vanishing probability, this strategy will not violate the expected cost constraint. It however achieves a strictly better exponent than the local test at the decision center. When the expected cost constraint is also imposed under $\mathcal{H} = 1$, then the sensor also needs to send the zero-cost symbol when its observation is marginal-typical according to the distribution under $\mathcal{H} = 1$, and otherwise it suffices to send it when it is typical according to the distribution under $\mathcal{H} = 0$.

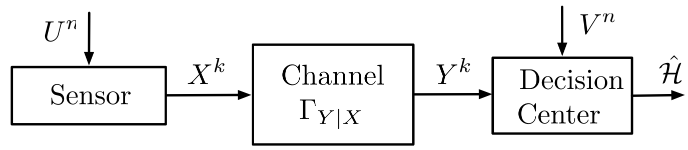

Fig. 1: Distributed Hypothesis testing with a sublinear number of channel uses.

A special case of our Scenario 1 was already considered in [26], where only the sensor observed data but not the decision center. Our proof technique in this special case differs from the one proposed in [26]. In particular, our converse proof in the fully-connected case is simpler and only employs basic bounding steps, while the converse in [26] requires the blowing-up lemma. Notice however that the proof and results in [26] also characterize the Stein-exponent under traditional assumptions on the communication, i.e., without cost constraints and when the number of channel uses is linear in the number of observations.

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. We use the function $w_{\mathrm{H}}(\cdot)$ to indicate the Hamming weight. We abbreviate *independent and identically distributed* as *i.i.d.* and *probability mass function* as *pmf*. Also, we denote by $\pi_{u^n v^n}$ the joint type of the sequences (u^n, v^n):

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

and $\pi_{u^n}(a)$ the corresponding marginal type of the u^n sequence. We use $\mathcal{T}_\mu^{(n)}(P_U)$ to denote the jointly strongly-typical set as in [27, Definition 2.9], i.e., the set of all sequences u^n with type π_{u^n} satisfying $\pi_{u^n}(a) = 0$ if $P_U(a) = 0$ and $|\pi_{u^n}(a) - P_U(a)| \leq \mu$, otherwise. We also use Landau notation $o(1)$ to indicate any function that tends to 0 for blocklengths $n \to \infty$.

II. Setup and Results

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 $\mathcal{H} \in \{0, 1\}$:

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

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

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 [18], we assume that the support of P_{UV} is included in the support of Q_{UV}.

Communication from the sensor to the 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 $\Gamma_{Y|X}$. 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:

$$X^k = f^{(n)}(U^n) \in \mathcal{X}^k, \tag{3}$$

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

$$\hat{\mathcal{H}} = g^{(n)}(V^n, Y^k) \in \{0,1\}. \tag{4}$$

The goal is to design encoding and decision functions such that the type-I (false alarm) error probability

$$\alpha_n \triangleq \Pr\left[\hat{\mathcal{H}} = 1 | \mathcal{H} = 0\right] \tag{5}$$

stays below given threshold $\epsilon > 0$ and the type-II (miss-detection) error probability

$$\beta_n \triangleq \Pr\left[\hat{\mathcal{H}} = 0 | \mathcal{H} = 1\right] \tag{6}$$

decays to 0 with largest possible exponential decay.

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:

$$\overline{\lim_{n\to\infty}}\, \alpha_n \leq \epsilon \tag{7a}$$

$$\underline{\lim_{n\to\infty}}\, -\frac{1}{n}\log \beta_n \geq \theta. \tag{7b}$$

The supremum over all ϵ-achievable miss-detection error exponents θ is denoted $\theta^\star(\epsilon)$ and called the *Stein-exponent*.

We will consider three different communication scenarios, that we explain in the following three subsections:

1) Scenario 1: Sublinear number of channel uses: The number of channel uses grows sublinearly in n:

$$\lim_{n\to\infty} k(n) = \infty \tag{8a}$$

$$\lim_{n\to\infty} \frac{k(n)}{n} = 0. \tag{8b}$$

2) Scenario 2: Sublinear Costs: The channel can be used n times, i.e.,

$$k = n \tag{9}$$

but a stringent cost constraint is imposed.

Given a bounded and nonnegative cost function $c(\cdot)\colon \mathcal{X} \to \mathbb{R}_0^+$, for which there exists a unique zero-symbol:

$$0 \in \mathcal{X} \tag{10}$$

$$c(x) = 0 \quad \text{ if, and only if, } x = 0. \tag{11}$$

The stringent cost constraint is described by condition

$$\sum_{i=1}^{n} c(X_i) \leq C_n, \text{ with prob. } 1, \tag{12}$$

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

$$\lim_{n\to\infty} C_n = \infty \tag{13a}$$

$$\lim_{n\to\infty} \frac{C_n}{n} = 0. \tag{13b}$$

The Stein-exponent for this scenario is denoted $\theta^\star_{\text{str-cost}}(\epsilon, \{C_n\})$.

3) Scenario 3: Sublinear Expected Costs: Reconsider the same setup and definitions from the previous subsection, but where the almost-sure cost constraint (12) is replaced by an expected cost constraint:

$$\sum_{i=1}^{n} \mathbb{E}[c(X_i)|\mathcal{H} = \mathsf{H}] \leq C_n. \tag{14}$$

We shall impose above expected cost constraint either under both hypotheses or only under the null hypothesis $\mathcal{H} = 0$. In the former case the Stein-exponent is denoted $\theta^\star_{\text{Exp-cost, both}}(\epsilon, \{C_n\})$, and in the latter case by $\theta^\star_{\text{Exp-cost},\mathcal{H}=0}(\epsilon, \{C_n\})$.

B. Results

We determine the Stein-exponent for all the introduced scenarios. Our results show a dichotomy of this exponent depending on the connectivity of the channel. We therefore make the following definition.

Definition 2: A DMC $\Gamma_{Y|X}$ is said *partially-connected* if there exist two inputs $x, x' \in \mathcal{X}$ and an output $y^* \in \mathcal{Y}$ satisfying the two conditions:

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

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

Otherwise it is said *fully-connected.*

To express our results in this section, we define the following three exponents:

$$E_1 \triangleq \min_{\substack{\pi_{UV}:\\ \pi_U = P_U\\ \pi_V = P_V}} D(\pi_{UV}\|Q_{UV}) \tag{16a}$$

$$E_2 \triangleq D(P_V\|Q_V) + \max_{\hat{x}\in\mathcal{X}\setminus\{0\}} D(\Gamma_{Y|X=0}\|\Gamma_{Y|X=\hat{x}}) \tag{16b}$$

$$E_3 \triangleq \min_{\substack{\pi_{UV}:\\ \pi_U = Q_U\\ \pi_V = P_V}} D(\pi_{UV}\|Q_{UV}), \tag{16c}$$

where notice that the difference between E_1 and E_3 only lies in the minimization constraints $\pi_U = P_U$ or $\pi_U = Q_U$. Moreover, E_1 is the Stein-exponent of a distributed hypothesis testing setup where communication is over a noise-free communication link that can be used a sublinear number of times [18].

The following theorem shows that for partially-connected channels, in all three scenarios the Stein-exponent $\theta^\star(\epsilon)$ coincides with the Stein-exponent when communication takes place over a noise-free channel that can be used a sublinear number of times [18]. For fully-connected DMCs, the Stein-exponent collapses to the exponent of the local test at the decision center in Scenarios 1 and 2, and is thus obtained by a simple test at the decision center without any communication. In these scenarios, fully-connected DMCs are thus not useful in terms of Stein-exponent. In Scenario 3, fully-connected DMCs are still useful and improve the Stein-exponent over the local test at the decision center. Our results further show that imposing the expected cost constraint under both hypotheses is more stringent compared to imposing it only under the null hypothesis whenever $P_U \neq Q_U$.

Theorem 1: Fix $\epsilon \in [0,1)$ and let $\{C_n\}$ be any sequence of sublinear cost constraints satisfying (13). Then:

1) For partially-connected DMCs $\Gamma_{Y|X}$:

$$\theta^\star_{\text{sublin}}(\epsilon) = \theta^\star_{\text{str-cost}}(\epsilon, \{C_n\}) \\ = \theta^\star_{\text{Exp-cost},\mathcal{H}=0}(\epsilon, \{C_n\}) = \theta^\star_{\text{Exp-cost, both}}(\epsilon, \{C_n\}) = E_1. \tag{17a}$$

2) For fully-connected DMCs $\Gamma_{Y|X}$:

$$\theta^\star_{\text{sublin}}(\epsilon) = \theta^\star_{\text{str-cost}}(\epsilon, \{C_n\}) = D(P_V \| Q_V), \tag{18}$$

and

$$\theta^\star_{\text{Exp-cost},\mathcal{H}=0}(\epsilon, \{C_n\}) = \min\{E_1, E_2\}, \tag{19}$$

$$\theta^\star_{\text{Exp-cost,both}}(\epsilon, \{C_n\}) = \min\{E_1, E_2, E_3\}. \tag{20}$$

Proof: Achievability of (17) can be proved based on the following scheme. Fix a small number $\mu > 0$ and let x, x', y^* be so that Condition (15) is satisfied. Also, in Scenario 1, let $k' = k(n)$, and in Scenario 2, let $k' = \frac{C_n}{\min_{x \neq 0} c(x)}$, which by Assumptions (13) grows sublinearly in n.

Sensor: If $U^n \in \mathcal{T}_\mu^{(n)}(P_U)$, send $X^{k'} = x'^{k'}$. Otherwise, send $X^{k'} = x^{k'}$. For Scenarios 2 and 3, send the zero-cost symbol 0 during the remaining channel uses $k'+1, \ldots, n$.

Decision Center: If at least one of the channel outputs among $y^{k'}$ is y^* and if $V^n \in \mathcal{T}_\mu^{(n)}(P_V)$, then declare $\hat{\mathcal{H}} = 0$. Otherwise, declare $\hat{\mathcal{H}} = 1$. The analysis of this scheme in Appendix A concludes the achievability proof of (17).

The converse to (17) for $\theta^\star_{\text{sublin}}(\epsilon)$ follows directly from [18]. In fact, [18] shows that E_1 is the Stein-exponent over a noiseless link, and any DMC can be simulated over a noiseless link. To prove the converse to (17) for the other exponents, we first show that (with high probability) the weight of the input sequence grows only sublinearly in n and thus the number of possible input sequences grows only subexponentially in n. Subsequently, then the result in [18] can be invoked. Details of these proofs are provided in Appendix B.

Achievability of (18) can be proved via a local test at the decision center, without any communication. The converse to (18) is proved in Appendix C.

Achievability and converse proofs for (19) are presented in Appendices D and E, respectively. Achievability is based on a similar scheme as described above. The sensor sends the all-0 sequence when its observation is typical according to P_U (and also when it is typical according to Q_U in case the expected cost constraint is imposed also under $\mathcal{H} = 1$) but it sends only non-zero inputs for all other observations. The decision center decides on $\hat{\mathcal{H}} = 0$ only when the channel output sequence y^n is typical according to $\Gamma_{Y|X=0}$ and its own observation v^n is typical according to P_V. Notice that in this scheme, the probability of not sending the all-0 sequence is exponentially-vanishing in the blocklength n, and the inputs thus satisfy the expected cost constraints. ∎

Remark 1: The two exponents $\theta^\star_{\text{Exp-cost},\mathcal{H}=0}(\epsilon, \{C_n\})$ and $\theta^\star_{\text{Exp-cost, both}}(\epsilon, \{C_n\})$ coincide when $P_V = Q_V$, because then $E_1 = E_3$. In our numerical simulations with randomly chosen P_{UV} and Q_{UV}, for U and V ternary, the exponent E_1 was always strictly larger than E_3 and thus for DMCs that are not too noisy (in the sense that E_2 is not too small) we observed $\theta^\star_{\text{Exp-cost},\mathcal{H}=0}(\epsilon, \{C_n\}) > \theta^\star_{\text{Exp-cost, both}}(\epsilon, \{C_n\})$ with a strict inequality. See also our Example 1 ahead.

Remark 2 (Finite Values of k): The above theorem for $\theta^\star_{\text{sublin}}(\epsilon)$ holds under the assumption that $k \to \infty$. The result for fully-connected channels trivially remains valid even for finite values of k. For partially-connected channels, trivially again, the converse remains valid for finite k. By inspecting the proof in Appendix A, we see that achievability continues to hold for finite k when the allowed type-I error probability $\epsilon \geq (1 - \Gamma_{Y|X}(y^*|x'))^k$.

Remark 3: Inspecting the proof, one remakrs that Result (18) holds also when the support of P_{UV} is not included in the support of Q_{UV}.

Corollary 2 (When communication never helps): Consider pmfs P_{UV} and Q_{UV} satisfying

$$\sum_v P_V(v) Q_{U|V}(u|v) = P_U(u), \quad \forall u \in \mathcal{U}. \tag{21}$$

Then, for any $\epsilon \in [0,1)$, and irrespective of the DMC $\Gamma_{Y|X}$:

$$\theta^\star_{\text{sublin}}(\epsilon) = \theta^\star_{\text{str-cost}}(\epsilon, \{C_n\}) = \theta^\star_{\text{Exp-cost},\mathcal{H}=0}(\epsilon, \{C_n\}) \\ = \theta^\star_{\text{Exp-cost, both}}(\epsilon, \{C_n\}) = D(P_V \| Q_V). \tag{22}$$

In particular, for testing against independence ($Q_{UV} = P_U \cdot P_V$) the Stein-exponent is zero in all three scenarios. The following-up work [28] characterized the optimal scaling of the type-II error probability for our Scenario 1, which decreases exponentially in k.

Inspecting the proof of the corollary, we can in fact conclude that the dichotomy observed in Theorem 1 disappears for all three scenarios, i.e., irrespective of the employed DMC (sublinear) communication does not improve the Stein-exponent, if, and only if, Condition (21) is satisfied.

Proof: Observe that under Condition (21):

$$E_1 = \min_{\substack{\pi_{UV}: \\ \pi_U = P_U \\ \pi_V = P_V}} D(\pi_{UV} \| Q_{UV}) \tag{23}$$

$$= D(P_V \| Q_V) + \min_{\substack{\pi_{V|U}: \\ \pi_V = P_V}} \mathbb{E}_{P_V}[D(\pi_{U|V} \| Q_{U|V})] \tag{24}$$

$$\geq D(P_V \| Q_V), \tag{25}$$

where the inequality holds with equality if, and only if, $\pi_{U|V} = Q_{U|V}$ is a permissible choice in the minimization. (This is the case if, and only if, Condition (21) holds.) Since all exponents in (22) are lower bounded $D(P_V \| Q_V)$ and upper bounded by E_1, this proves the corollary. ∎

Example 1: Let the source distributions be

$$P_{UV} = \begin{bmatrix} 0.065 & 0.145 & 0.090 \\ 0.010 & 0.080 & 0.110 \\ 0.025 & 0.075 & 0.400 \end{bmatrix} \tag{26}$$

and

$$Q_{UV} = \begin{bmatrix} 0.057803 & 0.115607 & 0.138728 \\ 0.184971 & 0.069364 & 0.115607 \\ 0.173410 & 0.063584 & 0.080925 \end{bmatrix}, \tag{27}$$

and let $\Gamma_{Y|X}$ be a binary symmetric channel BSC(p).

Figure 2 illustrates the exponents $\theta^\star_{\text{sublin}}(\epsilon)$, $\theta^\star_{\text{str-cost}}(\epsilon, \{C_n\})$, $\theta^\star_{\text{Exp-cost},\mathcal{H}=0}(\epsilon, \{C_n\})$, and $\theta^\star_{\text{Exp-cost, both}}(\epsilon, \{C_n\})$ in function of the BSC cross-over probability $p \in [0,35,0,65]$. The exponents are also compared to their counterparts over a partially-connect DMC. (Recall that all exponents coincide for partially-connected DMCs and they do not depend on the exact channel law $\Gamma_{Y|X}$.) We observe that when the BSC is not too noisy, the exponent $\theta^{\star}_{\text{Exp-cost},\mathcal{H}=0}(\epsilon,\{C_n\})$ can reach the same error exponent as over a partially-connected DMC. This is not the case for $\theta^{\star}_{\text{Exp-cost},\mathcal{H}=0}(\epsilon,\{C_n\})$, which saturates at a lower value. The exponents $\theta^{\star}_{\text{sublin}}(\epsilon)$ and $\theta^{\star}_{\text{str-cost}}(\epsilon,\{C_n\})$ are generally even lower and do not depend on the BSC crossover probability p.

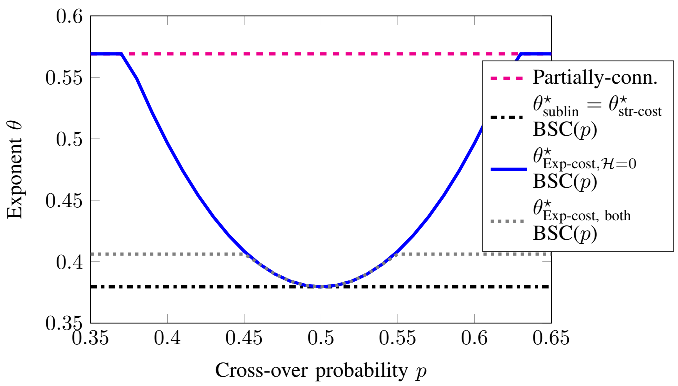

Fig. 2: Error exponents over a BSC(p) in function of the crossover parameter p. The exponents are also compared to their counterparts over a partially-connected DMC.

III. Summary

This article presents a dichotomy in distributed detection under zero-rate and highly resource-limited communication over discrete memoryless channels (DMCs), based on the structure of the channel's transition law. When the DMC is fully-connected (i.e., every output symbol can be induced by any input with positive probability), the maximum achievable Stein-exponent (largest type-II error exponent) collapses compared to cases where the channel is not fully-connected, i.e. only partially-connected.

In the partially-connected case, we showed that the Stein-exponent matches that of a zero-rate noiseless link, in all three communication settings that we consider: 1) the number of channel uses of the DMC is sublinear in the observation length n; 2) the number of channel uses is n but a block-input cost constraint C_n is imposed with probability 1 and C_n is sublinear in n; 3) the block-input constraint C_n is imposed only on expectation over the source observations.

In contrast, when the DMC is fully-connected, then we showed for cases 1) and 2) that the Stein-exponent degrades to that of a purely local test. In these cases, the remote sensor provides no benefit and the decision center should rely only on its own observations. Situation improves in case 3), where communication from the sensor can improve performance and the decision center should base its decision on both its local observations and the outputs of the DMC. The optimal strategy for the sensor is to transmit zero symbols for typical source observations and non-zero symbols for atypical ones. We distinghuished two subcases of 3), where the expected cost constraint is either imposed under both hypotheses or just under $\mathcal{H}=0$. We determined the Stein-exponent for both subcases, and showed that they differ in general, except when the local observations at the sensor follow the same distributions under both marginals.

Appendix A
Achievability Proof of Equation (17) (Partially-Connected DMCs)

We analyze the coding scheme proposed in Section II-B.

Define $\gamma_{x'} \triangleq \Gamma_{Y|X}(y^*|x')$.

<u>Analysis of α_n:</u>

$$
\begin{aligned}
1-\alpha_n &\\
&= \Pr\left[\hat{\mathcal{H}}=0|\mathcal{H}=0\right] && (28)\\
&= \Pr\left[\exists i\in\{1,\ldots,k'\}\colon Y_i=y^* \text{ and } V^n\in\mathcal{T}^{(n)}_{\mu}(P_V)|\mathcal{H}=0\right] && (29)\\
&\overset{(a)}{=} \Pr\big[\exists i\in\{1,\ldots,k'\}\colon Y_i=y^* \text{ and } V^n\in\mathcal{T}^{(n)}_{\mu}(P_V) \text{ and}\\
&\qquad X^{k'}=x'^{k'} \text{ and } U^n\in\mathcal{T}^{(n)}_{\mu}(P_U)|\mathcal{H}=0\big] && (30)\\
&= \Pr\left[V^n\in\mathcal{T}^{(n)}_{\mu}(P_V) \text{ and } U^n\in\mathcal{T}^{(n)}_{\mu}(P_U)|\mathcal{H}=0\right]\\
&\quad\cdot\underbrace{\Pr\left[X^{k'}=x'^{k'}|U^n\in\mathcal{T}^{(n)}_{\mu}(P_U)\right]}_{=1}\\
&\quad\cdot\Pr\left[\exists i\in\{1,\ldots,k'\}\colon Y_i=y^*|X^{k'}=x'^{k'},\mathcal{H}=0\right] && (31)\\
&= \Pr\left[V^n\in\mathcal{T}^{(n)}_{\mu}(P_V) \text{ and } U^n\in\mathcal{T}^{(n)}_{\mu}(P_U)|\mathcal{H}=0\right]\\
&\quad\cdot(1-(1-\gamma_{x'})^{k'}), && (32)
\end{aligned}
$$

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

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

$$\lim_{n\to\infty}\Pr\left[V^n\in\mathcal{T}^{(n)}_{\mu}(P_V) \text{ and } U^n\in\mathcal{T}^{(n)}_{\mu}(P_U)|\mathcal{H}=0\right]=1. \tag{33}$$

Plugging these limits into (32), we can conclude that the type-I error probability of our scheme vanishes:

$$\lim_{n\to\infty}\alpha_n=0. \tag{34}$$

<u>Analysis of β_n and θ:</u> Similarly to above, we have:

$$
\begin{aligned}
\beta_n &= \Pr\left[\hat{\mathcal{H}}=0|\mathcal{H}=1\right] && (35)\\
&= \Pr\big[\exists i\in\{1,\ldots,k'\}\colon Y_i=y^* \text{ and } V^n\in\mathcal{T}^{(n)}_{\mu}(P_V)\\
&\qquad \text{and } X^k=x'^k \text{ and } U^n\in\mathcal{T}^{(n)}_{\mu}(P_U)|\mathcal{H}=1\big] && (36)\\
&= \Pr\left[V^n\in\mathcal{T}^{(n)}_{\mu}(P_V) \text{ and } U^n\in\mathcal{T}^{(n)}_{\mu}(P_U)|\mathcal{H}=1\right]\\
&\quad\cdot\underbrace{\Pr\left[X^{k'}=x'^{k'}|U^n\in\mathcal{T}^{(n)}_{\mu}(P_U)\right]}_{=1}\\
&\quad\cdot\Pr\left[\exists i\in\{1,\ldots,k'\}\colon Y_i=y^*|X^{k'}=x'^{k'}|\mathcal{H}=1\right] && (37)\\
&\le \Pr\left[V^n\in\mathcal{T}^{(n)}_{\mu}(P_V) \text{ and } U^n\in\mathcal{T}^{(n)}_{\mu}(P_U)|\mathcal{H}=1\right] && (38)\\
&\le (n+1)^{|\mathcal{U}||\mathcal{V}|}2^{-n\min D(\pi_{UV}\|Q_{UV})} && (39)
\end{aligned}
$$

We can conclude that

$$\varliminf_{n\to\infty}-\frac{1}{n}\log\beta_n\ge\min D(\pi_{UV}\|Q_{UV}), \tag{40}$$

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 (17) can be shown to be achievable.

Appendix B
Converse Proof to Equation (17) under Stringent Cost Constraints (Partially-Connected DMCs)

We start by proving the converse to (17) under Scenario 2 with stringent cost constraints. Define

$$k'(n) \triangleq \frac{C_n}{c_{\min}}, \tag{41}$$

where $c_{\min}$ denotes the minimum cost of any non-zero input:

$$c_{\min} \triangleq \min_{x \neq 0} c(x) \tag{42}$$

with $c_{\min} > 0$. 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

$$\tilde{\mathcal{X}}^n \triangleq \{x^n \in \mathcal{X}^n : w_{\mathrm{H}}(x^n) \leq k'(n)\}. \tag{43}$$

We show in the following that the cardinality of $\tilde{\mathcal{X}}^n$ grows sublinearly in n, which immediately establishes the converse result to (17) because the type-II error exponent in this setup cannot exceed the type-II error exponent in the setup of [18] where communication consists in sending a sublinear number of noiseless bits.

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:

$$\left|\tilde{\mathcal{X}}^n\right| \leq (n+1)^{|\mathcal{X}|} 2^{n \max_{\boldsymbol{\pi}} H(\boldsymbol{\pi})}, \tag{44}$$

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

$$\begin{aligned}&\lim_{n\to\infty} \frac{1}{n} \log \left|\tilde{\mathcal{X}}^n\right| \\ &\leq \lim_{n\to\infty} \left[\frac{|\mathcal{X}|}{n} \log(n+1) + \max_{\substack{\boldsymbol{\pi}: \\ \boldsymbol{\pi}(0) \geq 1 - \frac{k'(n)}{n}}} H(\boldsymbol{\pi}) \right] = 0. \end{aligned}\tag{45}$$

This concludes the converse proof for Scenario 2.

To prove the converse to (17) in Scenario 3, define $k'(n)$ an increasing function of positive integers satisfying:

$$\varliminf_{n\to\infty} \frac{k'(n)}{C_n} = \infty \tag{46a}$$

$$\lim_{n\to\infty} \frac{k'(n)}{n} = 0, \tag{46b}$$

and the corresponding set of low-weight inputs:

$$\tilde{\mathcal{X}}^n \triangleq \{x^n \in \mathcal{X}^n : w_{\mathrm{H}}(x^n) \leq k'(n)\}, \tag{47}$$

Under expected resource constraints, we have:

$$\begin{aligned} C_n &\geq \sum_{i=1}^{n} \mathbb{E}[c(X_i)|\mathcal{H} = 0] &(48)\\ &\geq c_{\min} \cdot \mathbb{E}[w_{\mathrm{H}}(X^n)|\mathcal{H} = 0] &(49)\\ &\geq c_{\min} \cdot k'(n) \cdot \Pr[X^n \notin \tilde{\mathcal{X}}^n|\mathcal{H} = 0], &(50)\end{aligned}$$

where $c_{\min} = \min_{x\neq 0} c(x) > 0$. Then, the following simple bound holds:

$$\Pr[X^n \notin \tilde{\mathcal{X}}|\mathcal{H} = 0] \leq \frac{C_n}{c_{\min} \cdot k'(n)}. \tag{51}$$

By our assumptions on $k'(n)$ in (46), the set $\tilde{\mathcal{X}}$ thus occurs asymptotically with probability 1:

$$\lim_{n\to\infty} \Pr\left[X^n \notin \tilde{\mathcal{X}}^n \,\middle|\, \mathcal{H} = 0\right] = 0. \tag{52}$$

Repeating the arguments above (Equations (44)–(45)), one can show that also for this new set $\tilde{\mathcal{X}}^n$ it holds that:

$$\lim_{n\to\infty} \frac{1}{n} \log \left|\tilde{\mathcal{X}}^n\right| = 0 \tag{53}$$

The desired converse proof is then obtained by restricting to the set $\tilde{\mathcal{X}}^n$ and invoking the result and steps in [18]. In more detail, one can follow the steps after (36) until (65) in [15] specialized to a constant S.

Appendix C
Converse Proof to Equation (18) (Fully-Connected DMC, Sublinear k or Stringent Cost Constraints)

Even though the converse to $\theta^{\star}_{\text{str-cost}}(\epsilon, \{C_n\})$ directly implies a converse for $\theta^{\star}_{\text{sublin}}(\epsilon)$, we find it instructive to present converse proofs for both scenarios.

A. Converse proof for $\theta^{\star}_{\text{sublin}}(\epsilon)$

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

$$\mathcal{A}_V(y^k) \triangleq \{v^n \in \mathcal{V}^n : g^{(n)}(v^n, y^k) = 0\}, \quad y^k \in \mathcal{Y}^k, \tag{54}$$

and

$$\mathcal{A}_V \triangleq \bigcup_{y^k \in \mathcal{Y}^k} \mathcal{A}_V(y^k). \tag{55}$$

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

$$\begin{aligned}&\beta_n \\ &= \Pr\left[\hat{\mathcal{H}} = 0|\mathcal{H} = 1\right] &(56)\\ &= \sum_{y^k} \Pr\left[Y^k = y^k, V^n \in \mathcal{A}_V(y^k)|\mathcal{H} = 1\right] &(57)\\ &= \sum_{y^k} \sum_{v^n \in \mathcal{A}_V(y^k)} \sum_{u^n} \Pr\left[Y^k = y^k, V^n = v^n, U^n = u^n|\mathcal{H} = 1\right] &(58)\\ &= \sum_{y^k} \sum_{v^n \in \mathcal{A}_V(y^k)} \sum_{u^n} \Pr\left[V^n = v^n, U^n = u^n|\mathcal{H} = 1\right] \\ &\qquad \cdot \Pr\left[Y^k = y^k|U^n = u^n\right]. &(59)\end{aligned}$$

We continue to bound the second probability, by noticing that

$$\Pr\left[Y^k = y^k | U^n = u^n\right] \geq \gamma_{\min}^k, \tag{60}$$

where we define

$$\gamma_{\min} \triangleq \min_{x,y} \Gamma_{Y|X}(y|x), \tag{61}$$

which is strictly positive by assumption. Thus:

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

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

$$\varlimsup_{n\to\infty} -\frac{1}{n}\log \beta_n \leq \varlimsup_{n\to\infty} -\frac{1}{n} \log \Pr[V^n \in \mathcal{A}_V | \mathcal{H} = 1], \tag{65}$$

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

$$\Pr[V^n \in \mathcal{A}_V | \mathcal{H} = 0] \geq \Pr[V^n \in \mathcal{A}_V(Y^k) | \mathcal{H} = 0] = 1 - \alpha_n, \tag{66}$$

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

$$\varlimsup_{n\to\infty} \Pr[V^n \notin \mathcal{A}_V | \mathcal{H} = 0] \leq \epsilon < 1. \tag{67}$$

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

$$\varlimsup_{n\to\infty} -\frac{1}{n} \log \Pr[V^n \in \mathcal{A}_V | \mathcal{H} = 1] \leq D(P_V \| Q_V), \tag{68}$$

which concludes the proof.

B. *Converse proof for* $\theta^\star_{\text{str-cost}}(\epsilon, \{C_n\})$

We reuse the notations

$$k'(n) \triangleq \frac{C_n}{c_{\min}}, \tag{69}$$

and

$$\tilde{\mathcal{X}}^n \triangleq \{x^n \in \mathcal{X}^n \colon w_{\mathrm{H}}(x^n) \leq k'(n)\}, \tag{70}$$

and recall that in Scenario 2 all transmitted input sequences x^n lie in $\tilde{\mathcal{X}}^n$. Therefore, for any two input sequences $x_1^n, x_2^n \in \tilde{\mathcal{X}}^n$ and any output sequence $y^n \in \mathcal{Y}^n$ we have:

$$\gamma_Q^{2k'(n)} \leq \frac{\Pr\left[Y^n = y^n | X^n = x_1^n\right]}{\Pr\left[Y^n = y^n | X^n = x_2^n\right]} \leq \gamma_Q^{-2k'(n)}, \tag{71}$$

where

$$\gamma_Q \triangleq \min_{\substack{x_1, x_2, y\colon\\ x_1 \neq x_2}} \frac{\Gamma_{Y|X}(y|x_1)}{\Gamma_{Y|X}(y|x_2)} \in (0, 1]. \tag{72}$$

Trivially, this implies also the bounds:

$$\gamma_Q^{2k'(n)} \leq \frac{\Pr\left[Y^n = y^n | U^n = u_1^n\right]}{\Pr\left[Y^n = y^n | U^n = u_2^n\right]} \leq \gamma_Q^{-2k'(n)}, \tag{73}$$

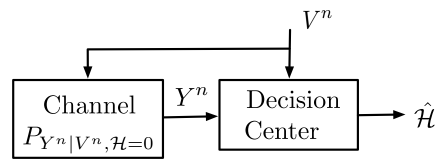

Fig. 3: Derived binary hypothesis test with channel $P_{Y^n|V^n,\mathcal{H}=0}$ used under both hypotheses.

for any two source sequences $u_1^n, u_2^n \in \mathcal{U}^n$ and output sequence $y^n \in \mathcal{Y}^n$, and in particular for any two measures R and R' on $\mathcal{U}^n$:

$$\begin{aligned}
&\mathbb{E}_R\left[\Pr\left[Y^n = y^n | U^n\right]\right] \\
&\geq \gamma_Q^{2k'(n)} \mathbb{E}_{R'}\left[\Pr\left[Y^n = y^n | U^n\right]\right]. &(74)
\end{aligned}$$

Define the acceptance regions

$$\mathcal{A}_V(y^n) \triangleq \{v^n \in \mathcal{V}^n \colon g^{(n)}(v^n, y^n) = 0\}, \quad y^n \in \mathcal{Y}^n. \tag{75}$$

Similarly to the converse proof to (18)—but where y^k needs to be replaced by y^n—we have:

$$\begin{aligned}
\beta_n &= \sum_{y^n} \sum_{v^n \in \mathcal{A}_V(y^n)} \Pr\left[V^n = v^n | \mathcal{H} = 1\right] \\
&\qquad \cdot \Pr\left[Y^n = y^n | V^n = v^n, \mathcal{H} = 1\right] &(76)\\
&\overset{(a)}{\geq} \sum_{y^n} \sum_{v^n \in \mathcal{A}_V(y^n)} \Pr\left[V^n = v^n | \mathcal{H} = 1\right] \\
&\qquad \cdot \Pr\left[Y^n = y^n | V^n = v^n, \mathcal{H} = 0\right] \gamma_Q^{2k'(n)}, &(77)
\end{aligned}$$

where Inequality (a) holds by Inequality (74) because both $\Pr\left[Y^n = y^n | V^n = v^n, \mathcal{H} = 1\right]$ and $\Pr\left[Y^n = y^n | V^n = v^n, \mathcal{H} = 0\right]$ are expectations of the probabilities $\Pr\left[Y^n = y^n | U^n\right]$ with respect to the two measures $P_{U|V}^{\otimes n}$ and $Q_{U|V}^{\otimes n}$.

By assumption,

$$\lim_{n\to\infty} -\frac{1}{n} \log \gamma_Q^{2k'(n)} = 0, \tag{78}$$

and thus

$$\begin{aligned}
&\varlimsup_{n\to\infty} -\frac{1}{n} \log \beta_n \\
&\leq \varlimsup_{n\to\infty} -\frac{1}{n} \log \sum_{y^n} \sum_{v^n \in \mathcal{A}_V(y^n)} \Pr\left[V^n = v^n | \mathcal{H} = 1\right] \\
&\qquad \cdot \Pr\left[Y^n = y^n | V^n = v^n, \mathcal{H} = 0\right]. &(79)
\end{aligned}$$

Notice however that the right-hand side of the above equation corresponds to the miss-detection error exponent of a local detection problem of the form in Figure 3, where the decision center observes V^n, which is i.i.d. according to P_V or Q_V depending on the two hypotheses, and the outcome Y^n, which is generated from V^n according to the law $\Pr\left[Y^n = y^n | V^n = v^n, \mathcal{H} = 0\right]$ irrespective of the hypothesis. This hypothesis testing problem is however a special case of a randomized decision rule for a binary hypothesis test with observation V^n. Moreover, we know that under $\mathcal{H} = 0$ for the randomized acceptance region $\mathcal{A}_V(Y^n)$:

$$1 - \alpha_n \triangleq \sum_{y^n} \sum_{v^n \in \mathcal{A}_V(y^n)} \Pr\left[V^n = v^n | \mathcal{H} = 0\right]$$

$$\cdot \Pr\left[Y^n = y^n | V^n = v^n, \mathcal{H}=0\right], \quad (80)$$

which by assumption is bounded from below by $1-\epsilon > 0$. By standard Stein-exponent arguments, we can thus deduce that the Limit on the right-hand side of (79) is bounded above by $D(P_V \| Q_V)$, which concludes the proof.

Appendix D
Achievability Proofs for Equations (19) and (20) (Fully-Connected DMC, Expected Cost Constraints)

A. Achievability Proof for Equation (19)

Consider the following scheme. Let $\hat{x}$ be the maximizer in (16b) and define

$$\mu_n \triangleq \sqrt{\frac{|\mathcal{U}| c(\hat{x})}{4C_n}}. \quad (81)$$

Notice that by Assumptions (13):

$$\lim_{n\to\infty} \mu_n = 0 \quad (82a)$$

$$\lim_{n\to\infty} \mu_n^2 n = \infty. \quad (82b)$$

Sensor: If $U^n \in \mathcal{T}_{\mu_n}(P_U)$, send $X^n = 0^n$. Else, send $X^n = \hat{x}^n$.

Decision Center: If $Y^n \in \mathcal{T}_{\mu_n}^{(n)}(\Gamma_{Y|X}(\cdot|x=0))$ and $V^n \in \mathcal{T}_{\mu}^{(n)}(P_V)$, declare $\hat{\mathcal{H}}=0$. Otherwise, declare $\hat{\mathcal{H}}=1$.
Analysis of cost constraint: The expected cost is:

$$\sum_{i=1}^{n} \mathbb{E}[c(X_i)|\mathcal{H}=0] = \Pr[U^n \notin \mathcal{T}_{\mu_n}^{(n)}(P_U)|\mathcal{H}=0] \cdot nc(\hat{x}) \quad (83)$$

$$\leq \frac{|\mathcal{U}|}{4\mu_n^2 n} nc(\hat{x}) = C_n, \quad (84)$$

where in the inequality we used the well-known inequality on the probability of the typical set from [27, Remark to Lemma 2.12].
Analysis of α_n: We have:

$$\begin{aligned}
&1-\alpha_n \\
&= \Pr\left[\hat{\mathcal{H}}=0|\mathcal{H}=0\right] && (85)\\
&= \Pr\left[Y^n \in \mathcal{T}_{\mu_n}^{(n)}(\Gamma_{Y|X}(\cdot|0)) \text{ and } V^n \in \mathcal{T}_{\mu}^{(n)}(P_V)|\mathcal{H}=0\right] && (86)\\
&\geq \Pr\left[Y^n \in \mathcal{T}_{\mu_n}^{(n)}(\Gamma_{Y|X}(\cdot|0)) \text{ and } \right.\\
&\qquad \left. U^n \in \mathcal{T}_{\mu_n}^{(n)}(P_U) \text{ and } V^n \in \mathcal{T}_{\mu}^{(n)}(P_V)|\mathcal{H}=0\right] && (87)\\
&= \Pr[Y^n \in \mathcal{T}_{\mu_n}^{(n)}(\Gamma_{Y|X}(\cdot|0))|X^n=0^n] \\
&\quad \cdot \Pr\left[U^n \in \mathcal{T}_{\mu_n}^{(n)}(P_U) \text{ and } V^n \in \mathcal{T}_{\mu_n}^{(n)}(P_V)|\mathcal{H}=0\right]. && (88)
\end{aligned}$$

Notice that by the weak law of large numbers and because the sequence μ_n satisfies (82), both probabilities in (88) tend to 1 and thus

$$\lim_{n\to\infty} \alpha_n = 0. \quad (89)$$

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

$$\begin{aligned}
\beta_n &= \Pr\left[\hat{\mathcal{H}}=0|\mathcal{H}=1\right] && (90)\\
&= \Pr\left[Y^n \in \mathcal{T}_{\mu_n}^{(n)}(\Gamma_{Y|X}(\cdot|0)) \text{ and } \right.\\
&\qquad \left. U^n \in \mathcal{T}_{\mu_n}^{(n)}(P_U) \text{ and } V^n \in \mathcal{T}_{\mu_n}^{(n)}(P_V) \;\middle|\; \mathcal{H}=1\right]\\
&\quad + \Pr\left[Y^n \in \mathcal{T}_{\mu_n}^{(n)}(\Gamma_{Y|X}(\cdot|0)) \text{ and } \right.\\
&\qquad \left. U^n \notin \mathcal{T}_{\mu_n}^{(n)}(P_U) \text{ and } V^n \in \mathcal{T}_{\mu_n}^{(n)}(P_V) \;\middle|\; \mathcal{H}=1\right] && (91)\\
&\leq \Pr\left[U^n \in \mathcal{T}_{\mu_n}^{(n)}(P_U) \text{ and } V^n \in \mathcal{T}_{\mu_n}^{(n)}(P_V) \;\middle|\; \mathcal{H}=1\right]\\
&\quad + \Pr\left[Y^n \in \mathcal{T}_{\mu_n}^{(n)}(\Gamma_{Y|X}(\cdot|0)) \;\middle|\; X^n=\hat{x}^n\right]\\
&\quad \cdot \Pr\left[V^n \in \mathcal{T}_{\mu_n}^{(n)}(P_V)|\mathcal{H}=1\right] && (92)\\
&\leq (n+1)^{|\mathcal{U}||\mathcal{V}|} 2^{-n \min D(\pi_{UV}\|Q_{UV})}\\
&\quad + (n+1)^{|\mathcal{Y}|} 2^{-n \min D(\pi_Y\|\Gamma_{Y|X=\hat{x}})}\\
&\quad \cdot (n+1)^{|\mathcal{V}|} 2^{-n\min D(\pi_V\|Q_V)}, && (93)
\end{aligned}$$

where the three minima are over types π_{UV}, π_Y, and π_V satisfying the respective typicality constraints. We next notice that since $\mu_n \to 0$ as $n \to \infty$, the following limits hold as $n \to \infty$:

$$\lim_{n\to\infty} \min D(\pi_Y\|\Gamma_{Y|X=\hat{x}}) = D(\Gamma_{Y|X=0}\|\Gamma_{Y|X=\hat{x}}) \quad (94)$$

$$\lim_{n\to\infty} \min D(\pi_V\|Q_V) = D(P_V\|Q_V) \quad (95)$$

$$\lim_{n\to\infty} \min D(\pi_{UV}\|Q_{UV}) = E_1. \quad (96)$$

We can thus conclude that for the proposed scheme:

$$\lim_{n\to\infty} -\frac{1}{n}\log \beta_n \geq \min\{E_1, E_2\}. \quad (97)$$

B. Achievability Proof for Equation (20)

The scheme is similar to the scheme in the previous section, except that the sensor also sends $X^n = 0^n$ when $U^n \in \mathcal{T}_{\mu_n}^{(n)}(Q_U)$.
Analysis of cost constraint: The expected cost under hypothesis $\mathcal{H}=0$ satisfies

$$\sum_{i=1}^{n} \mathbb{E}[c(X_i)|\mathcal{H}=0] \leq \Pr[U^n \notin \mathcal{T}_{\mu_n}^{(n)}(P_U)|\mathcal{H}=0] \cdot nc(\hat{x}) \quad (98)$$

$$\leq \frac{|\mathcal{U}|}{4\mu_n^2 n} nc(\hat{x}) = C_n, \quad (99)$$

and similarly, under $\mathcal{H}=1$:

$$\sum_{i=1}^{n} \mathbb{E}[c(X_i)|\mathcal{H}=1] \leq \Pr[U^n \notin \mathcal{T}_{\mu_n}^{(n)}(Q_U)|\mathcal{H}=1] \cdot nc(\hat{x}) \quad (100)$$

$$\leq \frac{|\mathcal{U}|}{4\mu_n^2 n} nc(\hat{x}) = C_n, \quad (101)$$

Analysis of α_n: One can follow exactly the same analysis steps (85)–(88) as in the previous section.
Analysis of β_n: We have:

$$\begin{aligned}
&\beta_n \\
&= \Pr\left[\hat{\mathcal{H}}=0|\mathcal{H}=1\right] && (102)\\
&= \Pr\left[Y^n \in \mathcal{T}_{\mu_n}^{(n)}(\Gamma_{Y|X}(\cdot|0)) \text{ and } V^n \in \mathcal{T}_{\mu}^{(n)}(P_V)\right.\\
&\qquad \left. \text{and } U^n \in \mathcal{T}_{\mu_n}^{(n)}(P_U)|\mathcal{H}=1\right]\\
&\quad + \Pr\left[Y^n \in \mathcal{T}_{\mu_n}^{(n)}(\Gamma_{Y|X}(\cdot|0)) \text{ and } V^n \in \mathcal{T}_{\mu}^{(n)}(P_V)\right.\\
&\qquad \left. \text{and } U^n \in \mathcal{T}_{\mu_n}^{(n)}(Q_U)|\mathcal{H}=1\right]
\end{aligned}$$

$$
\begin{aligned}
&+\Pr\Big[Y^n \in \mathcal{T}^{(n)}_{\mu_n}(\Gamma_{Y|X}(\cdot|0)) \text{ and } V^n \in \mathcal{T}^{(n)}_{\mu}(P_V) \\
&\qquad\qquad \text{and } U^n \notin (\mathcal{T}^{(n)}_{\mu_n}(P_U) \cup \mathcal{T}^{(n)}_{\mu_n}(Q_U))|\mathcal{H}=1\Big] \\
&\le \Pr\Big[U^n \in \mathcal{T}^{(n)}_{\mu_n}(P_U) \text{ and } V^n \in \mathcal{T}^{(n)}_{\mu}(P_V)|\mathcal{H}=1\Big] \\
&\quad +\Pr\Big[U^n \in \mathcal{T}^{(n)}_{\mu_n}(Q_U) \text{ and } V^n \in \mathcal{T}^{(n)}_{\mu}(P_V)|\mathcal{H}=1\Big] \\
&\quad +\Pr[Y^n \in \mathcal{T}^{(n)}_{\mu_n}(\Gamma_{Y|X}(\cdot|0))|X^n=\hat{x}^n] \\
&\qquad \cdot \Pr\Big[V^n \in \mathcal{T}^{(n)}_{\mu}(P_V)|\mathcal{H}=1\Big] \qquad (103) \\
&\le (n+1)^{|\mathcal{U}||\mathcal{V}|}2^{-n\min D(\pi_{UV}\|Q_{UV})} \\
&\quad +(n+1)^{|\mathcal{U}||\mathcal{V}|}2^{-n\min D(\pi_{UV}\|Q_{UV})} \\
&\quad +2^{-n\min D(\pi_Y\|\Gamma_{Y|X=\hat{x}})}(n+1)^{|\mathcal{V}|+|\mathcal{Y}|}2^{-n\min D(\pi_V\|Q_V)},
\end{aligned}
\tag{104}
$$

where the first two minimima are over types π_{UV} with marginals that are μ_n-close to P_U and P_V or to Q_U and P_V, respectively, and the latter two minima are over types π_Y and π_V that are μ_n-close to $\Gamma_{Y|X=0}$ and P_V, respectively. Letting $n \to \infty$, by standard arguments we have:

$$
\varliminf_{n\to\infty} -\frac{1}{n}\log\beta_n \ge \min\{E_1, E_2, E_3\}. \tag{105}
$$

Appendix E
Converse Proofs for Equations (19) and (20) (Fully-Connected DMC, Expected Cost Constraints)

The following considerations will be used throughout this appendix. Define $k'(n)$ an increasing function of positive integers satisfying:

$$
\varliminf_{n\to\infty} \frac{k'(n)}{C_n} = \infty \tag{106a}
$$

$$
\lim_{n\to\infty} \frac{k'(n)}{n} = 0. \tag{106b}
$$

Also, define the set of u^n-sequences that are mapped to "low-weight inputs"

$$
\tilde{\mathcal{U}}^n \triangleq \{u^n \in \mathcal{U}^n : w_{\mathrm{H}}(f^{(n)}(u^n)) \le k'(n)\}, \tag{107}
$$

Under an expected resource constraint under hypothesis H, we have:

$$
\begin{aligned}
C_n &\ge \sum_{i=1}^n \mathbb{E}[c(X_i)|\mathcal{H}=\mathsf{H}] && (108)\\
&\ge c_{\min}\cdot \mathbb{E}[w_{\mathrm{H}}(X^n)|\mathcal{H}=\mathsf{H}] && (109)\\
&\ge c_{\min}\cdot k'(n)\cdot \Pr[U^n \notin \tilde{\mathcal{U}}^n|\mathcal{H}=\mathsf{H}], && (110)
\end{aligned}
$$

where $c_{\min} = \min_{x\neq 0} c(x) > 0$. Then, the following simple bound holds:

$$
\Pr[U^n \notin \tilde{\mathcal{U}}|\mathcal{H}=\mathsf{H}] \le \frac{C_n}{c_{\min}\cdot k'(n)}. \tag{111}
$$

By our assumptions on $k'(n)$ in (106a), the set $\tilde{\mathcal{U}}$ thus occurs asymptotically with probability 1:

$$
\lim_{n\to\infty} \Pr\Big[U^n \notin \tilde{\mathcal{U}}^n \,\Big|\mathcal{H}=\mathsf{H}\Big] = 0, \tag{112}
$$

under any of the hypotheses H for which the expected resource constraint is imposed.

Let $\{\mu_n\}_n$ be any sequence satisfying

$$
\lim_{n\to\infty} \mu_n = 0 \tag{113a}
$$

$$
\lim_{n\to\infty} \mu_n^2 n = \infty. \tag{113b}
$$

Since the probability of the set $\mathcal{T}^{(n)}_{\mu_n/2}(P_U)$ asymptotically tends to 1 as $n\to\infty$ under hypothesis $\mathsf{H}=0$, we conclude that if Limit (112) holds for $\mathsf{H}=0$, then

$$
\lim_{n\to\infty} \Pr\Big[U^n \in \Big(\mathcal{T}^{(n)}_{\mu_n/2}(P_U)\cap\tilde{\mathcal{U}}^n\Big) \,\Big|\, \mathcal{H}=0\Big] = 1. \tag{114}
$$

Similarly, we can conclude that if Limit (112) holds for $\mathsf{H}=1$, then

$$
\lim_{n\to\infty} \Pr\Big[U^n \in \Big(\mathcal{T}^{(n)}_{\mu_n/2}(Q_U)\cap\tilde{\mathcal{U}}^n\Big) \,\Big|\, \mathcal{H}=1\Big] = 1, \tag{115}
$$

and because all sequences in $\mathcal{T}^{(n)}_{\mu_n/2}(Q_U)$ have almost equal probability:

$$
\lim_{n\to\infty} \frac{\Big|\mathcal{T}^{(n)}_{\mu_n/2}(Q_U)\cap\tilde{\mathcal{U}}^n\Big|}{\Big|\mathcal{T}^{(n)}_{\mu_n/2}(Q_U)\Big|} = 1. \tag{116}
$$

A. Upper Bounds E_1 and E_3

The region $\mathcal{F}_n$ and its probability: Let F_{UV} be an arbitrary pmf with marginals $F_V = P_V$ and $F_U = P_U$ if the expected cost constraint is only imposed under $\mathcal{H}=0$, and marginals $F_V = P_V$ and $F_U \in \{P_U, Q_U\}$ if the expected cost constraint is imposed under both hypotheses.

Define the region

$$
\begin{aligned}
\mathcal{F}_n \triangleq \{(u^n, v^n, y^n) \colon (u^n, v^n) \in \mathcal{T}^{(n)}_{\mu_n}(F_{UV}),\\
u^n \in (\tilde{\mathcal{U}}^n \cap \mathcal{T}^{(n)}_{\mu_n/2}(F_U)),\ g(v^n, y^n) = 0\}.
\end{aligned}
\tag{117}
$$

For the probability of this set under $\mathcal{H}=0$ we have:

$$
\begin{aligned}
\Delta_n &\triangleq \Pr[(U^n, V^n, Y^n) \in \mathcal{F}_n|\mathcal{H}=0] \qquad (118)\\
&= \Pr\Big[U^n \in (\tilde{\mathcal{U}}^n \cap \mathcal{T}^{(n)}_{\mu_n/2}(F_U))|\mathcal{H}=0\Big] \\
&\quad \cdot \Pr\Big[(U^n, V^n) \in \mathcal{T}^{(n)}_{\mu_n}(F_{UV})|U^n \in (\tilde{\mathcal{U}}^n \cap \mathcal{T}^{(n)}_{\mu_n/2}(F_U)),\\
&\qquad\qquad \mathcal{H}=0\Big] \\
&\quad \cdot \Pr\Big[g^{(n)}(V^n, Y^n) = 0|U^n \in (\tilde{\mathcal{U}}^n \cap \mathcal{T}^{(n)}_{\mu_n/2}(F_U)),\\
&\qquad\qquad (U^n, V^n) \in \mathcal{T}^{(n)}_{\mu_n}(F_{UV}), \mathcal{H}=0\Big].
\end{aligned}
\tag{119}
$$

We bound each of these probabilities individually. Notice first that by (112) and (114) for $F_U = P_U$ we have:

$$
\Pr\Big[U^n \in (\tilde{\mathcal{U}}^n \cap \mathcal{T}^{(n)}_{\mu_n/2}(P_U))|\mathcal{H}=0\Big] \ge 1 - o(1). \tag{120}
$$

If (116) holds, then because all sequences of $\mathcal{T}^{(n)}_{\mu_n/2}(F_U))$ have approximately equal probability, we have for $F_U = Q_U$:

$$
\Pr\Big[U^n \in (\tilde{\mathcal{U}}^n \cap \mathcal{T}^{(n)}_{\mu_n/2}(F_U))|\mathcal{H}=0\Big] \ge 2^{-n(D(F_U\|P_U)+o(1))}. \tag{121}
$$

We further observe that by well-known arguments, if $F_{V|U} \neq P_{V|U}$:

$$
\Pr\Big[(U^n, V^n) \in \mathcal{T}^{(n)}_{\mu_n}(F_{UV})|U^n \in (\tilde{\mathcal{U}}^n \cap \mathcal{T}^{(n)}_{\mu_n/2}(F_U)),\ \mathcal{H}=0\Big]
$$

$$\geq 2^{-n(D(F_{UV}\|F_U P_{V|U})+o(1))}. \tag{122}$$

If $F_{V|U} = P_{V|U}$:

$$\Pr\Big[(U^n,V^n)\in\mathcal{T}^{(n)}_{\mu_n}(F_{UV})|U^n\in(\tilde{\mathcal{U}}^n\cap\mathcal{T}^{(n)}_{\mu_n/2}(F_U)),\ \mathcal{H}=0\Big] \geq 1-o(1). \tag{123}$$

It remains to analyze the third probabilty in (119). For ease of notation, let $T_{U^nV^n}(u^n,v^n)$ denote the pmf of the tuples (u^n,v^n) conditional on $\mathcal{H}=0$ and on the conditions $(U^n,V^n)\in\mathcal{T}^{(n)}_{\mu_n}(F_{UV})$ and $U^n\in(\tilde{\mathcal{U}}^n\cap\mathcal{T}^{(n)}_{\mu_n/2}(F_U))$:

$$\begin{aligned}&T_{U^nV^n}(u^n,v^n)\\ &\triangleq \Pr[U^n=u^n,V^n=v^n|U^n\in(\tilde{\mathcal{U}}^n\cap\mathcal{T}^{(n)}_{\mu_n/2}(F_U)),\\ &\qquad (U^n,V^n)\in\mathcal{T}^{(n)}_{\mu_n}(F_{UV}),\mathcal{H}=0].\end{aligned} \tag{124}$$

Similarly, let $S_{U^nV^n}(u^n,v^n)$ denote the pmf of the tuples (u^n,v^n) conditional on $\mathcal{H}=0$ and on the condition $U^n\in\tilde{\mathcal{U}}^n$:

$$S_{U^nV^n}(u^n,v^n)\triangleq\Pr[U^n=u^n,V^n=v^n|U^n\in\tilde{\mathcal{U}}^n,\mathcal{H}=0]. \tag{125}$$

By the definition of the set $\tilde{\mathcal{U}}^n$, we can write:

$$\begin{aligned}&\Pr\Big[g^{(n)}(V^n,Y^n)=0|U^n\in(\tilde{\mathcal{U}}^n\cap\mathcal{T}^{(n)}_{\mu_n/2}(F_U)),\\ &\qquad (U^n,V^n)\in\mathcal{T}^{(n)}_{\mu_n}(F_{UV}),\mathcal{H}=0\Big]\\ &=\sum_{y^n}\sum_{v^n\in A_V(y^n)}\cdot\sum_{\substack{u^n\in\tilde{\mathcal{U}}^n\cap\mathcal{T}^{(n)}_{\mu_n/2}(F_U):\\ (u^n,v^n)\in\mathcal{T}^{(n)}_{\mu_n}(F_{UV})}}\Pr[Y^n=y^n|U^n=u^n]\cdot T_{U^nV^n}(u^n,v^n) &(126)\\ &\geq\sum_{y^n}\sum_{v^n\in A_V(y^n)}\cdot\gamma_Q^{2k'(n)}\sum_{u^n\in\tilde{\mathcal{U}}^n}\Pr[Y^n=y^n|U^n=u^n]S_{U^nV^n}(u^n,v^n) &(127)\\ &=\gamma_Q^{2k'(n)}\Pr\Big[g^{(n)}(V^n,Y^n)=0|U^n\in\tilde{\mathcal{U}}^n,\mathcal{H}=0\Big], &(128)\end{aligned}$$

where the inequality holds because $u^n\in\tilde{\mathcal{U}}^n$ and thus the corresponding inputs all have weights bounded by $k'(n)$, see also (74) and the definition of γ_Q in (72).

Since for any events A and B it holds that $\Pr[A|B]\geq\Pr[A\cap B]=\Pr[A]-\Pr[A\cap B^c]\geq\Pr[A]-\Pr[B^c]$, we can further bound:

$$\begin{aligned}&\Pr\Big[g^{(n)}(V^n,Y^n)=0|U^n\in\tilde{\mathcal{U}}^n,\mathcal{H}=0\Big]\\ &\geq\Pr\Big[g^{(n)}(V^n,Y^n)=0|\mathcal{H}=0\Big]-\Pr\Big[U^n\notin\tilde{\mathcal{U}}^n|\mathcal{H}=0\Big] &(129)\\ &\geq 1-\epsilon-o(1), &(130)\end{aligned}$$

where the last inequality holds by the assumption on the type-I error probability and by (114).

Combining all these observations, we obtain:

- *If* $F_U=P_U$ *and* $F_{V|U}=P_{V|U}$,
$$\overline{\lim_{n\to\infty}}-\frac{1}{n}\log\Delta_n=0. \tag{131}$$
- *If* $F_U=P_U$ *and* $F_{V|U}\neq P_{V|U}$,
$$\overline{\lim_{n\to\infty}}-\frac{1}{n}\log\Delta_n\leq D(F_{UV}\|P_UP_{V|U}). \tag{132}$$
- *If* $F_U=Q_U$ *and* $F_{V|U}=P_{V|U}$,
$$\overline{\lim_{n\to\infty}}-\frac{1}{n}\log\Delta_n\leq D(Q_U\|P_U)=D(F_{UV}\|P_UP_{V|U}). \tag{133}$$
- *If* $F_U=Q_U$ *and* $F_{V|U}\neq P_{V|U}$
$$\begin{aligned}\overline{\lim_{n\to\infty}}-\frac{1}{n}\log\Delta_n&\leq D(Q_U\|P_U)+D(F_{UV}\|Q_UP_{V|U}) &(134)\\ &=D(F_{UV}\|P_UP_{V|U}). &(135)\end{aligned}$$

So, in all cases we have:

$$\overline{\lim_{n\to\infty}}-\frac{1}{n}\log\Delta_n\leq D(F_{UV}\|P_{UV}). \tag{136}$$

<u>*Change of measure on $\mathcal{F}_n$:*</u>
Define now the tuple $(\tilde{U}^n,\tilde{V}^n,\tilde{Y}^n)$ to be of joint pmf

$$P_{\tilde{U}^n\tilde{V}^n\tilde{Y}^n}(u^n,v^n,y^n)=\frac{P^{\otimes n}_{UV}(u^n,v^n)\Gamma^{\otimes n}_{Y|X}(y^n|f^{(n)}(u^n))}{\Delta_n}\cdot\mathbb{1}\{(u^n,v^n,y^n)\in\mathcal{F}_n\}. \tag{137}$$

We obtain:

$$\begin{aligned}&-\frac{1}{n}\log\beta_n\\ &=-\frac{1}{n}\log\Pr[\hat{\mathcal{H}}=0|\mathcal{H}=1]\\ &=-\frac{1}{n}\log\left(\sum_{\substack{u^n,v^n,y^n:\\ g(v^n,y^n)=0}}Q^{\otimes n}_{UV}(u^n,v^n)\cdot\Pr[Y^n=y^n|U^n=u^n]\right) &(138)\\ &\leq-\frac{1}{n}\log\left(\sum_{\substack{(u^n,v^n,y^n)\\ \in\mathcal{F}_n}}Q^{\otimes n}_{UV}(u^n,v^n)\cdot\Pr[Y^n=y^n|U^n=u^n]\right) &(139)\\ &=-\frac{1}{n}\log Q_{U^nV^nY^n}(\mathcal{F}_n) &(140)\\ &\overset{(a)}{\leq}\frac{1}{n}D\left(P_{\tilde{U}^n\tilde{V}^n\tilde{Y}^n}(\mathcal{F}_n)\|Q_{U^nV^nY^n}(\mathcal{F}_n)\right) &(141)\\ &\overset{(b)}{\leq}\frac{1}{n}D\left(P_{\tilde{U}^n\tilde{V}^n\tilde{Y}^n}\|Q^{\otimes n}_{UV}P_{Y^n|U^n}\right) &(142)\\ &\overset{(c)}{=}\frac{1}{n}\sum_{(u^n,v^n)}P_{\tilde{U}^n\tilde{V}^n}(u^n,v^n)\log\frac{P^{\otimes n}_{UV}(u^n,v^n)}{Q^{\otimes n}_{UV}(u^n,v^n)}-\frac{1}{n}\log\Delta_n &(143)\\ &=\frac{1}{n}\sum_{i=1}^n\sum_{(u^n,v^n)}P_{\tilde{U}^n\tilde{V}^n}(u^n,v^n)\log\frac{P_{UV}(u_i,v_i)}{Q_{UV}(u_i,v_i)}-\frac{1}{n}\log\Delta_n &(144)\\ &=\frac{1}{n}\sum_{i=1}^n\sum_{u_i,v_i}P_{\tilde{U}_i\tilde{V}_i}(u_i,v_i)\log\frac{P_{UV}(u_i,v_i)}{Q_{UV}(u_i,v_i)}\end{aligned}$$

$$-\frac{1}{n}\log\Delta_n \tag{145}$$

$$= \sum_{u,v} P_{\tilde{U}_T\tilde{V}_T}(u,v)\log\frac{P_{UV}(u,v)}{Q_{UV}(u,v)} - \frac{1}{n}\log\Delta_n, \tag{146}$$

where T is a uniform random variable over $\{1,\ldots,n\}$ independent of $(\tilde{U}^n,\tilde{V}^n)$. In above inequalities, (a) holds because $P_{\tilde{U}^n\tilde{V}^n\tilde{Y}^n}$ is only defined on $\mathcal{F}_n$ and thus $P_{\tilde{U}^n\tilde{V}^n\tilde{Y}^n}(\mathcal{F}_n)=1$; (b) holds by the data processing inequality; and (c) holds by the definition of $P_{\tilde{U}^n\tilde{V}^n\tilde{Y}^n}$.

We use Limit (136) and the fact that $P_{\tilde{U}_T\tilde{V}_T}(u,v) = F_{UV}(u,v)+o(1)$ because $\mathcal{F}_n$ only contains (u^n,v^n)-sequences that are F_{UV}-typical. Then, taking the limit $n\to\infty$ on (146), we obtain:

$$\varlimsup_{n\to\infty} -\frac{1}{n}\log\beta_n$$

$$\leq \sum_{u,v} F_{UV}(u,v)\log\frac{P_{UV}(u,v)}{Q_{UV}(u,v)} + D(F_{UV}\|P_{UV}) \tag{147}$$

$$= D(F_{UV}\|Q_{UV}). \tag{148}$$

Concluding the Proof:

Recall that we have derived the bound (148) for any pmf F_{UV} with marginal $F_V=P_V$ and $F_U=P_U$ if the expected cost constraint is imposed only under hypothesis $\mathcal{H}=0$ and marginals $F_V=P_V$ and $F_U\in\{P_U,Q_U\}$ if the cost constraint is imposed under both hypotheses.

By Limits (114) and (148), we thus obtain the desired converse result E_1 on both $\theta_{\text{Exp-cost},\mathcal{H}=0}$ and $\theta_{\text{Exp-cost,both}}$, and by also using Limit (115), we establish the converse result E_3 on $\theta_{\text{Exp-cost,both}}$.

B. Upper Bound E_2

We again use change of measure arguments to obtain upper bound E_2, but now combined with the blowing up lemma.

Define for each sequence u^n, the set of indices where this sequence induces input zero:

$$\mathcal{I}_{u^n} \triangleq \{i\in\{1,\ldots,n\}\colon f_i(u^n)=0\} \tag{149}$$

for $f_i(u^n)$ the i-th component of $f^{(n)}(u^n)$.

Let $\mathcal{A}_n$ denote the acceptance region:

$$\mathcal{A}_n \triangleq \{(v^n,y^n)\colon g(v^n,y^n)=0\}. \tag{150}$$

Further, let $\{\ell_n\}$ be a sequence with $\lim_{n\to\infty}\ell_n/\sqrt{n}=\infty$ and $\lim_{n\to\infty}\ell_n/n=0$. Define then for each v^n the ℓ_n-blown-up: acceptance region

$$\hat{\mathcal{A}}_n^{\ell_n}(v^n) \triangleq \{\tilde{y}^n\colon \exists y^n\in\mathcal{A}_n(v^n,y^n) \text{ s.t. } d_{\mathrm{H}}(\tilde{y}^n,y^n)\leq\ell_n\}. \tag{151}$$

We next define the set on which we will be employing the change of measure. Fix $\eta\in(0,(1-\epsilon))$ and define

$$\mathcal{D}_n \triangleq \{(u^n,v^n)\colon u^n\in\tilde{\mathcal{U}},\ v^n\in\mathcal{T}_{\mu_n}^{(n)}(P_V),$$
$$\Pr[(v^n,Y^n)\in\mathcal{A}_n|U^n=u^n,V^n=v^n]\geq\eta\}, \tag{152}$$

where notice that the last probability is not conditioned on the hypothesis as it only depends on the encoding function and the channel transition law.

We continue to apply the union bound to obtain:

$$\Delta_n \triangleq \Pr\left[(U^n,V^n)\in\mathcal{D}_n \,\middle|\, \mathcal{H}=0\right] \tag{153}$$

$$\geq 1-\Pr[U\notin\tilde{\mathcal{U}}^n|\,\mathcal{H}=0] - \Pr[v^n\notin\mathcal{T}_{\mu_n}^{(n)}(P_V)|\mathcal{H}=0]$$
$$-\Pr[\,\Pr[(V^n,Y^n)\in\mathcal{A}_n|U^n,V^n]<\eta] \tag{154}$$

$$\geq 1-\frac{C_n}{c_{\min}k'(n)} - \frac{|\mathcal{V}|}{4\mu_n^2 n} - \frac{1-\epsilon}{\eta}. \tag{155}$$

To prove the desired converse result, we define the new pmf

$$\tilde{P}_{U^nV^nY^n\mathcal{I}_{U^n}}(u^n,v^n,y^n,\mathcal{I})$$
$$= \frac{P_{UV}^{\otimes n}(u^n,v^n)}{\Delta_n}\cdot\mathbb{1}\{(u^n,v^n)\in\mathcal{D}_n\}\cdot\Gamma_{Y|X}^{\otimes n}(y^n|f^{(n)}(u^n))$$
$$\cdot\mathbb{1}\{\mathcal{I}=\mathcal{I}_{u^n}\} \tag{156}$$

and notice that this new measure satisfies:

$$\tilde{P}(Y^n\in\hat{\mathcal{A}}^{\ell_n}(V^n))$$
$$= \sum_{(u^n,v^n)\in\mathcal{D}_n}\tilde{P}_{U^nV^n}(u^n,v^n)$$
$$\cdot\Pr[Y^n\in\hat{\mathcal{A}}^{\ell_n}(v^n)|U^n=u^n,V^n=v^n]$$
$$\geq \sum_{(u^n,v^n)\in\mathcal{D}_n}\tilde{P}_{U^nV^n}(u^n,v^n)\left(1-\frac{\sqrt{n\ln 1/\eta}}{\ell_n}\right)$$
$$= 1-\lambda_n, \tag{157}$$

where we defined $\lambda_n \triangleq \frac{\sqrt{n\ln 1/\eta}}{\ell_n}$, which by our assumptions tends to 0 for $n\to\infty$; and the inequality is obtained by applying the blowing-up lemma [25, remark p. 446] on each pair (u^n,v^n).

We next examine how the blow-up of the acceptance region influences the type-II error probability:

$$Q\left(Y^n\in\hat{\mathcal{A}}^{\ell_n}(V^n)\right)$$
$$\triangleq \sum_{(u^n,v^n)} Q_{UV}^{\otimes n}(u^n,v^n)$$
$$\cdot\Pr[Y^n\in\hat{\mathcal{A}}^{\ell_n}(v^n)|U^n=u^n,V^n=v^n] \tag{158}$$

$$\leq \sum_{(u^n,v^n)} Q_{UV}^{\otimes n}(u^n,v^n)$$
$$\cdot\Pr[Y^n\in\mathcal{A}(v^n)|U^n=u^n,V^n=v^n]$$
$$\cdot e^{nh_{\mathrm{b}}(\ell_n/n)}\cdot|\mathcal{Y}|^{\ell_n}\cdot\gamma_{\min}^{\ell_n} \tag{159}$$

$$= \beta_n\cdot e^{nh_{\mathrm{b}}(\ell_n/n)}\cdot|\mathcal{Y}|^{\ell_n}\cdot\gamma_{\min}^{\ell_n}, \tag{160}$$

where recall that $\gamma_{\min}\triangleq\min_{(x,y)}\Gamma_{y|x}(y|x)$, and where the inequality can be obtaiend following the steps in [27, the Proof of Lemma 5.1].

Let now

$$\tilde{P}_{\mathcal{H}}(0) \triangleq \tilde{P}\left(Y^n\in\hat{\mathcal{A}}^{\ell_n}(V^n)\right) \tag{161}$$

$$Q_{\mathcal{H}}(0) \triangleq Q\left(Y^n\in\hat{\mathcal{A}}^{\ell_n}(V^n)\right) \tag{162}$$

and $\tilde{P}_{\mathcal{H}}(1)$ and $Q_{\mathcal{H}}(1)$ their complements. By the data processing inequality:

$$D(\tilde{P}_{V^nY^n}\|Q_{V^nY^n}) \geq D(\tilde{P}_{\mathcal{H}}\|Q_{\mathcal{H}}) \tag{163}$$

$$\geq -1+\tilde{P}_{\mathcal{H}}(0)\log\frac{1}{Q_{\mathcal{H}}(0)}, \tag{164}$$

where the second inequality is obtained by writing out the divergence term and bounding binary entropy by 1.

Combining (157) and (160) with (164), we can then write:

$$\begin{aligned}
&-\frac{1}{n}\log\beta_n \\
&\leq \frac{1}{(1-\lambda_n)n}\left(1+D\left(\tilde{P}_{V^nY^n}\|Q_{V^nY^n}\right)\right) \\
&\quad +h_{\mathfrak{b}}(\ell_n/n)+\frac{\ell_n}{n}\log(\gamma_{\min}\cdot|\mathcal{Y}|) && (165)\\
&=\frac{1}{n(1-\lambda_n)}\left(1+D\left(\tilde{P}_{V^n}\|Q_V^{\otimes n}\right)\right) \\
&\quad +\frac{1}{n(1-\lambda_n)}\mathbb{E}_{\tilde{P}_{V^n}}\left[D\left(\tilde{P}_{Y^n|V^n}\|Q_{Y^n|V^n}\right)\right]\Bigg) \\
&\quad +h_{\mathfrak{b}}(\ell_n/n)+\frac{\ell_n}{n}\log(\gamma_{\min}\cdot|\mathcal{Y}|), && (166)
\end{aligned}$$

where the inequality holds by the data-processing inequality.

Following previous arguments, we have

$$\begin{aligned}
&\frac{1}{n}D\left(\tilde{P}_{V^n}\|Q_V^{\otimes n}\right) \\
&\quad\leq \frac{1}{n}\sum_{v^n}\tilde{P}_{V^n}(v^n)\log\left(\frac{P_V^{\otimes n}(v^n)}{Q_V^{\otimes n}(v^n)}\right)-\log\Delta_n, \qquad (167)
\end{aligned}$$

which tends to $D(P_V\|Q_V)$ by our definition of $\tilde{P}_{V^n}$ and because Δ is bounded away from 0 and smaller than 1, see (155).

Before considering the second term in the decomposition (166), we observe that by basic probability:

$$\tilde{P}_{Y^n|V^n}(y^n|v^n)=\mathbb{E}_{\tilde{P}_{\mathcal{I}_{U^n}}}\left[\tilde{P}_{Y^n|V^n\mathcal{I}_{U^n}}(y^n|v^n,\mathcal{I}_{U^n})\right] \qquad (168)$$

$$Q_{Y^n|V^n}(y^n|v^n)=\mathbb{E}_{\tilde{P}_{\mathcal{I}_{U^n}}}\left[Q_{Y^n|V^n}(y^n|v^n)\right], \qquad (169)$$

and thus by the convexity of KL divergence:

$$\begin{aligned}
&\frac{1}{n}\mathbb{E}_{\tilde{P}_{V^n}}\left[D\left(\tilde{P}_{Y^n|V^n}\|Q_{Y^n|V^n}\right)\right] \\
&\quad\leq \frac{1}{n}\mathbb{E}_{\tilde{P}_{V^n\mathcal{I}_{U^n}}}\left[D\left(\tilde{P}_{Y^n|V^n\mathcal{I}_{U^n}}\|Q_{Y^n|V^n}\right)\right]. \qquad (170)
\end{aligned}$$

Moreover, for given realizations $\mathcal{I}_{U^n}=\mathcal{I}$ and $V^n=v^n$, we can decompose the divergence as:

$$\begin{aligned}
&\frac{1}{n}D\left(\tilde{P}_{Y^n|V^n=v^n,\mathcal{I}_{U^n}=\mathcal{I}}\|Q_{Y^n|V^n=v^n}\right) \\
&=\frac{1}{n}D\left(\tilde{P}_{Y^n_{\mathcal{I}}|V^n=v^n,\mathcal{I}_{U^n}=\mathcal{I}}\|Q_{Y^n_{\mathcal{I}}|v^n}\right) \\
&\quad+\frac{1}{n}\mathbb{E}\left[D\left(\tilde{P}_{Y^n_{\mathcal{I}^c}|V^n=v^n,\mathcal{I}_{U^n}=\mathcal{I}}\|Q_{Y^n_{\mathcal{I}^c}|V^n=v^n}\right)\right] && (171)\\
&\leq\frac{1}{n}D\left(\tilde{P}_{Y^n_{\mathcal{I}}|V^n=v^n,\mathcal{I}_{U^n}=\mathcal{I}}\|Q_{Y^n_{\mathcal{I}}|V^n=v^n}\right) \\
&\quad-\frac{n-|\mathcal{I}|}{n}\log\gamma_{\min} && (172)\\
&=\frac{1}{n}D\left(\Gamma_{Y|X}(\cdot|0)^{\otimes|\mathcal{I}|}\|Q_{Y^n_{\mathcal{I}}|V^n=v^n}\right)-\frac{n-|\mathcal{I}|}{n}\log\gamma_{\min}. && (173)
\end{aligned}$$

Notice that the second term in (173) vanishes as $n\to\infty$ by our assumption that $\gamma_{\min}$ is bounded away from zero and because under $\tilde{P}$ the number of non-zero input symbols is upper bounded by $C_n/c_{\min}$ and thus sublinear in n.

The first term in (173) can be bounded using the concavity of the log-function:

$$\begin{aligned}
&\frac{1}{n}D\left(\Gamma_{Y|X}(\cdot|0)^{\otimes|\mathcal{I}|}\|Q_{Y^n_{\mathcal{I}}|v^n}\right) \\
&\leq\frac{1}{n}\sum_{x^n}Q_{X^n|v^n}(x^n)D\left(\Gamma_{Y|X}(\cdot|0)^{\otimes|\mathcal{I}|}\|Q_{Y^n_{\mathcal{I}}|x^n_{\mathcal{I}}}\right) && (174)\\
&=\frac{|\mathcal{I}|}{n}\sum_{x^n}Q_{X^n|v^n}(x^n) \\
&\qquad\sum_{x'\in\mathcal{X}}\pi_{x^n}(x')D\left(\Gamma_{Y|X}(\cdot|0)\|\Gamma_{Y|X}(\cdot|x')\right) && (175)\\
&\leq\sum_{x^n}Q_{X^n|v^n}(x^n)\max_{x'\in\mathcal{X}}D\left(\Gamma_{Y|X}(\cdot|0)\|\Gamma_{Y|X}(\cdot|x')\right) && (176)\\
&=\max_{x'\in\mathcal{X}}D\left(\Gamma_{Y|X}(\cdot|0)\|\Gamma_{Y|X}(\cdot|x')\right). && (177)
\end{aligned}$$

Combining these observations with (166) and (167) and using that $\ell_n/n\to 0$ as $n\to\infty$, we can conclude;

$$\begin{aligned}
\overline{\lim_{n\to\infty}}\,-\frac{1}{n}\log\beta_n &\leq D(P_V\|Q_V) \\
&\quad+\max_{x'\in\mathcal{X}}D\left(\Gamma_{Y|X}(\cdot|0)\|\Gamma_{Y|X}(\cdot|x')\right), \qquad (178)
\end{aligned}$$

which concludes the proof.

REFERENCES

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

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

[3] H. Shimokawa, T. Han, and S. I. Amari, “Error bound for hypothesis testing with data compression,” in *1994 IEEE Proc. ISIT*, p. 114, Jul. 1994.

[4] 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.

[5] E. Haim and Y. Kochman, “Binary distributed hypothesis testing via korner-marton coding,” in *2016 IEEE Proc. ITW*, 2016.

[6] N. Weinberger and Y. Kochman, “On the reliability function of distributed hypothesis testing under optimal detection,” *IEEE Trans. Inf. Theory*, vol. 65, no. 8, pp. 4940–4965, 2019.

[7] Y. Kochman and L. Wang, “Improved random-binning exponent for distributed hypothesis testing,” *IEEE Trans. Inf. Theory*, pp. 1–1, 2025.

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

[9] S. Salehkalaibar, M. Wigger, and R. Timo, “On hypothesis testing against independence with multiple decision centers,” *IEEE Trans. Comm.*, vol. 66, pp. 2409–2420, Jan. 2018.

[10] M. Mhanna and P. Piantanida, “On secure distributed hypothesis testing,” in *2015 IEEE Proc. ISIT*, pp. 1605–1609, 2015.

[11] J. Liao, L. Sankar, F. P. Calmon, and V. Y. F. Tan, “Hypothesis testing under maximal leakage privacy constraints,” in *2017 IEEE International Symposium on Information Theory (ISIT)*, pp. 779–783, 2017.

[12] A. Gilani, S. Belhadj Amor, S. Salehkalaibar, and V. Y. F. Tan, “Distributed hypothesis testing with privacy constraints,” *Entropy*, vol. 21, p. 478, May 2019.

[13] S. Sreekumar and D. Gündüz, “Testing against conditional independence under security constraints,” in *2018 IEEE Proc. ISIT*, pp. 181–185, 2018.

[14] S. Faour, M. Hamad, M. Sarkiss, and M. Wigger, “Testing against independence with an eavesdropper,” in *2023 IEEE Proc. ITW*, pp. 277–282, 2023.

[15] A. Bounhar, M. Sarkiss, and M. Wigger, “Covert distributed detection over discrete memoryless channels,” in *2024 IEEE Proc. ISIT*, pp. 172–177, 2024.

[16] M. Hamad, M. Wigger, and M. Sarkiss, “Cooperative multi-sensor detection under variable-length coding,” in *2020 IEEE Information Theory Workshop (ITW)*, pp. 1–5, 2021.

[17] S. Salehkalaibar and M. Wigger, “Distributed hypothesis testing with variable-length coding,” *IEEE J. Sel. Areas Inf. Th.*, vol. 1, no. 3, pp. 681–694, 2020.

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

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

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

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

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

[23] 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.

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

[25] K. Marton, "A simple proof of the blowing-up lemma," *IEEE Trans. Inf. Theory*, vol. 32, pp. 445–446, May 1986.

[26] S. Sreekumar and D. Gündüz, "Hypothesis testing over a noisy channel," in *2019 IEEE Proc. ISIT*, pp. 2004–2008, 2019.

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

[28] N. LeGouic and M. Wigger, "Independence testing under zero-rate communication: Beyond the stein-regime," in *2026 IEEE Proc. ISIT*, June 28–July 3 2026.