Distributed Hypothesis Testing Over Noisy Channels

Sreejith Sreekumar and Deniz Gündüz

Imperial College London, UK

Email: {s.sreekumar15, d.gunduz}@imperial.ac.uk

Abstract—A distributed binary hypothesis testing problem, in which multiple observers transmit their observations to a detector over noisy channels, is studied. Given its own side information, the goal of the detector is to decide between two hypotheses for the joint distribution of the data. Single-letter upper and lower bounds on the optimal type 2 error exponent (T2-EE), when the type 1 error probability vanishes with the block-length are obtained. These bounds coincide and characterize the optimal T2-EE when only a single helper is involved. Our result shows that the optimal T2-EE depends on the marginal distributions of the data and the channels rather than their joint distribution. However, an operational separation between HT and channel coding does not hold, and the optimal T2-EE is achieved by generating channel inputs correlated with observed data.

I. Introduction

Statistical inference and learning have assumed prime importance in the fields of machine learning, data analytics and communications applications. An important problem arising in these scenarios is that of discerning the statistics of the available data. This leads to the formulation of a hypothesis testing (HT) problem, in which the objective is to identify the underlying probability distribution of the data samples, from among a set of candidate distributions. With the increasing adaption of distributed sensing technologies and the Internet of Things (IoT) paradigm, the data is often collected from multiple remote locations and communicated to the detector over noisy communication links. This naturally leads to the problem of distributed statistical inference over noisy communication channels.

In this paper, we study the problem of distributed binary HT over noisy channels depicted in Fig. 1. The detector is interested in determining whether the data $(U_1, \ldots, U_L, V, Z)$ is distributed according to $P_{U_1 \ldots U_L V Z}$ or $Q_{U_1 \ldots U_L V Z}$ under hypotheses H_0 and H_1, respectively. Each encoder l, $l = 1, \ldots, L$, observes k samples independent and identically distributed (i.i.d) according to P_{U_l}, and communicates its observation to the detector by n uses of the discrete memoryless channel (DMC), characterized by the conditional distribution $P_{Y_l | X_l}$. The detector decides between the two hypotheses H_0 and H_1 based on the channel outputs $Y_1^n, \ldots, Y_L^n$ as well as its own observations V^k and Z^k. Our goal is to characterize the optimal type 2 error exponent (T2-EE) for this model as a function of the bandwidth ratio, $\tau = \frac{n}{k}$, under the constraint that the type 1 error probability is less than a specified value. We will focus mostly on the special case in which $P_{U_1 \ldots U_L V Z} = P_{U_1 \ldots U_L V | Z} P_Z$ and $Q_{U_1 \ldots U_L V Z} =$

$P_{U_1 \ldots U_L | Z} P_{V | Z} P_Z$, known as the *testing against conditional independence (TACI)* problem.

Distributed statistical inference under communication constraints was originally formulated by Berger in [1]. A simplified version of this is considered in [2], which studies binary HT for the model in Fig. 1 when $L = 1$, Z is absent and the channel between the encoder and the detector is a noise-free channel of rate R. Ahlswede and Csiszár establish a single-letter characterization of the optimal T2-EE for the testing against independence (TAI) problem (including a strong converse), along with single-letter lower bounds for the general HT problem in [2]. For the same model, [3] provides a tighter lower bound on the T2-EE, which coincides with that of [2] for the TAI problem. An improved lower bound for the same problem is obtained in [4] by introducing "binning" at the encoder. HT for the model in Fig. 1 with noise free rate-limited channels is studied in [5], and the authors establish the optimality of binning for the TACI problem. A single-letter characterization of the optimal T2-EE for the multi-terminal TAI problem is obtained in [6] under a certain Markovian condition. In a slightly different setting with two decision centers, the optimal T2-EE for a three terminal dependence testing problem is characterized in [7]. The optimal T2-EE, when multiple interactions between the encoder and detector are allowed, is studied in [8],[9]. We remark here that all the above works consider rate-limited bit-pipes from the observers to the detector, and to the best of our knowledge, HT over noisy channels has not been studied previously.

Notations: The support of a random variable (r.v.) is denoted by calligraphic letters, e.g., $\mathcal{X}$ for r.v. X. The cardinality of $\mathcal{X}$ is denoted by $|\mathcal{X}|$. The joint distribution of r.v.'s X and Y is denoted by P_{XY} and its marginals by P_X and P_Y. $X - Y - Z$ denotes that X, Y, Z form a Markov chain. For $m \in \mathbb{Z}^+$, X^m denotes the sequence $X_1, \ldots, X_m$, while X_l^m denotes $X_{l,1}, \ldots, X_{l,m}$ associated with observer l. The group of m r.v's $X_{l,((j-1)m+1)}, \ldots, X_{l,((jm)}$ is denoted by $X_l^m(j)$, and the infinite sequence $X_l^m(1), X_l^m(2), \ldots$ is denoted by $\{X_l^m(j)\}_{j \in \mathbb{Z}^+}$. Similarly, for a subset $S = \{l_1, \ldots, l_s\}$ of observers, $\{X_{l_1}^m, \ldots, X_{l_s}^m\}$, $\{X_{l_1}^m(j), \ldots, X_{l_s}^m(j)\}$ and $\left\{\{X_{l_1}^m(j)\}_{j \in \mathbb{Z}^+}, \ldots, \{X_{l_s}^m(j)\}_{j \in \mathbb{Z}^+}\right\}$ are denoted by X_S^m, $X_S^m(j)$ and $\{X_S^m(j)\}_{j \in \mathbb{Z}^+}$, respectively. Following the notation in [10], T_P and $T_{[X]_\delta}^m$ (or T_δ^m when there is no ambiguity) denote the set of sequences of type P and the set of P_X-typical sequences of length m, respectively. $D(P||Q)$ denotes

Fig. 1: Illustration of a distributed hypothesis testing system over noisy channels

the Kullback-Leibler (KL) divergence between distributions P and Q [10]. All logarithms are to the base 2. $\mathbb{1}$ denotes the indicator function.

II. SYSTEM MODEL

All the r.v.'s considered henceforth are discrete with finite support. Let $k, n \in \mathbb{Z}^+$ be arbitrary. Let $\mathcal{L} = \{1, \ldots, L\}$ denote the set of observers which communicate to the detector over orthogonal noisy channels, as shown in Fig. 1. For $l \in \mathcal{L}$, encoder l observes U_l^k and transmits $X_l^n = f_l^{(k,n)}(U_l^k)$, where $f_l^{(k,n)} : \mathcal{U}_l^k \to \mathcal{X}_l^n$ is a stochastic mapping. Let $\tau \triangleq \frac{n}{k}$ denote the *bandwidth ratio*. The channel output $Y_{\mathcal{L}}^n$ is given by the probability law $P_{Y_{\mathcal{L}}^n|X_{\mathcal{L}}^n}(y_{\mathcal{L}}^n|x_{\mathcal{L}}^n) = \prod_{l=1}^{L} \prod_{j=1}^{n} P_{Y_l|X_l}(y_{l,j}|x_{l,j})$, i.e., the channels between the observers and the detector are orthogonal and discrete memoryless. Depending on the received symbols $Y_{\mathcal{L}}^n$ and samples (V^k, Z^k), the detector makes a decision between the two hypotheses $H_0 : P_{U_{\mathcal{L}}VZ}$ or $H_1 : Q_{U_{\mathcal{L}}VZ}$ according to the decision rule $g^{(k,n)} : \mathcal{Y}_{\mathcal{L}}^n \times \mathcal{V}^k \times \mathcal{Z}^k \to \{0,1\}$ given by $g^{(k,n)}(y_{\mathcal{L}}^n, v^k, z^k) = \mathbb{1}\left((y_{\mathcal{L}}^n, v^k, z^k) \in A^c\right)$, where A denotes the acceptance region for H_0. It is assumed that the r.v's $U_{\mathcal{L}}, V$ and Z have the same marginal distributions under both H_0 and H_1 and that $Q_{U_{\mathcal{L}}VZ}(u_{\mathcal{L}}, v, z) > 0$ for all $(u_{\mathcal{L}}, v, z) \in \mathcal{U}_{\mathcal{L}} \times \mathcal{V} \times \mathcal{Z}$. In this paper, we focus mostly on the special case when $H_0 : P_{U_{\mathcal{L}}V|Z}P_Z$ and $H_1 : P_{U_{\mathcal{L}}|Z}P_{V|Z}P_Z$, i.e, TACI between V and $U_{\mathcal{L}}$ given Z.

Let $\bar{\alpha}\left(k, n, f_1^{(k,n)}, \ldots, f_L^{(k,n)}, g^{(k,n)}\right) \triangleq P_{Y_{\mathcal{L}}^n V^k Z^k}(A^c)$ and $\bar{\beta}\left(k, n, f_1^{(k,n)}, \ldots, f_L^{(k,n)}, g^{(k,n)}\right) \triangleq Q_{Y_{\mathcal{L}}^n V^k Z^k}(A)$ denote the type 1 and type 2 error probabilities, respectively. Define

$$\beta'\left(k, n, f_1^{(k,n)}, \ldots, f_L^{(k,n)}, \epsilon\right) \triangleq$$
$$\inf_{g^{(k,n)}} \bar{\beta}\left(k, n, f_1^{(k,n)}, \ldots, f_L^{(k,n)}, g^{(k,n)}\right) \quad (1)$$

such that

$$\bar{\alpha}\left(k, n, f_1^{(k,n)}, \ldots, f_L^{(k,n)}, g^{(k,n)}\right) \leq \epsilon, \quad (2a)$$

$$(Z^k, V^k) - U_l^k - X_l^n = f_l^{(k,n)}(U_l^k) - Y_l^n, \ l \in \mathcal{L}, \quad (2b)$$

and let

$$\beta(k, \tau, \epsilon) \triangleq \inf_{\substack{f_1^{(k,n)}, \ldots, f_L^{(k,n)}, \\ n \leq \tau k}} \beta'\left(k, n, f_1^{(k,n)}, \ldots, f_L^{(k,n)}, \epsilon\right). \quad (3)$$

Note that $\beta(k, \tau, \epsilon)$ is a non-increasing function of k and ϵ. A T2-EE κ' is said to be (τ, ϵ) achievable if there exists a sequence of integers k, encoding functions $f_l^{(k, n_k)} : \mathcal{U}^k \to \mathcal{X}^{n_k}$, $l \in \mathcal{L}$ and decoding function $g^{(k,n)}$ such that $n_k \leq \tau k$, $\forall \ k$, and for any $\delta > 0$,

$$\limsup_{k \to \infty} \frac{\log\left(\beta(k, \tau, \epsilon)\right)}{k} \leq -(\kappa' - \delta). \quad (4)$$

Let $\kappa(\tau, \epsilon) \triangleq \sup\{\kappa' : \kappa' \text{ is } (\tau, \epsilon) \text{ achievable}\}$.

For $k \in \mathbb{Z}^+$, we define

$$\theta(k, \tau) \triangleq \sup_{\substack{f_1^{(k,n)}, \ldots, f_L^{(k,n)} \\ n \leq \tau k}} \frac{D\left(P_{Y_{\mathcal{L}}^n V^k Z^k} \| Q_{Y_{\mathcal{L}}^n V^k Z^k}\right)}{k}, \quad (5)$$

and

$$\theta(\tau) \triangleq \sup_k \theta(k, \tau). \quad (6)$$

In this paper, we obtain single-letter upper and lower bounds on $\kappa(\tau, \epsilon)$ for the TACI problem. It is shown that the two bounds coincide when $L = 1$. Our approach is similar to that in [2], where we first obtain bounds for $\kappa(\tau, \epsilon)$ in terms of θ, and then show that θ has a single-letter characterization in terms of information theoretic quantities. We establish this characterization by considering the joint source-channel coding (JSCC) problem with noisy helpers. The next lemma obtains the bounds for $\kappa(\tau, \epsilon)$ in terms of θ.

Lemma 1. *For any bandwidth ratio $\tau > 0$, we have*

(i) $\limsup_{k \to \infty} \frac{\log(\beta(k, \tau, \epsilon))}{k} \leq -\theta(\tau)$, $\epsilon \in (0, 1)$.

(ii) $\lim_{\epsilon \to 0} \liminf_{k \to \infty} \log\left(\frac{\beta(k, \tau, \epsilon)}{k}\right) \geq -\theta(\tau)$.

Proof: The proof is similar to that of Theorem 1 in [2]. We prove (i) and omit the proof of (ii) due to space limitations. Let $k \in \mathbb{Z}^+$ and $\tilde{\epsilon} > 0$ be arbitrary, and $\tilde{n}_k$, $\tilde{f}_l^{(k, \tilde{n}_k)}$, $l \in \mathcal{L}$, and $\tilde{Y}_{\mathcal{L}}^{\tilde{n}_k}$ be the channel block length, encoding functions and channel outputs respectively, such that $k\theta(k, \tau) - D\left(P_{Y_{\mathcal{L}}^{\tilde{n}_k} V^k Z^k} \| Q_{Y_{\mathcal{L}}^{\tilde{n}_k} V^k Z^k}\right) < k\tilde{\epsilon}$. For each $l \in \mathcal{L}$, $\left\{\tilde{Y}_l^{\tilde{n}_k}(j)\right\}_{j \in \mathbb{Z}^+}$ form an infinite sequence of i.i.d. r.v.'s indexed by j. Hence, by the application of Stein's Lemma [2] to the sequences $\left\{\tilde{Y}_{\mathcal{L}}^{\tilde{n}_k}(j), V^k(j), Z^k(j)\right\}_{j \in \mathbb{Z}^+}$, we have

$$\limsup_{j \to \infty} \frac{\log\left(\beta(kj, \tau, \epsilon)\right)}{kj} \leq -(\theta(k, \tau) - \tilde{\epsilon}). \quad (7)$$

For $m \geq kj$, $\beta(m, \tau, \epsilon) \leq \beta(kj, \tau, \epsilon)$. Hence,

$$\limsup_{m \to \infty} \frac{\log\left(\beta(m, \tau, \epsilon)\right)}{m} \leq \limsup_{j \to \infty} \frac{\log\left(\beta(kj, \tau, \epsilon)\right)}{kj} \leq -(\theta(k, \tau) - \tilde{\epsilon}).$$

Note that the left hand side (L.H.S) of the above equation does not depend on k. Taking supremum with respect to k on both sides of the equation and noting that $\tilde{\epsilon}$ is arbitrary, proves (i). ∎

Remark 2. *Part (ii) of Lemma 1 is known as the weak converse for the HT problem in the literature, since it holds only when type 1 error probability tends to zero. Also, (i) and (ii) together imply that $\theta(\tau)$ is the optimal T2-EE as $\epsilon \to 0$, i.e., $\lim_{\epsilon \to 0} \kappa(\tau, \epsilon) = \theta(\tau)$.*

Part (i) of Lemma 1 proves the achievability of the T2-EE $\theta(\tau)$ using Stein's Lemma. In Appendix A, we show an explicit proof of the achievability by computing the type 1 and type 2 errors for a block-memoryless stochastic encoding function at the observer and a joint typicality detector.

Note that for the TACI problem, the KL-divergence becomes mutual information, and we have

$$\theta(\tau) = \sup_{\substack{f_1^{(k,n)}, \dots, f_L^{(k,n)} \\ k, n \leq \tau k}} \frac{I(V^k; Y_{\mathcal{L}}^n | Z^k)}{k}$$

$$\text{s.t. } (Z^k, V^k) - U_l^k - X_l^n = f_l^{(k,n)}(U_l^k) - Y_l^n, \forall \, l \in \mathcal{L}$$

Although Lemma 1 implies that $\theta(\tau)$ is an achievable T2-EE, it is in general not computable as it is defined in terms of a multi-letter characterization. However, as we will show below, for the TACI problem, single-letter bounds for $\theta(\tau)$ can be obtained. By the memoryless property of the sequences V^k and Z^k, we can write

$$\theta(\tau) = H(V|Z) - \inf_{\substack{f_1^{(k,n)}, \dots, f_L^{(k,n)} \\ k, n \leq \tau k}} \frac{H(V^k|Y_{\mathcal{L}}^n, Z^k)}{k} : \qquad (8)$$

$$(Z^k, V^k) - U_l^k - X_l^n = f_l^{(k,n)}(U_l^k) - Y_l^n, \, \forall \, l \in \mathcal{L}.$$

In the next section, we introduce the $L-$helper JSCC problem and show that the multi-letter characterization of this problem coincides with obtaining the infimum in (8). The computable characterization of the lower and upper bounds for (8) then follows from the single-letter characterization of the $L-$helper JSCC problem.

III. $L-$HELPER JSCC PROBLEM

Consider the model shown in Fig. 2 where there are $L+2$ correlated discrete memoryless sources $(U_{\mathcal{L}}, V, Z)$ i.i.d. with joint distribution $P_{U_{\mathcal{L}} V Z}$. For $1 \leq l \leq L$, encoder l observes the sequence U_l^k and transmits $X_l^n = f_l^{(k,n)}(U_l^k)$ over the corresponding noisy channel, where $f_l^{(k,n)} : \mathcal{U}_l^k \to \mathcal{X}_l^n$, whereas encoder $L+1$ observes V^k, and outputs $f_{L+1}^k(V^k)$, $f_{L+1}^k : \mathcal{V}^k \to \mathcal{M} = \{1, \dots, 2^{kR}\}$. The decoder has access to side-information Z^k, receives $f_{L+1}^k(V^k)$ error-free, and also observes $Y_{\mathcal{L}}^n$, the output of the DMCs $P_{Y_l|X_l}$, $l \in \mathcal{L}$. The output of the decoder is given by the mapping $g^{(k,n)} : (\mathcal{M}, \mathcal{Y}_{\mathcal{L}}^n, \mathcal{Z}^k) \to \hat{V}^k$. The decoder is interested in reconstructing V^k losslessly. For a given bandwidth ratio τ, a rate R is said to be achievable for the $L-$helper JSCC problem if for

Fig. 2: $L-$helper JSCC problem.

every $\lambda \in (0, 1]$, there exist a sequence of numbers $\delta_k \geq 0$ with $\lim_{k \to \infty} \delta_k = 0$, encoders $f_{L+1}^k(\cdot)$, $f_l^{(k,n_k)}(\cdot)$, $l \in \mathcal{L}$, and decoder $g^{(k,n_k)}(\cdot, \cdot, \cdot)$, such that $n_k \leq \tau k$ and

$$\Pr\left(g^{(k,n_k)}\left(f_{L+1}^k(V^k), Y_{\mathcal{L}}^{n_k}, Z^k \right) = V^k \right) \geq 1 - \lambda,$$

$$\text{and } \frac{\log(|\mathcal{M}|)}{k} \leq R + \delta_k.$$

The infimum of all achievable rates R for the $L-$helper JSCC problem with bandwidth ratio τ is denoted by $R(\tau)$.

Next, we show that the problem of obtaining the infimum in (8) coincides with the multi-letter characterization of $R(\tau)$ for the $L-$helper JSCC problem. Let

$$R_k \triangleq \inf_{\substack{f_1^{(k,n)}, \dots, f_L^{(k,n)} \\ n \leq \tau k}} \frac{H(V^k|Y_{\mathcal{L}}^n, Z^k)}{k} \qquad (9)$$

$$\text{s.t. } (Z^k, V^k) - U_l^k - X_l^n = f_l^{(k,n)}(U_l^k) - Y_l^n, \, l \in \mathcal{L}.$$

Theorem 3. *For the $L-$helper JSCC problem,*

$$R(\tau) = \inf_k R_k.$$

Proof: The proof is given in Appendix B. ∎

Having shown the equivalence between the multi-letter characterizations of $\theta(\tau)$ for the TACI problem over noisy channels and $R(\tau)$ for the $L-$helper JSCC problem, our next step is to obtain computable single-letter lower and upper bounds on $R(\tau)$, which can then be used to obtain bounds on $\theta(\tau)$. For this purpose, we use the *source-channel separation theorem* [11, Th. 2.4] for orthogonal multiple access channels. The theorem states that all achievable average distortion-cost tuples in a multi-terminal JSCC (MT-JSCC) problem over an orthogonal multiple access channel (MAC) can be obtained by the intersection of the rate-distortion region and the MAC region. We need a slight generalization of this result when there is side information Z at the decoder, which can be proved similar to [11]. Note that the $L-$helper JSCC problem is a special case of the MT-JSCC problem with $L+1$ correlated sources $P_{U_{\mathcal{L}} V}$ and side information Z available at the decoder, where the objective is to reconstruct V losslessly. Although the above theorem proves that separation holds, a single-letter expression is not available in general for the multi-terminal rate distortion problem [12]. However, single-letter inner and outer bounds have been given in [12], which enable

us to obtain single-letter upper and lower bounds on $R(\tau)$ as follows.

Theorem 4. *Let* $C_l \triangleq \max_{P_{X_l}} I(X_l; Y_l)$, $l \in \mathcal{L}$ *denote the capacity of the channel* $P_{Y_l|X_l}$, *and* τ *the bandwidth ratio for the* $L-$*helper JSCC problem. Define*

$$R^i(\tau) \triangleq \inf_{W_{\mathcal{L}}} \max_{S \subseteq \mathcal{L}} F_S, \qquad (10)$$

where

$$F_S = H(V|W_{S^c}, Z) + I(U_S; W_S|W_{S^c}, V, Z) - \tau \sum_{l \in S} C_l$$

for some auxiliary r.v.'s W_l, $l \in \mathcal{L}$, *such that*

$$(Z, V, U_{l^c}, W_{l^c}) - U_l - W_l, \qquad (11)$$

$|\mathcal{W}_l| \leq |\mathcal{U}_l| + 4$, *and for all subsets* $S \subseteq \mathcal{L}$,

$$I(U_S; W_S|V, W_{S^c}, Z) \leq \tau \left(\sum_{l \in S} C_l \right). \qquad (12)$$

Similarly, let $R^o(\tau)$ *denote the right hand side (R.H.S) of* (10), *when the auxiliary r.v.'s* W_l, $l \in \mathcal{L}$, *satisfy* (12), $|\mathcal{W}_l| \leq |\mathcal{U}_l| + 4$ *and*

$$(V, U_{l^c}, Z) - U_l - W_l. \qquad (13)$$

Then,

$$R^o(\tau) \leq R(\tau) \leq R^i(\tau), \qquad (14)$$
$$H(V|Z) - R^i(\tau) \leq \theta(\tau) \leq H(V|Z) - R^o(\tau). \qquad (15)$$

Proof: From the source-channel separation theorem, an upper bound on $R(\tau)$ can be obtained by the intersection of the Berger-Tung (BT) inner bound [12, Th. 12.1] with the capacity region $(C_1, \ldots, C_L, C_{L+1})$, where C_{L+1} is the rate available over the noiseless link from the encoder of source V to the decoder. Writing the BT inner bound [1] explicitly, we obtain that for all $S \subseteq \mathcal{L}$ (including the null-set),

$$I(U_S; W_S|V, W_{S^c}, Z) \leq \sum_{l \in S} \tau C_l,$$

$$I(U_S; W_S|V, W_{S^c}, Z) + H(V|W_{S^c}, Z) \leq \sum_{l \in S} \tau C_l + C_{L+1},$$

where the auxiliary r.v.'s $W_{\mathcal{L}}$ satisfy (11) and $|\mathcal{W}_l| \leq |\mathcal{U}_l| + 4$. Taking the infimum of C_{L+1} over all such $W_{\mathcal{L}}$ and denoting it by $R^i(\tau)$, we obtain the second inequality in (14). The other direction in (14) is obtained similarly by using the BT outer bound [12, Th. 12.2]. Since $R(\tau)$ is equal to the infimum in (8), substituting (14) in (8) proves (15). ∎

The BT inner bound is tight for the two terminal case, when one of the distortion requirements is zero (lossless) [12, Ch.12]. Thus, we have the following result (for convenience, we drop the index 1 from the associated variables).

---

[1] $R^i(\tau)$ can be improved by introducing a time sharing r.v. T (independent of all the other r.v.'s) in the BT inner bound, but it is omitted here for simplicity.

Lemma 5. *For the TACI problem with* $L = 1$ *and bandwidth ratio* τ,

$$\theta(\tau) = \sup_W I(V; W|Z) \qquad (16)$$

such that $I(U; W|Z) \leq \tau C$, $\qquad (17)$

$$(Z, V) - U - W, \quad |\mathcal{W}| \leq |\mathcal{U}| + 4 \qquad (18)$$

Proof: Note that the Markov chain conditions in (11) and (13) are identical for $L = 1$. Hence,

$$R^i(\tau) = R^o(\tau) = R(\tau). \qquad (19)$$

Using the BT inner bound in [12, Ch.12], we obtain $R(\tau)$ as the infimum of R' such that

$$H(V|Z, W) \leq R' \qquad (20)$$
$$I(U; W|V, Z) \leq \tau C \qquad (21)$$
$$H(V|Z, W) + I(U; W|Z) \leq \tau C + R' \qquad (22)$$

for some auxiliary r.v. W satisfying (18). Hence,

$$R(\tau) = \inf_W \max \big(H(V|W, Z), \ H(V|W, Z) + I(U; W|Z) - \tau C \big) \qquad (23)$$

such that (18) and (21) hold. We next prove that (23) can be simplified as

$$R(\tau) = \inf_W H(V|Z, W) \qquad (24)$$

such that (17) and (18) are satisfied. This is done by showing that, for every r.v. W for which $I(U; W|Z) > \tau C$, there exists a r.v. $\bar{W}$ such that $I(U; \bar{W}|Z) = \tau C$, $H(V|\bar{W}, Z) \leq H(V|W, Z) + I(U; W|Z) - \tau C$ and (18) and (21) are satisfied with W replaced by $\bar{W}$. Setting

$$\bar{W} = \begin{cases} W, & \text{with probability 1-p,} \\ \text{constant,} & \text{with probability p,} \end{cases}$$

suffices, where we choose p such that $I(U; \bar{W}|Z) = \tau C$. The details can be found in [13, Lemma 5]. Eqn. (16) now follows from (15), (19) and (24). ∎

Remark 6. *We note here that the single-letter T2-EE characterization in Lemma 5 exhibits a separation between the distributions of the data sources* U, V, Z *and the channel distribution* $P_{Y|X}$. *Together with the fact that the optimal* $R(\tau)$ *in the* $L-$*helper JSCC problem is achieved by separate source and channel coding, one might be inclined to assume that* $\theta(\tau)$ *for the TACI problem over noisy channels can also be achieved by a communication scheme that performs independent HT and channel coding, and the optimal T2-EE can be obtained by simply replacing the rate constraints in the TACI T2-EE expressions in [5] with the corresponding channel capacity values. Although such a scheme is intuitively pleasing, the T2-EE analysis for such a scheme would involve a tradeoff between two competing error exponents, one being the T2-EE assuming that an error does not occur in channel decoding, and the other being the reliability function* E_r *of the channel* $P_{Y|X}$ *[10]. The details of the analysis can be found in [13].*

IV. Conclusions

We have studied the T2-EE for the distributed HT problem over orthogonal noisy channels with side information available at the detector. For the special case of TACI, single-letter upper and lower bounds are obtained for the T2-EE, which are shown to be tight when there is a single observer in the system. It is interesting to note that the reliability function of the channel does not play a role in the T2-EE, and a strict operational separation between HT and channel coding does not apply in general, even though the optimal T2-EE can be evaluated using the marginal distributions of the data sources and the channels, rather than their joint distributions. Obtaining single-letter bounds for the general HT problem, and analyzing the error exponents for the weighted sum of the type 1 and type 2 errors in the Bayesian setting are some of the interesting problems for future research.

Appendix A
T2-EE using joint typicality detector

Here, we provide the proof for the case $L = 1$. For given arbitrary integers k and n such that $n \leq k\tau$, fix $f_1^{(k,n)} = P_{X_1^n|U_1^k}$. For any integer j and sequence u_1^{kj}, the observer transmits $X_1^{nj} = f_1^{(kj,nj)}(u_1^{kj})$ generated i.i.d. according to $\prod_{j'=1}^{j} P_{X_1^n|U_1^k=u_1^k(j')}$. The detector declares $H_0 : P_{U_1VZ}$ if $\left(Y_1^{nj}, V^{kj}, Z^{kj}\right) \in T_{[Y_1^nV^kZ^k]_{\delta_j}}^j$ (here $\delta_j \to 0$ as $j \to \infty$) and $H_1 : Q_{U_1VZ}$ otherwise. To simplify the exposition, we denote (Y_1^n, V^k, Z^k) and $T_{[Y_1^nV^kZ^k]_{\delta_j}}^j$ by $W_{k,n}$ and $T_{[W_{k,n}]_{\delta_j}}^j$, respectively. By the Markov lemma [12], the type 1 error probability tends to zero as $j \to \infty$. The type 2 error probability is bounded by

$$
\beta'\left(kj, nj, f_1^{(kj,nj)}, \epsilon\right) \leq Q_{Y_1^{nj}V^{kj}Z^{kj}}\left(T_{[Y_1^nV^kZ^k]_{\delta_j}}^j\right)
$$
$$
\leq \sum_{\tilde{P} \in T_{[W_{k,n}]_{\delta_j}}^j} \sum_{w_{k,n}^j \in T_{\tilde{P}}} Q_{W_{k,n}^j}(w_{k,n}^j)
$$
$$
\overset{(a)}{=} \sum_{\tilde{P} \in T_{[W_{k,n}]_{\delta_j}}^j} \sum_{w_{k,n}^j \in T_{\tilde{P}}} 2^{-j\left(H(\tilde{P})+D(\tilde{P}||Q_{W_{k,n}})\right)}
$$
$$
\overset{(b)}{=} \sum_{\tilde{P} \in T_{[W_{k,n}]_{\delta_j}}^j} 2^{-jD(\tilde{P}||Q_{W_{k,n}})} \overset{(c)}{\leq} (j+1)^{|\mathcal{W}_{k,n}|} 2^{-jB(k,n)}
$$

where

$$
B_{k,n}(j) \triangleq \min_{\tilde{P} \in T_{[W_{k,n}]_{\delta_j}}^j} D(\tilde{P}||Q_{W_{k,n}}).
$$

(a), (b) and (c) follow from Lemma's 2.3, 2.6 and 2.2 in [10], respectively. Hence,

$$
\frac{\log\left(\beta'\left(kj, nj, f_1^{(kj,nj)}, \epsilon\right)\right)}{kj} \leq -\frac{B_{k,n}(j)}{k} + \delta'_{k,n}(j),
$$

where $\delta'_{k,n}(j) \triangleq \frac{|\mathcal{W}_{k,n}|\log(j+1)}{kj}$ and $|\mathcal{W}_{k,n}| \leq |\mathcal{Y}|^n|\mathcal{V}|^k|\mathcal{Z}|^k$. Note that for any k and n, $\delta'_{k,n}(j) \to 0$ as $j \to \infty$. Also,

since δ_j is chosen such that it tends to 0 as $j \to \infty$, $B_{k,n}(j)$ converges to $D(P_{W_{k,n}}||Q_{W_{k,n}})$ by the continuity of $D(\tilde{P}||Q_{W_{k,n}})$ in $\tilde{P}$ for fixed $Q_{W_{k,n}}$. Since k, n and $f_1^{(k,n)}$ are arbitrary, it follows from (4) and (6) that $\theta(\tau)$ is an achievable T2-EE for any upper bound ϵ on the type 1 error probability. It is easy to see that this scheme can be generalized to $L > 1$.

Appendix B
Proof of Theorem 3

For the achievability part, consider the following scheme.

Encoding: Fix $k, n \in \mathbb{Z}^+$ and $P_{X_l^n|U_l^k}$ at encoder l, $l \in \mathcal{L}$. Let $j \in \mathbb{Z}^+$. On observing u_l^{kj}, encoder l transmits $X_l^{nj} = f_l^{(kj,nj)}(U_l^{kj})$ generated i.i.d. according to $\prod_{j'=1}^{j} P_{X_l^n|U_l^k=u_l^k(j')}$. Encoder $L+1$ performs uniform random binning on V^k, i.e, $f_{L+1}^{kj} : \mathcal{V}^{kj} \to \mathcal{M} = \{1, 2, \cdots, 2^{kjR}\}$. By uniform random binning, we mean that $f_{L+1}^{kj}(V^{kj}) = m$, where m is selected uniformly at random from the set $\mathcal{M}$.

Decoding: Let M denote the received bin index, and $\delta > 0$ be an arbitrary number. If there exists a unique sequence $\hat{V}^{kj}$ such that $f_{L+1}^{kj}(\hat{V}^{kj}) = M$ and $(\hat{V}^{kj}, Y_{\mathcal{L}}^{nj}, Z^{kj}) \in T_{[V^kY_{\mathcal{L}}^nZ^k]_{\delta}}^j$, then the decoder outputs $g^{(kj,nj)}(M, Y_{\mathcal{L}}^{nj}, Z^{kj}) = \hat{V}^{kj}$. Else, an error is declared.

It can be shown that the probability of decoding error tends to 0 as $j \to \infty$, if $R > H(V^k|Y_{\mathcal{L}}^n, Z^k) + \delta$. The details can be found in [13, Appendix B], along with the proof of the converse.

References

[1] T. Berger, "Decentralized estimation and decision theory," in *IEEE 7th. Spring Workshop on Inf. Theory*, Mt. Kisco, NY, Sep. 1979.

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

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

[4] 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.

[5] 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.

[6] W. Zhao and L. Lai, "Distributed testing against independence with multiple terminals," in *52nd Annual Allerton Conference on Communication, Control and Computing*, Monticello (IL), USA, Oct. 2014.

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

[8] G. Katz, P. Piantanida, and M. Debbah, "Collaborative distributed hypothesis testing," *arXiv:1604.01292 [cs.IT]*, Apr. 2016.

[9] Y. Xiang and Y. H. Kim, "Interactive hypothesis testing against independence," in *Proc. IEEE Int. Symp. Inf. Theory*, Istanbul, Turkey, Nov. 2013.

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

[11] J.-J. Xiao and Z.-Q. Luo, "Multiterminal source-channel communication over an orthogonal multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 53, no. 9, pp. 3255–3264, Sep. 2007.

[12] A. E. Gamal and Y.-H. Kim, *Network Information theory*. Cambridge University Press, 2011.

[13] S. Sreekumar and D. Gündüz, "Distributed hypothesis testing over noisy channels," *[online]. Available. http://www.iis.ee.ic.ac.uk/dgunduz/Papers/Conference/SG_ISIT17.pdf* .