Hypothesis Testing over a Noisy Channel

Sreejith Sreekumar and Deniz Gündüz
Dept. of Electrical and Electronic Engineering, Imperial College London
{s.sreekumar15,d.gunduz}@imperial.ac.uk

***Abstract*—A point to point hypothesis testing problem involving two parties, one referred to as the observer and the other as the detector, is studied. The observer observes a discrete memoryless source and communicates its observations to the detector over a discrete memoryless channel. The detector performs a binary hypothesis test on the probability distribution of the observer's observation. The trade-off between the type 1 error probability and the type 2 error exponent is explored. We obtain a single-letter characterization of the optimal type 2 error exponent for a given constraint on the type 1 error probability. We also show that a strong converse holds, in the sense that, the optimal type 2 error exponent is independent of the constraint on the type 1 error probability.**

I. Introduction

Consider the scenario depicted in Fig. 1, in which, a statistician, or *detector*, wants to ascertain the underlying probability distribution of the data observed at a remote node, referred to as the *observer*. The data samples observed at the observer, $U_1, \ldots, U_k$, denoted by U^k, are communicated to the detector over a noisy communication channel. Based on the data received from the observer, the detector performs a hypothesis test to determine the probability distribution of U^k. Assuming that U^k is generated independent and identically according to a fixed distribution, the simplest case of such a test is a binary hypothesis test with the following null and alternate hypothesis:

$$\text{Null hypothesis } H_0 : U^k \sim \prod_{i=1}^{k} P_U, \tag{1a}$$

$$\text{Alternate hypothesis } H_1 : U^k \sim \prod_{i=1}^{k} Q_U. \tag{1b}$$

Our aim is to characterize the performance of the above hypothesis test as measured by the type 1 and type 2 error probabilities. More specifically, denoting by $\beta(\epsilon)$, the minimum value of the type 2 error probability subject to a fixed constraint ϵ on the type 1 error probability, we are interested in characterizing the optimal type 2 error exponent (T2EE) given by $-\frac{1}{k}\log(\beta(\epsilon))$, $\epsilon \in (0,1)$.

It is well known that the optimal trade-off between the type 1 and type 2 error probabilities for the hypothesis test given in (1) is achieved by the Neyman-Pearson test [1]. Also, when the observations U^k are available directly at the detector, a single-letter characterization of the optimal T2EE for a given constraint ϵ on the type 1 error probability is known. Denoting the optimal T2EE by $\kappa(\epsilon)$, it is given by

$$\kappa(\epsilon) = D(P_U||Q_U), \ \forall \ \epsilon \in (0,1), \tag{2}$$

where, $D(P_U||Q_U)$ denotes the Kullback-Leibler divergence between probability distributions P_U and Q_U [2]. Notice that a *strong converse* holds in this case, in the sense that, $\kappa(\epsilon)$ is independent of ϵ.

If the detector and the observer are connected with a noise-free link of capacity $R > 0$, it is easy to see that the T2EE in (2) can be achieved by performing the Neyman-Pearson test locally at the observer and transmitting the decision to the detector over the noiseless link. We refer to this as the *local decision scheme* since the test is performed locally at the observer.

The trade-off between the T2EE and the type 1 error probability has also been explored in a related distributed hypothesis testing (HT) setting, where, the detector has access to an independent and identically distributed (i.i.d.) side information sequence V^k correlated with U^k, and the hypothesis test is performed on the joint distribution of U^k and V^k. Although a single-letter characterization of the optimal T2EE for this setting remains open, special cases have been solved. In [3], the case when the communication between the observer and detector is over a noiseless channel subject to a rate constraint of R bits per source sample, is studied. A single letter characterization of the optimal T2EE for a special case of HT known as *testing against independence* is obtained in [3]. Also, a single-letter lower bound on the optimal T2EE is established for the general case and a *strong converse* result is shown to hold. The lower bound obtained in [3] is subsequently improved in [4] and [5].

In [6], we proved a single-letter characterization of the optimal T2EE for testing against independence, when the noiseless channel in [3] is replaced by a noisy channel. Extensions of this problem to the case of general HT is studied in [7], where lower bounds on the optimal T2EE are obtained by using a separation based scheme that performs independent HT and channel coding, and, a joint HT and channel coding scheme

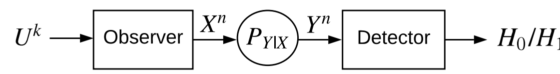

Fig. 1: Hypothesis testing over a noisy channel.

This work is supported in part by the European Research Council (ERC) through Starting Grant BEACON (agreement #677854).

that utilizes hybrid coding [8]. While the above mentioned works study the most basic setting, the trade-off between the T2EE and type 1 error probability has also been explored in several other distributed HT settings, see [9], [10], [11], [12], [13].

When the communication channel from the observer to the detector is noisy, it is unclear whether performing a hypothesis test at the observer and communicating the decision to the detector is optimal. More specifically, since the reliability of the transmitted messages depends on the communication rate employed, there is a trade-off between transmitting less information more reliably versus transmitting more information less reliably, to the detector. In the sequel, we show that making decisions locally at the observer, and communicating it to the detector is indeed optimal. We also provide a single-letter characterization of the optimal T2EE, and show that the strong converse holds for HT over a noisy channel.

A. Notations

Random variables (r.v.'s) and their realizations are denoted by upper and lower case letters (e.g., X and x), respectively. Sets are denoted by calligraphic letters, e.g., the alphabet of a r.v. X is denoted by $\mathcal{X}$. Following the notation in [2], T_P and $T^m_{[P_X]_\delta}$ denotes the set of sequences of type P and the set of P_X-typical sequences of length m, respectively. For $a \in \mathbb{R}^+$, $[a]$ denotes the set of integers $\{1, 2, \ldots, \lceil a \rceil\}$. $a_k \xrightarrow{(k)} b$ indicates that $\lim_{k\to\infty} a_k = b$. For functions $f_1 : \mathcal{A} \to \mathcal{B}$ and $f_2 : \mathcal{B} \to \mathcal{C}$, $f_2 \circ f_1$ denotes function composition. Finally, $\mathbb{1}(\cdot)$, $O(\cdot)$ and $o(\cdot)$ denote the indicator function, the Big-o and the Little-o notation of Landau, respectively.

B. Problem formulation

All the r.v.'s considered henceforth are discrete with finite support, and, all logarithms considered in this paper are with respect to base e. Let $k, n \in \mathbb{Z}^+$ be arbitrary. The encoder observes U^k, and transmits codeword $X^n = f^{(k,n)}(U^k)$, where $f^{(k,n)} : \mathcal{U}^k \to \mathcal{X}^n$ represents the encoding function (possibly stochastic) of the observer. Let $\tau := \frac{n}{k}$ denote the *bandwidth ratio*. Let P_{X^n} and Q_{X^n} denote the probability distribution of the channel input under the null and alternate hypothesis, respectively. The channel output Y^n given X^n is generated according to the probability law

$$P_{Y^n|X^n}(y^n|x^n) = \prod_{j=1}^{n} P_{Y|X}(y_j|x_j), \tag{3}$$

i.e., the channel $P_{Y^n|X^n}$ is memoryless. Let P_{Y^n} and Q_{Y^n} denote the probability distribution of the channel output under the null and alternate hypothesis, respectively. Depending on the received symbols Y^n, the detector makes a decision between the two hypotheses H_0 and H_1 given in (1). Let $H \in \{0, 1\}$ denote the actual hypothesis and $\hat{H} \in \{0, 1\}$ denote the output of the hypothesis test, where 0 and 1 denote H_0 and H_1, respectively. Let $\mathcal{A}^{(n)} \subseteq \mathcal{Y}^n$ denote the acceptance region for H_0. Then, the decision rule $g^{(n)} : \mathcal{Y}^n \to \{0, 1\}$ is given by

$$g^{(n)}(y^n) = 1 - \mathbb{1}\left(y^n \in \mathcal{A}^{(n)}\right).$$

Let

$$\alpha\left(k, n, f^{(k,n)}, g^{(n)}\right) := 1 - P_{Y^n}\left(\mathcal{A}^{(n)}\right), \tag{4}$$

$$\text{and } \beta\left(k, n, f^{(k,n)}, g^{(n)}\right) := Q_{Y^n}\left(\mathcal{A}^{(n)}\right), \tag{5}$$

denote the type 1 and type 2 error probabilities for the encoding function $f^{(k,n)}$ and decision rule $g^{(n)}$, respectively.

Definition 1. *A T2EE κ is (τ, ϵ) achievable if there exists sequences of integers k and n_k, corresponding sequence of encoding functions $f^{(k,n_k)}$ and decoding functions $g^{(n_k)}$ such that*

$$\limsup_{k\to\infty} \frac{n_k}{k} \leq \tau, \tag{6a}$$

$$\liminf_{k\to\infty} -\frac{1}{k}\log\left(\beta\left(k, n_k, f^{(k,n_k)}, g^{(n_k)}\right)\right) \geq \kappa, \tag{6b}$$

$$\textit{and } \limsup_{k\to\infty} \alpha\left(k, n_k, f^{(k,n_k)}, g^{(n_k)}\right) \leq \epsilon. \tag{6c}$$

For $(\tau, \epsilon) \in \mathbb{R}^+ \times [0, 1]$, let

$$\kappa(\tau, \epsilon) := \sup\{\kappa' : \kappa' \text{ is } (\tau, \epsilon) \text{ achievable}\}. \tag{7}$$

We are interested in obtaining a computable characterization of $\kappa(\tau, \epsilon)$.

II. Optimal type 2 error exponent

In this section, we prove a single-letter characterization of $\kappa(\tau, \epsilon)$. Let

$$E_c := E_c(P_{Y|X}) := D(P_{Y|X=a}||P_{Y|X=b}), \tag{8}$$

where

$$(a, b) := \underset{(x,x')\in\mathcal{X}\times\mathcal{X}}{\arg\max}\, D(P_{Y|X=x}||P_{Y|X=x'}). \tag{9}$$

Let

$$\kappa_0 := \kappa_0(\tau, P_U, Q_U, P_{Y|X}) := \begin{cases} D(P_U||Q_U), & \text{if } \tau = 0 \text{ and } E_c = \infty, \\ \min\left(D(P_U||Q_U), \tau E_c\right), & \text{otherwise.} \end{cases}$$

The next theorem states that there exists a single-letter characterization for $\kappa(\tau, \epsilon)$ given by κ_0, and furthermore, that a strong converse holds. Before proceeding to the proof details, we provide a brief sketch of it as follows. We first prove the achievability for arbitrary $\epsilon \in (0, 1)$, and converse for the case when $\epsilon \to 0$, known as a *weak converse*. Subsequently, the strong converse is shown using a constructive method that starts with any coding scheme that satisfies the given type 1 error probability constraint ϵ, and constructs a modified encoding and decoding scheme such that the type 1 error probability tends to zero, asymptotically, and the T2EE is decreased infinitesimally from that achieved by the original coding scheme. The blowing up lemma [14][2] is the key tool used for this purpose. The weak converse along with the strong converse for HT when U^k is observed directly by the detector then implies the desired strong converse.

Theorem 2. $\kappa(\tau,\epsilon)=\kappa_0,\ \forall\ \epsilon\in(0,1),\ \tau\geq 0.$

Proof: The proof consists of three steps as follows:

(i) $\kappa(\tau,\epsilon)\geq\kappa_0,\ \forall\ \epsilon\in(0,1)$.
(ii) $\lim_{\epsilon\to 0}\kappa(\tau,\epsilon)=\kappa_0$.
(iii) $\kappa(\tau,\epsilon)\leq\kappa_0,\ \forall\ \epsilon\in(0,1)$.

Assume that $E_c<\infty$. First, we prove (i). Let $k\in\mathbb{Z}^+$ and $n_k=\lfloor\tau k\rfloor$. We define $f^{(k,n_k)}$ as the composition of two functions $f_s^{(k)}$ and $f_c^{(k,n_k)}$, i.e., $f^{(k,n_k)}=f_c^{(k,n_k)}\circ f_s^{(k)}$, where

$$f_s^{(k)}(u^k)=\begin{cases}0, \text{ if } P_{u^k}\in T^k_{[P_U]_\delta},\\ 1, \text{ otherwise,}\end{cases}$$

and

$$f_c^{(k,n_k)}\left(f_s^{(k)}(u^k)\right)=\begin{cases}a^{n_k}, \text{ if } f_s^{(k)}(u^k)=0,\\ b^{n_k}, \text{ otherwise.}\end{cases}$$

Here, $\delta>0$ is an arbitrarily small number, and a^{n_k} and b^{n_k} denote the codewords formed by repeating the symbols a and b from the channel input alphabet $\mathcal{X}$, which are chosen such that (9) is satisfied. Let the decision rule be defined as

$$g^{(n_k)}(y^{n_k})=\begin{cases}0, \text{ if } y^{n_k}\in T^{n_k}_{[P_{Y|X=a}]_{\delta'}},\\ 1, \text{ otherwise,}\end{cases}$$

where $\delta'>\delta$. By the law of large numbers, the type 1 error probability tends to zero asymptotically, since

$$\lim_{k\to\infty}\mathbb{P}(U^k\in T^k_{[P_U]_\delta}|H=0)=1, \tag{10}$$
$$\text{and }\lim_{k\to\infty}\mathbb{P}(Y^{n_k}\in T^{n_k}_{[P_{Y|X=a}]_{\delta'}}|H=0)=1. \tag{11}$$

Next, we focus on the events that could possibly lead to a type 2 error. Note that a type 2 error may occur only under the following two events:

$$\mathcal{E}_1:=\{U^k\in T^k_{[P_U]_\delta}\text{ and }Y^{n_k}\in T^{n_k}_{[P_{Y|X=a}]_{\delta'}}\},$$
$$\mathcal{E}_2:=\{U^k\notin T^k_{[P_U]_\delta}\text{ and }Y^{n_k}\in T^{n_k}_{[P_{Y|X=a}]_{\delta'}}\}.$$

It follows from Lemma 2.6 [2] and the fact that the number of types is at most polynomial in k, that, for any $\gamma>0$,

$$\begin{aligned}\mathbb{P}(\mathcal{E}_1|H=0)&\leq\mathbb{P}(U^k\in T^k_{[P_U]_\delta}|H=0)\\&\leq e^{-k(D(P_U||Q_U)-\gamma-O(\delta))}\end{aligned} \tag{12}$$

for sufficiently large k. The probability of the second event is upper bounded (asymptotically with k) similarly by

$$\begin{aligned}\mathbb{P}(\mathcal{E}_2|H=1)&\leq\mathbb{P}\left(Y^{n_k}\in T^{n_k}_{[P_{Y|X=a}]_{\delta'}}|U^k\notin T^k_{[P_U]_\delta}\right)\\&=\mathbb{P}(Y^{n_k}\in T^{n_k}_{[P_{Y|X=a}]_{\delta'}}|X^{n_k}=b^{n_k})\\&\leq e^{-n_k(E_c-\gamma-O(\delta'))}\end{aligned} \tag{13}$$

where (13) follows again from [2, Lemma 2.6]. By the union bound, it follows that

$$\beta(k,n_k,f^{(k,n_k)},g^{(n_k)})\leq\mathbb{P}(\mathcal{E}_1|H=1)+\mathbb{P}(\mathcal{E}_2|H=1),$$

which in turn implies in the limit δ and δ' tending to zero (subject to delta-convention given in [2]), that

$$\kappa(\tau,\epsilon)\geq\min\left(D(P_U||Q_U)-\gamma),\tau(E_c-\gamma)\right),\ \forall\ \epsilon\in(0,1).$$

The proof of achievability is completed by noting that $\gamma>0$ is arbitrary.

Next, we show part (ii). We will prove the weak converse

$$\lim_{\epsilon\to 0}\kappa(\tau,\epsilon)\leq\kappa_0(\tau), \tag{14}$$

which combined with part (i) proves part (ii). Similarly to [3, Theorem 1 (b)], it follows that for any sequence of encoding functions $f^{(k,n_k)}$ and decision rules $g^{(n_k)}$ such that $\limsup\frac{n_k}{k}\leq\tau$ and

$$\limsup_{k\to\infty}\alpha\left(k,n_k,f^{(k,n_k)},g^{(n_k)}\right)\leq\epsilon_k\xrightarrow{(k)}0, \tag{15}$$

we have

$$\limsup_{k\to\infty}\frac{-1}{k}\log\left(\beta\left(k,n_k,f^{(k,n_k)},g^{(n_k)}\right)\right)\leq\frac{1}{k}D(P_{Y^{n_k}}||Q_{Y^{n_k}}). \tag{16}$$

The right hand side (R.H.S.) of (16) can be upper bounded as follows.

$$\begin{aligned}&D(P_{Y^{n_k}}||Q_{Y^{n_k}})\\&=\sum_{i=1}^{n_k}D(P_{Y_i|Y^{i-1}}||Q_{Y_i|Y^{i-1}}|P_{Y^{i-1}}) \quad (17)\\&=\sum_{i=1}^{n_k}\sum_{y^{i-1}\in\mathcal{Y}^{i-1}}P_{Y^{i-1}}(y^{i-1})\Bigg[\sum_{y_i\in\mathcal{Y}}P_{Y_i|Y^{i-1}}(y_i|y^{i-1})\\&\qquad\log\left(\frac{P_{Y_i|Y^{i-1}}(y_i|y^{i-1})}{Q_{Y_i|Y^{i-1}}(y_i|y^{i-1})}\right)\Bigg]\\&=\sum_{i=1}^{n_k}\sum_{\substack{y^{i-1}\\\in\mathcal{Y}^{i-1}}}P_{Y^{i-1}}(y^{i-1})\ D\left(P_{Y_i|Y^{i-1}=y^{i-1}}||Q_{Y_i|Y^{i-1}=y^{i-1}}\right).\end{aligned}$$

Since

$$P_{Y_i|Y^{i-1}}(y_i|y^{i-1})=\sum_{x_i\in\mathcal{X}}P_{X_i|Y^{i-1}}(x_i|y^{i-1})P_{Y_i|X_i}(y_i|x_i), \tag{18}$$

$$\text{and }Q_{Y_i|Y^{i-1}}(y_i|y^{i-1})=\sum_{x_i\in\mathcal{X}}Q_{X_i|Y^{i-1}}(x_i|y^{i-1})P_{Y_i|X_i}(y_i|x_i), \tag{19}$$

we can write

$$D(P_{Y^{n_k}}||Q_{Y^{n_k}})\leq\sum_{i=1}^{n_k}\sum_{\substack{y^{i-1}\\\in\mathcal{Y}^{i-1}}}P_{Y^{i-1}}(y^{i-1})\Bigg[\sup_{P_{X_i|Y^{i-1}=y^{i-1}}}D\left(P_{Y_i|Y^{i-1}=y^{i-1}}||Q_{Y_i|Y^{i-1}=y^{i-1}}\right)\Bigg]. \tag{20}$$

It follows from (18), (19) and the convexity of $D(P_X||Q_X)$

in (P_X, Q_X) that, $D\left(P_{Y_i|Y^{i-1}=y^{i-1}}||Q_{Y_i|Y^{i-1}=y^{i-1}}\right)$ is a convex function of $(P_{X_i|Y^{i-1}=y^{i-1}}, Q_{X_i|Y^{i-1}=y^{i-1}})$ for any $y^{i-1} \in \mathcal{Y}^{i-1}$. It is well-known that the maximum of a convex function over a convex feasible set is achieved at the extreme points of the feasible set. Since the extreme points of the probability simplex $P_{\mathcal{X}}$ are probability distributions of the form

$$P_{X_i}(x) = \mathbb{1}(x = x'),\ x \in \mathcal{X}, \tag{21}$$

for some $x' \in \mathcal{X}$, it follows that for some functions $h_{1i} : \mathcal{Y}^{i-1} \rightarrow \mathcal{X}$ and $h_{2i} : \mathcal{Y}^{i-1} \rightarrow \mathcal{X}$, $i \in [1:n]$, we can write

$$\begin{aligned} &\sup_{P_{X_i|Y^{i-1}=y^{i-1}}} D\left(P_{Y_i|Y^{i-1}=y^{i-1}}||Q_{Y_i|Y^{i-1}=y^{i-1}}\right) \\ &= D\left(P_{Y_i|X_i=h_{1i}(y^{i-1})}||P_{Y_i|X_i=h_{2i}(y^{i-1})}\right) \\ &\leq \max_{(x,x')\in\mathcal{X}\times\mathcal{X}} D\left(P_{Y|X=x}||P_{Y|X=x'}\right) = E_c. \end{aligned} \tag{22}$$

Thus, it follows from (20) and (22) that

$$\frac{1}{k} D(P_{Y^{n_k}}||Q_{Y^{n_k}}) \leq \frac{n_k}{k} E_c. \tag{23}$$

Also, the data processing inequality for Kullback-Leibler divergence applied to Markov chain $U^k - X^{n_k} - Y^{n_k}$ yields

$$\frac{1}{k} D(P_{Y^{n_k}}||Q_{Y^{n_k}}) \leq \frac{1}{k} D(P_{U^k}||Q_{U^k}) = D(P_U||Q_U). \tag{24}$$

Hence, from (6a), (16), (23) and (24), it follows that

$$\begin{aligned} \limsup_{k\rightarrow\infty} \frac{-1}{k} \log\left(\beta\left(k, n_k, f^{(k,n_k)}, g^{(n_k)}\right)\right) \\ \leq \min\left(D(P_U||Q_U), \tau E_c\right). \end{aligned} \tag{25}$$

Noting that the R.H.S. of (25) is independent of $\left(f^{(k,n_k)}, g^{(n_k)}\right)$, the proof of (14) is completed by taking supremum with respect to $\left(f^{(k,n_k)}, g^{(n_k)}\right)$.

Finally, we prove part (iii), i.e.,

$$\kappa(\tau, \epsilon) \leq \kappa_0(\tau),\ \forall\ \epsilon \in (0,1). \tag{26}$$

For $k \in \mathbb{Z}^+$, let $\{n_k, f^{(k,n_k)}, g^{(n_k)}\}$ be any sequence such that (6a) and (6c) are satisfied. Let $\mathcal{A}^{(n_k)}$ denote the acceptance region corresponding to $g^{(n_k)}$. For fixed $\gamma > 0$ and $\delta > 0$, let

$$\begin{aligned} &\mathcal{B}_{\gamma,\delta}^{(k,n_k)} \\ &= \left\{u^k \in T_{[P_U]_\delta}^k :\ \mathbb{P}\left(Y^{n_k} \in \mathcal{A}^{(n_k)}|U^k = u^k, H = 0\right) \geq \gamma\right\}. \end{aligned}$$

By the weak law of large numbers, for $\gamma' > 0$ and sufficiently large k, we have that

$$\mathbb{P}\left(U^k \in T_{[P_U]_\delta}^k|H = 0\right) \geq 1 - \gamma'. \tag{27}$$

Then, it follows from (6c) and (27) that

$$\mathbb{P}\left(U^k \in \mathcal{B}_{\gamma,\delta}^{(k,n_k)}|H = 0\right) \geq \frac{1 - \epsilon - \gamma}{1 - \gamma} - \gamma'. \tag{28}$$

Taking $\gamma = \frac{1-\epsilon}{2}$ and $\gamma' \in \left(0, \frac{1-\epsilon}{2(1+\epsilon)}\right)$, we have that

$$\mathbb{P}\left(U^k \in \mathcal{B}_{\gamma,\delta}^{(k,n_k)}|H = 0\right) \geq \frac{1 - \epsilon}{2(1+\epsilon)}. \tag{29}$$

For arbitrary $u^k \in \mathcal{B}_{\gamma,\delta}^{(k,n_k)}$, let $\bar{x}^{n_k}$ be such that

$$P_{Y^{n_k}|X^{n_k}}\left(\mathcal{A}^{(n_k)}|\bar{x}^{n_k}\right) \geq \gamma, \tag{30}$$

$$\text{and } P_{X^{n_k}|U^k}\left(\bar{x}^{n_k}|u^k\right) > 0. \tag{31}$$

The existence of such a $\bar{x}^{n_k}$ follows by definition of $\mathcal{B}_{\gamma,\delta}^{(k,n_k)}$.

For any set $\mathcal{D} \subset \mathcal{X}^n$, let $\Gamma^l(\mathcal{D})$ denote the Hamming $l-$neighbourhood of $\mathcal{D}$, i.e.,

$$\Gamma^l(\mathcal{D}) := \{\tilde{x}^n \in \mathcal{X}^n : d_H(x^n, \tilde{x}^n) \leq l \text{ for some } x^n \in \mathcal{D}\}.$$

Due to (29), it follows by the application of the blowing-up lemma [2] that there exists sequences of non-negative numbers, $\{\lambda_k\}_{k\in\mathcal{Z}^+}$ and $\{l_k\}_{k\in\mathcal{Z}^+}$ such that, $\lambda_k \xrightarrow{(k)} 0$, $\frac{l_k}{k} \xrightarrow{(k)} 0$ and

$$P_{Y^{n_k}|X^{n_k}}\left(\Gamma^{l_k}\left(\mathcal{A}^{(n_k)}\right)|\bar{x}^{n_k}\right) \geq 1 - \lambda_k. \tag{32}$$

Let $\bar{\mathcal{A}}^{(n_k)} := \Gamma^{l_k}\left(\mathcal{A}^{(n_k)}\right)$. Note that $E_c < \infty$, if and only if $P_{Y|X}(y|x) > 0,\ \forall\ (x,y) \in \mathcal{X} \times \mathcal{Y}$. Let

$$\underline{v} := \min_{(x,y)\in\mathcal{X}\times\mathcal{Y}} P_{Y|X}(y|x) > 0. \tag{33}$$

For each $\bar{y}^{n_k} \in \bar{\mathcal{A}}^{(n_k)}$, there exists a $y^{n_k} \in \mathcal{A}^{(n_k)}$ such that $d_H(\bar{y}^{n_k}, y^{n_k}) \leq l_k$. Hence, for each such $\bar{y}^{n_k}$ and arbitrary $x^{n_k} \in \mathcal{X}^{n_k}$, we have

$$P_{Y^{n_k}|X^{n_k}}(\bar{y}^{n_k}|x^{n_k})\ \underline{v}^{l_k} \leq P_{Y^{n_k}|X^{n_k}}(y^{n_k}|x^{n_k}). \tag{34}$$

Also, for each $y^{n_k} \in \mathcal{A}^{(n_k)}$, the number of $\bar{y}^{n_k} \in \bar{\mathcal{A}}^{(n_k)}$ is $|\mathcal{Y}|^{l_k}$. Hence, from (34), we have

$$P_{Y^{n_k}|X^{n_k}}\left(\bar{\mathcal{A}}^{(n_k)}|x^{n_k}\right) \leq |\mathcal{Y}|^{l_k} P_{Y^{n_k}|X^{n_k}}\left(\mathcal{A}^{(n_k)}|x^{n_k}\right) \underline{v}^{-l_k}.$$

This implies that

$$P_{Y^{n_k}|U^k}\left(\bar{\mathcal{A}}^{(n_k)}|u^k\right) \leq |\mathcal{Y}|^{l_k} P_{Y^{n_k}|U^k}\left(\mathcal{A}^{(n_k)}|u^k\right) \underline{v}^{-l_k}. \tag{35}$$

Let the new encoding function $\tilde{f}^{(k,n_k)} : \mathcal{U}^k \rightarrow \mathcal{X}^{n_k}$ and decision rule $\tilde{g}^{(n_k)} : \mathcal{Y}^{n_k} \rightarrow \{0,1\}$ be defined as follows:

$$\tilde{f}^{(k,n_k)}(u^k) = \begin{cases} \bar{x}^{n_k}, & \forall\ u^k \in T_{[P_U]_\delta}^k \\ f^{(k,n_k)}(u^k), & \text{otherwise,} \end{cases} \tag{36a}$$

$$\text{and } \tilde{g}^{(n_k)}(y^{n_k}) := 1 - \mathbb{1}\left(y^{n_k} \in \bar{\mathcal{A}}^{(n_k)}\right). \tag{36b}$$

From (27), (32) and (36), it follows that

$$\alpha\left(k, n_k, \tilde{f}^{(k,n_k)}, \tilde{g}^{(n_k)}\right) \leq 1 - (1 - \lambda_k)(1 - \gamma') \xrightarrow{(k)} \gamma'.$$

Also,

$$\begin{aligned} \beta\left(k, n_k, \tilde{f}^{(k,n_k)}, \tilde{g}^{(n_k)}\right) \leq \sum_{u^k\in T_{[P_U]_\delta}^k} Q_{U^k}(u^k) \\ + \sum_{u^k\notin T_{[P_U]_\delta}^k} Q_{U^k}(u^k) P_{Y^{n_k}|U^k}\left(\bar{\mathcal{A}}^{(n_k)}|u^k\right) \end{aligned}$$

$$\leq \sum_{u^k \in T^k_{[P_U]_\delta}} Q_{U^k}(u^k)$$

$$+ \underline{v}^{-l_k} |\mathcal{Y}|^{l_k} \sum_{u^k \notin T^k_{[P_U]_\delta}} Q_{U^k}(u^k) P_{Y^{n_k}|U^k} \left(\mathcal{A}^{(n_k)} | u^k\right) \quad (37)$$

$$\leq \sum_{u^k \in T^k_{[P_U]_\delta}} Q_{U^k}(u^k) + \underline{v}^{-l_k} |\mathcal{Y}|^{l_k} \sum_{u^k \in \mathcal{U}^k} Q_{U^k}(u^k) P_{Y^{n_k}|U^k} \left(\mathcal{A}^{(n_k)} | u^k\right)$$

$$= \sum_{u^k \in T^k_{[P_U]_\delta}} Q_{U^k}(u^k) + \underline{v}^{-l_k} |\mathcal{Y}|^{l_k} \beta\left(k, n_k, f^{(k,n_k)}, g^{(n_k)}\right)$$

$$\leq e^{-k(D(P_U||Q_U) - O(\delta))} + \underline{v}^{-l_k} |\mathcal{Y}|^{l_k} \beta\left(k, n_k, f^{(k,n_k)}, g^{(n_k)}\right),$$

where (37) follows from (35). Thus, it follows from the facts $\frac{l_k}{k} \xrightarrow{(k)} 0$ and $\underline{v} > 0$ that, for any $\gamma'' > 0$,

$$-\frac{1}{k} \log\left(\beta\left(k, n_k, \tilde{f}^{(k,n_k)}, \tilde{g}^{(n_k)}\right)\right)$$
$$\geq \min\left(D(P_U||Q_U) - O(\delta), -\frac{1}{k} \log\left(\beta\left(k, n_k, f^{(k,n_k)}, g^{(n_k)}\right)\right) - \gamma''\right),$$

provided k is sufficiently large. Since $D(P_U||Q_U)$ is the maximum T2EE achievable for any type 1 error probability constraint $\epsilon \in (0,1)$, when U^k is directly observed at the detector, it follows by taking $\delta, \gamma'' \to 0$ that

$$\liminf_{k \to \infty} -\frac{1}{k} \log\left(\beta\left(k, n_k, \tilde{f}^{(k,n_k)}, \tilde{g}^{(n_k)}\right)\right)$$
$$\geq \liminf_{k \to \infty} -\frac{1}{k} \log\left(\beta\left(k, n_k, f^{(k,n_k)}, g^{(n_k)}\right)\right).$$

This completes the proof for the case when $E_c < \infty$.

Next, consider the case $E_c = \infty$. Then, there exist y such that $P_{Y|X}(y|b) = 0$ and $P_{Y|X}(y|a) > 0$, where a and b are defined as in (8). Assume $\tau > 0$. Since the proofs of steps (i) and (ii) given above carry over, it follows that $\lim_{\epsilon \to 0} \kappa(\tau, \epsilon) = \kappa_0(\tau) = D(P_U||Q_U)$. On the other hand, if $\tau = 0$, then taking $n_k = k^{\frac{2}{3}}$ and using the same encoding and decision rule as in step (i) above, it follows from (13) that $\mathbb{P}(\mathcal{E}_2|H = 1) = 0$. From (10) and (11), the type 1 error probability tends to zero, asymptotically with k. Also, notice that (12) holds and that the above choice of n_k satisfies (6). Hence, $\kappa(0, \epsilon) \geq D(P_U||Q_U)$. Noting again that $D(P_U||Q_U)$ is the maximum T2EE achievable for any type 1 error probability constraint $\epsilon \in (0,1)$, when U^k is directly observed at the detector, it follows that

$$\kappa(0, \epsilon) = D(P_U||Q_U), \ \forall \ \epsilon \in (0,1).$$

This completes the proof of the theorem. ∎

III. Conclusion

In this paper, we obtained a single-letter characterization of the optimal T2EE for HT over a noisy channel. The achievability scheme shows that the optimal T2EE is achieved by a scheme, in which the observer makes a decision about the hypothesis locally, and communicates it to the detector using a channel code of maximum reliability. This implies that "separation" holds, in the sense that, there is no loss in optimality incurred by separating the tasks of HT and channel coding. It is also interesting to note that the optimal T2EE is independent of the constraint on the type 1 error probability.

References

[1] J. Neyman and E. Pearson, "On the problem of the most efficient tests of statistical hypotheses," *Philos. Trans. of the Royal Society of London*, vol. 231, pp. 289–337, Feb. 1933.

[2] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge University Press, 2011.

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

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

[5] H. Shimokawa, T. S. Han, and S. Amari, "Error bound of hypothesis testing with data compression," in *Proc. IEEE Int. Symp. Inf. Theory*, Trondheim, Norway, 1994.

[6] S. Sreekumar and D. Gündüz, "Distributed hypothesis testing over noisy channels," in *Proc. IEEE Int. Symp. Inf. Theory*, Aachen, Germany, Jun. 2017.

[7] ——, "Distributed hypothesis testing over discrete memoryless channels," *arXiv:1802.07665 [cs.IT]*.

[8] P. Minero, S. H. Lim, and Y.-H. Kim, "A unified approach to hybrid coding," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 1509–1523, Apr. 2015.

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

[10] G. Katz, P. Piantanida, and M. Debbah, "Distributed binary detection with lossy data compression," *IEEE Trans. Inf. Theory*, vol. 63, no. 8, pp. 5207–5227, Mar. 2017.

[11] M. Wigger and R. Timo, "Testing against independence with multiple decision centers," in *Int. Conf. on Signal Processing and Communication*, Bengaluru, India, Jun. 2016.

[12] S. Salehkalaibar, M. Wigger, and L. Wang, "Hypothesis testing in multi-hop networks," *arXiv:1708.05198*.

[13] S. Salehkalaibar and M. Wigger, "Distributed hypothesis testing based on unequal-error protection codes," *arXiv:1806.05533*.

[14] R. Ahlswede, P. Gács, and J. Körner, "Bounds on conditional probabilities with applications in multi-user communication," *Zeitschrift fr Wahrscheinlichkeitstheorie und verwandte Gebiete*, vol. 34, no. 2, pp. 157–177, 1976.