Distributed Hypothesis Testing Over Orthogonal Discrete Memoryless 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 is studied in which multiple helpers transmit their observations to a remote detector over orthogonal discrete memoryless channels. The detector uses the received samples from the helpers along with its own observations to test for the joint distribution of the data. Single-letter inner and outer bounds for the type 2 error exponent (T2EE) is established for the special case of testing against conditional independence (TACI). Specializing this result for the one-helper problem, a single-letter characterization of the optimal T2EE is obtained. Finally, for the general hypothesis testing problem, a lower bound on the T2EE is established by using a separation based scheme that performs independent channel coding and hypothesis testing. It is shown that this separation based scheme recovers the optimal T2EE for the TACI problem.

I. Introduction

Given data samples, statistical hypothesis testing (HT) deals with the problem of ascertaining the true assumption, that is, the true hypothesis, about the data from among a set of hypotheses. In modern communication networks (like in sensor networks, cloud computing and Internet of things (IoT)), data is gathered at multiple remote nodes, referred to as *observers*, and transmitted over noisy links to another node for further processing. Often, there is some prior statistical knowledge available about the data, for example, that the joint probability distribution of the data belongs to a certain prescribed set. In such scenarios, it is of interest to identify the true underlying probability distribution, and this naturally leads to the problem of distributed HT over noisy channels, which is depicted in Fig. 1. Each encoder l, $l = 1, \dots, L$, 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 $Y_1^n, \dots, Y_L^n$ as well as its own observations V^k and Z^k, where H_0 (resp. H_1) is the hypothesis that the data $(U_1, \dots, U_L, V, Z)$ is distributed according to the joint distribution $P_{U_1 \dots U_L V Z}$ (resp. $Q_{U_1 \dots U_L V Z}$). Our goal is to characterize the set of achievable type 2 error exponents (T2EE's) for a prescribed constraint on the type 1 error probability. We will refer to this problem as the *general hypothesis testing with side-information* (GHTS) problem. The instance of the GHTS problem in which $H_0 : P_{U_1 \dots U_L V Z}$ and $H_1 : P_{U_1 \dots U_L|Z} \times P_{V|Z} \times P_Z$ will be referred to as the *testing against conditional independence* (TACI) problem. The special cases of the GHTS and the TACI problem when the side information Z is absent is referred to as the *testing against independence* (TAI) and the *general hypothesis testing* (GHT) problem, respectively.

The study of distributed statistical inference under communication constraints was conceived by Berger in [2]. In [2], and in the follow up literature summarized below, communication from the observers to the detector are assumed to be over rate-limited error-free channel. Ahlswede and Csiszár studied the GHT problem for the case of a single observer ($L = 1$) [3]. They proved a tight single-letter characterization of the optimal T2EE for the TAI problem and also established a lower bound for the GHT problem. Furthermore, they also proved a *strong converse* which states that the optimal achievable T2EE is independent of the constraint on the type 1 error probability. A more general lower bound for the TAI problem with a single observer is established by Han [4], which recovers the corresponding lower bound in [3]. Han also considered complete data compression in a related setting where either U_1, or V, or both (also referred to as two-sided compression setting) are compressed and communicated to the detector using a message set of cardinality two. It is shown that, asymptotically, the optimal T2EE achieved in these three settings are equal. In contrast, even the TAI problem with two-sided compression and general rate constraints remains open till date. Shalaby et. al [5] extended the complete data compression result of Han to show that the optimal T2EE is not improved even if the rate constraint is relaxed to that of zero-rate compression (sub-exponential message set with respect to block-length). Shimokawa et. al [6] obtained a better lower bound for the GHT problem by considering quantization and binning at the encoder along with a minimum empirical-entropy decoder. Rahman and Wagner [7] established inner bound for the TACI problem with L observers, by performing quantization and binning at the encoders. This quantize-bin-test bound is then shown to be tight, and also to coincide with the one achieved by the Shimokawa-Han-Amari scheme in [6] for the case of a single observer, thereby implying the optimality of both these schemes. The optimal T2EE for the TAI problem with two decision centers is obtained in [8], where the encoder communicates to both detectors via a common bit-pipe in addition to

This work was presented in part at the International Symposium on Information theory (ISIT), Aachen, 2017 [1].

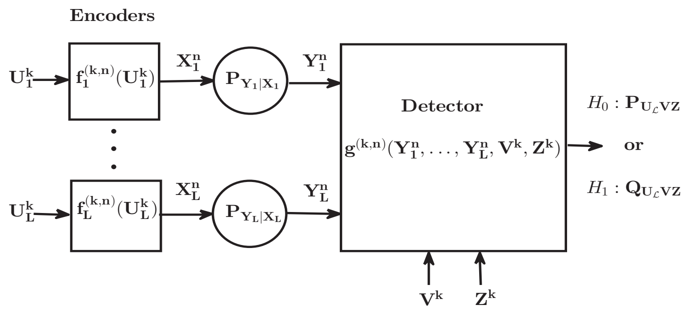

Fig. 1: Illustration of the system model.

individual private bit-pipes to each. The TACI problem with multiple observers is still open, although a special case has been solved in [9] when the observed data follows a certain Markovian condition. The T2EE for more complex settings involving interaction between two observers, where one of the observer also acts as the detector has also been studied in [10], [11]. The observers exchange messages over a noiseless link for K rounds of interaction under a constraint on the total exchange rate. On completion of K rounds, the HT decision is taken at the observer which receives the last message. The optimal T2EE for TAI in this model with $K = 1$ and $K > 1$ is obtained in [10] and [11], respectively. Recently, a lower bound for the general HT case in this setting is established in [12]. The authors also prove a single-letter expression for the optimal T2EE in the zero-rate compression regime, analogous to that of [5]. When the detector also performs lossy source reconstruction in addition to hypothesis testing, the set of all simultaneously achievable T2EE-distortion pairs for the GHT problem is studied in [13]. Therein, the authors also prove a single-letter characterization of the exponent-distortion region for the special case of TAI.

In contrast, hypothesis testing in distributed settings that involve communication over noisy channels has received relatively less attention in the past. In noiseless rate-limited settings, the encoder can reliably communicate its observation subject to a rate constraint. However, this is no longer the case in noisy settings, which complicates the study of error exponents in HT. A measure of the noisiness of the channel is the so-called reliability function $E(R)$ (function of the communication rate R) of the channel [14]. $E(R)$ denotes the exponent (first order) of best asymptotic decay of the probability of error achievable in channel coding when the rate of messages is R. It is reasonable to expect that $E(R)$ plays a role in the characterization of the achievable T2EE. The problem of designing a channel codebook that achieves an error probability decreasing with exponent equal to $E(R)$ is an open problem in general. However, it is well known that $E(R) \geq E_r(R)$, where $E_r(R)$ is the random coding exponent [14]. As the name suggests, the existence of a channel codebook achieving $E_r(R)$ can be shown by the standard random coding method.

The goal of this paper is to study the best attainable T2EE for the GHTS problem, and obtain a computable characterization of the same. Although a complete solution is not to be expected (since even the corresponding noiseless case is still open), the aim is to provide an achievable scheme for the general problem, and to discuss special cases in which a tight characterization can be obtained. The main contributions can be summarized as follows. We establish single-letter lower and upper bounds on the achievable T2EE for the TACI problem with multiple observers. This is done by first mapping the problem to an equivalent joint source channel coding (JSCC) problem with helpers. The Berger-Tung bounds [15] and the source- channel separation theorem in [16] are then used to obtain the desired bounds. Subsequently, these bounds are shown to be tight for the special case of a single observer. This tight single-letter characterization for TACI reveals that the optimal T2EE depends only on the marginal distributions of the observed data and the channel rather than on their joint distribution. Motivated by this, we obtain a lower bound on the T2EE for the GHTS problem by using a separation based scheme that performs independent hypothesis testing and channel coding. This scheme recovers the optimal T2EE derived earlier for TACI.

The rest of the paper is organized as follows. In Section II, we introduce the system model, definitions and supporting lemmas. Following this, we introduce the $L-$helper JSCC problem and obtain lower and upper bounds for the achievable T2EE for the TACI problem in Section III. Finally in Section IV, we describe the separation based scheme for the GHTS problem, and analyze the T2EE attained by this scheme.

A. Notations

Random variables (r.v.'s) are denoted by capital letters (e.g., X), their realizations by the corresponding lower case letters (e.g., x), and their support by calligraphic letters (e.g., $\mathcal{X}$). 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 and 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 the sequence $X_{l,1}, \ldots, X_{l,m}$. 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 any $\mathcal{G}=\{l_1,\ldots,l_g\}$, $\left\{X_{l_1}^m,\ldots,X_{l_g}^m\right\}$, $\left\{X_{l_1}^m(j),\ldots,X_{l_g}^m(j)\right\}$ and $\left\{\left\{X_{l_1}^m(j)\right\}_{j\in\mathbb{Z}^+},\ldots,\left\{X_{l_g}^m(j)\right\}_{j\in\mathbb{Z}^+}\right\}$ are denoted by $X_{\mathcal{G}}^m$, $X_{\mathcal{G}}^m(j)$ and $\left\{X_{\mathcal{G}}^m(j)\right\}_{j\in\mathbb{Z}^+}$, respectively. Following the notation in [14], 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. $D(P||Q)$, $I(X;Y)$ and $H(X)$ denote, respectively, the Kullback-Leibler (KL) divergence between distributions P and Q, mutual information between r.v.'s X and Y, and the entropy of X, respectively, defined as follows

$$D(P||Q)\triangleq\sum_{x\in\mathcal{X}}P(x)\log\left(\frac{P(x)}{Q(x)}\right), \tag{1}$$

$$I(X;Y)\triangleq D(P_{XY}||P_XP_Y), \tag{2}$$

$$H(X)\triangleq I(X;X). \tag{3}$$

For $a\in\mathbb{R}^+$, $[a]$ denotes the set of integers $\{1,2,\ldots,\lceil a\rceil\}$. All logarithms are to the base 2. For any set $\mathcal{G}$, $\mathcal{G}^c$ denotes the set complement. $X\perp Y$ denotes that r.v.'s X and Y are statistically independent. Finally, $\mathbb{1}(\cdot)$ and $o(\cdot)$ denote the indicator function and the Little-o notation of Landau, respectively.

II. System Model and Definitions

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 codeword $X_l^n=f_l^{(k,n)}(U_l^k)$, where $f_l^{(k,n)}:\mathcal{U}_l^k\rightarrow\mathcal{X}_l^n$ 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 independent of each other and memoryless. Depending on the received symbols $Y_{\mathcal{L}}^n$ and its own observations (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 map $g^{(k,n)}:\mathcal{Y}_{\mathcal{L}}^n\times\mathcal{V}^k\times\mathcal{Z}^k\rightarrow\{H_0,H_1\}$. Let $\mathcal{A}$ denote the acceptance region for H_0, i.e., $g^{(k,n)}(y_{\mathcal{L}}^n,v^k,z^k)=H_0$ if $(y_{\mathcal{L}}^n,v^k,z^k)\in\mathcal{A}$ and $g^{(k,n)}(y_{\mathcal{L}}^n,v^k,z^k)=H_1$ otherwise. It is assumed that $P_{U_{\mathcal{L}}VZ}<<Q_{U_{\mathcal{L}}VZ}$, i.e., the joint distribution of the data under H_0 is absolutely continuous[1] with respect to that under H_1.

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

$$\beta'\left(k,n,f_1^{(k,n)},\ldots,f_L^{(k,n)},\epsilon\right)\triangleq\inf_{g^{(k,n)}}\bar{\beta}\left(k,n,f_1^{(k,n)},\ldots,f_L^{(k,n)},\ g^{(k,n)}\right) \tag{4}$$

such that

$$\bar{\alpha}\left(k,n,f_1^{(k,n)},\ldots,f_L^{(k,n)},\ g^{(k,n)}\right)\le\epsilon,$$

and

$$(Z^k,\ V^k,\ U_{l^c}^k)-U_l^k-X_l^n-Y_l^n,\ l\in\mathcal{L},$$

where $X_l^n=f_l^{(k,n)}(U_l^k)$, $l^c\triangleq\mathcal{L}\backslash l$ and

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

Definition 1. *A T2EE κ is said to be (τ,ϵ) achievable for the GHT problem if there exists a sequence of integers k, corresponding sequences of encoding functions $f_1^{(k,n_k)},\ \ldots,f_L^{(k,n_k)}$ and decoding functions $g^{(k,n_k)}$ such that $n_k\le\tau k$, $\forall\ k$, and for any $\delta>0$,*

$$\limsup_{k\rightarrow\infty}\frac{\log\left(\beta(k,\tau,\epsilon)\right)}{k}\le-(\kappa-\delta). \tag{6}$$

Let

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

$$\theta(\tau)\triangleq\sup_{k\in\mathbb{Z}^+}\theta(k,\tau), \tag{8}$$

[1] This assumption is necessary for the Kullback-Leibler divergence quantities like $D(P_{Y^nV^kZ^k}||Q_{Y^nV^kZ^k})$ that characterize the T2EE to be finite.

where

$$\theta(k,\tau) \triangleq \sup_{\substack{f_1^{(k,n)},\ldots,f_L^{(k,n)} \\ n\le \tau k}} \frac{D\big(P_{Y_{\mathcal{L}}^n V^k Z^k}||Q_{Y_{\mathcal{L}}^n V^k Z^k}\big)}{k}. \tag{9}$$

Next, we obtain single-letter inner and outer bounds on $\kappa(\tau,\epsilon)$ for the problem of TACI over noisy channels. Our approach is similar to that in [3], in that we first obtain bounds on $\kappa(\tau,\epsilon)$ in terms of $\theta(\tau)$, and subsequently show that $\theta(\tau)$ has a single-letter characterization in terms of information theoretic quantities. We establish this characterization by considering an equivalent joint source-channel coding (JSCC) problem with noisy helpers.

Lemma 2. *For the GHT problem with any bandwidth ratio $\tau > 0$, we have*

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

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

Proof: The proof is similar to that of Theorem 1 in [3]. Here, we prove (i), which states that a T2EE of $\theta(\tau)$ is achievable. The proof of (ii) follows in a straightforward manner from the proof given in [3] and is omitted here. For $k, n_k \in \mathbb{Z}^+$ such that $n_k \le \tau k$, consider a sequence of encoding and decoding functions, $f_l^{(k,n_k)},\ l \in \mathcal{L}$ and $g^{(k,n_k)}$, respectively. Fix k and n_k. For each $l \in \mathcal{L}$, consider block-encoding with channel inputs $\{X_l^{n_k}(j)\}_{j\in\mathbb{Z}^+}$, where $X_l^{n_k}(j) = f_l^{(k,n_k)}(U^k(j))$. Let $\{Y_l^{n_k}(j)\}_{j\in\mathbb{Z}^+}$, $l \in \mathcal{L}$ denote the corresponding channel outputs. Note that $\{Y_l^{n_k}(j)\}_{j\in\mathbb{Z}^+}$ is an infinite sequence of i.i.d. r.v.'s indexed by j. Hence, by the application of Stein's Lemma [3] 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

$$\limsup_{j\to\infty} \frac{\log\left(\beta'\left(kj, n_k j, f_1^{(kj,n_kj)},\ldots,f_L^{(kj,n_kj)},\epsilon\right)\right)}{kj} \le \frac{-D\big(P_{Y_{\mathcal{L}}^n V^k Z^k}||Q_{Y_{\mathcal{L}}^n V^k Z^k}\big)}{k}. \tag{10}$$

Next, note that the following property holds for any infinite sequence a_n,

$$\limsup_{n\to\infty} \liminf_{n\to\infty} a_n \le \liminf_{n\to\infty} \limsup_{n\to\infty} a_n. \tag{11}$$

Taking infimum over all sequences of integers $n_k \le \tau k$ and encoding functions $f_l^{(k,n_k)},\ l \in \mathcal{L}$ in (10) and using (11) to interchange the limits on the left hand side (L.H.S) of (10), we obtain,

$$\limsup_{j\to\infty} \frac{\log\left(\beta(kj,\tau,\epsilon)\right)}{kj} \le -\theta(k,\tau). \tag{12}$$

For $m \ge kj$, $\beta(m,\tau,\epsilon) \le \beta(kj,\tau,\epsilon)$. Hence,

$$\begin{aligned}\limsup_{m\to\infty} \frac{\log\left(\beta(m,\tau,\epsilon)\right)}{m} &\le \limsup_{j\to\infty} \frac{\log\left(\beta(kj,\tau,\epsilon)\right)}{kj} \\ &\le -\theta(k,\tau). \end{aligned} \tag{13}$$

Noting that the L.H.S of (13) does not depend on k, the proof of (i) is completed by taking supremum with respect to k on both sides of the equation. ■

Remark 3. *We remark here that the tightness of the characterizations obtained in this paper is claimed only as ϵ tends to zero, unlike the results in [3] and other related papers. The reason is that it is not known whether a strong-converse property proved in [3] holds when rate-limited noiseless channels are replaced by orthogonal DMC's. Part (ii) of Lemma 2 is known as the weak converse for the HT problem in the literature and (i) and (ii) together imply that $\theta(\tau)$ is the optimal T2EE as $\epsilon \to 0$, i.e., $\lim_{\epsilon\to 0} \kappa(\tau,\epsilon) = \theta(\tau)$.*

Part (i) of Lemma 2 proves the achievability of the T2EE $\theta(\tau)$ 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 in (9) becomes mutual information, and we have

$$\begin{aligned}\theta(\tau) = &\sup_{\substack{f_1^{(k,n)},\ldots,f_L^{(k,n)} \\ k,n\le \tau k}} \frac{I(V^k; Y_{\mathcal{L}}^n|Z^k)}{k} \text{ s.t.} \\ &(Z^k,\, V^k) - U_l^k - X_l^n - Y_l^n,\ l \in \mathcal{L}.\end{aligned}$$

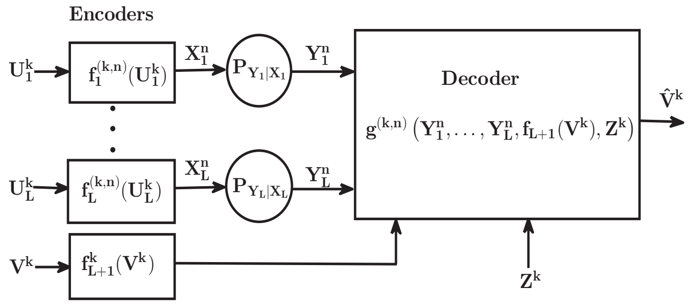

Fig. 2: $L-$helper JSCC problem.

Although Lemma 2 implies that $\theta(\tau)$ is an achievable T2EE, it is in general not computable as it is characterized in terms of a multi-letter expression. 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

$$\theta(\tau) = H(V|Z) - \inf_{\substack{f_1^{(k,n)},\dots,f_L^{(k,n)} \\ k, n \leq \tau k}} \frac{H(V^k|Y_{\mathcal{L}}^n, Z^k)}{k} \text{ s.t.} \quad (14)$$
$$(Z^k,\, V^k) - U_l^k - X_l^n - Y_l^n,\ l \in \mathcal{L}.$$

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 (14). The computable characterization of the lower and upper bounds for (14) then follows from the single-letter outer and inner bounds available for 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 $P_{U_{\mathcal{L}}VZ}$. For $1 \leq l \leq L$, encoder $f_l^{(k,n)} : \mathcal{U}_l^k \rightarrow \mathcal{X}_l^n$ observes the sequence U_l^k and transmits $X_l^n = f_l^{(k,n)}(U_l^k)$ over the corresponding DMC $P_{Y_l|X_l}$, while encoder $f^k : \mathcal{V}^k \rightarrow \mathcal{M} = \{1, \dots, 2^{kR}\}$ observes V^k, and outputs $M = f^k(V^k)$. Decoder $g^{(k,n)}$ has access to side-information Z^k, receives $f_{L+1}^k(V^k)$ error-free, observes $Y_{\mathcal{L}}^n$ and outputs an $\hat{V}^k$ according to the map $g^{(k,n)} : \mathcal{Y}_{\mathcal{L}}^n \times \mathcal{M} \times \mathcal{Z}^k \rightarrow \hat{\mathcal{V}}^k$. The goal of $g^{(k,n)}$ is to reconstruct V^k losslessly. We will first establish the multi-letter characterization of the rate region of the $L-$ helper JSCC. Prior to stating the result, we require some new definitions.

Definition 4. *For a given bandwidth ratio τ, a rate R is said to be achievable for the L-helper JSCC problem if for every $\lambda \in (0, 1]$, there exist sequences of positive numbers δ_k tending to 0 as $k \rightarrow \infty$, encoders $f^k(\cdot)$, $f_l^{(k,n_k)}(\cdot)$, and decoder $g^{(k,n_k)}(\cdot,\cdot,\cdot)$ such that $n_k \leq \tau k$ and*

$$\Pr\left(g_h^{(k,n_k)}\left(Y^{n_k}, M, Z^k\right) = V^k\right) \geq 1 - \lambda \textit{ and}$$
$$\frac{\log(|\mathcal{M}|)}{k} \leq R + \delta_k.$$

Let

$$R_{TE}(\tau) \triangleq \inf\{R : R \textit{ is achievable.}\} \quad (15)$$

We next show that the problem of obtaining the infimum in (14) coincides with the multi-letter characterization of $R_{TE}(\tau)$ in (15). Define

$$R_k \triangleq \inf_{\substack{f_1^{(k,n)},\dots,f_L^{(k,n)} \\ n \leq \tau k}} \frac{H(V^k|Y_{\mathcal{L}}^n, Z^k)}{k} \text{ s.t} \quad (16)$$
$$(Z^k,\, V^k) - U_l^k - X_l^n - Y_l^n,\ l \in \mathcal{L}.$$

Theorem 5. *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* [16, 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 similarly to [16]. 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 [15]. However, single-letter inner and outer bounds have been given in [15], which enable us to obtain single-letter upper and lower bounds on $R(\tau)$ as follows.

Theorem 6. *For $\mathcal{G} \subseteq \mathcal{L}$, let $C_{\mathcal{G}} \triangleq \sum_{l \in \mathcal{G}} C_l$, where $C_l \triangleq \max_{P_{X_l}} I(X_l;Y_l)$ denotes the capacity of the channel $P_{Y_l|X_l}$. For the $L-$helper JSCC problem with bandwidth ratio τ, define*

$$R^i(\tau) \triangleq \inf_{W_{\mathcal{L}}} \max_{\mathcal{G} \subseteq \mathcal{L}} F_{\mathcal{G}}, \tag{17}$$

where

$$F_{\mathcal{G}} = H(V|W_{\mathcal{G}^c}, Z) + I(U_{\mathcal{G}}; W_{\mathcal{G}}|W_{\mathcal{G}^c}, V, Z) - \tau \sum_{l \in \mathcal{G}} C_l$$

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

$$(Z,\ V,\ U_{l^c},\ W_{l^c}) - U_l - W_l, \tag{18}$$

$|\mathcal{W}_l| \leq |\mathcal{U}_l| + 4$*, and* $\forall\ \mathcal{G} \subseteq \mathcal{L}$*,*

$$I(U_{\mathcal{L}}; W_{\mathcal{G}}|V, W_{\mathcal{G}^c}, Z) \leq \tau C_{\mathcal{G}}. \tag{19}$$

Similarly, let $R^o(\tau)$ denote the right hand side (R.H.S) of (17)*, when the auxiliary r.v.'s $W_l,\ l \in \mathcal{L}$ satisfy $|\mathcal{W}_l| \leq |\mathcal{U}_l| + 4$, Eqn.*(19) *and*

$$(V, U_{l^c}, Z) - U_l - W_l. \tag{20}$$

Then,

$$R^o(\tau) \leq\ R(\tau) \leq R^i(\tau),\ \text{and} \tag{21}$$

$$H(V|Z) - R^i(\tau) \leq\ \theta(\tau) \leq H(V|Z) - R^o(\tau). \tag{22}$$

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 [15, 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 [2] explicitly, we obtain that for all $\mathcal{G} \subseteq \mathcal{L}$ (including the null-set),

$$I(U_{\mathcal{G}}; W_{\mathcal{G}}|V, W_{\mathcal{G}^c}, Z) \leq \sum_{l \in \mathcal{G}} \tau C_l,$$

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

where the auxiliary r.v.'s $W_{\mathcal{L}}$ satisfy (18) 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 (21). The other direction in (21) is obtained similarly by using the BT outer bound [15, Th. 12.2]. Since $R(\tau)$ is equal to the infimum in (14), substituting (21) in (14) proves (22). ■

The BT inner bound is tight for the two terminal case, when one of the distortion requirements is zero (lossless) [15, Ch.12]. Thus, we have the following result (for convenience, we drop the index 1 from the associated variables).

Lemma 7. *For the TACI problem with $L = 1$ and bandwidth ratio τ,*

$$\theta(\tau) = \sup_{W} I(V; W|Z) \tag{23}$$

$$\text{such that } I(U; W|Z) \leq \tau C, \tag{24}$$

$$(Z, V) - U - W,\ |\mathcal{W}| \leq |\mathcal{U}| + 4. \tag{25}$$

[2] $R^i(\tau)$ 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.

Proof: Note that the Markov chain conditions in (18) and (20) are identical for $L=1$. Hence,

$$R^i(\tau) = R^o(\tau) = R(\tau). \tag{26}$$

Using the BT inner bound in [15, Ch.12], we obtain $R(\tau)$ as the infimum of R' such that

$$H(V|Z,W) \leq R' \tag{27}$$
$$I(U;W|V,Z) \leq \tau C \tag{28}$$
$$H(V|Z,W) + I(U;W|Z) \leq \tau C + R', \tag{29}$$

for some auxiliary r.v. W satisfying (25). Hence,

$$\begin{aligned} R(\tau) = \inf_W \max \big(&H(V|W,Z),\ H(V|W,Z) \\ &+ I(U;W|Z) - \tau C \big), \end{aligned} \tag{30}$$

such that (25) and (28) hold. We next prove that (30) can be simplified as

$$R(\tau) = \inf_W H(V|Z,W) \tag{31}$$

such that (24) and (25) 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

$$I(U;\bar{W}|Z) = \tau C \tag{32}$$
$$H(V|\bar{W},Z) \leq H(V|W,Z) + I(U;W|Z) - \tau C \tag{33}$$

and (25) and (28) are satisfied with W replaced by $\bar{W}$. Setting

$$\bar{W} = \begin{cases} W, & \text{with probability 1-p,} \\ \text{constant}, & \text{with probability p,} \end{cases} \tag{34}$$

suffices, where p is chosen such that $I(U;\bar{W}|Z) = \tau C$. To see this, first note that $H(V|\bar{W},Z)$ is an increasing function of p, while $I(U;\bar{W}|Z)$ and $I(U;\bar{W}|V,Z)$ are decreasing functions of p. Hence, it is possible to choose p such that (32) and (28) are satisfied with $\bar{W}$ in place of W. It is clear that such a choice of $\bar{W}$ also satisfies (25). To complete the proof of (31), it remains to be shown that for such a $\bar{W}$, (33) holds. We can write,

$$H(V|\bar{W},Z) = (1-p)H(V|W,Z) + pH(V|Z) \tag{35}$$

Taking derivative with respect to p, we obtain

$$\frac{d}{dp} H(V|\bar{W},Z) = I(V;W|Z) \tag{36}$$

Similarly,

$$\frac{d}{dp} H(U|\bar{W},Z) = I(U;W|Z) \tag{37}$$

For the Markov chain in (25), we have $I(V;W|Z) \leq I(U;W|Z)$ by the data processing inequality [14]. Hence, we have that

$$\frac{d}{dp} H(V|\bar{W},Z) \leq \frac{d}{dp} H(U|\bar{W},Z) \tag{38}$$

From (38), it follows that

$$F(p) \triangleq H(V|\bar{W},Z) + I(U;\bar{W}|Z) - \tau C \tag{39}$$

is a decreasing function of p. Together with the fact that $H(V|\bar{W},Z)$ is increasing with p, it follows that (33) is satisfied for $\bar{W}$ chosen in (34). Having shown (31), (23) now follows from (22) and (26). This completes the proof. ∎

IV. T2EE FOR THE GHTS PROBLEM

Lemma 7 provides a single-letter characterization of the optimal T2EE for the TACI problem. As shown in Appendix A, the optimal T2EE for the GHT problem is achieved when the input X^n is generated correlated with the observed source sequence U^k, according to some conditional distribution $P_{X^n|U^k}$ and the optimal test achieving the Stein's exponent, i.e., the Neyman-Pearson test is performed on blocks of data (received and observed) at the detector. The encoder and the detector for such a scheme would be computationally complex to implement from a practical viewpoint. In [6], Shimokawa et.al. obtained the T2EE for the GHT problem with a single observer in the rate-limited noiseless channel setting using a quantization and binning scheme. In this scheme, the type of the observed sequence or a message indicating whether the observed sequence is typical is sent by the encoder to the detector, which aids in the HT. Since the number of types is polynomial in the block-length, these can be communicated error-free at asymptotically zero-rate. Intuitively, it is desirable to do the same in the noisy channel setting as well, however this is not possible in general. In this section, we study the T2EE achieved by a *Separate Hypothesis Testing and Channel Coding* (SHTCC) *scheme* for the case of a single observer ($L=1$). In the SHTCC scheme, the encoding and decoding functions are restricted to be of the form $f^{(k,n)} = f_C^{(k,n)} \circ f_S^{(k)}$ and $g^{(k,n)} = g_S^{(k)} \circ g_C^{(k,n)}$, respectively. The source encoder $f_S^{(k)} : \mathcal{U}^k \rightarrow \mathcal{M} = \{0, 1, \cdots, \lceil 2^{kR} \rceil\}$ generates an index $M \in \mathcal{M}$ based on the observed sequence U^k and the channel encoder $f_C^{(k,n)} : \mathcal{M} \rightarrow \tilde{\mathcal{C}} = \{X^n(j),\ j \in [0 : \lceil 2^{kR} \rceil]\}$ maps M into the codeword $X^n(M)$ from the channel codebook $\tilde{\mathcal{C}}$. Note that the actual rate of channel transmission is $\frac{kR}{n} = \frac{R}{\tau}$. The channel decoder $g_C^{(k,n)} : \mathcal{Y}^n \rightarrow \mathcal{M}$ maps the received sequence into an index M and $g_S^{(k)} : \mathcal{M} \times \mathcal{V}^k \rightarrow \{H_0, H_1\}$ outputs the result of the HT. The codewords $X^n(j)$, $j \in [1 : \lceil 2^{kR} \rceil]$ in $\tilde{C}$ are generated i.i.d. according to the distribution P_X that achieves the exponent $E_r(\frac{R}{\tau})$ for decoding at the channel decoder. $M = 0$ denotes a special error message indicating that the observed sequence U^k is not typical and $X^n(0)$ has to be chosen such that it achieves the best exponent for the probability of error at the channel decoder. By generating all the $\lceil 2^{kR} \rceil + 1$ codewords in $\tilde{C}$ i.i.d according to P_X, it is clear that this exponent is at least $E_r(\frac{R}{\tau})$, but it can be higher in some cases.

We state the main result of this section below. To simplify the exposition, the case when the side information Z is absent is considered. Also, we omit the subscript denoting the observer index from all the relevant variables since $L = 1$.

Theorem 8. *Consider the GHT problem with $L = 1$, bandwidth ratio τ and channel $P_{Y|X}$ with capacity C. Then, $\kappa(\tau, \epsilon) \geq \kappa_s$ for $0 < \epsilon \leq 1$, where*

$$\kappa_s = \sup_{P_{W|U} \in \mathcal{B}} \min \left(E_1(P_{W|U}),\ E_2(R, P_{W|U}),\ E_3(R, P_{W|U}, \tau),\ E_4(R, P_{W|U}, \tau) \right), \tag{40}$$

where

$$\mathcal{B} = \{P_{W|U} : I(U;W|V) \leq R < \tau C,\ V - U - W\},$$
$$E_1(P_{W|U}) = \min_{P_{\tilde{U}\tilde{V}\tilde{W}} \in \mathcal{T}_1(P_{UW}, P_{VW})} D(P_{\tilde{U}\tilde{V}\tilde{W}} || Q_{UV} P_{W|U}), \tag{41}$$

$$E_2(R, P_{W|U}) = \begin{cases} \min_{P_{\tilde{U}\tilde{V}\tilde{W}} \in \mathcal{T}_2(P_{UW}, P_V)} D(P_{\tilde{U}\tilde{V}\tilde{W}} || Q_{UV} P_{W|U}) + (R - I(U;W|V)), & \text{if } I(U;W) > R \\ \infty, & \text{otherwise} \end{cases} \tag{42}$$

$$E_3(R, P_{W|U}, \tau) = \begin{cases} \min_{P_{\tilde{U}\tilde{V}\tilde{W}} \in \mathcal{T}_3(P_{UW}, P_V)} D(P_{\tilde{U}\tilde{V}\tilde{W}} || Q_{UV} P_{W|U}) + (R - I(U;W|V)) + \tau E_r\left(\frac{R}{\tau}\right), & \text{if } I(U;W) > R \\ \min_{P_{\tilde{U}\tilde{V}\tilde{W}} \in \mathcal{T}_3(P_{UW}, P_V)} D(P_{\tilde{U}\tilde{V}\tilde{W}} || Q_{UV} P_{W|U}) + I(V;W) + \tau E_r\left(\frac{R}{\tau}\right), & \text{otherwise} \end{cases} \tag{43}$$

$$E_4(R, P_{W|U}, \tau) = \begin{cases} \min_{P_{\tilde{U}\tilde{V}\tilde{W}} \in \mathcal{T}_4(P_V)} D(P_{\tilde{V}} || Q_V) + (R - I(U;W|V)) + E_s(R, \tau), & \text{if } I(U;W) > R \\ \min_{P_{\tilde{U}\tilde{V}\tilde{W}} \in \mathcal{T}_4(P_V)} D(P_{\tilde{V}} || Q_V) + I(V;W) + E_s(R, \tau). & \text{otherwise} \end{cases} \tag{44}$$

$$E_s(R, \tau) \geq \tau E_r\left(\frac{R}{\tau}\right), \tag{45}$$
$$\mathcal{T}_1(P_{UW}, P_{VW}) = \{P_{\tilde{U}\tilde{V}\tilde{W}} \in \mathcal{T}(\mathcal{U} \times \mathcal{V} \times \mathcal{W}) : P_{\tilde{U}\tilde{W}} = P_{UW},\ P_{\tilde{V}\tilde{W}} = P_{VW}\},$$
$$\mathcal{T}_2(P_{UW}, P_V) = \{P_{\tilde{U}\tilde{V}\tilde{W}} \in \mathcal{T}(\mathcal{U} \times \mathcal{V} \times \mathcal{W}) : P_{\tilde{U}\tilde{W}} = P_{UW},\ P_{\tilde{V}} = P_V,\ H(W|V) \leq H(\tilde{W}|\tilde{V})\},$$
$$\mathcal{T}_3(P_{UW}, P_V) = \{P_{\tilde{U}\tilde{V}\tilde{W}} \in \mathcal{T}(\mathcal{U} \times \mathcal{V} \times \mathcal{W}) : P_{\tilde{U}\tilde{W}} = P_{UW},\ P_{\tilde{V}} = P_V\},$$
$$\mathcal{T}_4(P_V) = \{P_{\tilde{U}\tilde{V}\tilde{W}} \in \mathcal{T}(\mathcal{U} \times \mathcal{V} \times \mathcal{W}) : P_{\tilde{V}} = P_V\},$$

and $I(U;W|V)$, $I(U;W)$ and $I(V;W)$ above are computed with respect to the joint distribution $P_{UV}P_{W|U}$.

The proof of Theorem 8 is given in Appendix C. Although the expression κ_s in Theorem 8 appears complicated, the terms $E_1(\cdot)$ to $E_4(\cdot)$ can be understood to correspond to exponents caused due to distinct events that can possibly lead to a type 2 error. Note that $E_1(\cdot)$ and $E_2(\cdot)$ are the same terms appearing in the exponent achieved by the Shimokawa-Han-Amari scheme [6] for the noiseless channel setting, while $E_3(\cdot)$ and $E_4(\cdot)$ are additional terms introduced due to the noisiness of the channel. $E_3(\cdot)$ corresponds to the event when the encoder transmits the message $M \neq 0$, the channel decoder outputs $\hat{M} \neq M$ and $g_S^{(k)}(\hat{M}, V^k) = H_0$. $E_4(\cdot)$ is due to the type 2 error event that occurs when the error message $M = 0$ is transmitted, $\hat{M} \neq M$ and $g_S^{(k)}(\hat{M}, V^k) = H_0$. The term $E_s(R, \tau)$ in (45) requires further explanation. This corresponds to the best error exponent in channel coding that can be achieved for the error message $M = 0$ when the codewords for the remaining $\lceil 2^{kR} \rceil$ messages in $\tilde{\mathcal{C}}$ achieve the random coding error exponent $\tau E_r(\frac{R}{\tau})$. As mentioned above, $E_s(R,\tau)$ is at least equal to $\tau E_r(\frac{R}{\tau})$. However, $E_s(R,\tau)$ may be significantly higher in some cases, for instance, when the channel has positive zero error capacity [14] and it is possible to choose one error-free channel codeword while simultaneously achieving $\tau E_r(\frac{R}{\tau})$ for the remaining $\lceil 2^{kR} \rceil$ codewords, in which case $E_s(R,\tau) = \infty$.

Remark 9. *It is easy to extend the above result to the case when the side information Z is available at the detector. Let $\hat{\kappa}_s$ denote the value of κ_s when,*

(i) $D(P_{\tilde{U}\tilde{V}\tilde{W}}||Q_{UV}P_{W|U})$ and $D(P_{\tilde{V}}||Q_V)$ are replaced by $D(P_{\tilde{U}\tilde{V}\tilde{W}\tilde{Z}}||Q_{UVZ}P_{W|U})$ and $D(P_{\tilde{V}\tilde{Z}}||Q_{VZ})$, respectively.
(ii) $I(U;W|V)$, I(U;W) and $I(V;W)$ are replaced by $I(U;W|V,Z)$, $I(U;W|Z)$ and $I(V;W|Z)$, respectively.
(iii) the set $\mathcal{B}$ is replaced by $\{P_{W|U} : I(U;W|V,Z) \leq R < \tau C,\ (V,Z) - U - W\}$.
(iv) the minimization over sets $\mathcal{T}_1(P_{UW},P_{VW})$, $\mathcal{T}_2(P_{UW},P_V)$, $\mathcal{T}_3(P_{UW},P_V)$ and $\mathcal{T}_4(P_V)$ are replaced, respectively, by that over the sets

$$\begin{aligned}
\mathcal{T}_1(P_{UW},P_{VWZ}) &= \{P_{\tilde{U}\tilde{V}\tilde{W}\tilde{Z}} \in \mathcal{T}(\mathcal{U}\times\mathcal{V}\times\mathcal{W}\times\mathcal{Z}) : P_{\tilde{U}\tilde{W}} = P_{UW},\ P_{\tilde{V}\tilde{W}\tilde{Z}} = P_{VWZ}\},\\
\mathcal{T}_2(P_{UW},P_{VZ}) &= \{P_{\tilde{U}\tilde{V}\tilde{W}\tilde{Z}} \in \mathcal{T}(\mathcal{U}\times\mathcal{V}\times\mathcal{W}\times\mathcal{Z}) : P_{\tilde{U}\tilde{W}} = P_{UW},\ P_{\tilde{V}\tilde{Z}} = P_{VZ},\ H(W|VZ) \leq H(\tilde{W}|\tilde{V}\tilde{Z})\},\\
\mathcal{T}_3(P_{UW},P_{VZ}) &= \{P_{\tilde{U}\tilde{V}\tilde{W}\tilde{Z}} \in \mathcal{T}(\mathcal{U}\times\mathcal{V}\times\mathcal{W}\times\mathcal{Z}) : P_{\tilde{U}\tilde{W}} = P_{UW},\ P_{\tilde{V}\tilde{Z}} = P_{VZ}\},\\
\mathcal{T}_4(P_{VZ}) &= \{P_{\tilde{U}\tilde{V}\tilde{W}\tilde{Z}} \in \mathcal{T}(\mathcal{U}\times\mathcal{V}\times\mathcal{W}\times\mathcal{Z}) : P_{\tilde{V}\tilde{Z}} = P_{VZ}\}.
\end{aligned}$$

It can be shown along similar lines to the proof of theorem 8 that $\hat{\kappa}_s$ is an achievable T2EE for the GHTS problem.

We will next show that the SHTCC scheme when specialized to the case of testing against conditional independence recovers the result of Lemma 7. Let $\mathcal{B}' \triangleq \{P_{W|U} : I(U;W|Z) \leq R < \tau C\}$. Note that $\mathcal{B}' \subseteq \mathcal{B}$ since $I(U;W|V,Z) \leq I(U;W|Z)$, which holds due to the Markov chain relation $(Z,V) - U - W$. Now, consider $P_{W|U} \in \mathcal{B}'$. For such W, we have

$$\begin{aligned}
E_1(R,P_{W|U}) &= \min_{P_{\tilde{U}\tilde{V}\tilde{Z}\tilde{W}} \in \mathcal{T}_1(P_{UW},P_{VWZ})} D(P_{\tilde{U}\tilde{V}\tilde{Z}\tilde{W}}||Q_{UVZ}P_{W|U})\\
&= I(V;W|Z)\\
E_2(R,P_{W|U}) &= \infty\\
E_3(R,P_{W|U},\tau) &= \min_{P_{\tilde{U}\tilde{V}\tilde{Z}\tilde{W}} \in \mathcal{T}_3(P_{UW},P_{VZ})} D(P_{\tilde{U}\tilde{V}\tilde{Z}\tilde{W}}||Q_{UVZ}P_{W|U}) + I(V;W|Z) + \tau E_r\left(\frac{R}{\tau}\right)\\
&\geq I(V;W|Z) \qquad (46)\\
E_4(R,P_{W|U},\tau) &= \min_{P_{\tilde{U}\tilde{V}\tilde{Z}\tilde{W}} \in \mathcal{T}_4(P_{VZ})} D(P_{\tilde{U}\tilde{V}\tilde{Z}\tilde{W}}||Q_{UVZ}P_{W|U}) + I(V;W|Z) + E_s(R,\tau)\\
&\geq I(V;W|Z) \qquad (47)
\end{aligned}$$

where in (46) and (47), we used the non-negativity of $D(\cdot||\cdot)$, $E_r(\cdot)$ and $E_s(R,\tau)$. Hence, we obtain that

$$\begin{aligned}
\kappa(\tau,\epsilon) &\geq \sup_{P_{W|U} \in \mathcal{B}} \min\{E_1(R,P_{W|U}), E_2(R,P_{W|U}), E_3(R,P_{W|U},\tau), E_4(R,P_{W|U},\tau)\}\\
&\geq \sup_{P_{W|U} \in \mathcal{B}} I(V;W|Z)\\
&\geq \sup_{P_{W|U} \in \mathcal{B}'} I(V;W|Z) \qquad (48)\\
&= \sup_{P_{W|U} : I(W;U|Z) \leq \tau C} I(V;W|Z) \qquad (49)
\end{aligned}$$

where in (48), we used the fact that $\mathcal{B}' \subseteq \mathcal{B}$ and (49) follows since $I(V;W|Z)$ and $I(U;W|Z)$ are continuous functions of $P_{W|U}$.

V. Conclusions

In this paper, we studied the T2EE achievable 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 T2EE, 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 T2EE for TACI. We also established single-letter lower bounds for the GHT problem with a single observer using the SHTCC scheme that performs independent HT and channel coding.

Appendix A
T2EE using joint typicality detector

Here, we provide the proof for the case $L=1$. For given arbitrary positive 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 $\left(Y_1^{nj}, V^{kj}, Z^{kj}\right) \in T^j_{[\tilde{Y}_1^n \tilde{V}^k \tilde{Z}^k]_{\delta_j}}$ (here $\delta_j \to 0$ as $j \to \infty$) where $(\tilde{Y}_1^n, \tilde{U}_1^k, \tilde{V}^k, \tilde{Z}^k) \sim P_{Y_1^n|U_1^k} P_{U_1VZ}^{\bigotimes k}$ and $H_1 : Q_{U_1VZ}$ otherwise. To simplify the exposition, we denote (Y_1^n, V^k, Z^k) and $T^j_{[\tilde{Y}_1^n \tilde{V}^k \tilde{Z}^k]_{\delta_j}}$ by $W_{k,n}$ and $T^j_{[\tilde{W}_{k,n}]_{\delta_j}}$, respectively. By the Markov lemma [15], type 1 error probability tends to zero as $j \to \infty$. The type 2 error probability is bounded by

$$\begin{aligned}
&\beta'\left(kj, nj, f_1^{(kj,nj)}, \epsilon\right) \leq Q_{Y_1^{nj} V^{kj} Z^{kj}}\left(T^j_{[\tilde{W}_{k,n}]_{\delta_j}}\right) \\
&\leq \sum_{\tilde{P} \in T^j_{[\tilde{W}_{k,n}]_{\delta_j}}} \sum_{w^j_{k,n} \in T_{\tilde{P}}} Q_{W^j_{k,n}}(w^j_{k,n}) \\
&\overset{(a)}{=} \sum_{\tilde{P} \in T^j_{[\tilde{W}_{k,n}]_{\delta_j}}} \sum_{w^j_{k,n} \in T_{\tilde{P}}} 2^{-j\left(H(\tilde{P}) + D\left(\tilde{P}||Q_{W_{k,n}}\right)\right)} \\
&\overset{(b)}{=} \sum_{\tilde{P} \in T^j_{[\tilde{W}_{k,n}]_{\delta_j}}} 2^{-jD\left(\tilde{P}||Q_{W_{k,n}}\right)} \overset{(c)}{\leq} (j+1)^{|\mathcal{W}_{k,n}|} 2^{-jB_{k,n}(j)}
\end{aligned}$$

where

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

and (a), (b) and (c) follow from Lemmas 2.3, 2.6 and 2.2 in [14], 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 $|\mathcal{W}_{k,n}| = |\mathcal{Y}|^n |\mathcal{V}|^k |\mathcal{Z}|^k$. 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 (6) and (8) that $\theta(\tau)$ is an achievable T2EE 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 5

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, $l \in \mathcal{L}$. For $j \in \mathbb{Z}^+$, upon 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 $(\hat{V}^{kj},\ Y_{\mathcal{L}}^{nj}, Z^{kj}) \in T^j_{[V^k Y_{\mathcal{L}}^n Z^k]_\delta}$, then the decoder outputs $g^{(kj,nj)}(M, Y_{\mathcal{L}}^{nj}, Z^{kj}) = \hat{V}^{kj}$. Else, an error is declared.

Analysis of the probability of error: The possible error events under the above encoding and decoding rules are: $\mathcal{E}_1 = \{(V^{kj}, Y_{\mathcal{L}}^{nj}, Z^{kj}) \notin T^j_{[V^k Y_{\mathcal{L}}^n, Z^k]_\delta}\}$ and

$$\mathcal{E}_2 = \begin{Bmatrix} \exists\ \tilde{V}^{kj} \neq V^{kj},\ f_{L+1}^{kj}(\tilde{V}^{kj}) = f_{L+1}^{kj}(V^{kj}) \\ (\tilde{V}^{kj}, Y_{\mathcal{L}}^{nj}, Z^{kj}) \in T^j_{[V^k Y_{\mathcal{L}}^n, Z^k]_\delta} \end{Bmatrix}$$

By the joint typicality lemma [15], $Pr(\mathcal{E}_1) \to 0$ as $j \to \infty$. Also,

$$\begin{aligned}
\Pr(\mathcal{E}_2) &= \sum_{v^{kj}, y_{\mathcal{L}}^{nj}, z^{kj}} \Pr(v^{kj}, y_{\mathcal{L}}^{nj}, z^{kj}) \times \Pr\left(f_{L+1}^{kj}(\tilde{V}^{kj}) = f_{L+1}^{kj}(v^{kj}),\ (\tilde{V}^{kj}, y_{\mathcal{L}}^{nj}, z^{kj}) \in T^j_{[V^k Y^n_{\mathcal{L}} Z^k]_\delta}\right) \\
&= \sum_{v^{kj}, y_{\mathcal{L}}^{nj}, z^{kj}} \Pr(v^{kj}, y_{\mathcal{L}}^{nj}, z^{kj}) \sum_{v^{kj} \in T^j_{[V^k Y^n_{\mathcal{L}} Z^k]_\delta}} e^{-kjR} \\
&\leq e^{j(H(V^k|Y^n_{\mathcal{L}}, Z^k)+\delta)} e^{-kjR} \\
&= e^{kj\left(\frac{H(V^k|Y^n_{\mathcal{L}}, Z^k)+\delta}{k} - R\right)}.
\end{aligned}$$

Hence, $Pr(\mathcal{E}_2) \to 0$ as $j \to \infty$ if $R > H(V^k|Y^n_{\mathcal{L}}, Z^k) + \delta$, $(Z^k, V^k) - U^k_l - X^n_l - Y^n_l,\ l \in \mathcal{L}$. Since $\delta > 0$ is arbitrary, this proves that $R > \frac{H(V^k|Y^n_{\mathcal{L}}, Z^k)}{k}$ is an achievable rate.

For the converse, we have by Fano's inequality that $H(V^k|f^k_{L+1}(V^k), Y^n_{\mathcal{L}}, Z^k) \leq \gamma_k$, where $\gamma_k \to 0$ as $k \to \infty$. Hence, we obtain

$$\begin{aligned}
kR = \log(|\mathcal{M}|) &\geq H(M|Y^n_{\mathcal{L}}, Z^k) \\
&= H(M|Y^n_{\mathcal{L}}, Z^k) - H(V^k|M, Y^n_{\mathcal{L}},, Z^k) \\
&\quad + H(V^k|M, Y^n_{\mathcal{L}}, Z^k) \\
&\geq H(V^k, M|Y^n_{\mathcal{L}}, Z^k) - \gamma_k \\
&= H(V^k|Y^n_{\mathcal{L}}, Z^k) + H(M|V^k, Y^n_{\mathcal{L}}, Z^k) - \gamma_k \\
&= H(V^k|Y^n_{\mathcal{L}}, Z^k) - \gamma_k
\end{aligned}$$

This proves the converse by noting that $(Z^k, V^k) - U^k_l - X^n_l - Y^n_l,\ l \in \mathcal{L}$ holds for any communication scheme.

APPENDIX C
PROOF OF THEOREM 8

We first describe the encoding and decoding operations in the SHTCC scheme. The random coding method is used to analyze the type 1 and type 2 error probabilities achieved by this scheme, averaged over the ensemble of randomly generated codebooks. This guarantees the existence of at least one deterministic codebook that achieves same or lower type 1 and type 2 error probabilities.

Codebook Generation:
Fix $0 \leq R < \tau C$, conditional distribution $P_{W|U}$ and positive numbers $\delta',\ \delta'' > 0$ with $\delta' > \delta''$. Generate [3] $2^{k(I(U;W)+\delta')}$ independent sequences $W^k(j),\ j \in \left[2^{k(I(U;W)+\delta')}\right]$ randomly according to the distribution $\prod_{i=1}^k P_W(w_i)$ where

$$P_W = \sum_{u \in \mathcal{U}} \sum_{w \in \mathcal{W}} P_U(u) P_{W|U}(w|u).$$

Denote this codebook by $\mathcal{C}$ which is referred to as the *source codebook*. Next, the *channel codebook* used by $f_C^{(k,n)}$ is obtained by independently generating $\lceil 2^{kR} \rceil$ codewords $X^n(m),\ m \in \left[1 : \lceil 2^{kR} \rceil\right]$ i.i.d. according to the distribution $\prod_{i=1}^n P_X(x_i)$, which achieves the random coding exponent $E_r(\frac{R}{\tau})$ of the channel $P_{Y|X}$ [17]. The codeword for $X^n(0)$ is chosen such that it achieves an error probability with exponent $E_s(R, \tau)$ at the channel decoder, when $M = 0$. Denote this collection of codewords by $\tilde{\mathcal{C}}$.

Encoding: If $I(U;W) + \delta' > R$, i.e., the number of codewords in the source codebook is larger than the number of codewords in the channel codebook, then the encoder $f^{(k,n)}$ performs uniform random binning on the sequences $W^k(j),\ j \in \left[2^{k(I(U;W)+\delta')}\right]$ in $\mathcal{C}$, i.e., for each codeword in $\mathcal{C}$, it selects an index uniformly at random from the set $[2^{kR}]$. Denote the bin index selected for $W^k(j)$ by $f_B(j)$. If the observed sequence U^k is typical, i.e., $U^k \in T^k_{[U]_{\delta''}}$, the source encoder $f_S^{(k)}$ first looks for a sequence $W^k(J)$ such that $(U^k, W^k(J)) \in T^k_{[UW]_\delta}$, $\delta > \delta''$. If there exists multiple such codewords, it chooses one of the index J among them uniformly at random and outputs the bin-index $M = f_B(J)$, $M \in [1 : 2^{kR}]$ or $M = J$ depending on whether $I(U;W) + \delta' > R$ or otherwise. If $U^k \notin T^k_{[U]_{\delta''}}$ or such an index J does not exist, $f_s^{(k)}$ outputs $M = 0$, which is also referred to as the *error* message. The channel encoder $f_C^{(k,n)}$ outputs the codeword $X^n(M)$ from the codebook $\tilde{\mathcal{C}}$.

[3] We specifically mention that unless specified otherwise, the mutual information and entropy terms appearing in the proof, like for example, $I(U;W|V)$, $H(W|V)$ etc. are computed with respect to the joint distribution $P_{UV}P_{W|U}$.

Decoding: At the decoder, $g_C^{(k,n)}$ maps the received channel output Y^n to an estimate $\hat{M}$ of the transmitted message M. If $\hat{M}=0$, H_1 is declared. Else, given the side information sequence V^k and estimated bin-index $\hat{M}$, $g_S^{(k,n)}$ searches for a sequence $\hat{W}^k = W^k(\hat{l})$ in the codebook such that

$$\hat{l} = \underset{l:\ \hat{M}=f_B(l)}{\arg\min}\ H_e(W^k(l)|V^k),$$

where $H_e(W^k(l)|V^k) \triangleq H(P_{W^k(l)}|P_{V^k})$ is the conditional empirical entropy of the type $P_{W^k(l)}$ given type P_{V^k} [14]. The decoder declares H_0 if $(\hat{W}^k, V^k) \in T^k_{[WV]_\delta}$, else it declares H_1.

We next analyze the Type 1 and Type 2 error probabilities achieved by the above scheme.

Analysis of Type 1 error: The following events can possibly cause Type 1 error.

$$\begin{aligned}
\mathcal{E}_{CE} &= \left\{ g_C^{(k,n)}(Y^n) \neq X^n(M) \right\}\\
\mathcal{E}_{TE} &= \left\{ (U^k, V^k) \notin T^k_{[UV]_\delta} \right\}\\
\mathcal{E}_{EE} &= \left\{ \nexists\ j \in \left[1 : 2^{k(I(U;W)+\delta')}\right] : (U^k, W^k(j)) \in T^k_{[UW]_\delta} \right\}\\
\mathcal{E}_{DE} &= \Bigg\{ \exists\ l \in \left[1 : 2^{k(I(U;W)+\delta')}\right],\ l \neq J : f_B(l) = f_B(J)\\
&\qquad\qquad \text{and } H_e(W^k(l)|V^k) \leq H_e(W^k(J)|V^k) \Bigg\}
\end{aligned}$$

Note that a Type 1 error happens only if one of the events among $\mathcal{E}_{CE}$, $\mathcal{E}_{TE}$, $\mathcal{E}_{EE}$ or $\mathcal{E}_{DE}$ happens. The probability of the event $\mathcal{E}_{CE}$, that an error occurs at the channel decoder $g_C^{(k,n)}$ tends to 0 as $n \to \infty$ since $E_r(\frac{R}{\tau})$ is positive for $R < \tau C$. $\mathcal{E}_{TE}$ tends to 0 asymptotically by the weak law of large numbers. By the covering lemma [14, Lemma 9.1], it is well known that $\mathcal{E}_{EE} \cap \mathcal{E}^c_{TE}$ tends to 0 doubly exponentially for any $\delta' > 0$. Next, we consider the probability of the event $\mathcal{E}_{DE}$. Given that $\mathcal{E}^c_{EE} \cap \mathcal{E}^c_{TE}$ holds, it follows from the Markov chain relation $V - U - W$ and the Markov lemma [15] that $(V^k, W^k(J)) \in T^k_{[VW]_{\tilde{\delta}}}$ for $\tilde{\delta} > \delta$ and hence the empirical entropy tends to $H_e(W^k(J)|V^k) \to H(W|V)$ as $k \to \infty$ and $\delta \to 0$.

$$\begin{aligned}
&\mathbb{P}(\mathcal{E}_{DE}|\mathcal{E}^c_{EE} \cap \mathcal{E}^c_{TE}, W^k(J)=w^k, V^k=v^k)\\
&\leq \sum_{\substack{l=1,\\ l\neq J}}^{2^{k(I(U;W)+\delta')}} \mathbb{P}\left(f_B(l)=f_B(J),\ H_e(W^k(l)|v^k) \leq H_e(w^k|v^k)|\ \mathcal{E}^c_{EE} \cap \mathcal{E}^c_{TE}, W^k(J)=w^k, V^k=v^k\right)\\
&= \sum_{\substack{l=1,\\ l\neq J}}^{2^{k(I(U;W)+\delta')}} \mathbb{P}\left(H_e(W^k(l)|V^k) \leq H_e(w^k|v^k)|\ \mathcal{E}^c_{EE} \cap \mathcal{E}^c_{TE}, W^k(J)=w^k, V^k=v^k\right)\ \frac{1}{2^{kR}}\\
&= \sum_{\substack{l=1,\\ l\neq J}}^{2^{k(I(U;W)+\delta')}} \sum_{\substack{w^k \in T^k_{[W]_\delta}:\\ H_e(w^k|v^k) \leq H(W|V)}} \mathbb{P}(W^k(l)=w^k|\ \mathcal{E}^c_{EE} \cap \mathcal{E}^c_{TE}, W^k(J)=w^k, V^k=v^k)\ \frac{1}{2^{kR}}\\
&\leq \sum_{\substack{l=1,\\ l\neq J}}^{2^{k(I(U;W)+\delta')}} \sum_{\substack{w^k \in T^k_{[W]_\delta}:\\ H_e(w^k|v^k) \leq H(W|V)}} 2 \times 2^{-kR} 2^{-k(H(W)-\delta)} \qquad (50)\\
&\leq \sum_{\substack{l=1,\\ l\neq J}}^{2^{k(I(U;W)+\delta')}} (k+1)^{|\mathcal{V}||\mathcal{W}|}\ 2^{kH_e(w^k|v^k)} \times 2 \times 2^{-kR} 2^{-k(H(W)-\delta)} \qquad (51)\\
&\leq 2^{-k(R-I(U;W|V)-\delta_2)}
\end{aligned}$$

where $\delta_2 \to 0$ as $k \to \infty$ and $\delta \to 0$. To obtain (50), we used the fact that

$$\mathbb{P}(W^k(l)=w^k|\ \mathcal{E}^c_{EE} \cap \mathcal{E}^c_{TE}, W^k(J)=w^k, V^k=v^k) \leq 2 \times \mathbb{P}(W^k(l)=w^k|\ \mathcal{E}^c_{EE} \cap \mathcal{E}^c_{TE}, V^k=v^k) \qquad (52)$$

This follows similar to (64), which is discussed in the type 2 error analysis subsection given below. In order to obtain the expression in (51), we first summed over the types $\tilde{P}_W$ of sequences within the typical set $T^k_{[W]_\delta}$ that have empirical entropy less than $H(W|V)$ and used the facts that the number of sequences within such a type is upper bounded by $2^{k(H(W|V))}$ and the total number of types is upper bounded by $(k+1)^{|\mathcal{V}||\mathcal{W}|}$ [14]. Summing over all $(w^k, v^k) \in T^k_{[VW]_{\tilde{\delta}}}$, we obtain

$$\begin{aligned}
\mathbb{P}(\mathcal{E}_{DE}|\mathcal{E}^c_{EE} \cap \mathcal{E}^c_{TE}) &\leq \sum_{(w^k,v^k)\in T^k_{[WV]_{\tilde{\delta}}}} \mathbb{P}(W^k(J) = w^k, V^k = v^k|\mathcal{E}^c_{EE} \cap \mathcal{E}^c_{TE})\ 2^{-k(R-I(U;W|V)-\delta_2)} \\
&\leq 2^{-k(R-I(U;W|V)-\delta_2)} \qquad (53)
\end{aligned}$$

Hence, if $I(U;W|V) < R < \tau C$, the probability of the events causing Type 1 error tends to zero asymptotically.

Analysis of Type 2 error: Denote the event that a type 2 error happens by $\mathcal{D}_0$. The type 2 error probability can be written as

$$\begin{aligned}
\beta(k,\tau,\epsilon) &= \mathbb{P}(\mathcal{E}_{EE} \cap \mathcal{E}^c_{TE})\ \mathbb{P}(\mathcal{D}_0|\mathcal{E}_{EE} \cap \mathcal{E}^c_{TE}) + \mathbb{P}(\mathcal{E}^c_{EE} \cap \mathcal{E}^c_{TE})\ \mathbb{P}(\mathcal{D}_0|\mathcal{E}^c_{EE} \cap \mathcal{E}^c_{TE}) + \mathbb{P}(\mathcal{E}_{TE})\ \mathbb{P}(\mathcal{D}_0|\mathcal{E}_{TE}) \\
&\leq \mathbb{P}(\mathcal{E}_{EE} \cap \mathcal{E}^c_{TE}) + \mathbb{P}(\mathcal{D}_0|\mathcal{E}^c_{EE} \cap \mathcal{E}^c_{TE}) + \mathbb{P}(\mathcal{D}_0|\mathcal{E}_{TE}) \qquad (54)
\end{aligned}$$

As mentioned above, $\mathbb{P}(\mathcal{E}_{EE} \cap \mathcal{E}^c_{TE})$ decreases doubly exponential with k where the probability is averaged over the ensemble of randomly generated codebooks, provided δ' and δ are chosen appropriately. Note that since $P_{W|U}$ is the same under H_0 and H_1, the double exponential decay holds under both these hypotheses. Hence, the contribution of the event $\mathcal{E}_{EE} \cap \mathcal{E}^c_{TE}$ to the type 2 error can be safely ignored from the T2EE analysis as the factor in the T2EE due to this event will eventually become larger than the other factors (as will become evident later). Henceforth, we analyze the type 2 error focusing on the events $\mathcal{E}_{TE}$ or $\mathcal{E}_{NE} \triangleq \mathcal{E}^c_{EE} \cap \mathcal{E}^c_{TE}$.

First, assume that $\mathcal{E}_{NE}$ holds. Then, $\mathcal{D}_0$ may occur in two possible ways. The first case is when the channel decoder makes an error and the sequence retrieved from the bin is jointly typical with V^k and the other case is when an unintended wrong sequence is retrieved from the correct bin that is jointly typical with V^k. We refer to these events as the *channel error* and *binning error* respectively and denote it by $\mathcal{E}_{CE}$ and $\mathcal{E}_{BE}$ respectively. More specifically,

$$\begin{aligned}
\mathbb{P}(\mathcal{D}_0|\ \mathcal{E}_{NE}) = &\sum_{\substack{(u^k,v^k)\\ \in\ \mathcal{U}^k\times\mathcal{V}^k}} \sum_{j=1}^{2^{k(I(U;W)+\delta')}} \sum_{m=1}^{2^{nR}} \mathbb{P}(U^k = u^k, V^k = v^k, J = j, f_B(J) = m|\ \mathcal{E}_{NE}) \\
&\mathbb{P}(\mathcal{D}_0|U^k = u^k, V^k = v^k, J = j, f_B(J) = m,\ \mathcal{E}_{NE}) \qquad (55)
\end{aligned}$$

By the symmetry of the codebook generation, encoding and decoding procedure, the term $\mathbb{P}(\mathcal{D}_0|U^k = u^k, V^k = v^k, J = j, f_B(J) = m,\ \mathcal{E}_{NE})$ in (55) is independent of the value of J and $f_B(J)$. Hence, w.l.o.g. assuming $J = 1$ and $f_B(J) = 1$, we can write

$$\begin{aligned}
\mathbb{P}(\mathcal{D}_0|\ \mathcal{E}_{NE}) = &\sum_{\substack{(u^k,v^k)\\ \in\ \mathcal{U}^k\times\mathcal{V}^k}} \sum_{j=1}^{2^{k(I(U;W)+\delta')}} \sum_{m=1}^{2^{nR}} \mathbb{P}(U^k = u^k, V^k = v^k, J = j, f_B(J) = m|\ \mathcal{E}_{NE}) \\
&\qquad \mathbb{P}(\mathcal{D}_0|U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1,\ \mathcal{E}_{NE}) \\
= &\ \mathbb{P}(\mathcal{D}_0|J = 1, f_B(J) = 1,\ \mathcal{E}_{NE}) \\
= &\sum_{\substack{(u^k,v^k,w^k)\\ \in\ \mathcal{U}^k\times\mathcal{V}^k\times\mathcal{W}^k}} \mathbb{P}(U^k = u^k, V^k = v^k, W^k(1) = w^k|J = 1, f_B(J) = 1,\ \mathcal{E}_{NE}) \\
&\qquad \mathbb{P}(\mathcal{D}_0|U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, W^k(1) = w^k,\ \mathcal{E}_{NE}) \qquad (56)
\end{aligned}$$

Define the events

$$\begin{aligned}
\mathcal{F} &= \{U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, W^k(1) = w^k,\ \mathcal{E}_{NE}\} \\
\mathcal{F}_1 &= \{U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, W^k(1) = w^k,\ \mathcal{E}_{NE},\ \mathcal{E}_{CE}\} \\
\mathcal{F}_2 &= \{U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, W^k(1) = w^k,\ \mathcal{E}_{NE},\ \mathcal{E}^c_{CE}\} \\
\mathcal{F}_{21} &= \{U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, W^k(1) = w^k,\ \mathcal{E}_{NE},\ \mathcal{E}^c_{CE},\ \mathcal{E}_{BE}\} \\
\mathcal{F}_{22} &= \{U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, W^k(1) = w^k,\ \mathcal{E}_{NE},\ \mathcal{E}^c_{CE},\ \mathcal{E}^c_{BE}\}
\end{aligned}$$

The last term in (56) can be expressed as follows.

$$\mathbb{P}(\mathcal{D}_0|\mathcal{F}) = \mathbb{P}(\mathcal{E}_{CE}|\mathcal{F})\ \mathbb{P}(\mathcal{D}_0|\mathcal{F}_1) + \mathbb{P}(\mathcal{E}_{CE}^c|\mathcal{F})\ \mathbb{P}(\mathcal{D}_0|\mathcal{F}_2)$$

where

$$\mathbb{P}(\mathcal{D}_0|\mathcal{F}_2) = \mathbb{P}(\mathcal{E}_{BE}|\mathcal{F}_2)\ \mathbb{P}(\mathcal{D}_0|\mathcal{F}_{21}) + \mathbb{P}(\mathcal{E}_{BE}^c|\mathcal{F}_2)\ \mathbb{P}(\mathcal{D}_0|\mathcal{F}_{22}). \tag{57}$$

Since the channel encoder and decoder uses randomly generated codewords achieving the best random coding error exponent $E_r(\cdot)$, we have

$$\mathbb{P}(\mathcal{E}_{CE}|\mathcal{F}) = 2^{-nE_r\left(\frac{R}{\tau}\right)} = 2^{-k\tau E_r\left(\frac{R}{\tau}\right)} \tag{58}$$

Next, consider the type 2 error event which happens when H_0 is declared in spite of an error at the channel decoder. We need to consider two separate cases (i) $I(U;W) > R$ and $I(U;W) \leq R$. Note that in the former case, binning is performed and type 2 error happens at the decoder only if a sequence $W^k(l)$ exists in the wrong bin $\hat{M} \neq M = f_B(J)$ such that $(V^k, W^k(l)) \in T^k_{[VW]_\delta}$. However as noted in [18], the calculation of probability of this event does not follow using the standard random coding argument usually encountered in achievability proofs due to the fact that the codeword $W^k(J)$ chosen depends on the entire codebook. Following techniques similar to [18], we analyze the probability of this event (averaged over the codebooks $\mathcal{C}$, $\tilde{\mathcal{C}}$ and random binning) as follows. We first consider the case when $I(U;W) > R$.

$$\begin{aligned}
&\mathbb{P}(\mathcal{D}_0|\mathcal{F}_1) \leq \mathbb{P}(\ \exists\ W^k(l):\ f_B(l) = \hat{M} \neq 1,\ (W^k(l), v^k) \in T^k_{[WV]_\delta}|\mathcal{F}_1) \\
&\leq \sum_{l=2}^{2^{k(I(U:W)+\delta')}} \sum_{\hat{m}\neq 1} \mathbb{P}(\hat{M} = \hat{m}|\mathcal{F}_1)\ \mathbb{P}((W^k(l), v^k) \in T^k_{[WV]_\delta}:\ f_B(l) = \hat{m}|\mathcal{F}_1) \qquad (59)\\
&= \sum_{l=2}^{2^{k(I(U:W)+\delta')}} \sum_{\hat{m}\neq 1} \mathbb{P}(\hat{M} = \hat{m}|\mathcal{F}_1) \sum_{\tilde{w}^k \in T^k_{[W|V]_\delta}} \mathbb{P}(W^k(l) = \tilde{w}^k:\ f_B(l) = \hat{m}|\mathcal{F}_1) \\
&= \sum_{l=2}^{2^{k(I(U:W)+\delta')}} \sum_{\hat{m}\neq 1} \mathbb{P}(\hat{M} = \hat{m}|\mathcal{F}_1) \sum_{\tilde{w}^k \in T^k_{[W|V]_\delta}} \mathbb{P}(W^k(l) = \tilde{w}^k|\mathcal{F}_1)\ \frac{1}{2^{kR}} \qquad (60)
\end{aligned}$$

$$= \sum_{l=2}^{2^{k(I(U:W)+\delta')}} \sum_{\tilde{w}^k \in T^k_{[W|V]_\delta}} \mathbb{P}(W^k(l) = \tilde{w}^k|\mathcal{F}_1)\ \frac{1}{2^{kR}} \tag{61}$$

Let $\mathcal{C}_l^- = \mathcal{C}\backslash\{W^k(1), W^k(l)\}$. Then,

$$\mathbb{P}(W^k(l) = \tilde{w}^k|\mathcal{F}_1) = \sum_{\mathcal{C}_l^- = c} \mathbb{P}(\mathcal{C}_l^- = c)\mathbb{P}(W^k(l) = \tilde{w}^k|\mathcal{F}_1, \mathcal{C}_l^- = c) \tag{62}$$

The last term in (62) can be upper bounded as follows.

$$\begin{aligned}
&\mathbb{P}(W^k(l) = \tilde{w}^k|\mathcal{F}_1,\ \mathcal{C}_l^- = c) \\
&= \frac{\mathbb{P}(W^k(l) = \tilde{w}^k, U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, W^k(1) = w^k,\ \mathcal{E}_{NE},\ \mathcal{E}_{CE},\ \mathcal{C}_l^- = c)}{\mathbb{P}(U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, W^k(1) = w^k,\ \mathcal{E}_{NE},\ \mathcal{E}_{CE},\ \mathcal{C}_l^- = c)} \\
&= \mathbb{P}(W^k(l) = \tilde{w}^k|U^k = u^k, V^k = v^k,\ \mathcal{E}_{NE},\ \mathcal{E}_{CE}, \mathcal{C}_l^- = c) \\
&\quad \frac{\mathbb{P}(J = 1, f_B(J) = 1, W^k(1) = w^k|W^k(l) = \tilde{w}^k, U^k = u^k, V^k = v^k,\ \mathcal{E}_{NE},\ \mathcal{E}_{CE},\ \mathcal{C}_l^- = c)}{\mathbb{P}(J = 1, f_B(J) = 1, W^k(1) = w^k|U^k = u^k, V^k = v^k,\ \mathcal{E}_{NE},\ \mathcal{E}_{CE},\ \mathcal{C}_l^- = c)} \\
&= \mathbb{P}(W^k(l) = \tilde{w}^k|U^k = u^k, V^k = v^k,\ \mathcal{E}_{NE},\ \mathcal{E}_{CE},\ \mathcal{C}_l^- = c) \\
&\quad \frac{\mathbb{P}(J = 1, f_B(J) = 1, W^k(1) = w^k|W^k(l) = \tilde{w}^k, U^k = u^k, V^k = v^k,\ \mathcal{E}_{NE},\ \mathcal{E}_{CE},\ \mathcal{C}_l^- = c)}{\mathbb{P}(J = 1, f_B(J) = 1, W^k(1) = w^k|U^k = u^k, V^k = v^k,\ \mathcal{E}_{NE},\ \mathcal{E}_{CE},\ \mathcal{C}_l^- = c)}
\end{aligned}$$

$$= \mathbb{P}(W^k(l) = \tilde{w}^k | U^k = u^k, V^k = v^k,\ \mathcal{E}_{NE},\ \mathcal{E}_{CE},\ \mathcal{C}_l^- = c)$$
$$\frac{\mathbb{P}(W^k(1) = w^k | W^k(l) = \tilde{w}^k, U^k = u^k, V^k = v^k,\ \mathcal{E}_{NE},\ \mathcal{E}_{CE},\ \mathcal{C}_l^- = c)}{\mathbb{P}(W^k(1) = w^k | U^k = u^k, V^k = v^k,\ \mathcal{E}_{NE},\ \mathcal{E}_{CE},\ \mathcal{C}_l^- = c)}$$
$$\frac{\mathbb{P}(J = 1 | W^k(1) = w^k, W^k(l) = \tilde{w}^k, U^k = u^k, V^k = v^k,\ \mathcal{E}_{NE},\ \mathcal{E}_{CE},\ \mathcal{C}_l^- = c)}{\mathbb{P}(J = 1 | W^k(1) = w^k, U^k = u^k, V^k = v^k,\ \mathcal{E}_{NE},\ \mathcal{E}_{CE},\ \mathcal{C}_l^- = c)} \tag{63}$$
$$\frac{\mathbb{P}(f_B(J) = 1 | J = 1, W^k(1) = w^k, W^k(l) = \tilde{w}^k, U^k = u^k, V^k = v^k,\ \mathcal{E}_{NE},\ \mathcal{E}_{CE},\ \mathcal{C}_l^- = c)}{\mathbb{P}(f_B(J) = 1 | J = 1, W^k(1) = w^k, U^k = u^k, V^k = v^k,\ \mathcal{E}_{NE},\ \mathcal{E}_{CE},\ \mathcal{C}_l^- = c)}$$

Since the codewords are generated independently of each other and the binning operation is done independent of the codebook generation, we have

$$\mathbb{P}(W^k(1) = w^k | W^k(l) = \tilde{w}^k, U^k = u^k, V^k = v^k, \mathcal{E}_{NE},\ \mathcal{E}_{CE},\ \mathcal{C}_l^- = c)$$
$$= \mathbb{P}(W^k(1) = w^k | U^k = u^k, V^k = v^k, \mathcal{E}_{NE},\ \mathcal{E}_{CE},\ \mathcal{C}_l^- = c)$$

and

$$\mathbb{P}(f_B(J) = 1 | J = 1, W^k(1) = w^k, W^k(l) = \tilde{w}^k, U^k = u^k, V^k = v^k, \mathcal{E}_{NE},\ \mathcal{E}_{CE},\ \mathcal{C}_l^- = c)$$
$$= \mathbb{P}(f_B(J) = 1 | J = 1, W^k(1) = w^k, U^k = u^k, V^k = v^k, \mathcal{E}_{NE},\ \mathcal{E}_{CE},\ \mathcal{C}_l^- = c)$$

Consider the term in (63). Let $N(u^k, \mathcal{C}_l^-) = |\{w^k(l') \in \mathcal{C}_l^- : l' \neq 1,\ l' \neq l,\ (w^k(l'), u^k) \in T^k_{[WU]_\delta}\}|$. Recall that if there are multiple sequences in the codebook $\mathcal{C}$ that are typical with the observed sequence U^k, then the encoder selects one of them uniformly at random. Thus if $(\tilde{w}^k, u^k) \in T^k_{[WU]_\delta}$, then

$$\frac{\mathbb{P}(J = 1 | W^k(1) = w^k, W^k(l) = \tilde{w}^k, U^k = u^k, V^k = v^k, \mathcal{E}_{NE},\ \mathcal{E}_{CE},\ \mathcal{C}^- = c)}{\mathbb{P}(J = 1 | W^k(1) = w^k, U^k = u^k, V^k = v^k, \mathcal{E}_{NE},\ \mathcal{E}_{CE},\ \mathcal{C}^- = c)}$$
$$= \left[\frac{1}{N(u^k, \mathcal{C}^-) + 2}\right] \frac{1}{\mathbb{P}(J = 1 | W^k(1) = w^k, U^k = u^k, V^k = v^k, \mathcal{E}_{NE},\ \mathcal{E}_{CE},\ \mathcal{C}^- = c)}$$
$$\leq \frac{N(u^k, \mathcal{C}^-) + 1}{N(u^k, \mathcal{C}^-) + 2} \leq 1$$

If $(\tilde{w}^k, u^k) \notin T^k_{[WU]_\delta}$, then

$$\frac{\mathbb{P}(J = 1 | W^k(1) = w^k, W^k(l) = \tilde{w}^k, U^k = u^k, V^k = v^k, \mathcal{E}_{NE},\ \mathcal{E}_{CE},\ \mathcal{C}^- = c)}{\mathbb{P}(J = 1 | W^k(1) = w^k, U^k = u^k, V^k = v^k, \mathcal{E}_{NE},\ \mathcal{E}_{CE},\ \mathcal{C}^- = c)}$$
$$= \left[\frac{1}{N(u^k, \mathcal{C}^-) + 1}\right] \frac{1}{\mathbb{P}(J = 1 | W^k(1) = w^k, U^k = u^k, V^k = v^k, \mathcal{E}_{NE},\ \mathcal{E}_{CE},\ \mathcal{C}^- = c)}$$
$$\leq \frac{N(u^k, \mathcal{C}^-) + 2}{N(u^k, \mathcal{C}^-) + 1} \leq 2$$

Hence, the term in (62) can be upper bounded as

$$\mathbb{P}(W^k(l) = \tilde{w}^k | \mathcal{F}_1)$$
$$\leq \sum_{\mathcal{C}^- = c} \mathbb{P}(\mathcal{C}^- = c)\ 2\ \mathbb{P}(W^k(l) = \tilde{w}^k | U^k = u^k, V^k = v^k, \mathcal{E}_{NE},\ \mathcal{E}_{CE},\ \mathcal{C}^- = c)$$
$$= 2\ \mathbb{P}(W^k(l) = \tilde{w}^k | U^k = u^k, V^k = v^k, \mathcal{E}_{NE},\ \mathcal{E}_{CE}) \tag{64}$$

Substituting (64) in (61), we obtain

$$\mathbb{P}(\mathcal{D}_0 | \mathcal{F}_1)$$
$$\leq 2 \sum_{l=1}^{2^{k(I(U:W)+\delta')}} \sum_{\tilde{w}^k \in T^k_{[W|V]_\delta}} \mathbb{P}(W^k(l) = \tilde{w}^k | U^k = u^k, V^k = v^k, \mathcal{E}_{NE},\ \mathcal{E}_{CE})\ \frac{1}{2^{kR}}$$
$$= 2 \sum_{l=1}^{2^{k(I(U:W)+\delta')}} \sum_{\tilde{w}^k \in T^k_{[W|V]_\delta}} 2^{-k(H(W)-\delta_1)}\ \frac{1}{2^{kR}}$$

$$
\begin{aligned}
&= 2\ 2^{k(I(U:W)+\delta')}\ 2^{k(H(W|V)+\delta_2)}\ 2^{-k(H(W)-\delta_1)}\ \frac{1}{2^{kR}} \\
&\leq 2^{-k(R-I(U;W|V)-\delta_3^{(k)})}
\end{aligned} \tag{65}
$$

where $\delta_3^{(k)} \triangleq \delta' + \delta_1 + \delta_2 + \frac{1}{k} \xrightarrow{k} 0$ as $\delta,\ \delta' \to 0$.

For the case $I(U;W) \leq R$ (when binning is not done), the terms can be bounded similarly using (64) as follows.

$$
\begin{aligned}
&\mathbb{P}(\mathcal{D}_0|\mathcal{F}_1) \\
&= \sum_{\hat{m}\neq 1} \mathbb{P}(\hat{M}=\hat{m}|\mathcal{F}_1)\ \mathbb{P}((W^k(\hat{m}), v^k) \in T^k_{[WV]_\delta}|\mathcal{F}_1) \\
&\leq \sum_{\hat{m}\neq 1} \mathbb{P}(\hat{M}=\hat{m}|\mathcal{F}_1) \sum_{\tilde{w}^k \in T^k_{[W|V]_\delta}} 2\ \mathbb{P}(W^k(\hat{m}) = \tilde{w}^k | U^k=u^k, V^k=v^k, \mathcal{E}_{NE}, \mathcal{E}_{CE}) \\
&\leq 2 \sum_{\hat{m}\neq 1} \mathbb{P}(\hat{M}=\hat{m}|\mathcal{F}_1) \sum_{\tilde{w}^k \in T^k_{[W|V]_\delta}} 2^{-k(H(W)-\delta_1)} \\
&\leq \sum_{\hat{m}\neq 1} \mathbb{P}(\hat{M}=\hat{m}|\mathcal{F}_1)\ 2^{-k(I(V;W)-\delta_4)} \\
&= 2^{-k(I(V;W)-\delta_4^{(k)})}
\end{aligned} \tag{66}
$$

where $\delta_4^{(k)} \triangleq \delta_1 + \delta_2 + \frac{1}{k} \xrightarrow{k} 0$ as $\delta \to 0$.

Next, consider the event when the encoding and channel error does not happen, i.e., $\mathcal{E}_{NE} \cap \mathcal{E}^c_{CE}$ holds. For the case $I(U;W) > R$, the binning error event denoted by $\mathcal{E}_{BE}$ happens when a wrong codeword $W^k(l),\ l \neq J$ is retrieved from the bin with index M by the empirical entropy decoder such that $(W^k(l), V^k) \in T^k_{[WV]_\delta}$. Let $\tilde{U}, \tilde{V}, \tilde{W}$ denote the r.v.'s with marginal distributions equal to the types $P_{U^k},\ P_{V^k}$ and $P_{W^k(J)}$ respectively. Note that $P_{\tilde{U}\tilde{W}} \in \mathcal{T}^k_{[UW]_\delta}$ when $\mathcal{E}_{NE}$ holds. If $H(\tilde{W}|\tilde{V}) < H(W|V)$, there exists a codeword in the bin with index M having empirical entropy strictly less than $H(W|V)$. Hence, the decoded codeword $\hat{W}^k \notin T^k_{[WV]_\delta}$(asymptotically) since $(\hat{W}^k,\ V^k) \in T^k_{[WV]_\delta}$ necessarily implies that $H_e(\hat{W}^k|V^k) \triangleq H(P_{\hat{W}^k}|P_{V^k}) \to H(W|V)$ as $\delta \to 0$. Consequently, a Type 2 error can happen under the event $\mathcal{E}_{BE}$ only when $H(\tilde{W}|\tilde{V}) \geq H(W|V)$. The probability of the event $\mathcal{E}_{BE}$ can be upper bounded under this condition as follows.

$$
\begin{aligned}
&\mathbb{P}(\mathcal{E}_{BE}|\mathcal{F}_2) \\
&\leq \mathbb{P}\left(\exists\ l \neq 1,\ l \in [1:2^{k(I(U:W)+\delta')}]:\ f_B(l)=1 \text{ and } W^k(l) \in T^k_{[W|V]_\delta}|\mathcal{F}_2\right) \\
&\leq \sum_{l=2}^{2^{k(I(U;W)+\delta')}} \mathbb{P}\left(W^k(l) \in T^k_{[W|V]_\delta}|\mathcal{F}_2\right)\ \mathbb{P}\left(f_B(l)=1|\mathcal{F}_2, W^k(l) \in T^k_{[W|V]_\delta}\right) \\
&= \sum_{l=2}^{2^{k(I(U;W)+\delta')}} \mathbb{P}\left(W^k(l) \in T^k_{[W|V]_\delta}|\mathcal{F}_2\right)\ 2^{-kR} \\
&\leq \sum_{l=2}^{2^{k(I(U;W)+\delta')}} 2\ \mathbb{P}(W^k(l) = \tilde{w}^k|U^k=u^k, V^k=v^k, \mathcal{E}_{NE}, \mathcal{E}^c_{CE})\ 2^{-kR} && (67) \\
&= 2^{-k(R-I(U;W|V)-\delta_3^{(k)})} && (68)
\end{aligned}
$$

where $\delta_3^{(k)} \xrightarrow{(k)} 0$ as $\delta, \delta' \to 0$. In (67), we used the fact that

$$
\mathbb{P}\left(W^k(l) \in T^k_{[W|V]_\delta}|\mathcal{F}_2\right) \leq 2\ \mathbb{P}(W^k(l) = \tilde{w}^k|U^k=u^k, V^k=v^k, \mathcal{E}_{NE}, \mathcal{E}^c_{CE}) \tag{69}
$$

which follows in a similar way as (64). Also, note that, by definition, $\mathbb{P}(\mathcal{D}_0|\mathcal{F}_{21}) = 1$. Rewriting the summation in (55) as a sum over the types and sequences within a type, we obtain

$$
\mathbb{P}(\mathcal{D}_0|\ \mathcal{E}_{NE}) = \sum_{\substack{P_{\tilde{U}\tilde{V}\tilde{W}} \in \\ \mathcal{T}(\mathcal{U}\times\mathcal{V}\times\mathcal{W})}} \sum_{\substack{(u^k,v^k,w^k) \\ \in T_{P_{\tilde{U}\tilde{V}\tilde{W}}}}} \mathbb{P}(U^k=u^k, V^k=v^k, W^k(1)=w^k|J=1, f_B(J)=1, \mathcal{E}_{NE})\ \mathbb{P}(\mathcal{D}_0|\mathcal{F}) \tag{70}
$$

Noting that H_1 is the true underlying hypothesis for the T2EE analysis, we can write

$$\begin{aligned}
&\mathbb{P}(U^k = u^k, V^k = v^k, W^k(1) = w^k | J = 1, f_B(J) = 1, \mathcal{E}_{NE}) \\
&= Q^k_{UV}(u^k, v^k)\mathbb{P}(W^k(1) = w^k | U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, \mathcal{E}_{NE}) \\
&\leq Q^k_{UV}(u^k, v^k)\frac{1}{|T_{P_{\tilde{W}|\tilde{U}}}|} \\
&= Q^k_{UV}(u^k, v^k)2^{-kH(\tilde{W}|\tilde{U})} \\
&= 2^{-k(H(\tilde{U}\tilde{V})+D(P_{\tilde{U}\tilde{V}}||Q_{UV})+H(\tilde{W}|\tilde{U}))} \qquad (71)
\end{aligned}$$

where $P_{\tilde{U}\tilde{V}\tilde{W}}$ denotes the type of the sequence (u^k, v^k, w^k).

With (58), (65), (66), (68) and (71), we have the necessary machinery to analyze (70). First, consider that the event $\mathcal{E}_{NE} \cap \mathcal{E}^c_{CE} \cap \mathcal{E}^c_{BE}$ holds. In this case,

$$\begin{aligned}
\mathbb{P}(\mathcal{D}_0|\mathcal{F}_{22}) &= \mathbb{P}(\mathcal{D}_0|U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, W^k(1) = w^k, \mathcal{E}_{NE}, \mathcal{E}^c_{CE}, \mathcal{E}^c_{BE}) \\
&= \begin{cases} 1, & \text{if } P_{u^k w^k} \in \mathcal{T}^k_{[UW]_\delta} \\ & \text{and } P_{v^k w^k} \in\in \mathcal{T}^k_{[VW]_\delta}, \\ 0, & \text{otherwise.} \end{cases} \qquad (72)
\end{aligned}$$

Thus, the following terms in (70) can be simplified (in the limit $\delta \to 0$) as

$$\begin{aligned}
&\sum_{\substack{P_{\tilde{U}\tilde{V}\tilde{W}} \in \\ \mathcal{T}(\mathcal{U}\times\mathcal{V}\times\mathcal{W})}} \sum_{\substack{(u^k,v^k,w^k) \\ \in T_{P_{\tilde{U}\tilde{V}\tilde{W}}}}} \Big[\mathbb{P}(U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, W^k(1) = w^k | \mathcal{E}_{NE}) \ \mathbb{P}(\mathcal{E}^c_{CE}|\mathcal{F}) \ \mathbb{P}(\mathcal{E}^c_{BE}|\mathcal{F}_2) \ \mathbb{P}(\mathcal{D}_0|\mathcal{F}_{22})\Big] \\
&\leq \sum_{\substack{P_{\tilde{U}\tilde{V}\tilde{W}} \in \\ \mathcal{T}(\mathcal{U}\times\mathcal{V}\times\mathcal{W})}} \sum_{\substack{(u^k,v^k,w^k) \\ \in T_{P_{\tilde{U}\tilde{V}\tilde{W}}}}} \Big[\mathbb{P}(U^k = u^k, V^k = v^k, W^k(1) = w^k | J = 1, f_B(J) = 1, \mathcal{E}_{NE}) \ \mathbb{P}(\mathcal{E}^c_{CE}|\mathcal{F}) \ \mathbb{P}(\mathcal{E}^c_{BE}|\mathcal{F}_2) \ \mathbb{P}(\mathcal{D}_0|\mathcal{F}_{22})\Big] \\
&\leq \sum_{\substack{P_{\tilde{U}\tilde{V}\tilde{W}} \in \\ \mathcal{T}_1(P_{UW}, P_{VW})}} 2^{H(\tilde{U}\tilde{V}\tilde{W})} 2^{-k(H(\tilde{U}\tilde{V})+D(P_{\tilde{U}\tilde{V}}||Q_{UV})+H(\tilde{W}|\tilde{U}))} \qquad (73) \\
&\leq (k+1)^{|\mathcal{U}||\mathcal{V}||\mathcal{W}|} \max_{\substack{P_{\tilde{U}\tilde{V}\tilde{W}} \in \\ \mathcal{T}_1(P_{UW}, P_{VW})}} 2^{kH(\tilde{U}\tilde{V}\tilde{W})} 2^{-k(H(\tilde{U}\tilde{V})+D(P_{\tilde{U}\tilde{V}}||Q_{UV})+H(\tilde{W}|\tilde{U}))} = 2^{-k\tilde{E}} \qquad (74)
\end{aligned}$$

where

$$\begin{aligned}
\tilde{E} &= \min_{\substack{P_{\tilde{U}\tilde{V}\tilde{W}} \in \\ \mathcal{T}_1(P_{UW}, P_{VW})}} H(\tilde{U}\tilde{V}) + D(P_{\tilde{U}\tilde{V}}||Q_{UV}) + H(\tilde{W}|\tilde{U}) - H(\tilde{U}\tilde{V}\tilde{W}) - \frac{|\mathcal{U}||\mathcal{V}||\mathcal{W}|\log(k+1)}{k} \\
&= \min_{\substack{P_{\tilde{U}\tilde{V}\tilde{W}} \in \\ \mathcal{T}_1(P_{UW}, P_{VW})}} \sum P_{\tilde{U}\tilde{V}\tilde{W}} \log\left(\frac{P_{\tilde{U}\tilde{V}}}{Q_{UV}}\frac{1}{P_{\tilde{U}\tilde{V}}}\frac{P_{\tilde{U}}}{P_{\tilde{U}\tilde{W}}}P_{\tilde{U}\tilde{V}\tilde{W}}\right) - \delta^{(k)}_5 \\
&= \min_{P_{\tilde{U}\tilde{V}\tilde{W}} \in \mathcal{T}_1(P_{UW}, P_{VW})} D(P_{\tilde{U}\tilde{V}\tilde{W}}||Q_{UVW}) - \delta^{(k)}_5, \qquad (75)
\end{aligned}$$

$$Q_{UVW} = Q_{UV}P_{W|U}, \quad \delta^{(k)}_5 \triangleq \frac{|\mathcal{U}||\mathcal{V}||\mathcal{W}|\log(k+1)}{k} \xrightarrow{(k)} 0.$$

Eqn. (73) follows from (71) and (72). This results in the term $E_1(P_{W|U})$ in (41).

Next, consider the terms corresponding to the event $\mathcal{E}_{NE} \cap \mathcal{E}^c_{CE} \cap \mathcal{E}_{BE}$ in (70). Note that given the event $\mathcal{F}_{21} = \{U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, W^k(1) = w^k,\ \mathcal{E}_{NE},\ \mathcal{E}^c_{CE},\ \mathcal{E}_{BE}\}$ occurs, $P_{u^k w^k} \in T^k_{[UW]_\delta}$, $H_e(w^k|v^k) \geq H(W|V) - \gamma(\delta)$ for some $\gamma(\delta) \to 0$ as $\delta \to 0$ and $\mathcal{D}_0$ occurs only if $P_{v^k} \in T^k_{[V]_\delta}$. Using these facts to simplify the terms corresponding to the event $\mathcal{E}_{NE} \cap \mathcal{E}_{CE}^c \cap \mathcal{E}_{BE}$ in (70), we obtain

$$\sum_{\substack{P_{\tilde{U}\tilde{V}\tilde{W}} \in \\ \mathcal{T}(\mathcal{U}\times\mathcal{V}\times\mathcal{W})}} \sum_{\substack{(u^k,v^k,w^k) \\ \in T_{P_{\tilde{U}\tilde{V}\tilde{W}}}}} \left[\mathbb{P}(U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, W^k(1) = w^k | \mathcal{E}_{NE})\ \mathbb{P}(\mathcal{E}_{CE}^c|\mathcal{F})\ \mathbb{P}(\mathcal{E}_{BE}|\mathcal{F}_2)\ \mathbb{P}(\mathcal{D}_0|\mathcal{F}_{21})\right]$$

$$\leq \sum_{\substack{P_{\tilde{U}\tilde{V}\tilde{W}} \in \\ \mathcal{T}(\mathcal{U}\times\mathcal{V}\times\mathcal{W})}} \sum_{\substack{(u^k,v^k,w^k) \\ \in T_{P_{\tilde{U}\tilde{V}\tilde{W}}}}} \left[\mathbb{P}(U^k = u^k, V^k = v^k, W^k(1) = w^k | J = 1, f_B(J) = 1, \mathcal{E}_{NE})\ \mathbb{P}(\mathcal{E}_{CE}^c|\mathcal{F})\ \mathbb{P}(\mathcal{E}_{BE}|\mathcal{F}_2)\ \mathbb{P}(\mathcal{D}_0|\mathcal{F}_{21})\right]$$

$$\leq \sum_{\substack{P_{\tilde{U}\tilde{V}\tilde{W}} \in \\ \mathcal{T}_2(P_{UW},P_V)}} 2^{H(\tilde{U}\tilde{V}\tilde{W})} 2^{-k(H(\tilde{U}\tilde{V}) + D(P_{\tilde{U}\tilde{V}}||Q_{UV}) + H(\tilde{W}|\tilde{U}))} 2^{-k(R - I(U;W|V) - \delta_3^{(k)})}$$

$$\leq (k+1)^{|\mathcal{U}||\mathcal{V}||\mathcal{W}|} \max_{\substack{P_{\tilde{U}\tilde{V}\tilde{W}} \in \\ \mathcal{T}_2(P_{UW},P_V)}} 2^{kH(\tilde{U}\tilde{V}\tilde{W})} 2^{-k(H(\tilde{U}\tilde{V}) + D(P_{\tilde{U}\tilde{V}}||Q_{UV}) + H(\tilde{W}|\tilde{U}) + R - I(U;W|V) - \delta_3^{(k)})}$$

Note that the $\mathcal{E}_{BE}$ occurs only when $I(U;W) > R$. This leads to the term $E_2(P_{W|U})$ in (42).

Now, assume that the event $\mathcal{E}_{NE} \cap \mathcal{E}_{CE}$ holds. As in the case above, note that given $\mathcal{F}_1 = \{U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, W^k(1) = w^k,\ \mathcal{E}_{NE},\ \mathcal{E}_{CE}\}$, $P_{u^k w^k} \in T^k_{[UW]_\delta}$ and $\mathcal{D}_0$ occurs only if $P_{v^k} \in T^k_{[V]_\delta}$. Using these facts and eqns. (65), (66) and (58), it can be shown that the terms corresponding to this event in (70) result in the factor $E_3(R,\tau)$ given in (43).

Finally, we analyze the case when the event $\mathcal{E}_{TE}$ occurs. Since the encoder declares H_1 if $\hat{M} = 0$, it is clear that $\mathcal{D}_0$ occurs only when the channel error event $\mathcal{E}_{CE}$ happens. Thus, we have

$$\mathbb{P}(\mathcal{D}_0|\ \mathcal{E}_{TE}) = \mathbb{P}(\mathcal{E}_{CE}|\ \mathcal{E}_{TE})\ \mathbb{P}(\mathcal{D}_0|\ \mathcal{E}_{TE} \cap\ \mathcal{E}_{CE}) \tag{76}$$

From the coding scheme, it follows that

$$\mathbb{P}(\mathcal{E}_{CE}|\ \mathcal{E}_{TE}) \leq 2^{-kE_s(R,\tau)} \tag{77}$$

It is readily seen that when binning is performed at the encoder, a type 2 error occurs only if there exists a sequence $\hat{W}^k$ in the bin $\hat{M} \neq 0$ such that $(\hat{W}^k, V^k) \in T^k_{[WV]_\delta}$. On the other hand, when binning is not performed, $\mathcal{D}_0$ occurs only if $(W^k(\hat{M}), V^k) \in T^k_{[WV]_\delta}$. Also, recall that the encoder sends the error message $M = 0$ independent of the source codebook $\mathcal{C}$. From standard arguments, it can be shown that the probability of these events can be upper bounded as $2^{-k(R - I(U;W|V))}$ and $2^{-k(I(V;W))}$, respectively. This results in the factor $E_4(R,\tau)$ in the T2EE. This completes the proof of the theorem.

References

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

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

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

[6] 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.

[7] 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.

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

[9] 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.

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

[11] ——, "Interactive hypothesis testing with communication constraints," in *50th Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, USA, Oct. 2012.

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

[13] ——, "Distributed binary detection with lossy data compression," *arXiv:1601.01152 [cs.IT]*, Jan. 2016.

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

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

[16] 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.

[17] R. G. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Trans. Inf. Theory*, vol. 11, pp. 3–18, Jan 1965.

[18] S. H. Lim, P. Minero, and Y.-H. Kim, "Lossy communication of correlated sources over multiple access channels," in *48th Annual Allerton Conference on Communication, Control and Computing*, Illinois, USA, Sep. 2010.