Distributed Hypothesis Testing Over 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 involving two parties, one referred to as the observer and the other as the detector is studied. The observer observes a discrete memoryless source (DMS) and communicates its observations to the detector over a discrete memoryless channel (DMC). The detector observes another DMS correlated with that at the observer, and performs a binary hypothesis test on the joint distribution of the two DMS's using its own observed data and the information received from the observer. The trade-off between the type 1 error probability and the type 2 error exponent (T2EE) of the hypothesis test is explored. Single-letter lower bounds on the optimal T2EE are obtained by using three different coding schemes, a separate hypothesis testing and channel coding scheme, a local decision scheme, and a joint hypothesis testing and channel coding scheme. Exact single-letter characterizations of the optimal T2EE are established for three special cases; (i) testing against conditional independence, (ii) distributed HT when the DMC has zero capacity, and (iii) HT over a DMC. Moreover, it is shown that a *strong converse* holds in cases (ii) and (iii). Single-letter lower bounds on the optimal T2EE are also obtained for testing against conditional independence with multiple observers communicating over orthogonal DMCs.

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

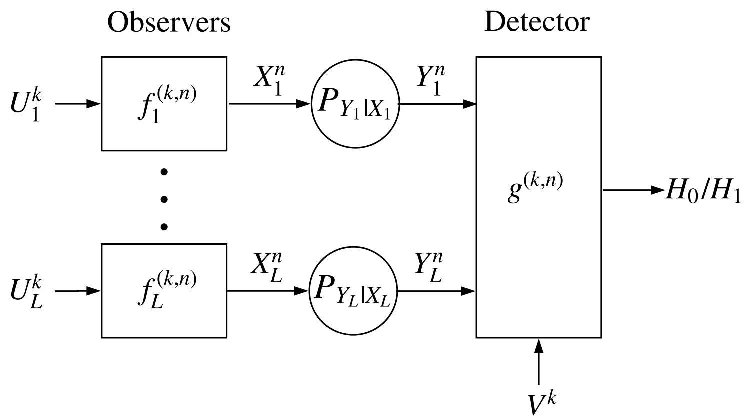

Fig. 1: Distributed HT over orthogonal DMCs.

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, $l = 1, \ldots, L$, 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 DMC, characterized by the conditional distribution $P_{Y_l|X_l}$. In the simplest case in which there are two possibilities $P_{U_1 \ldots U_L V}$ and $Q_{U_1 \ldots U_L V}$ for the joint distribution of the data, the detector performs a binary hypothesis test to decide between them based on the channel outputs $Y_1^n, \ldots, Y_L^n$ as well as its own observations V^k with the null and the alternate hypothesis given by

$$H_0 : (U_1^k, \ldots, U_L^k, V^k) \sim \prod_{i=1}^{k} P_{U_1 \ldots U_L V}, \tag{1a}$$

and

$$H_1 : (U_1^k, \ldots, U_L^k, V^k) \sim \prod_{i=1}^{k} Q_{U_1 \ldots U_L V}, \tag{1b}$$

respectively. Our goal is to characterize the optimal T2EE for a prescribed constraint on the type 1 error probability for the above hypothesis test.

In the centralized scenario, in which the detector performs a binary hypothesis test on the probability distribution of the data it observes directly, the optimal T2EE is characterized by the well-known lemma of Stein [2] (see also [3]). The study of distributed statistical inference under communication constraints was conceived by Berger in [4]. In [4], and in the follow up literature summarized below, communication from the observers to the detector are assumed to be over rate-limited error-free channels. Some of the fundamental results in this setting for the case of a single observer ($L = 1$) was established by Ahlswede and Csiszár in [5]. They obtained a tight single-letter characterization of the optimal T2EE for a special case of HT known as *testing against independence* (TAI), in which, $Q_{U_1V} = P_{U_1} \times P_V$. Furthermore, the authors established a lower bound on the optimal T2EE for the general HT case, and proved a *strong converse* result, which states that the optimal achievable T2EE is independent of the constraint on the type 1 error probability. A tighter lower bound for the general HT problem with a single observer is established by Han [6], which recovers the corresponding lower bound in [5]. 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 size two. It is shown that, asymptotically, the optimal T2EE achieved in these three settings are equal. In contrast, a single-letter characterization of the optimal T2EE for even the TAI with two-sided compression and general rate constraints remains open till date. Shalaby et. al [7] 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 blocklength k). Shimokawa et. al [8] obtained a tighter lower bound on the optimal T2EE for general HT by considering quantization and binning at the encoder along with a minimum empirical-entropy decoder. Rahman and Wagner [9] studied the setting with multiple observers, in which, they showed that for the case of a single-observer, the *quantize-bin-test* scheme achieves the optimal T2EE for *testing against conditional independence* (TACI), in which, $V = (E, Z)$ and $Q_{U_1EZ} = P_{U_1Z}P_{E|Z}$. Extensions of the distributed HT problem has also been considered in several other interesting scenarios involving multiple detectors [10], multiple observers [11], interactive HT [12], [13], collaborative HT [14], HT with lossy source reconstruction [15], HT over a multi-hop relay network [17], etc., in which, the authors obtain a single-letter characterization of the optimal T2EE in some special cases.

While the works mentioned above have studied the unsymmetric case of focusing on the T2EE for a constraint on the type 1 error probability, other works have analyzed the trade-off between the type 1 and type 2 error probabilities in the exponential sense. In this direction, the optimal trade-off between the type 1 and type 2 error exponents in the centralized scenario is obtained in [18]. The distributed version of this problem is first studied in [19], where inner bounds on the above trade-off are established. This problem has also been explored from an information-geometric perspective for the zero-rate compression scenario in [20] and [21], which provide further insights into the geometric properties of the optimal trade-off between the two exponents. A Neyman-Pearson like test in the zero-rate compression scenario is proposed in [22], which, in addition to achieving the optimal trade-off between the two exponents, also achieves the optimal second order asymptotic performance among all symmetric (type-based) encoding schemes. However, the optimal trade-off between the type 1 and type 2 error exponents for the general distributed HT problem remains open. Recently, an inner bound for this trade-off is obtained in [23], by using the reliability function of the optimal channel detection codes.

In contrast, HT in distributed settings that involve communication over noisy channels has not been considered until now. 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. Since the capacity of the channel $P_{Y|X}$, denoted by $C(P_{Y|X})$, quantifies the maximum rate of reliable communication over the channel, it is reasonable to expect that it plays a role in the characterization of the optimal T2EE similar to the rate-constraint R in the noiseless setting. Another measure of the noisiness of the channel is the so-called *reliability function* $E(R, P_{Y|X})$ [24], which is defined as the maximum achievable exponential decay rate of the probability of error (asymptotically) with respect to the blocklength for message rate of R. It appears natural that the reliability function plays a role in the characterization of the achievable T2EE for distributed HT over a noisy channel. Indeed, in Theorem 2 given below, we provide a lower bound on the optimal T2EE that depends on the *expurgated exponent* at rate R, $E_x(R, P_{Y|X})$, which is a lower bound on $E(R, P_{Y|X})$ [25]. However, surprisingly, it will turn out that the reliability function does not play a role in the characterization of the T2EE for TACI in the regime of vanishing type 1 error probability constraint.

The goal of this paper is to study the best attainable T2EE for distributed HT over a DMC and obtain a computable characterization of the same. Although a complete solution is not to be expected for this problem (since even the corresponding noiseless case is still open), the aim is to provide an achievable scheme for the general problem, and to identify special cases in which a tight characterization can be obtained. We will focus mostly on the case of a single observer in the system, but generalization to multiple observers will also be considered. Our main contributions can be summarized as follows.

(i) We propose three different coding schemes and analyze the T2EE achieved by these schemes.

(ii) We obtain an exact single-letter characterization of the optimal T2EE for three special:

 (a) TACI for the case of vanishing type 1 error probability constraint,

 (b) when $C(P_{Y|X}) = 0$, i.e., the communication channel has zero capacity,

 (c) HT over a DMC, i.e., when there is no side-information at the detector.

 We show that for the cases (b) and (c), the optimal T2EE is in fact independent of the constraint on the type 1 error probability, thus implying that a strong-converse holds.

(iii) We also obtain single-letter lower bounds on the T2EE for TACI when there are multiple observers in the system communicating over orthogonal DMCs.

In the sequel, we first introduce a separation based scheme that performs independent hypothesis testing and channel coding, which we refer to as the *separate hypothesis testing and channel coding* (SHTCC) scheme. This scheme combines the Shimokawa-Han-Amari scheme [8], which is the best known coding scheme till date for distributed HT over a rate-limited noiseless channel, with the channel coding scheme that achieves the expurgated exponent of the channel [25] [24]. A separation based scheme similar to SHTCC scheme has been proposed recently in [26], where the authors study the T2EE for distributed HT over a point to point, multiple-access and broadcast channels. Our second scheme is a zero-rate compression scheme referred to as the *local decision* (LD) scheme, in which, the observer makes a tentative guess on the true hypothesis based on its own observation, and communicates its one bit decision to the detector. The third scheme is a *joint hypothesis testing and channel coding* (JHTCC) scheme, in which, hybrid coding [27] is used to communicate from the observer to the detector. As we show later, the SHTCC scheme achieves the optimal T2EE for TACI, while the LD scheme achieves the optimal T2EE for general distributed HT when the channel has zero capacity, and, also for HT over a noisy channel of arbitrary capacity (i.e., no side-information at the detector). We also show that the JHTCC scheme recovers the optimal T2EE for TACI. Although the T2EE achieved by the SHTCC and JHTCC schemes are incomparable in general, we establish conditions under which the JHTCC scheme achieves a T2EE at least as good as the SHTCC scheme. Finally, we establish single-letter lower and upper bounds on the achievable T2EE for TACI with multiple observers, in which, $Q_{U_1 \ldots U_L E Z} = P_{U_1 \ldots U_L Z} P_{E|Z}$. This is done by first mapping the problem to an equivalent *joint source-channel coding* (JSCC) problem with helpers. The Berger-Tung (BT) bounds [28] [29] and the *source-channel separation theorem* in [30] are then used to obtain the desired bounds.

The rest of the paper is organized as follows. In Section II, we introduce the notations, detailed system model and definitions. Following this, we introduce the main results in Section III and IV focusing on the case of a single observer. The achievable schemes are presented in Section III and the optimality results for special cases are discussed in Section IV. In Section V, we obtain lower bounds on the optimal T2EE for distributed HT with multiple observers communicating to the detector over orthogonal DMCs. Finally, Section VI concludes the paper.

II. Preliminaries

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. Equality by definition is represented by $:=$. For $m, l \in \mathbb{Z}^+$, 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 $\mathcal{G} = \{l_1, \ldots, l_g\} \subseteq \mathbb{Z}^+$, $\left\{X_{l_1}^m, \ldots, X_{l_g}^m\right\}$, $\left\{X_{l_1}^m(j), \ldots, X_{l_g}^m(j)\right\}$ and $\left\{\{X_{l_1}^m(j)\}_{j\in\mathbb{Z}^+}, \ldots, \left\{X_{l_g}^m(j)\right\}_{j\in\mathbb{Z}^+}\right\}$ 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. $D(P_X||Q_X)$, $H_{P_X}(X)$, $H_{P_{XY}}(X|Y)$ and $I_{P_{XY}}(X;Y)$ represent the standard quantities of Kullback-Leibler (KL) divergence between distributions P_X and Q_X, the entropy of X with distribution P_X, the conditional entropy of X given Y and the mutual information between X and Y with joint distribution P_{XY}, respectively. When the distribution of the r.v.'s involved are clear from the context, the entropic and mutual information quantities are denoted simply by $I(X;Y)$, $H(X)$ and $H(X|Y)$, respectively. Following the notation in [24], T_P and $T^m_{[P_X]_\delta}$ (or $T^m_{[X]_\delta}$ or T^m_δ 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. The set of all types of k-length sequences of r.v.'s X^k and Y^k is denoted by $\mathcal{T}^k_{\mathcal{X}\mathcal{Y}}$ and $\cup_{k\in\mathbb{Z}^+}\mathcal{T}^k_{\mathcal{X}\mathcal{Y}}$ is denoted by $\mathcal{T}_{\mathcal{X}\mathcal{Y}}$. Given realizations $X^n = x^n$ and $Y^n = y^n$, $H_e(x^n|y^n)$ denote the conditional empirical entropy defined as

$$H_e(x^n|y^n) := H_{P_{\tilde{X}\tilde{Y}}}(\tilde{X}|\tilde{Y}), \tag{2}$$

where $P_{\tilde{X}\tilde{Y}}$ denote the joint type of (x^n, y^n). For $a \in \mathbb{R}^+$, $[a]$ denotes the set of integers $\{1, 2, \ldots, \lceil a \rceil\}$. All logarithms considered in this paper are with respect to the base e. For any set $\mathcal{G}$, $\mathcal{G}^c$ denotes the set complement. $a_k \xrightarrow{(k)} b$ indicates that $\lim_{k\to\infty} a_k = b$. For functions $f_1 : \mathcal{A} \to \mathcal{B}$ and $f_2 : \mathcal{B} \to \mathcal{C}$, $f_2 \circ f_1$ denotes function composition. Finally, $\mathbb{1}(\cdot)$ denotes the indicator function, and $O(\cdot)$ and $o(\cdot)$ denote the standard asymptotic notation.

B. Problem formulation

All the r.v.'s considered henceforth are discrete with finite support. Unless specified otherwise, we will denote the probability distribution of a r.v. Z under the null and alternate hypothesis by P_Z and Q_Z, respectively. 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 $f_l^{(k,n)} : \mathcal{U}_l^k \to \mathcal{X}_l^n$ represents the encoding function (possibly stochastic) of observer l. Let $\tau := \frac{n}{k}$ denote the *bandwidth ratio*. The channel output $Y_{\mathcal{L}}^n$ is given by the probability law

$$P_{Y_{\mathcal{L}}^n|X_{\mathcal{L}}^n}(y_{\mathcal{L}}^n|x_{\mathcal{L}}^n) = \prod_{l=1}^{L}\prod_{j=1}^{n} P_{Y_l|X_l}(y_{l,j}|x_{l,j}), \tag{3}$$

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, the detector makes a decision between the two hypotheses H_0 and H_1 given in (1). Let $H \in \{0, 1\}$ denote the actual hypothesis and $\hat{H} \in \{0, 1\}$ denote the output of the HT, where 0 and 1 denote H_0 and H_1, respectively, and $\mathcal{A}^{(k,n)} \subseteq \mathcal{Y}_{\mathcal{L}}^n \times \mathcal{V}^k$ denote the acceptance region for H_0. Then, the decision rule $g^{(k,n)} : \mathcal{Y}_{\mathcal{L}}^n \times \mathcal{V}^k \to \{0, 1\}$ is given by

$$g^{(k,n)}\left(y_{\mathcal{L}}^n, v^k\right) = 1 - \mathbb{1}\left(\left(y_{\mathcal{L}}^n, v^k\right) \in \mathcal{A}^{(k,n)}\right).$$

Let

$$\alpha\left(k, n, f_1^{(k,n)}, \ldots, f_L^{(k,n)}, g^{(k,n)}\right) := 1 - P_{Y_{\mathcal{L}}^n V^k}\left(\mathcal{A}^{(k,n)}\right), \tag{4}$$

$$\text{and } \beta\left(k, n, f_1^{(k,n)}, \ldots, f_L^{(k,n)}, g^{(k,n)}\right) := Q_{Y_{\mathcal{L}}^n V^k}\left(\mathcal{A}^{(k,n)}\right), \tag{5}$$

denote the type 1 and type 2 error probabilities for the encoding functions $f_1^{(k,n)}, \ldots, f_L^{(k,n)}$ and decision rule $g^{(k,n)}$, respectively.

Definition 1. *A T2EE κ is (τ,ϵ) achievable 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 decision rules $g^{(k,n_k)}$ such that $n_k \leq \tau k$, $\forall\ k$,*

$$\liminf_{k\to\infty} \frac{-1}{k}\log\left(\beta\left(k, n_k, f_1^{(k,n_k)},\ldots,f_L^{(k,n_k)}, g^{(k,n_k)}\right)\right) \geq \kappa, \tag{6a}$$

$$\textit{and}\ \limsup_{k\to\infty} \alpha\left(k, n_k, f_1^{(k,n_k)},\ldots,f_L^{(k,n_k)}, g^{(k,n_k)}\right) \leq \epsilon. \tag{6b}$$

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

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

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

It is well known that the Neyman-Pearson test [1] gives the optimal trade-off between the type 1 and type 2 error probabilities, and hence, also between the error exponents in a HT. From this, it follows that the optimal T2EE for distributed HT over DMC's is achieved when for each $l\in\mathcal{L}$, the channel-input X_l^n is generated correlated with U_l^k according to some optimal conditional distribution $P_{X_l^n|U_l^k}$, and the optimal Neyman-Pearson test is performed on the data available (both 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. Moreover, analysis of such a scheme is prohibitively complex as it involves optimization over large dimensional probability simplexes, when k and n are large. In the next section, we establish three different single-letter lower bounds on $\kappa(\tau,\epsilon)$ by using the SHTCC, LD and JHTCC schemes, respectively. We will limit the discussion to the case of a single observer, i.e., $L=1$, until Section V, and therefore, omit the subscript associated with the index of the observer, e.g., U_1 will be denoted as U.

III. Achievable schemes

In [8], Shimokawa et. al. obtained a lower bound on the optimal T2EE for distributed HT over a rate-limited noiseless channel by using a coding scheme that involves quantization and binning at the encoder. In this scheme, the type[1] of the observed sequence U^k is sent by the encoder to the detector, which aids in the HT. In fact, in order to achieve the T2EE proposed in [8], it is sufficient to send a message indicating whether U^k is typical or not, rather than sending the

[1]Since the number of types is polynomial in the blocklength, these can be communicated error-free at asymptotically zero-rate.

exact type of U^k. Although it is not possible to get perfect reliability for messages transmitted over a noisy channel, intuitively, it is desirable to protect the typicality information about the observed sequence as reliably as possible. Based on this intuition, we next propose the SHTCC scheme that performs independent HT and channel coding and protects the message indicating whether U^k is typical or not, as reliably as possible.

A. *SHTCC Scheme:*

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 $f_s^{(k)} : \mathcal{U}^k \to \mathcal{M} = \{0, 1, \cdots, \lceil e^{kR} \rceil\}$ generates an index $M = f_s^{(k)}(U^k)$ and the channel encoder $f_c^{(k,n)} : \mathcal{M} \to \tilde{\mathcal{C}} = \{X^n(j),\ j \in [0 : \lceil e^{kR} \rceil]\}$ generates the channel-input codeword $X^n = f_c^{(k,n)}(M)$. Note that the rate of this coding scheme is $\frac{kR}{n} = \frac{R}{\tau}$ bits per channel use. The channel decoder $g_c^{(k,n)} : \mathcal{Y}^n \to \mathcal{M}$ maps the channel-output Y^n into an index $\hat{M} = g_c^{(k,n)}(Y^n)$, and $g_s^{(k)} : \mathcal{M} \times \mathcal{V}^k \to \{0, 1\}$ outputs the result of the HT as $\hat{H} = g_s^{(k)}(\hat{M}, V^k)$. Note that $f_c^{(k,n)}$ depends on U^k only through the output of $f_s^{(k)}(U^k)$ and $g_c^{(k,n)}$ depends on V^k only through Y^n. Hence, the scheme is modular in the sense that $(f_c^{(k,n)}, g_c^{(k,n)})$ can be designed independent of $(f_s^{(k)}, g_s^{(k)})$. In other words, any good channel coding scheme may be used in conjunction with a good compression scheme. If U^k is not typical according to P_U, $f_s^{(k)}$ outputs a *special* message, referred to as the *error* message, denoted by $M = 0$, to inform the detector to declare $\hat{H} = 1$. There is obviously a trade-off between the reliability of the error message and the other messages in channel coding. The best known reliability for protecting a single *special* message when the other messages $M \in [e^{nR}]$ of rate R, referred to as *ordinary* messages, are required to be communicated reliably is given by the *red-alert exponent* in [31]. The red-alert exponent is defined as

$$E_m(R, P_{Y|X}) := \max_{\substack{P_{SX}:\ \mathcal{S}=\mathcal{X},\\ I(X;Y|S)=R,\\ S-X-Y}} \sum_{s \in \mathcal{S}} P_S(s)\ D\left(P_{Y|S=s} || P_{Y|X=s}\right). \tag{8}$$

Borade's scheme uses an appropriately generated codebook along with a two-stage decoding procedure. The first stage is a *joint-typicality* decoder to decide whether $X^n(0)$ is transmitted, while the second stage is a *maximum-likelihood decoder* to decode the ordinary message if the output of the first stage is not zero, i.e., $\hat{M} \neq 0$. On the other hand, it is well-known that if the rate of the messages is R, a channel coding error exponent equal to $E_x(R, P_{Y|X})$ is achievable, where

$$E_x(R, P_{Y|X}) := \max_{P_X} \max_{\rho \geq 1} \left\{ -\rho\ R - \rho\ \log \left(\sum_{x, \tilde{x}} P_X(x) P_X(\tilde{x}) \left(\sum_y \sqrt{P_{Y|X}(y|x) P_{Y|X}(y|\tilde{x})} \right)^{\frac{1}{\rho}} \right) \right\}, \quad (9)$$

is the *expurgated* exponent at rate R [25] [24]. Let

$$E_m(P_{SX}, P_{Y|X}) := \sum_{s \in \mathcal{S}} P_S(s)\ D\left(P_{Y|S=s} || P_{Y|X=s}\right), \quad (10)$$

where, $\mathcal{S} = \mathcal{X}$ and $S - X - Y$, and

$$E_x(R, P_{SX}, P_{Y|X}) := \max_{\rho \geq 1} \left\{ -\rho\ R - \rho\ \log \left(\sum_{s, x, \tilde{x}} P_S(s) P_{X|S}(x|s) P_{X|S}(\tilde{x}|s) \left(\sum_y \sqrt{P_{Y|X}(y|x) P_{Y|X}(y|\tilde{x})} \right)^{\frac{1}{\rho}} \right) \right\}.$$

Although Borade's scheme is concerned only with the reliability of the special message, it is not hard to see using the technique of *random-coding* that for a fixed distribution P_{SX}, there exists a codebook $\tilde{C}$, and encoder and decoder as in Borade's scheme, such that the rate is $I(X;Y|S)$ and the special message achieves a reliability equal to $E_m(P_{SX}, P_{Y|X})$, while the ordinary messages achieve a reliability equal to $E_x(I(X;Y|S), P_{SX}, P_{Y|X})$. Note that $E_m(P_{SX}, P_{Y|X})$ and $E_x(I(X;Y|S), P_{SX}, P_{Y|X})$ denote Borade's red-alert exponent and the expurgated exponent with fixed distribution P_{SX}, respectively, and that both are inter-dependent through P_{SX}. Thus, varying P_{SX} provides a trade-off between the reliability for the ordinary messages and the special message. We will use Borade's scheme for channel coding in the SHTCC scheme, such that the error message and the other messages correspond to the special and ordinary messages, respectively. The SHTCC scheme will be described in detail in Appendix A. We next state a lower bound on $\kappa(\tau, \epsilon)$ that is achieved by the SHTCC scheme. For brevity, we will use the shorter notations C, $E_m(P_{SX})$ and $E_x(R, P_{SX})$ instead of $C(P_{Y|X})$, $E_m(P_{SX}, P_{Y|X})$ and $E_x(R, P_{SX}, P_{Y|X})$, respectively.

Theorem 2. *For $\tau \geq 0$, $\kappa(\tau, \epsilon) \geq \kappa_s(\tau)$, $\forall\ \epsilon \in (0,1]$, where*

$$\kappa_s(\tau) := \sup_{\substack{(P_{W|U}, P_{SX}) \\ \in\ \mathcal{B}(\tau, C)}} \min\left(E_1(P_{W|U}),\ E_2(P_{W|U}, P_{SX}, \tau),\ E_3(P_{W|U}, P_{SX}, \tau),\ E_4(P_{W|U}, P_{SX}, \tau)\right), \tag{11}$$

where

$$\mathcal{B}(\tau, C) := \left\{ \begin{array}{c} (P_{W|U}, P_{SX}) :\ \mathcal{S} = \mathcal{X},\ P_{UVWSXY}(P_{W|U}, P_{SX}) := P_{UV} P_{W|U} P_{SX} P_{Y|X}, \\ I_P(U;W|V) < \tau I_P(X;Y|S) \leq \tau C \end{array} \right\}.$$

$$E_1(P_{W|U}) := \min_{P_{\tilde{U}\tilde{V}\tilde{W}} \in \mathcal{T}_1(P_{UW}, P_{VW})} D(P_{\tilde{U}\tilde{V}\tilde{W}} || Q_{UVW}), \tag{12}$$

$$E_2(P_{W|U}, P_{SX}, \tau) := \begin{cases} \min\limits_{P_{\tilde{U}\tilde{V}\tilde{W}} \in \mathcal{T}_2(P_{UW}, P_V)} D(P_{\tilde{U}\tilde{V}\tilde{W}} || Q_{UVW}) + \tau I_P(X;Y|S) & \\ \qquad -I_P(U;W|V), & \textit{if } I_P(U;W) > \tau I_P(X;Y|S), \\ \infty, & \textit{otherwise}, \end{cases} \tag{13}$$

$$E_3(P_{W|U}, P_{SX}, \tau) := \begin{cases} \min\limits_{P_{\tilde{U}\tilde{V}\tilde{W}} \in \mathcal{T}_3(P_{UW}, P_V)} D(P_{\tilde{U}\tilde{V}\tilde{W}} || Q_{UVW}) + \tau I_P(X;Y|S) - I_P(U;W|V) & \\ \qquad +\ \tau E_x\left(I_P(X;Y|S), P_{SX}\right), & \textit{if } I_P(U;W) > \tau I_P(X;Y|S), \\ \min\limits_{P_{\tilde{U}\tilde{V}\tilde{W}} \in \mathcal{T}_3(P_{UW}, P_V)} D(P_{\tilde{U}\tilde{V}\tilde{W}} || Q_{UVW}) + I_P(V;W) & \\ \qquad + \tau E_x\left(I_P(X;Y|S), P_{SX}\right), & \textit{otherwise}, \end{cases} \tag{14}$$

$$E_4(P_{W|U}, P_{SX}, \tau) := \begin{cases} D(P_V || Q_V) + \tau I_P(X;Y|S) - I_P(U;W|V) & \\ \qquad + \tau E_m\left(P_{SX}\right), & \textit{if } I_P(U;W) > \tau I_P(X;Y|S), \\ D(P_V || Q_V) + I_P(V;W) + \tau E_m\left(P_{SX}\right), & \textit{otherwise}, \end{cases} \tag{15}$$

$$Q_{UVW} := Q_{UV} P_{W|U},$$

$$\mathcal{T}_1(P_{UW}, P_{VW}) := \{P_{\tilde{U}\tilde{V}\tilde{W}} \in \mathcal{T}_{\mathcal{UVW}} : P_{\tilde{U}\tilde{W}} = P_{UW},\ P_{\tilde{V}\tilde{W}} = P_{VW}\},$$

$$\mathcal{T}_2(P_{UW}, P_V) := \{P_{\tilde{U}\tilde{V}\tilde{W}} \in \mathcal{T}_{\mathcal{UVW}} : P_{\tilde{U}\tilde{W}} = P_{UW},\ P_{\tilde{V}} = P_V,\ H(\tilde{W}|\tilde{V}) \geq H_P(W|V)\},$$

$$\mathcal{T}_3(P_{UW}, P_V) := \{P_{\tilde{U}\tilde{V}\tilde{W}} \in \mathcal{T}_{\mathcal{UVW}} : P_{\tilde{U}\tilde{W}} = P_{UW},\ P_{\tilde{V}} = P_V\}.$$

The proof of Theorem 2 is given in Appendix A. Although the expression $\kappa_s(\tau)$ in Theorem 2 appears complicated, the terms $E_1(P_{W|U})$ to $E_4(P_{W|U}, P_{SX}, \tau)$ can be understood to correspond to distinct events that can possibly lead to a type 2 error. Note that $E_1(P_{W|U})$ and $E_2(P_{W|U}, P_{SX}, \tau)$ are the same terms appearing in the T2EE achieved by the Shimokawa-Han-Amari scheme [8] for the noiseless channel setting, while $E_3(P_{W|U}, P_{SX}, \tau)$ and $E_4(P_{W|U}, P_{SX}, \tau)$ are additional terms introduced due to the channel. $E_3(P_{W|U}, P_{SX}, \tau)$ corresponds to the event when $M \neq 0$, $\hat{M} \neq M$ and $g_s^{(k)}(\hat{M}, V^k) = 0$, whereas $E_4(P_{W|U}, P_{SX}, \tau)$ is due to the event when $M = 0$, $\hat{M} \neq M$ and $g_s^{(k)}(\hat{M}, V^k) = 0$.

Remark 3. *Note that, in general, $E_m(P_{SX})$ can take the value of ∞ and when this happens, the term $\tau E_m(P_{SX})$ becomes undefined for $\tau = 0$. In this case, we define $\tau E_m(P_{SX}) := 0$. A similar rule applies for $\tau E_x(I_P(X;Y|S), P_{SX})$ when $\tau = 0$ and $E_x(I_P(X;Y|S), P_{SX}) = \infty$.*

Remark 4. *In the SHTCC scheme, we used Borade's scheme for channel coding, that is concerned specifically with the protection of a special message. Another scheme can be obtained by replacing Borade's scheme by a scheme such that the ordinary messages achieve an error exponent equal to the reliability function $E(R, P_{Y|X})$ [24] of the channel $P_{Y|X}$ at rate R, while the special message achieves the maximum reliability possible subject to this constraint. However, a computable characterization of the best reliability achievable for a single message when the ordinary messages achieve $E(R, P_{Y|X})$, or even a computable characterization of $E(R, P_{Y|X})$ for all values of R is unknown in general. Due to this reason, a comparison between κ_s and the T2EE achieved by the above mentioned scheme is not straighforward.*

B. Local Decision (LD) Scheme (Zero-Rate Compression Scheme)

The SHTCC scheme described above is a two stage scheme in which the observer communicates a compressed version W^k of U^k using a channel code of rate $\frac{R}{\tau}$ bits per channel use, where $R \leq \tau C$, while the detector makes the decision on the hypothesis using an estimate of W^k and side-information V^k. Now, suppose the observer makes the decision about the hypothesis locally using U^k and transmits its 1 bit decision to the detector using a channel code for two messages, while the detector makes the final decision based on its estimate of the 1 bit message and V^k. The encoder $f^{(k,n)} = f_c^{(k,n)} \circ f_s^{(k)}$ and decoder $g^{(k,n)} = g_s^{(k)} \circ g_c^{(k,n)}$ are thus specified by maps $f_s^{(k)} : \mathcal{U}^k \rightarrow \{0,1\}$, $f_c^{(k,n)} : \{0,1\} \rightarrow \mathcal{X}^n$, $g_c^{(k,n)} : \mathcal{Y}^n \rightarrow \{0,1\}$ and $g_s^{(k)} : \{0,1\} \times \mathcal{V}^k \rightarrow \{0,1\}$. We refer to this scheme as the LD scheme. Observe that the rate of communication over the channel for this scheme is $R = \frac{1}{n}$ bits per channel use, which tends to zero asymptotically.

We will next obtain a lower bound on $\kappa(\tau, \epsilon)$ using the LD scheme. Let

$$\beta_0 := \beta_0(P_U, P_V, Q_{UV}) := \min_{\substack{P_{\tilde{U}\tilde{V}}: \\ P_{\tilde{U}}=P_U,\ P_{\tilde{V}}=P_V}} D(P_{\tilde{U}\tilde{V}} || Q_{UV}), \tag{16}$$

$$\text{and } E_c := E_c(P_{Y|X}) := D(P_{Y|X=a} || P_{Y|X=b}), \tag{17}$$

where a and b denote channel input symbols that satisfy

$$(a,b) = \operatorname*{arg\,max}_{(x,x') \in \mathcal{X} \times \mathcal{X}} D(P_{Y|X=x} || P_{Y|X=x'}). \tag{18}$$

Note that β_0 denotes the optimal T2EE for distributed HT over a noiseless channel, when the communication rate-constraint is zero [6] [7]. We define

$$\kappa_0(\tau) := \begin{cases} D(P_V||Q_V) & , \text{ if } \tau = 0, \\ \min\left(\beta_0, \tau E_c + D(P_V||Q_V)\right), & \text{otherwise,} \end{cases} \tag{19}$$

We have the following result.

Theorem 5. *For* $\tau \geq 0$, $\kappa(\tau, \epsilon) \geq \kappa_0(\tau)$, $\forall\ \epsilon \in (0,1]$.

Proof: Let $k \in \mathbb{Z}^+$ and $n_k = \lfloor \tau k \rfloor$. For $\tau = 0$, Theorem 5 follows from Stein's lemma [5] applied to i.i.d. sequence V^k available at the detector. Assume $\tau > 0$. For a fixed $\delta > 0$ (a small number), we define the functions $f_s^{(k)}$ and $f_c^{(k,n_k)}$ for the encoder $f^{(k,n_k)}$ as follows:

$$f_s^{(k)}(u^k) = \begin{cases} 0, & \text{if } P_{u^k} \in T^k_{[P_U]_\delta}, \\ 1, & \text{otherwise,} \end{cases} \tag{20}$$

and

$$f_c^{(k,n_k)}\left(f_s^{(k)}(u^k)\right) = \begin{cases} a^n, \text{ if } f_s^{(k)}(u^k) = 0, \\ b^n, \text{ otherwise.} \end{cases} \tag{21}$$

Here, a^{n_k} and b^{n_k} denote the codewords formed by repeating the symbols a and b from the channel input alphabet $\mathcal{X}$, which are defined in (18). Let the functions $g_s^{(k)}$ and $g_c^{(k,n_k)}$ of the decision rule $g^{(k,n_k)}$ be defined by

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

for some $\delta' > 0$ (a small number), and

$$g_s^{(k)}\left(v^k, g_c^{(k,n_k)}(y^{n_k})\right) = \begin{cases} 0, \text{ if } P_{v^k} \in T^k_{[P_V]_\delta} \text{ and } g_c^{(k,n_k)}(y^{n_k}) = 0, \\ 1, \text{ otherwise.} \end{cases}$$

By the law of large numbers, the type 1 error probability tends to zero asymptotically, since

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

A type 2 error occurs only under the following two events:

$$\mathcal{E}_{1p} := \{U^k \in T^k_{[P_U]_\delta},\ V^k \in T^k_{[P_V]_\delta} \text{ and } Y^{n_k} \in T^{n_k}_{[P_{Y|X=a}]_{\delta'}}\},$$
$$\mathcal{E}_{2p} := \{U^k \notin T^k_{[P_U]_\delta},\ V^k \in T^k_{[P_V]_\delta} \text{ and } Y^{n_k} \in T^{n_k}_{[P_{Y|X=a}]_{\delta'}}\}.$$

More specifically, $\mathcal{E}_{1p}$ occurs when both U^k and V^k are typical and there is no error at the channel decoder, while $\mathcal{E}_{2p}$ occurs when U^k is not typical, V^k is typical and the channel decoder $g_c^{(k,n_k)}$ makes a decoding error. It follows from the zero-rate compression result in [6] that the probability of the first event is upper bounded by $e^{-k(\beta_0 - O(\delta) - \gamma)}$ for any $\gamma > 0$ and k sufficiently large. The probability of the second event is upper bounded for any $\gamma > 0$ and k sufficiently large as

$$\begin{aligned}\mathbb{P}(\mathcal{E}_{2p}|H=1) &\leq \mathbb{P}(V^k \in T^k_{[P_V]_\delta}|H=1)\ \mathbb{P}\left(Y^{n_k} \in T^{n_k}_{[P_{Y|X=a}]_{\delta'}}|U^k \notin T^k_{[P_U]_\delta}\right)\\ &= \mathbb{P}(V^k \in T^k_{[P_V]_\delta}|H=1)\ \mathbb{P}\left(Y^{n_k} \in T^{n_k}_{[P_{Y|X=a}]_{\delta'}}|X^{n_k}=b^{n_k}\right)\\ &\leq e^{-k(D(P_V||Q_V)-O(\delta)-\gamma)} \cdot e^{-n_k(E_c-O(\delta')-\gamma)}. \end{aligned} \tag{22}$$

Here, (22) follows from [24, Lemma 2.6] and the fact that the number of types in $\mathcal{V}^k$ and $\mathcal{Y}^{n_k}$ is upper bounded by a polynomial in k and n_k, respectively [24]. By the union bound, it follows that

$$\beta(k, n_k, f^{(k,n_k)}, g^{(k,n_k)}) \leq \mathbb{P}(\mathcal{E}_{1p}|H=1) + \mathbb{P}(\mathcal{E}_{2p}|H=1),$$

which in turn implies, in the limit δ and δ' tending to zero (subject to delta-convention given in [24]), that

$$\kappa(\tau,\epsilon) \geq \min\left(\beta_0-\gamma), \tau(E_c-\gamma)\right),\ \forall\ \epsilon \in (0,1).$$

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

■

The LD scheme would be particularly useful when the communication channel is very noisy, so that reliable communication is not possible at any positive rate. In Section IV, we will show that the LD scheme achieves the optimal T2EE in two important scenarios: (i) for distributed HT over a DMC when the channel capacity is zero, and (ii) for HT over a DMC, i.e., when the detector has no side-information. In fact, we will show a stronger result that the optimal T2EE is not improved if the type 1 error probability constraint is relaxed; and hence, that a strong converse holds.

The SHTCC and LD schemes introduced above are schemes that perform independent HT and channel coding, i.e., the channel encoder $f_c^{(k,n)}$ neglects U^k given the output M of source encoder $f_s^{(k)}$, and $g_s^{(k)}$ neglects Y^n given the output of the channel decoder $g_c^{(k,n)}$. The following scheme ameliorates these restrictions and uses hybrid coding to perform joint HT and channel coding.

C. JHTCC Scheme

Hybrid coding is a form of JSCC introduced in [32] for the lossy transmission of sources over noisy networks. As the name suggests, hybrid coding is a combination of the digital and analog (uncoded) transmission schemes. For simplicity, assume that $k = n$ ($\tau = 1$). In hybrid coding, the source U^n is first mapped to one of the codewords $\bar{W}^n$ within a compression codebook. Then, a symbol-by-symbol function (deterministic) of the $\bar{W}^n$ and U^n is transmitted as the channel codeword X^n. This procedure is reversed at the decoder, in which, the decoder first attempts to obtain an estimate $\hat{\bar{W}}^n$ of $\bar{W}^n$ using the channel output Y^n and its own correlated side information V^n. Then, the reconstruction $\hat{U}^n$ of the source is obtained as a symbol-by-symbol function of the reconstructed codeword, Y^n and V^n. In this subsection, we propose a lower bound on the optimal T2EE that is achieved by a scheme that utilizes hybrid coding for the communication between the observer and the detector, which we refer to as the JHTCC scheme. Post estimation of $\hat{\bar{W}}^n$, the detector performs the hypothesis test using $\hat{\bar{W}}^n$, Y^n and V^n, instead of estimating $\hat{U}^n$ as is done in JSCC problems. We will in fact consider a slightly generalized form of hybrid coding in that the encoder and detector is allowed to perform "time-sharing" according to a sequence S^n that is known a priori to both parties. Also, the input X^n is allowed to be generated according to an arbitrary memoryless stochastic function instead of a deterministic function. The JHTCC scheme will be described in detail in Appendix B. Next, we state a lower bound on $\kappa(\tau, \epsilon)$ that is achieved by the JHTCC scheme.

Theorem 6. $\kappa(1,\epsilon) \geq \kappa_h, \ \ \forall\ \epsilon \in (0,1]$, *where*

$$\kappa_h := \sup_{\mathbf{b}\ \in\ \mathcal{B}_h} \min\Big(E_1'(P_S, P_{\bar{W}|US}, P_{X|US\bar{W}}),\ E_2'(P_S, P_{\bar{W}|US}, P_{X|US\bar{W}}), \\ E_3'(P_S, P_{\bar{W}|US}, P_{X'|S}, P_{X|US\bar{W}})\Big), \tag{23}$$

$$\mathcal{B}_h := \left\{ \begin{matrix} \mathbf{b} = \big(P_S, P_{\bar{W}|US}, P_{X'|S}, P_{X|US\bar{W}}\big) : I_{\hat{P}}(U;\bar{W}|S) < I_{\hat{P}}(\bar{W};Y,V|S),\ \mathcal{X}' = \mathcal{X}, \\ \hat{P}_{UVS\bar{W}X'XY}\ \big(P_S, P_{\bar{W}|US}, P_{X'|S}, P_{X|US\bar{W}}\big) := P_{UV} P_S P_{\bar{W}|US} P_{X'|S} P_{X|US\bar{W}} P_{Y|X} \end{matrix} \right\},$$

$$E_1'\left(P_S, P_{\bar{W}|US}, P_{X|US\bar{W}}\right) := \min_{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} \in \mathcal{T}_1'\left(\hat{P}_{US\bar{W}}, \hat{P}_{VS\bar{W}Y}\right)} D\left(P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} || \hat{Q}_{UVS\bar{W}Y}\right), \tag{24}$$

$$E_2'\left(P_S, P_{\bar{W}|US}, P_{X|US\bar{W}}\right) := \min_{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} \in \mathcal{T}_2'\left(\hat{P}_{US\bar{W}}, \hat{P}_{VS\bar{W}Y}\right)} D\left(P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} || \hat{Q}_{UVS\bar{W}Y}\right) + I_{\hat{P}}(\bar{W}; V, Y|S) - I_{\hat{P}}(U; \bar{W}|S), \quad (25)$$

$$E_3'\left(P_S, P_{\bar{W}|US}, P_{X'|S}, P_{X|US\bar{W}}\right) := D(\hat{P}_{VSY}||\check{Q}_{VSY}) + I_{\hat{P}}(\bar{W}; V, Y|S) - I_{\hat{P}}(U; \bar{W}|S), \quad (26)$$

$$\hat{Q}_{UVS\bar{W}X'XY}(P_S, P_{\bar{W}|US}, P_{X'|S}, P_{X|US\bar{W}}) := Q_{UV}P_S P_{\bar{W}|US}P_{X'|S}P_{X|US\bar{W}}P_{Y|X}, \quad (27)$$

$$\check{Q}_{UVSX'XY}(P_S, P_{X'|S}) := Q_{UV}P_S P_{X'|S}\mathbb{1}(X = X')P_{Y|X}, \quad (28)$$

$$\mathcal{T}_1'(\hat{P}_{US\bar{W}}, \hat{P}_{VS\bar{W}Y}) := \{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} \in \mathcal{T}_{\mathcal{UVSWY}} : P_{\tilde{U}\tilde{S}\tilde{W}} = \hat{P}_{US\bar{W}},\ P_{\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} = \hat{P}_{VS\bar{W}Y}\},$$

$$\mathcal{T}_2'(\hat{P}_{US\bar{W}}, \hat{P}_{VS\bar{W}Y}) := \{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} \in \mathcal{T}_{\mathcal{UVSWY}} : P_{\tilde{U}\tilde{S}\tilde{W}} = \hat{P}_{US\bar{W}},\ P_{\tilde{V}\tilde{S}\tilde{Y}} = \hat{P}_{VSY}, H(\tilde{W}|\tilde{V}, \tilde{S}, \tilde{Y}) \geq H_{\hat{P}}(\bar{W}|V, S, Y)\}.$$

The proof of Theorem 6 is given in Appendix B. The different factors inside the minimum in (23) can be intuitively understood to be related to the various events that could possibly lead to a type 2 error. More specifically, let the event that the encoder is unsuccessful in finding a codeword $\bar{W}^n$ in the quantization codebook that is typical with U^n be referred to as the *encoding error*, and the event that a wrong codeword $\hat{\bar{W}}^n$ (unintended by the encoder) is reconstructed at the detector be referred to as the *decoding error*. Then, $E_1'(P_S, P_{\bar{W}|US}, P_{X|US\bar{W}})$ is related to the event that neither the encoding nor the decoding error occurs, while $E_2'(P_S, P_{\bar{W}|US}, P_{X|US\bar{W}})$ and $E_3'(P_S, P_{\bar{W}|US}, P_{X'|S}, P_{X|US\bar{W}})$ are related to the events that only the decoding error and both the encoding and decoding errors occur, respectively. From Theorem 2, 5 and 6, we have the following corollary.

Corollary 7.

$$\kappa(1, \epsilon) \geq \max\left(\kappa_h, \kappa_0(1), \kappa_s(1)\right),\ \forall \epsilon \in (0, 1]. \quad (29)$$

It is well-known that in the context of JSCC, hybrid coding recovers separate source-channel coding as a special case. Since the SHTCC scheme performs independent channel coding and HT, and the JHTCC scheme uses hybrid coding for communication over DMC, it is tempting to think that this implies that $\kappa_h \geq \kappa_s(1)$. However, the schemes are not comparable in general, due to fact that $E_2'\left(P_S, P_{\bar{W}|US}, P_{X|US\bar{W}}\right)$ is not comparable to $E_3(P_{W|U}, P_{SX}, 1)$. One may ask when does the JHTCC scheme perform better than the SHTCC scheme. Towards answering this question, in Theorem 8 below, we obtain conditions under which $\kappa_h \geq \kappa_s(1)$. As a byproduct of the proof of Theorem 8, we also show that the JHTCC scheme achieves the optimal T2EE for TACI over a DMC.

Let $\bar{W}=(W,X)$, $\mathcal{S}=\mathcal{X}$, $X\perp(X',W,U,V)$, $(W,U,V)\perp S$ and $(U,V,W,S,X')-X-Y$ in Theorem 6, so that

$$\hat{P}_{UVSWX'XY}(P_S,P_{W|U},P_{X'|S},P_{X|S})=P_{UV}P_SP_{W|U}P_{X'|S}P_{X|S}P_{Y|X}, \tag{30}$$

$$\hat{Q}_{UVSWX'XY}(P_S,P_{W|U},P_{X'|S},P_{X|S})=Q_{UV}P_SP_{W|U}P_{X'|S}P_{X|S}P_{Y|X}, \tag{31}$$

$$\check{Q}_{UVSX'XY}(P_S,P_{X'|S})=Q_{UV}P_SP_{X'|S}\mathbb{1}(X=X')P_{Y|X}. \tag{32}$$

$$\mathcal{T}_1'(\hat{P}_{USWX},\hat{P}_{VSWXY}):=\{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{X}\tilde{Y}}\in\mathcal{T}_{\mathcal{UVSWXY}}:P_{\tilde{U}\tilde{S}\tilde{W}\tilde{X}}=\hat{P}_{USWX},\ P_{\tilde{V}\tilde{S}\tilde{W}\tilde{X}\tilde{Y}}=\hat{P}_{VSWXY}\},$$

$$\mathcal{T}_2'(\hat{P}_{USWX},\hat{P}_{VSWXY}):=\{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{X}\tilde{Y}}\in\mathcal{T}_{\mathcal{UVSWXY}}:P_{\tilde{U}\tilde{S}\tilde{W}\tilde{X}}=\hat{P}_{USWX},\ P_{\tilde{V}\tilde{S}\tilde{Y}}=\hat{P}_{VSY},$$
$$H(\tilde{W},\tilde{X}|\tilde{V},\tilde{S},\tilde{Y})\geq H_{\hat{P}}(W,X|V,S,Y)\}.$$

Let $(P^*_{W|U},P^*_{SX})\in\ \mathcal{B}(1,C)$ achieve the supremum in (11). Define

$$\hat{P}^*_{UVSWX'XY}(P^*_S,P^*_{W|U},P_{X'|S},P^*_{X|S}):=P_{UV}P^*_SP^*_{W|U}P_{X'|S}P^*_{X|S}P_{Y|X}, \tag{33}$$

$$\hat{Q}^*_{UVSWX'XY}(P^*_S,P^*_{W|U},P_{X'|S},P^*_{X|S}):=Q_{UV}P^*_SP^*_{W|U}P_{X'|S}P^*_{X|S}P_{Y|X}, \tag{34}$$

$$\check{Q}^*_{UVSX'XY}(P^*_S,P_{X'|S}):=Q_{UV}P_SP_{X'|S}\mathbb{1}(X=X')P_{Y|X}, \tag{35}$$

$$E_h\left(P^*_S,P^*_{W|U},P_{X'|S},P^*_{X|S}\right):=\min_{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{X}\tilde{Y}}\in\mathcal{T}_2'\left(\hat{P}^*_{USWX},\hat{P}^*_{VSWXY}\right)}D(P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{X}\tilde{Y}}||\hat{Q}^*_{UVSWXY}),$$

$$E_s\left(P^*_S,P^*_{W|U},P_{X'|S},P^*_{X|S}\right):=\min_{P_{\tilde{U}\tilde{V}\tilde{W}}\in\mathcal{T}_3\left(\hat{P}^*_{UW},\hat{P}^*_{V}\right)}D(P_{\tilde{U}\tilde{V}\tilde{W}}||\hat{Q}^*_{UVW})$$
$$+E_x(I_{\hat{P}^*}(X;Y|S),\hat{P}^*_{SX}).$$

Then, we have the following result.

Theorem 8. *If* $E_h\left(P^*_S,P^*_{W|U},P_{X'|S},P^*_{X|S}\right)\geq E_s\left(P^*_S,P^*_{W|U},P_{X'|S},P^*_{X|S}\right)$, *then* $\kappa_h\geq\kappa_s(1)$.

The proof of Theorem 8 is given in Appendix C.

Thus far, we obtained lower bounds on the optimal T2EE for distributed HT over a DMC. However, obtaining tight computable outer bounds is a challenging open problem, and consequently, an exact computable characterization of the optimal T2EE is unknown (even when the communication channel is noiseless). However, as we show in the next section, the problem does admit single-letter characterization in some special cases. These special cases are motivated from analogous results for distributed HT over rate-limited noiseless channels.

IV. Optimality results

A. TACI over a DMC

Recall that for TACI, $V=(E,Z)$ and $Q_{UEZ}=P_{UZ}P_{E|Z}$. Let

$$\kappa(\tau)=\lim_{\epsilon\to 0}\kappa(\tau,\epsilon). \tag{36}$$

We will drop the subscript P from information theoretic quantities like mutual information, entropy, etc., as there is no ambiguity on the joint distribution involved, e.g., $I_P(U;W)$ will be denoted by $I(U;W)$. The following result holds.

Proposition 9. *For TACI over a DMC $P_{Y|X}$,*

$$\kappa(\tau)=\sup\left\{\begin{aligned} I(E;W|Z):\exists\ W\ \textit{s.t.}\ I(U;W|Z)\leq \tau C(P_{Y|X}),\\ (Z,E)-U-W,\ |\mathcal{W}|\leq|\mathcal{U}|+1.\end{aligned}\right\},\ \tau\geq 0. \tag{37}$$

Proof: For the proof of achievability, we will show that $\kappa_s(\tau)$ when specialized to TACI recovers (37). Let
$\mathcal{B}'(\tau,C)$

$$:=\left\{\begin{aligned}(P_{W|U},P_{SX}):\ \ \mathcal{S}=\mathcal{X},\ P_{UEZWSXY}(P_{W|U},P_{SX}):=P_{UEZ}P_{W|U}P_{SX}P_{Y|X},\\ I(U;W|Z)\leq\tau I(X;Y|S)<\tau C\end{aligned}\right\}. \tag{38}$$

Note that $\mathcal{B}'(\tau,C)\subseteq\mathcal{B}(\tau,C)$ since $I(U;W|E,Z)\leq I(U;W|Z)$, which holds due to the Markov chain $(Z,E)-U-W$. Now, consider $(P_{W|U},P_{SX})\in\mathcal{B}'(\tau,C)$. Then, we have

$$\begin{aligned} E_1(P_{W|U}) &= \min_{P_{\tilde{U}\tilde{E}\tilde{Z}\tilde{W}}\in\mathcal{T}_1(P_{UW},P_{EZW})} D(P_{\tilde{U}\tilde{E}\tilde{Z}\tilde{W}}||P_Z P_{U|Z}P_{E|Z}P_{W|U})\\ &\geq \min_{P_{\tilde{U}\tilde{E}\tilde{Z}\tilde{W}}\in\mathcal{T}_1(P_{UW},P_{EZW})} D(P_{\tilde{E}\tilde{Z}\tilde{W}}||P_Z P_{E|Z}P_{W|Z}) \qquad (39)\\ &= I(E;W|Z), \end{aligned}$$

where (39) follows from the log-sum inequality [24]. Also,

$$\begin{aligned} E_2(P_{W|U},P_{SX},\tau) &\geq \tau I(X;Y|S)-I(U;W|E,Z)\geq I(U;W|Z)-I(U;W|E,Z)\\ &= I(E;W|Z), \end{aligned}$$

$$\min_{P_{\tilde{U}\tilde{E}\tilde{Z}\tilde{W}} \in \mathcal{T}_3(P_{UW}, P_{EZ})} D(P_{\tilde{U}\tilde{E}\tilde{Z}\tilde{W}}||P_Z P_{U|Z} P_{E|Z} P_{W|U}) + \tau I(X;Y|S) - I(U;W|E,Z)$$
$$+ \tau E_x\left(I(X;Y|S), P_{SX}\right)$$
$$\geq I(U;W|Z) - I(U;W|E,Z) = I(E;W|Z), \tag{40}$$
$$\min_{P_{\tilde{U}\tilde{E}\tilde{Z}\tilde{W}} \in \mathcal{T}_3(P_{UW}, P_{EZ})} D(P_{\tilde{U}\tilde{E}\tilde{Z}\tilde{W}}||P_Z P_{U|Z} P_{E|Z} P_{W|U}) + I(E,Z;W)$$
$$+ \tau E_x\left(I(X;Y|S), P_{SX}\right) \geq I(E;W|Z), \tag{41}$$
$$D(P_{EZ}||P_{EZ}) + \tau I(X;Y|S) - I(U;W|E,Z) + \tau E_m\left(P_{SX}\right)$$
$$\geq I(U;W|Z) - I(U;W|E,Z) = I(E;W|Z), \tag{42}$$
$$D(P_{EZ}||P_{EZ}) + I(E,Z;W) + \tau E_m\left(P_{SX}\right) \geq I(E;W|Z), \tag{43}$$

where in (40)-(43), we used the non-negativity of KL-divergence, $E_x(\cdot,\cdot,\cdot)$ and $E_m(\cdot,\cdot)$. Hence, $E_2(P_{W|U}, P_{SX}, \tau) \geq I(E;W|Z)$ and $E_3(P_{W|U}, P_{SX}, \tau) \geq I(E;W|Z)$. Denoting $\mathcal{B}(\tau, C)$ and $\mathcal{B}'(\tau, C)$ by $\mathcal{B}$ and $\mathcal{B}'$, respectively, we obtain

$$\kappa(\tau,\epsilon)$$
$$\geq \sup_{(P_{W|U}, P_{SX}) \in \mathcal{B}} \min\left(E_1(P_{W|U}), E_2(P_{W|U}, P_{SX}, \tau), E_3(P_{W|U}, P_{SX}, \tau), E_4(P_{W|U}, P_{SX}, \tau)\right)$$
$$\geq \sup_{(P_{W|U}, P_{SX}) \in \mathcal{B}} I(E;W|Z)$$
$$\geq \sup_{(P_{W|U}, P_{SX}) \in \mathcal{B}'} I(E;W|Z) \tag{44}$$
$$= \sup_{P_{W|U}: I(W;U|Z) \leq \tau C} I(E;W|Z), \tag{45}$$

where (44) follows from the fact that $\mathcal{B}' \subseteq \mathcal{B}$; and (45) follows since $I(E;W|Z)$ and $I(U;W|Z)$ are continuous functions of $P_{W|U}$.

Converse:

For any sequence of encoding functions $f^{(k,n_k)}$, acceptance regions $\mathcal{A}^{(k,n_k)}$ for H_0 such that $n_k \leq \tau k$ and

$$\limsup_{k \to \infty} \alpha\left(k, n_k, f^{(k,n_k)}, g^{(k,n_k)}\right) = 0, \tag{46}$$

we have similar to [5, Theorem 1 (b)], that

$$\limsup_{k\to\infty} \frac{-1}{k} \log\left(\beta\left(k, n_k, f^{(k,n_k)}, g^{(k,n_k)}\right)\right) \leq \limsup_{k\to\infty} \frac{1}{k} D\left(P_{Y^{n_k}E^kZ^k} || Q_{Y^{n_k}E^kZ^k}\right) \tag{47}$$

$$= \limsup_{n\to\infty} \frac{1}{k} I(Y^{n_k}; E^k|Z^k) \tag{48}$$

$$= H(E|Z) - \liminf_{k\to\infty} \frac{1}{k} H(E^k|Y^{n_k}, Z^k), \tag{49}$$

where (48) follows since $Q_{Y^{n_k}E^kZ^k} = P_{Y^{n_k}Z^k}P_{E^k|Z^k}$. Now, let T be a r.v. uniformly distributed over $[k]$ and independent of all the other r.v.'s $(U^k, E^k, Z^k, X^{n_k}, Y^{n_k})$. Define an auxiliary r.v. $W := (W_T, T)$, where $W_i := (Y^{n_k}, E^{i-1}, Z^{i-1}, Z_{i+1}^k)$, $i \in [k]$. Then, the last term can be single-letterized as follows.

$$\begin{aligned} H(E^k|Y^{n_k}, Z^k) &= \sum_{i=1}^k H(E_i|E^{i-1}, Y^{n_k}, Z^k) \\ &= \sum_{i=1}^k H(E_i|Z_i, W_i) \\ &= kH(E_T|Z_T, W_T, T) \\ &= kH(E|Z, W). \end{aligned} \tag{50}$$

Substituting (50) in (49), we obtain

$$\limsup_{k\to\infty} \frac{-1}{k} \log\left(\beta\left(k, n_k, f_1^{(k,n_k)}, g^{(k,n_k)}\right)\right) \leq I(E; W|Z). \tag{51}$$

Next, note that the data processing inequality applied to the Markov chain $(Z^k, E^k) - U^k - X^n - Y^n$ yields $I(U^k; Y^{n_k}) \leq I(X^{n_k}; Y^{n_k})$ which implies that

$$I(U^k; Y^{n_k}) - I(U^k; Z^k) \leq I(X^{n_k}; Y^{n_k}). \tag{52}$$

The R.H.S. of (52) can be upper bounded due to the memoryless nature of the channel as

$$I(X^{n_k}; Y^{n_k}) \leq n_k \max_{P_X} I(X; Y) = n_k C(P_{Y|X}), \tag{53}$$

while the left hand side (L.H.S.) can be simplified as follows.

$$I(U^k; Y^{n_k}) - I(U^k; Z^k) = I(U^k; Y^{n_k}|Z^k) \tag{54}$$

$$= \sum_{i=1}^k I(Y^{n_k}; U_i|U^{i-1}, Z^k)$$

$$= \sum_{i=1}^{k} I(Y^{n_k}, U^{i-1}, Z^{i-1}, Z_{i+1}^{k}; U_i|Z_i) \tag{55}$$

$$= \sum_{i=1}^{k} I(Y^{n_k}, U^{i-1}, Z^{i-1}, Z_{i+1}^{k}, E^{i-1}; U_i|Z_i) \tag{56}$$

$$\geq \sum_{i=1}^{k} I(Y^{n_k}, Z^{i-1}, Z_{i+1}^{k}, E^{i-1}; U_i|Z_i)$$

$$= \sum_{i=1}^{k} I(W_i; U_i|Z_i) = kI(W_T; U_T|Z_T, T)$$

$$= kI(W_T, T; U_T|Z_T) \tag{57}$$

$$= kI(W; U|Z).$$

Here, (54) follows due to $Z^k - U^k - Y^{n_k}$; (55) follows since the sequences (U^k, Z^k) are memoryless; (56) follows since $E^{i-1} - (Y^{n_k}, U^{i-1}, Z^{i-1}, Z_{i+1}^{k}) - U_i$; (57) follows from the fact that T is independent of all the other r.v.'s. Finally, note that $(E, Z) - U - W$ holds and that the cardinality bound on W follows by standard arguments based on Caratheodory's theorem. This completes the proof of the converse, and of the proposition. ■

As the above result shows, TACI is an instance of distributed HT over a DMC, in which, the optimal T2EE is equal to that achieved over a noiseless channel of the same capacity. Hence, a noisy channel does not always degrade the achievable T2EE. Also, notice that a simple separation based coding scheme that performs independent HT and channel coding is sufficient to achieve the optimal T2EE for TACI. From (40)-(43), we observe that this happens due to the fact that $E_3(P_{W|U}, P_{SX}, \tau)$ and $E_4(P_{W|U}, P_{SX}, \tau)$ are both larger than $I(E; W|Z)$. This can be explained intuitively as follows. For the scheme discussed in Appendix A that achieves a T2EE of $\kappa_s(\tau)$, a type 2 error occurs only when the detector decodes a codeword $\hat{W}^k$ that is jointly typical with the side information sequence V^k. For the case of TACI, when H_1 is the true hypothesis, then with high probability, the codeword $W^k(J)$ chosen by the encoder is not jointly typical with V^k, i.e., $(V^k, W^k(J)) \notin T_{[P_{VW}]_\delta}^{k}$. Then, the above phenomenon corroborates the fact that given an error occurs at the channel decoder, the probability that two independently chosen sequences V^k and $\hat{W}^k$ are such that $(V^k, \hat{W}^k) \in T_{[P_{VW}]_\delta}^{k}$, decays as $e^{-kI(V;W)}$.

We can also show that the JHTCC scheme achieves the optimal T2EE for TACI. The proof of this claim is given in Appendix D.

B. Distributed HT over a DMC with zero capacity

Next, we show that the LD scheme achieves the optimal T2EE when $C(P_{Y|X}) = 0$. Note that when the channel has zero capacity, the reliability function of the channel is zero for any positive rate of transmission, i.e., when there are exponential number of messages $e^{n\delta}$, where $\delta > 0$ is bounded away from zero.

Theorem 10. *If* $C(P_{Y|X}) = 0$, *then* $\kappa(\tau,\epsilon) = D(P_V||Q_V),\ \forall\ \epsilon \in (0,1),\ \tau \geq 0$.

Proof: The achievability follows from Theorem 5 which states that for $\tau \geq 0$, $\kappa(\tau,\epsilon) \geq \kappa_0(\tau)$, $\forall\ \epsilon \in (0,1]$. Now, it is well-known (see [24]) that $C(P_{Y|X}) = 0$ only if

$$P_Y^* := P_{Y|X=x} = P_{Y|X=x'},\ \forall\ x, x' \in \mathcal{X}. \tag{58}$$

From (58), it follows that $E_c(P_{Y|X}) = 0$. Also,

$$\begin{aligned}\beta_0 &\geq D(P_V||Q_V) + \min_{\substack{P_{\tilde{U}\tilde{V}}: \\ P_{\tilde{U}}=P_U,\ P_{\tilde{V}}=P_V}} D(P_{\tilde{U}|\tilde{V}}||Q_{U|V}\big|P_{\tilde{V}}) \\ &\geq D(P_V||Q_V),\end{aligned}$$

which implies that $\kappa_0(\tau) \geq D(P_V||Q_V)$.

Converse: We first show the weak converse, i.e., $\kappa(\tau) \leq D(P_V||Q_V)$, where $\kappa(\tau)$ is as defined in (36). For any sequence of encoding functions $f^{(k,n_k)}$ and acceptance regions $\mathcal{A}^{(k,n_k)}$ for H_0 that satisfy $n_k \leq \tau k$ and (46), it follows similarly to (47), that

$$\limsup_{k\to\infty} \frac{-1}{k} \log\left(\beta\left(k, n_k, f^{(k,n_k)}, g^{(k,n_k)}\right)\right) \leq \limsup_{k\to\infty} \frac{1}{k} D\left(P_{Y^{n_k}V^k}||Q_{Y^{n_k}V^k}\right) \tag{59}$$

The terms in the R.H.S. of (59) can be expanded as

$$\begin{aligned}&\frac{1}{k} D\left(P_{Y^{n_k}V^k}||Q_{Y^{n_k}V^k}\right) \\ &= D(P_V||Q_V) + \frac{1}{k}\left(\sum_{(v^k,y^{n_k})\in\mathcal{V}^k\times\mathcal{Y}^{n_k}} P_{V^kY^{n_k}}(v^k,y^{n_k}) \log\left(\frac{P_{Y^{n_k}|V^k}(y^{n_k}|v^k)}{Q_{Y^{n_k}|V^k}(y^{n_k}|v^k)}\right)\right)\end{aligned} \tag{60}$$

Now, note that

$$\begin{aligned}P_{Y^{n_k}|V^k}(y^{n_k}|v^k) &= \sum_{(u^k,x^{n_k})\in\ \mathcal{U}^k\times\mathcal{X}^{n_k}} P_{U^k|V^k}(u^k|v^k)P_{X^{n_k}|U^k}(x^{n_k}|u^k)P_{Y^{n_k}|X^{n_k}}(y^{n_k}|x^{n_k}) \\ &= \left(\prod_{i=1}^{n_k} P_Y^*(y_i)\right) \sum_{(u^k,x^{n_k})\in\ \mathcal{U}^k\times\mathcal{X}^{n_k}} P_{U^k|V^k}(u^k|v^k)P_{X^{n_k}|U^k}(x^{n_k}|u^k)\end{aligned} \tag{61}$$

$$= \prod_{i=1}^{n_k} P_Y^*(y_i), \tag{62}$$

where, (61) follows from (3) and (58). Similarly, it follows that

$$Q_{Y^{n_k}|V^k}(y^{n_k}|v^k) = \prod_{i=1}^{n_k} P_Y^*(y_i). \tag{63}$$

From (59), (60), (62) and (63), we obtain that

$$\limsup_{k \to \infty} \frac{-1}{k} \log \left(\beta \left(k, n_k, f^{(k,n_k)}, g^{(k,n_k)} \right) \right) \leq D(P_V || Q_V). \tag{64}$$

This completes the proof of the weak converse.

Next, we proceed to show that $D(P_V||Q_V)$ is the optimal T2EE for every $\epsilon \in (0,1)$. For any fixed $\epsilon \in (0,1)$, let $f^{(k,n_k)}$ and $\mathcal{A}^{(k,n_k)}$ denote any encoding function and acceptance region for H_0, respectively, such that $n_k \leq \tau k$ and

$$\limsup_{k \to \infty} \alpha \left(k, n_k, f^{(k,n_k)}, g^{(k,n_k)} \right) \leq \epsilon. \tag{65}$$

The joint distribution of (V^k, Y^{n_k}) under the null and alternate hypothesis is given by

$$P_{V^k Y^{n_k}}(v^k, y^{n_k}) = \left(\prod_{i=1}^{k} P_V(v_i) \right) \left(\prod_{j=1}^{n_k} P_Y^*(y_j) \right), \tag{66}$$

$$\text{and } Q_{V^k Y^{n_k}}(v^k, y^{n_k}) = \left(\prod_{i=1}^{k} Q_V(v_i) \right) \left(\prod_{j=1}^{n_k} P_Y^*(y_j) \right), \tag{67}$$

respectively. By the weak law of large numbers, for any $\delta > 0$, (66) implies that

$$\lim_{k \to \infty} P_{V^k Y^{n_k}} \left(T^k_{[P_V]_\delta} \times T^{n_k}_{[P_Y^*]_\delta} \right) = 1. \tag{68}$$

Also, from (65), we have

$$\liminf_{k \to \infty} P_{V^k Y^{n_k}} \left(\mathcal{A}^{(k,n_k)} \right) \geq (1 - \epsilon). \tag{69}$$

From (68) and (69), it follows that

$$P_{V^k Y^{n_k}} \left(\mathcal{A}^{(k,n_k)} \cap T^k_{[P_V]_\delta} \times T^{n_k}_{[P_Y^*]_\delta} \right) \geq 1 - \epsilon', \tag{70}$$

for any $\epsilon' > \epsilon$ and k sufficiently large ($k \geq k_0(\delta, |\mathcal{V}|, |\mathcal{Y}|)$). Let

$$\mathcal{A}(v^k,\delta) := \left\{ y^{n_k} : (v^k, y^{n_k}) \in \mathcal{A}^{(k,n_k)} \cap T^k_{[P_V]_\delta} \times T^{n_k}_{[P_Y^*]_\delta} \right\}, \tag{71}$$

$$\text{and } \mathcal{D}(\eta,\delta) := \left\{ v^k \in T^k_{[P_V]_\delta} : P_{Y^{n_k}}(\mathcal{A}(v^k,\delta)) \geq \eta \right\}. \tag{72}$$

Fix $0 < \eta' < 1 - \epsilon'$. Then, we have from (70) that for any $\delta > 0$ and sufficiently large k,

$$P_{V^k}\left(\mathcal{D}(\eta',\delta)\right) \geq \frac{1-\epsilon'-\eta'}{1-\eta'}. \tag{73}$$

From [24, Lemma 2.14], (73) implies that $\mathcal{D}(\eta',\delta)$ should contain atleast $\frac{1-\epsilon'-\eta'}{1-\eta'}$ fraction (approx.) of sequences in $T^k_{[P_V]_\delta}$ and for each $v^k \in \mathcal{D}(\eta',\delta)$, (72) implies that $\mathcal{A}(v^k,\delta)$ should contain atleast η' fraction (approx.) of sequences in $T^{n_k}_{[P_Y^*]_\delta}$, asymptotically. Hence, for sufficiently large k, we have

$$Q_{V^kY^{n_k}}\left(\mathcal{A}^{(k,n_k)}\right) \geq \sum_{v^k \in \mathcal{D}(\eta',\delta)} Q_{V^k}(v^k) \sum_{y^{n_k} \in \mathcal{A}(v^k,\delta)} P_{Y^n}(y^{n_k}) \tag{74}$$

$$\geq e^{-k\left(D(P_V||Q_V) - \frac{\log\left(\frac{1-\epsilon'-\eta'}{1-\eta'}\right)}{k} - \frac{\log(\eta')}{k} - O(\delta)\right)}. \tag{75}$$

Here, (75) follows from [24, Lemma 2.6].

Let $\mathcal{A}_0^{(k,n_k)} := T^k_{[P_V]_\delta} \times T^{n_k}_{[P_Y^*]_\delta}$. Then, for sufficiently large k,

$$P_{V^kY^{n_k}}\left(\mathcal{A}_0^{(k,n_k)}\right) \xrightarrow{(k)} 1, \text{ and} \tag{76}$$

$$Q_{V^kY^{n_k}}\left(\mathcal{A}_0^{(k,n_k)}\right) \leq e^{-k(D(P_V||Q_V) - O(\delta))}, \tag{77}$$

where, (76) and (77) follows from weak law of large numbers and [24, Lemma 2.6], respectively. Together (75), (76) and (77) implies that

$$|\kappa(\tau,\epsilon) - \kappa(\tau)| \leq O(\delta), \tag{78}$$

and the theorem is proved since $\delta > 0$ is arbitrary. ∎

Remark 11. *Theorem 10 shows that when the capacity of a DMC is zero, then no communication from the observer to the detector helps in terms of the T2EE. To contrast this with the optimal*

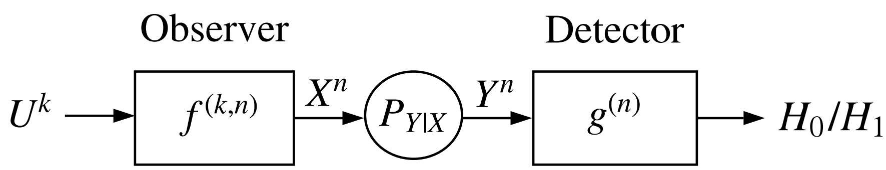

Fig. 2: Hypothesis testing over a noisy channel.

T2EE β_0 (see (16)) for the case when the channel is noiseless, note that

$$\beta_0 \geq D(P_V||Q_V) + \min_{\substack{P_{\tilde{U}\tilde{V}}: \\ P_{\tilde{U}}=P_U,\ P_{\tilde{V}}=P_V}} D(P_{\tilde{U}|\tilde{V}}||Q_{U|V}\big|P_{\tilde{V}}). \tag{79}$$

Since $P_{\tilde{U}\tilde{V}}$ achieving the minimum in (79) has to satisfy

$$\sum_{v \in \mathcal{V}} P_V(v) P_{\tilde{U}|\tilde{V}}(u|v) = P_U(u),\ \forall\ u \in \mathcal{U}, \tag{80}$$

and $D(P_1||P_2) > 0$ for probability distributions $P_1 \neq P_2$, it is clear that $\beta_0 > D(P_V||Q_V)$ if for some $u \in \mathcal{U}$,

$$\sum_{v \in \mathcal{V}} P_V(v) Q_{U|V}(u|v) \neq P_U(u). \tag{81}$$

Hence, in general, communication (even a single bit of information) between the observer and the detector helps to improve the T2EE compared to the scenario when there is no communication.

C. HT over a DMC

Consider now HT over a noisy channel as depicted in Fig. 2, in which, the side-information V^k is absent and the detector performs the following hypothesis test:

$$H_0 : U^k \sim \prod_{i=1}^{k} P_U, \tag{82a}$$

$$H_1 : U^k \sim \prod_{i=1}^{k} Q_U. \tag{82b}$$

When the observations U^k are available directly at the detector, a single-letter characterization of the optimal T2EE for a given constraint ϵ on the type 1 error probability is known, and given by

$$\kappa(\epsilon) = D(P_U||Q_U), \ \forall \ \epsilon \in (0,1). \tag{83}$$

Notice that a strong converse holds in this case, in the sense that, $\kappa(\epsilon)$ is independent of ϵ.

If the detector and the observer are connected with a noise-free link of capacity $R > 0$, it is easy to see that the T2EE in (83) can be achieved by performing the Neyman-Pearson test locally at the observer and transmitting the decision to the detector over the noiseless link. When the communication channel is noisy, however, it is unclear whether such a local decision scheme would still remain optimal. More specifically, since the reliability of the transmitted messages depends on the communication rate employed, there is a trade-off between transmitting less information more reliably versus transmitting more information less reliably, to the detector. In the sequel, we show that making a decision locally at the observer, and communicating it to the detector is indeed optimal. The optimal scheme is in fact an adaptation of the LD scheme to the case when the side information V^k is absent.

The problem formulation and definitions in Section II-B carry over as such without V^k (or by assuming V^k is a constant). We will denote the decision rule $g^{(k,n)}$ as $g^{(n)}$ since it is a function of Y^n only. Also, to differentiate between distributed HT and the current setting, we will denote the maximum achievable T2EE by $\kappa'(\tau, \epsilon)$. Let

$$\kappa_0'(\tau) := \begin{cases} 0 & , \text{ if } \tau = 0, \\ \min\left(D(P_U||Q_U), \tau E_c\right), & \text{otherwise,} \end{cases} \tag{84}$$

where E_c is as defined in (17). Note that E_c can take the value of ∞ in general. The following result provides a single-letter characterization of the optimal T2EE, and also shows that a strong converse holds.

Theorem 12. $\kappa'(\tau, \epsilon) = \kappa_0'(\tau), \ \forall \epsilon \in (0,1), \ \tau \geq 0.$

Proof: We prove the result in three steps as follows:

(i) $\kappa'(\tau, \epsilon) \geq \kappa_0'(\tau), \ \forall \ \epsilon \in (0,1).$

(ii) $\lim_{\epsilon \to 0} \kappa'(\tau, \epsilon) = \kappa_0'(\tau).$

(iii) $\kappa'(\tau, \epsilon) \leq \kappa_0'(\tau), \ \forall \ \epsilon \in (0,1).$

The proof of (i) follows from Theorem 5 by setting V^k equal to a constant under both hypotheses.

To show part (ii), we will prove the weak converse

$$\lim_{\epsilon\to 0}\kappa'(\tau,\epsilon)\leq\kappa'_0(\tau), \tag{85}$$

which combined with part (i) proves part (ii). Similarly to [5, Theorem 1(b)], it follows that for any sequence of encoding functions $f^{(k,n_k)}$ and decision rules $g^{(n_k)}$ satisfying $n_k \leq \tau k$ and (46), we have

$$\limsup_{k\to\infty}\frac{-1}{k}\log\left(\beta\left(k,n_k,f^{(k,n_k)},g^{(n_k)}\right)\right)\leq\frac{1}{k}D(P_{Y^{n_k}}||Q_{Y^{n_k}}). \tag{86}$$

If $\tau = 0$, then $n_k = 0$, and hence, the R.H.S. of (86) is zero, thus proving (85). Now, assume that $\tau > 0$. Then, the R.H.S. of (86) can be upper bounded as follows.

$$\begin{aligned}
D(P_{Y^{n_k}}||Q_{Y^{n_k}}) &= \sum_{i=1}^{n_k} D(P_{Y_i|Y^{i-1}}||Q_{Y_i|Y^{i-1}}|P_{Y^{i-1}}) \qquad (87)\\
&= \sum_{i=1}^{n_k}\sum_{y^{i-1}\in\mathcal{Y}^{i-1}} P_{Y^{i-1}}(y^{i-1})\left[\sum_{y_i\in\mathcal{Y}} P_{Y_i|Y^{i-1}}(y_i|y^{i-1})\log\left(\frac{P_{Y_i|Y^{i-1}}(y_i|y^{i-1})}{Q_{Y_i|Y^{i-1}}(y_i|y^{i-1})}\right)\right]\\
&= \sum_{i=1}^{n_k}\sum_{y^{i-1}\in\mathcal{Y}^{i-1}} P_{Y^{i-1}}(y^{i-1})\ D\left(P_{Y_i|Y^{i-1}=y^{i-1}}||Q_{Y_i|Y^{i-1}=y^{i-1}}\right).
\end{aligned}$$

Since

$$P_{Y_i|Y^{i-1}}(y_i|y^{i-1}) = \sum_{x_i\in\mathcal{X}} P_{X_i|Y^{i-1}}(x_i|y^{i-1})P_{Y_i|X_i}(y_i|x_i), \tag{88}$$

$$\text{and } Q_{Y_i|Y^{i-1}}(y_i|y^{i-1}) = \sum_{x_i\in\mathcal{X}} Q_{X_i|Y^{i-1}}(x_i|y^{i-1})P_{Y_i|X_i}(y_i|x_i), \tag{89}$$

we can write

$$\begin{aligned}
&D(P_{Y^{n_k}}||Q_{Y^{n_k}})\\
&\leq \sum_{i=1}^{n_k}\sum_{y^{i-1}\in\mathcal{Y}^{i-1}} P_{Y^{i-1}}(y^{i-1}) \sup_{\substack{\left(P_{X_i|Y^{i-1}=y^{i-1}},\right.\\ \left.Q_{X_i|Y^{i-1}=y^{i-1}}\right)}} D\left(P_{Y_i|Y^{i-1}=y^{i-1}}||Q_{Y_i|Y^{i-1}=y^{i-1}}\right). \qquad (90)
\end{aligned}$$

It follows from (88), (89), and the convexity of $D(P_X||Q_X)$ in (P_X, Q_X) that, $D\left(P_{Y_i|Y^{i-1}=y^{i-1}}||Q_{Y_i|Y^{i-1}=y^{i-1}}\right)$ is a convex function of $(P_{X_i|Y^{i-1}=y^{i-1}}, Q_{X_i|Y^{i-1}=y^{i-1}})$ for any $y^{i-1}\in\mathcal{Y}^{i-1}$. It is well-known that the maximum of a convex function over a convex feasible set is achieved at the extreme points of the feasible set. Since the extreme points of the probability simplex $P_{\mathcal{X}}$ are probability distributions of the form

$$P_{X_i}(x) = \mathbb{1}(x = x'),\ x \in \mathcal{X}, \tag{91}$$

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

$$\begin{aligned} \sup_{\substack{\left(P_{X_i|Y^{i-1}=y^{i-1}},\right.\\ \left.Q_{X_i|Y^{i-1}=y^{i-1}}\right)}} D\left(P_{Y_i|Y^{i-1}=y^{i-1}}||Q_{Y_i|Y^{i-1}=y^{i-1}}\right) &= D\left(P_{Y_i|X_i=h_{1i}(y^{i-1})}||P_{Y_i|X_i=h_{2i}(y^{i-1})}\right) \\ &\leq \max_{(x,x') \in \mathcal{X} \times \mathcal{X}} D\left(P_{Y|X=x}||P_{Y|X=x'}\right) = E_c. \end{aligned} \tag{92}$$

Thus, it follows from (90) and (92) that

$$\frac{1}{k} D(P_{Y^{n_k}}||Q_{Y^{n_k}}) \leq \frac{n_k}{k} E_c. \tag{93}$$

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

$$\frac{1}{k} D(P_{Y^{n_k}}||Q_{Y^{n_k}}) \leq \frac{1}{k} D(P_{U^k}||Q_{U^k}) = D(P_U||Q_U). \tag{94}$$

Hence, it follows from from (86), (93), (94), and the fact that $n_k \leq \tau k$, that,

$$\limsup_{k \to \infty} \frac{-1}{k} \log\left(\beta\left(k, n_k, f^{(k,n_k)}, g^{(n_k)}\right)\right) \leq \min\left(D(P_U||Q_U), \tau E_c\right). \tag{95}$$

Noting that the R.H.S. of (95) is independent of $(f^{(k,n_k)}, g^{(n_k)})$, the proof of (85) is completed by taking the supremum with respect to $(f^{(k,n_k)}, g^{(n_k)})$.

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

$$\kappa'(\tau, \epsilon) \leq \kappa_0'(\tau),\ \forall\ \epsilon \in (0,1). \tag{96}$$

If $\tau = 0$, then $n_k = 0$, and (96) holds. Now, assume $\tau > 0$. For $k \in \mathbb{Z}^+$, let $f^{(k,n_k)}$ and $g^{(n_k)}$ be any sequence of encoding functions and decision rules such that $n_k \leq \tau k$ and (65) is satisfied. Let $\mathcal{A}^{(n_k)}$ denote the acceptance region corresponding to $g^{(n_k)}$. For fixed $\gamma > 0$ and $\delta > 0$, let

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

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

$$\mathbb{P}\left(U^k \in T^k_{[P_U]_\delta}|H=0\right) \geq 1-\gamma'. \tag{97}$$

Then, it follows from (65) and (97) that

$$\mathbb{P}\left(U^k \in \mathcal{B}^{(k,n_k)}_{\gamma,\delta}|H=0\right) \geq \frac{1-\epsilon-\gamma}{1-\gamma} - \gamma'. \tag{98}$$

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

$$\mathbb{P}\left(U^k \in \mathcal{B}^{(k,n_k)}_{\gamma,\delta}|H=0\right) \geq \frac{1-\epsilon}{2(1+\epsilon)}. \tag{99}$$

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

$$P_{Y^{n_k}|X^{n_k}}\left(\mathcal{A}^{(n_k)}|\bar{x}^{n_k}\right) \geq \gamma, \tag{100}$$

$$\text{and } P_{X^{n_k}|U^k}\left(\bar{x}^{n_k}|u^k\right) > 0. \tag{101}$$

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

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

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

Due to (99), it follows by the application of the blowing-up lemma [24] that there exists sequences of non-negative numbers, $\{\lambda_k\}_{k \in \mathcal{Z}^+}$ and $\{l_k\}_{k \in \mathcal{Z}^+}$ such that, $\lambda_k \xrightarrow{(k)} 0$, $\frac{l_k}{k} \xrightarrow{(k)} 0$ and

$$P_{Y^{n_k}|X^{n_k}}\left(\Gamma^{l_k}\left(\mathcal{A}^{(n_k)}\right)|\bar{x}^{n_k}\right) \geq 1-\lambda_k. \tag{102}$$

Let $\bar{\mathcal{A}}^{(n_k)} := \Gamma^{l_k}\left(\mathcal{A}^{(n_k)}\right)$.

Suppose $E_c < \infty$. Then, $P_{Y|X}(y|x) > 0$, $\forall\ (x,y) \in \mathcal{X} \times \mathcal{Y}$. Let

$$\underline{v} := \min_{(x,y) \in \mathcal{X} \times \mathcal{Y}} P_{Y|X}(y|x) > 0. \tag{103}$$

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

$$P_{Y^{n_k}|X^{n_k}}(\bar{y}^{n_k}|x^{n_k})\ \underline{v}^{l_k} \leq P_{Y^{n_k}|X^{n_k}}(y^{n_k}|x^{n_k}). \tag{104}$$

Also, for each $y^{n_k} \in \mathcal{A}^{(n_k)}$, the number of $\bar{y}^{n_k} \in \bar{\mathcal{A}}^{(n_k)}$ is $|\mathcal{Y}|^{l_k}$. Hence, from (104), we have

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

This implies that

$$P_{Y^{n_k}|U^k}\left(\bar{\mathcal{A}}^{(n_k)}|u^k\right) \leq |\mathcal{Y}|^{l_k} P_{Y^{n_k}|U^k}\left(\mathcal{A}^{(n_k)}|u^k\right) \underline{v}^{-l_k}. \tag{105}$$

Let $\tilde{f}^{(k,n_k)} : \mathcal{U}^k \to \mathcal{X}^{n_k}$ and $\tilde{g}^{(n_k)} : \mathcal{Y}^{n_k} \to \{0,1\}$ be defined as follows:

$$\tilde{f}^{(k,n_k)}(u^k) = \begin{cases} \bar{x}^{n_k}, & \forall\ u^k \in T^k_{[P_U]_\delta} \\ f^{(k,n_k)}(u^k), & \text{otherwise,} \end{cases} \tag{106a}$$

$$\text{and } \tilde{g}^{(n_k)}(y^{n_k}) := 1 - \mathbb{1}\left(y^{n_k} \in \bar{\mathcal{A}}^{(n_k)}\right). \tag{106b}$$

From (97), (102) and (106), it follows that

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

Also,

$$\begin{aligned}
&\beta\left(k, n_k, \tilde{f}^{(k,n_k)}, \tilde{g}^{(n_k)}\right) \\
&\leq \sum_{u^k \in T^k_{[P_U]_\delta}} Q_{U^k}(u^k) + \sum_{u^k \notin T^k_{[P_U]_\delta}} Q_{U^k}(u^k) P_{Y^{n_k}|U^k}\left(\bar{\mathcal{A}}^{(n_k)}|u^k\right) \\
&\leq \sum_{u^k \in T^k_{[P_U]_\delta}} Q_{U^k}(u^k) + \underline{v}^{-l_k}|\mathcal{Y}|^{l_k} \sum_{u^k \notin T^k_{[P_U]_\delta}} Q_{U^k}(u^k) P_{Y^{n_k}|U^k}\left(\mathcal{A}^{(n_k)}|u^k\right) \qquad (107) \\
&\leq \sum_{u^k \in T^k_{[P_U]_\delta}} Q_{U^k}(u^k) + \underline{v}^{-l_k}|\mathcal{Y}|^{l_k} \sum_{u^k \in \mathcal{U}^k} Q_{U^k}(u^k) P_{Y^{n_k}|U^k}\left(\mathcal{A}^{(n_k)}|u^k\right) \\
&= \sum_{u^k \in T^k_{[P_U]_\delta}} Q_{U^k}(u^k) + \underline{v}^{-l_k}|\mathcal{Y}|^{l_k} \beta\left(k, n_k, f^{(k,n_k)}, g^{(n_k)}\right) \\
&\leq e^{-k(D(P_U||Q_U) - O(\delta))} + \underline{v}^{-l_k}|\mathcal{Y}|^{l_k} \beta\left(k, n_k, f^{(k,n_k)}, g^{(n_k)}\right),
\end{aligned}$$

where (107) follows from (105). Thus, it follows from the facts $\frac{l_k}{k} \xrightarrow{(k)} 0$ and $\underline{v} > 0$ that, for any $\gamma'' > 0$,

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

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

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

Since $f^{(k,n_k)}$ and $g^{(n_k)}$ is arbitrary, (96) follows.

Now, suppose $E_c=\infty$. Then, $\kappa_0'(\tau)=D(P_U||Q_U)$. Noting that $D(P_U||Q_U)$ is the maximum T2EE achievable for any type 1 error probability constraint $\epsilon \in (0,1)$, when U^k is directly observed at the detector, it follows that (96) holds. This completes the proof of the theorem. ■

V. Distributed HT with multiple observers

Thus far, we have focused on distributed HT with a single observer communicating to the detector over a DMC. In this section, we will extend our results to the distributed hypothesis test given in (1), where, there are multiple observers communicating their observations to the detector over orthogonal DMCs that satisfy the probability law given in (3). We will focus on TACI, and obtain a lower and upper bound on the optimal T2EE. To do this, we follow the method in [5], and first show an equivalence between the above problem and a *JSCC problem in the presence of noisy helpers* that will be introduced below. The desired bounds are then obtained indirectly via the best known inner and outer bounds for the equivalent problem. As a corollary, we provide yet another proof of the single-letter characterization of the optimal T2EE.

Let

$$\theta(\tau) := \sup_{k\in\mathbb{Z}^+} \theta(k,\tau), \tag{108}$$

$$\text{where } \theta(k,\tau) := \sup_{\substack{f_1^{(k,n)},\ldots,f_L^{(k,n)}\\ n\leq \tau k}} \frac{D\big(P_{Y_{\mathcal{L}}^n V^k}||Q_{Y_{\mathcal{L}}^n V^k}\big)}{k}. \tag{109}$$

We have the following multi-letter characterization of the optimal T2EE in terms of $\theta(\tau)$ whose proof follows along similar lines to [5, Theorem 1].

Lemma 13. *For $\tau \in \mathbb{R}^+$,*

(i) $\kappa(\tau,\epsilon) \geq \theta(\tau),\ \forall\ \epsilon \in (0,1]$.

(ii) $\lim_{\epsilon \to 0} \kappa(\tau,\epsilon) \leq \theta(\tau)$.

Proof: The proof can be found in Appendix E. ■

Part (i) and (ii) of Lemma 13 together imply that $\theta(\tau)$ is the optimal T2EE as $\epsilon \to 0$, i.e., $\kappa(\tau) = \theta(\tau)$. Recall that for TACI with multiple observers, $V = (E,Z)$ and $Q_{U_1 \ldots U_L EZ} = P_{U_1 \ldots U_L Z} P_{E|Z}$. In this case, the KL-divergence in (109) becomes mutual information, and we have

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

By the memoryless property of E^k and Z^k, we can write

$$\theta(\tau) = H(E|Z) - \inf_{\substack{f_1^{(k,n)},\ldots,f_L^{(k,n)}\\ k,n \leq \tau k}} \frac{H(E^k|Y_{\mathcal{L}}^n, Z^k)}{k} \tag{110}$$
$$\text{s.t. } (Z^k,E^k) - U_l^k - X_l^n - Y_l^n,\ l \in \mathcal{L}.$$

The last term in the R.H.S. of (110) can be identified as the multi-letter characterization of the source coding rate in an $L-$helper JSCC problem, as we show next.

A. $L-$helper JSCC problem

Consider the model shown in Fig. 3 where there are $L+2$ correlated DMS's $(U_{\mathcal{L}}, E, Z)$ with joint distribution $P_{U_1 \ldots U_L EZ}$. For $1 \leq l \leq L$, encoder $f_l^{(k,n)} : \mathcal{U}_l^k \to \mathcal{X}_l^n$ of helper l 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 the main encoder $f_s^k : \mathcal{E}^k \to \mathcal{M}$ observes E^k, and outputs an index $M = f_s^k(E^k)$. Decoder $g_s^{(k,n)} : \mathcal{Y}_{\mathcal{L}}^n \times \mathcal{M} \times \mathcal{Z}^k \to \hat{\mathcal{E}}^k$ observes $Y_{\mathcal{L}}^n$, receives M error-free and has access to side-information Z^k, based on which, it outputs $\hat{E}^k$. The goal of the decoder is to reconstruct E^k losslessly. Before establishing the multi-letter characterization of the rate region of the $L-$helper JSCC, we require a few definitions.

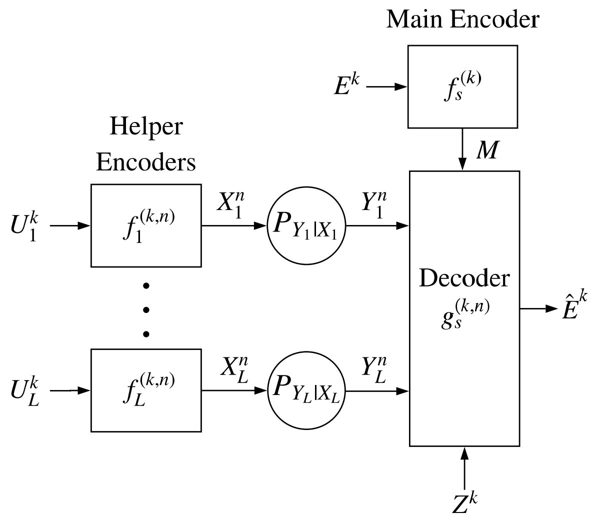

Fig. 3: $L-$helper JSCC problem.

Definition 14. *For a given bandwidth ratio τ, a rate R is said to be λ-achievable for the L-helper JSCC problem if there exist encoders f_s^k, $f_l^{(k,n_k)}$, $1 \leq l \leq L$, and decoder $g_s^{(k,n_k)}$ such that $n_k \leq \tau k$ and*

$$\limsup_{k\to\infty} \mathbb{P}\left(g_s^{(k,n_k)}\left(Y_{\mathcal{L}}^{n_k}, M, Z^k\right) \neq E^k\right) \leq \lambda,$$

$$\textit{and } \limsup_{k\to\infty} \frac{\log(|\mathcal{M}|)}{k} \leq R.$$

Let

$$R(\tau) := \inf\{R : R \textit{ is } \lambda\textit{-achievable for every } \lambda \in (0,1].\} \tag{111}$$

Define

$$R_k := \inf_{\substack{f_1^{(k,n)},\ldots,f_L^{(k,n)},\\ n\leq\tau k}} \frac{H(E^k|Y_{\mathcal{L}}^n, Z^k)}{k} \tag{112}$$

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

The equivalence between the multi-letter characterizations of $\theta(\tau)$ and $R(\tau)$ follows from (110) and the theorem stated below.

Theorem 15. *For the $L-$helper JSCC problem,*

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

Proof: The proof is given in Appendix F. ■

To obtain computable single-letter lower and upper bounds on $R(\tau)$, we use the *source-channel separation theorem* [30, 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 rate-region. We need a slight generalization of this result when there is side information Z at the decoder, which can be proved similarly to [30]. Note that the $L-$helper JSCC problem is a special case of the MT-JSCC problem with $L+1$ correlated sources $U_1, \ldots, U_L, E$ and side information Z available at the decoder, where the objective is to reconstruct source E losslessly. Although the source-channel separation theorem proves that separation holds, a single-letter expression is not available in general for the multi-terminal rate distortion problem [33]. However, single-letter inner and outer bounds are known. For simplicity, we will use the well-known BT bounds [28] [29] for our purpose. However, as will be apparent, these bounds may be replaced by any known inner and outer bound available in the literature. In particular, it is well-known that in some cases, the BT inner and outer bounds are strictly outperformed by the bounds in [34], [35] and [36], [37], respectively, and hence, tighter bounds on the optimal T2EE can be obtained by replacing the BT bounds with these bounds. Next, we present our result.

For $\mathcal{G} \subseteq \mathcal{L}$, let

$$C_{\mathcal{G}} := C_{\mathcal{G}}(P_{Y_1|X_1}, \ldots, P_{Y_L|X_L}) := \sum_{l \in \mathcal{G}} C_l(P_{Y_l|X_l}), \tag{113}$$

where $C_l := C_l(P_{Y_l|X_l}) := \max_{P_{X_l}} I(X_l; Y_l)$, $l \in \mathcal{L}$, denotes the capacity of the channel $P_{Y_l|X_l}$. For $\tau \in \mathbb{R}^+$, let

$$R^i(\tau) := \inf_{W_{\mathcal{L}}} \max_{\mathcal{G} \subseteq \mathcal{L}} F_{\mathcal{G}}, \tag{114}$$

where

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

for some auxiliary r.v.'s W_l, $l \in \mathcal{L}$, such that $|\mathcal{W}_l| \leq |\mathcal{U}_l| + 4$,

$$(Z,\ E,\ U_{l^c},\ W_{l^c}) - U_l - W_l, \tag{115}$$

$$\text{and } I(U_{\mathcal{L}}; W_{\mathcal{G}}|E, W_{\mathcal{G}^c}, Z) \leq \tau C_{\mathcal{G}},\ \forall\ \mathcal{G} \subseteq \mathcal{L}. \tag{116}$$

Similarly, let $R^o(\tau)$ denote the right hand side (R.H.S) of (114), when the auxiliary r.v.'s $W_l,\ l \in \mathcal{L}$ satisfy $|\mathcal{W}_l| \leq |\mathcal{U}_l| + 4$, (116) and

$$(E, U_{l^c}, Z) - U_l - W_l. \tag{117}$$

The following theorem combined with Lemma 13 provides a lower and upper bound on $\kappa(\tau, \epsilon)$.

Theorem 16.

$$R^o(\tau) \leq\ R(\tau) \leq R^i(\tau), \tag{118}$$

$$\textit{and } H(E|Z) - R^i(\tau) \leq\ \theta(\tau) \leq H(E|Z) - R^o(\tau). \tag{119}$$

Proof: The proof is presented in Appendix G. ■

The BT inner bound is tight for the two terminal case, when one of the distortion measure is the Hamming distortion measure and the corresponding average distortion requirement is zero (lossless) [33, Ch.12]. Using this fact, an alternate proof of Proposition 9 can be given. The details are given in Appendix H.

VI. Concluding remarks

In this paper, we have studied the T2EE achievable for a distributed HT problem over orthogonal DMCs with side information available at the detector. We obtained single-letter lower bounds on the optimal T2EE for general HT, and exact single-letter characterizations in some important special cases. It is interesting to note from our results that the reliability function of the channel does not play a role in the characterization of the optimal T2EE for TACI, and only the channel capacity matters. We also showed that the strong converse holds in two special scenarios, namely, when the channel has zero capacity and for HT over a DMC. While the strong converse holds for distributed HT over a rate-limited noiseless channel [5], it remains an open question whether this result carries over to noisy channels in general. While we assume that

$n_k \leq \tau k$ for all k, the results remain the same for $\tau > 0$, if this constraint is relaxed to

$$\limsup_{k \to \infty} \frac{n_k}{k} \leq \tau. \tag{120}$$

For $\tau = 0$, slight modifications are required for some of the results, which is due to the fact that it is possible to transmit some information to the detector (at asymptotically zero-rate) under the constraint in (120), while the same is not possible under the constraint $n_k \leq \tau k$. For instance, the choice $n_k = k^a$ for any fixed number $a < 1$ satisfies $\limsup_{k \to \infty} \frac{n_k}{k} = 0$. It can be shown that under constraint (120), Theorem 5 and Theorem 12 hold with

$$\kappa_0(\tau) := \begin{cases} \beta_0 & , \text{ if } \tau = 0 \text{ and } E_c = \infty, \\ \min\left(\beta_0, \tau E_c + D(P_V||Q_V)\right), & \text{otherwise}, \end{cases} \tag{121}$$

and

$$\kappa_0'(\tau) := \begin{cases} D(P_U||Q_U), & \text{if } \tau = 0 \text{ and } E_c = \infty, \\ \min\left(D(P_U||Q_U), \tau E_c\right), & \text{otherwise}, \end{cases} \tag{122}$$

respectively. Also, Theorem 2 hold with $\tau E_m(P_{SX})$ and $\tau E_x(I_P(X;Y|S), P_{SX})$ set to ∞ whenever $\tau = 0$, $E_m(P_{SX}) = \infty$ and $\tau = 0$, $E_x(I_P(X;Y|S), P_{SX}) = \infty$, respectively, as opposed to Remark 3. While we did not discuss the complexity of the schemes considered in this paper, it is an important factor that needs to be considered in any practical implementation of these schemes. In this regard, it is evident that the JHTCC, SHTCC and local decision schemes are in a decreasing order of complexity.

APPENDIX A
PROOF OF THEOREM 2

The proof outline is as follows. We first describe the encoding and decoding operations of 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 the same or lower type 1 and type 2 error probabilities. For brevity, in the proof below, we denote the information theoretic quantities like $I_P(U;W)$, $T^k_{[P_{UW}]_\delta}$, etc., that are computed with respect to joint distribution $P_{UVWSXY}(P_{W|U}, P_{SX}) := P_{UV}P_{W|U}P_{SX}P_{Y|X}$ by $I(U;W)$, $T^k_{[UW]_\delta}$, etc.

Codebook Generation: Let $k \in \mathbb{Z}^+$ and $n = \lfloor \tau k \rfloor$. Fix distributions $P_{W|U}$ and P_{SX}, positive numbers $\mu, \delta', \delta'', \delta''', \delta, \tilde{\delta}$ (arbitrarily small subject to the delta-convention [24] and certain other constraints that will be specified in the course of the proof), and R such that $0 \leq R = \tau I(X;Y|S) - \mu \leq \tau C$. The *source codebook* $\mathcal{C}$ used by the source encoder $f_s^{(k)}$ is obtained by generating $e^{k(I(U;W)+\delta')}$ sequences $w^k(j),\ j \in \left[e^{k(I(U;W)+\delta')}\right]$, independently at random according to the distribution $\prod_{i=1}^{k} P_W(w_i)$, where

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

The *channel codebook* $\tilde{\mathcal{C}}$ used by $f_c^{(k,n)}$ is obtained as follows. The codeword length n is divided into $|\mathcal{S}| = |\mathcal{X}|$ blocks, where the length of the first block is $\lceil P_S(s_1)n \rceil$, the second block is $\lceil P_S(s_2)n \rceil$, so on so forth, and the length of the last block is chosen such that the total length is n. The codeword $x^n(0) = s^n$ corresponding to $M = 0$ is obtained by repeating the letter s_i in block i. The remaining $\left\lceil e^{kR} \right\rceil$ ordinary codewords $x^n(m),\ m \in \left[e^{kR}\right]$, are obtained by blockwise i.i.d. random coding, i.e., the symbols in the i^{th} block of each codeword are generated i.i.d. according to $P_{X|S=s_i}(x_i)$. The sequence s^n is revealed to the detector.

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, the encoder performs uniform random binning on the sequences $w^k(i),\ i \in \left[e^{k(I(U;W)+\delta')}\right]$ in $\mathcal{C}$, i.e., for each codeword in $\mathcal{C}$, it selects an index uniformly at random from the set $[e^{kR}]$. Denote the bin index selected for $w^k(i)$ by $f_B(i)$. If the observed sequence $U^k = u^k$ is typical, i.e., $u^k \in T^k_{[U]_{\delta''}}$, the source encoder $f_s^{(k)}$ first looks for a sequence $w^k(j)$ in $\mathcal{C}$ such that $(u^k, w^k(j)) \in T^k_{[UW]_\delta}$, $\delta > \delta''$. If there exist multiple such codewords, it chooses an index j among them uniformly at random, and outputs the bin-index $M = m = f_B(j)$, $m \in [e^{kR}]$ or $M = m = j$ 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 the *error* message $M = 0$. The channel encoder $f_c^{(k,n)}$ transmits the codeword $x^n(m)$ from codebook $\tilde{\mathcal{C}}$.

Decoding: At the decoder, $g_c^{(k,n)}$ outputs $\hat{M} = 0$ if for some $1 \leq i \leq |\mathcal{S}|$, the channel outputs corresponding to the i^{th} block does not belong to $T^n_{[P_{Y|S=s_i}]_{\delta'''}}$. Otherwise, $\hat{M}$ is set as the index of the codeword corresponding to the maximum-likelihood candidate among the ordinary codewords. If $\hat{M} = 0$, H_1 is declared. Else, given the side information sequence $V^k = v^k$ and estimated bin-index $\hat{M} = \hat{m}$, $g_s^{(k,n)}$ searches for a typical sequence $\hat{w}^k = w^k(\hat{j}) \in T^k_{[W]_{\hat{\delta}}}$,

$\hat{\delta} = |\mathcal{U}|\delta$, in codebook $\mathcal{C}$ such that

$$\hat{j} = \underset{\substack{l:\ f_B(l)=\hat{m},\\ w^k(l)\in T^k_{[W]_{\hat{\delta}}}}}{\arg\min}\ H_e(w^k(l)|v^k),\ \text{if } I(U;W)+\delta' > R,$$

$$\hat{j} = \hat{m},\ \text{otherwise.}$$

The decoder declares H_0 if $(\hat{w}^k, v^k) \in T^k_{[WV]_{\tilde{\delta}}}$, for $\tilde{\delta} > \delta$, else, H_1 is declared.

We next analyze the type 1 and type 2 error probabilities achieved by the above scheme (in the limit $\delta, \delta', \tilde{\delta} \to 0$).

Analysis of Type 1 error: A type 1 error occurs only if one of the following events happen.

$$\mathcal{E}_{TE} = \left\{ (U^k, V^k) \notin T^k_{[UV]_{\bar{\delta}}}, \bar{\delta} = \frac{\delta''}{|\mathcal{V}|} \right\}$$

$$\mathcal{E}_{EE} = \left\{ \nexists\ j \in \left[e^{k(I(U;W)+\delta')}\right] : (U^k, W^k(j)) \in T^k_{[UW]_\delta} \right\}$$

$$\mathcal{E}_{ME} = \left\{ (V^k, W^k(J)) \notin T^k_{[VW]_{\tilde{\delta}}} \right\}$$

$$\mathcal{E}_{DE} = \Bigg\{ \exists\ l \in \left[e^{k(I(U;W)+\delta')}\right],\ l \neq J : f_B(l) = f_B(J),\ W^k(l) \in T^k_{[W]_{\hat{\delta}}},$$
$$H_e(W^k(l)|V^k) \leq H_e(W^k(J)|V^k) \Bigg\}$$

$$\mathcal{E}_{CD} = \left\{ g_c^{(k,n)}(Y^n) \neq M \right\}$$

$\mathbb{P}(\mathcal{E}_{TE}|H=0)$ tends to 0 asymptotically by the weak law of large numbers. Conditioned on $\mathcal{E}^c_{TE}$, $U^k \in T_{[U]_{\delta''}}$ and by the covering lemma [24, Lemma 9.1], it is well known that $\mathbb{P}(\mathcal{E}_{EE}|\mathcal{E}^c_{TE})$ tends to 0 doubly exponentially for $\delta > \delta''$ and δ' appropriately chosen. Given $\mathcal{E}^c_{EE} \cap \mathcal{E}^c_{TE}$ holds, it follows from the Markov chain relation $V - U - W$ and the Markov lemma [33], that $\mathbb{P}(\mathcal{E}_{ME}|\mathcal{E}^c_{TE} \cap \mathcal{E}^c_{EE})$ tends to zero as $k \to \infty$ for $\tilde{\delta} > \delta$ (appropriately chosen). Next, we consider $\mathbb{P}(\mathcal{E}_{DE})$. Given that $\mathcal{E}^c_{ME} \cap \mathcal{E}^c_{EE} \cap \mathcal{E}^c_{TE}$ holds, note that $\lim_{k\to\infty} H_e(W^k(J)|V^k) \to H(W|V)$ as $\tilde{\delta} \to 0$. Thus, we have

$$\mathbb{P}(\mathcal{E}_{DE}|\ \ V^k = v^k, W^k(J) = w^k, \mathcal{E}^c_{ME} \cap \mathcal{E}^c_{EE} \cap \mathcal{E}^c_{TE})$$
$$\leq \sum_{\substack{l=1,\\ l\neq J}}^{e^{k(I(U;W)+\delta')}} \sum_{\substack{\tilde{w}^k \in T^k_{[W]_{\hat{\delta}}}:\\ H_e(\tilde{w}^k|v^k) \leq H_e(w^k|v^k)}} \mathbb{P}\Big(f_B(l) = f_B(J),\ \ \ W^k(l) = \tilde{w}^k|\ V^k = v^k, W^k(J) = w^k,$$

$$\mathcal{E}_{ME}^c \cap \mathcal{E}_{EE}^c \cap \mathcal{E}_{TE}^c \Big)$$

$$= \sum_{\substack{l=1,\\ l\neq J}}^{e^{k(I(U;W)+\delta')}} \sum_{\substack{\tilde{w}^k \in T^k_{[W]_{\hat{\delta}}}:\\ H_e(\tilde{w}^k|v^k)\\ \leq H_e(w^k|v^k)}} \mathbb{P}(W^k(l)=\tilde{w}^k|\ V^k=v^k, W^k(J)=w^k, \mathcal{E}_{ME}^c \cap \mathcal{E}_{EE}^c \cap \mathcal{E}_{TE}^c)\ \frac{1}{e^{kR}}$$

$$\leq \sum_{\substack{l=1,\\ l\neq J}}^{e^{k(I(U;W)+\delta')}} \sum_{\substack{\tilde{w}^k \in T^k_{[W]_{\hat{\delta}}}:\\ H_e(\tilde{w}^k|v^k)\leq H_e(w^k|v^k)}} 2\cdot e^{-kR} e^{-k(H(W)-\delta_1)} \tag{123}$$

$$\leq \sum_{\substack{l=1,\\ l\neq J}}^{e^{k(I(U;W)+\delta')}} (k+1)^{|\mathcal{V}||\mathcal{W}|}\ e^{k(H(W|V)+\gamma_1(k))}\cdot 2\cdot e^{-kR} e^{-k(H(W)-\delta_1)} \tag{124}$$

$$\leq e^{-k(R-I(U;W|V)-\delta_e^{(k)})},$$

where $\delta_1 = O(\hat{\delta})$, $\gamma_1(k) = |H_e(w^k|v^k) - H(W|V)|$, and

$$\delta_e^{(k)} = \delta_1 + \frac{1}{k}|\mathcal{V}||\mathcal{W}|\log(k+1) + \frac{\log(2)}{k} + \delta' + \gamma_1(k) \xrightarrow{(k)} 0,$$

as $\tilde{\delta}, \delta', \delta \to 0$. To obtain (123), we used the fact that

$$\mathbb{P}(W^k(l)=\tilde{w}^k|\ \mathcal{E}_{ME}^c \cap \mathcal{E}_{EE}^c \cap \mathcal{E}_{TE}^c, W^k(J)=w^k, V^k=v^k) \leq 2\cdot \mathbb{P}(W^k(l)=\tilde{w}^k). \tag{125}$$

This follows similarly to (147), which is discussed in the type 2 error analysis section below. In order to obtain the expression in (124), we first summed over the types $P_{\tilde{W}}$ of sequences within the typical set $T^k_{[W]_\delta}$ that have empirical entropy less than $H_e(w^k|v^k)$; and used the facts that the number of sequences within such a type is upper bounded by $e^{k(H(W|V)+\gamma_1(k))}$, and the total number of types is upper bounded by $(k+1)^{|\mathcal{V}||\mathcal{W}|}$ [24]. Summing over all $(w^k, v^k) \in T^k_{[VW]_{\tilde{\delta}}}$, we obtain

$$\begin{aligned}
&\mathbb{P}(\mathcal{E}_{DE}|\mathcal{E}_{ME}^c \cap \mathcal{E}_{EE}^c \cap \mathcal{E}_{TE}^c)\\
&\leq \sum_{(w^k,v^k)\in T^k_{[WV]_{\tilde{\delta}}}} \mathbb{P}(W^k(J)=w^k, V^k=v^k|\mathcal{E}_{ME}^c \cap \mathcal{E}_{EE}^c \cap \mathcal{E}_{TE}^c)\ e^{-k(R-I(U;W|V)-\delta_e^{(k)})}\\
&\leq e^{-k(R-I(U;W|V)-\delta_e^{(k)})}.
\end{aligned} \tag{126}$$

Finally, we consider the event $\mathcal{E}_{CD}$. Denoting by $\mathcal{E}_{CT}$, the event that the channel outputs corresponding to the i^{th} block does not belong to $T^n_{[P_{Y|S=s_i}]_{\delta'''}}$ for some $1 \leq i \leq |\mathcal{S}|$, it follows from the weak law of large numbers and the union bound, that

$$\mathbb{P}(\mathcal{E}_{CT}|\mathcal{E}_{EE}^c) \xrightarrow{(k)} 0. \tag{127}$$

Also, it follows from [24, Exercise 10.18, 10.24] that

$$\mathbb{P}\left(\mathcal{E}_{CD}|\mathcal{E}_{EE}^c \cap \mathcal{E}_{CT}^c\right) \leq e^{-nE_x(I_P(X;Y|S)-\mu, P_{SX}, P_{Y|X})} \tag{128}$$

asymptotically. This implies that the probability that an error occurs at the channel decoder $g_c^{(k,n)}$ tends to 0 as $n \to \infty$ since $E_x(I(X;Y|S), P_{SX}, P_{Y|X}) > 0$ for $I(X;Y|S) < C(P_{Y|X})$. Thus, if $I(U;W|V) < R = \tau I(X;Y|S) - \mu \leq \tau C(P_{Y|X})$, the probability of the events causing type 1 error tends to zero asymptotically.

Analysis of Type 2 error: First, note that a type 2 error occurs only if $V^k \in T^k_{[V]_{|\mathcal{W}|\bar{\delta}}}$, and hence we can restrict the type 2 error analysis to only such V^k. Denote the event that a type 2 error happens by $\mathcal{D}_0$. Let

$$\mathcal{E}_0 = \left\{U^k \notin T^k_{[U]_{\delta''}}\right\}. \tag{129}$$

Type 2 error probability can be written as

$$\begin{aligned}
&\beta\left(k, n, f^{(k,n)}, g^{(k,n)}\right) \\
&= \sum_{(u^k,v^k)\in\mathcal{U}^k\times\mathcal{V}^k} \mathbb{P}(U^k = u^k, V^k = v^k|H=1)\ \mathbb{P}(\mathcal{D}_0|U^k = u^k, V^k = v^k).
\end{aligned} \tag{130}$$

Let $\mathcal{E}_{NE} := \mathcal{E}_{EE}^c \cap \mathcal{E}_0^c$. The last term in (130) can be upper bounded as follows.

$$\begin{aligned}
&\mathbb{P}(\mathcal{D}_0|U^k = u^k, V^k = v^k) \\
&= \mathbb{P}(\mathcal{E}_{NE}|U^k = u^k, V^k = v^k)\ \mathbb{P}(\mathcal{D}_0|U^k = u^k, V^k = v^k, \mathcal{E}_{NE}) \\
&\qquad + \mathbb{P}(\mathcal{E}_{NE}^c|U^k = u^k, V^k = v^k)\ \mathbb{P}(\mathcal{D}_0|U^k = u^k, V^k = v^k, \mathcal{E}_{NE}^c) \\
&\leq \mathbb{P}(\mathcal{D}_0|U^k = u^k, V^k = v^k, \mathcal{E}_{NE}) + \mathbb{P}(\mathcal{D}_0|U^k = u^k, V^k = v^k, \mathcal{E}_{NE}^c).
\end{aligned}$$

Thus, we have

$$\begin{aligned}
&\beta\left(k, n, f^{(k,n)}, g^{(k,n)}\right) \\
&\leq \sum_{(u^k,v^k)\in\mathcal{U}^k\times\mathcal{V}^k} \mathbb{P}(U^k = u^k, V^k = v^k|H=1)\Big[\mathbb{P}(\mathcal{D}_0|U^k = u^k, V^k = v^k, \mathcal{E}_{NE})
\end{aligned}$$

$$+ \mathbb{P}(\mathcal{D}_0|U^k = u^k, V^k = v^k, \mathcal{E}_{NE}^c)\Big]. \tag{131}$$

First, we assume that $\mathcal{E}_{NE}$ holds. Then,

$$\mathbb{P}(\mathcal{D}_0|\ U^k = u^k, V^k = v^k, \mathcal{E}_{NE}) = \sum_{j=1}^{e^{k(I(U;W)+\delta')}} \sum_{m=1}^{e^{kR}} \mathbb{P}(J = j, f_B(J) = m|\ U^k = u^k, V^k = v^k, \mathcal{E}_{NE})$$
$$\mathbb{P}(\mathcal{D}_0|U^k = u^k, V^k = v^k, J = j, f_B(J) = m,\ \mathcal{E}_{NE}). \tag{132}$$

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 (132) 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

$$\begin{aligned}
&\mathbb{P}(\mathcal{D}_0|\ U^k = u^k, V^k = v^k, \mathcal{E}_{NE}) \\
&= \sum_{j=1}^{e^{k(I(U;W)+\delta')}} \sum_{m=1}^{e^{kR}} \mathbb{P}(J = j, f_B(J) = m|\ U^k = u^k, V^k = v^k, \mathcal{E}_{NE}) \\
&\qquad \mathbb{P}(\mathcal{D}_0|U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1,\ \mathcal{E}_{NE}) \\
&= \mathbb{P}(\mathcal{D}_0|U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1,\ \mathcal{E}_{NE}) \\
&= \sum_{w^k \in \mathcal{W}^k} \mathbb{P}(W^k(1) = w^k|U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1,\ \mathcal{E}_{NE}) \\
&\qquad \mathbb{P}(\mathcal{D}_0|U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, W^k(1) = w^k,\ \mathcal{E}_{NE}).
\end{aligned} \tag{133}$$

Given $\mathcal{E}_{NE}$ holds, $\mathcal{D}_0$ may occur in three possible ways: (i) when $\hat{M} \neq 0$, i.e., $\mathcal{E}_{CT}^c$ occurs, the channel decoder makes an error and the codeword retrieved from the bin is jointly typical with V^k; (ii) when an unintended wrong codeword is retrieved from the correct bin that is jointly typical with V^k; and (iii) when there is no error at the channel decoder and the correct codeword is retrieved from the bin, that is also jointly typical with V^k. We refer to the event in case (i) as the *channel error event* $\mathcal{E}_{CE}$, and the one in case (ii) as the *binning error event* $\mathcal{E}_{BE}$. More specifically,

$$\mathcal{E}_{CE} = \{\mathcal{E}_{CT}^c \text{ and } \hat{M} = g_c^{(k,n)}(Y^n) \neq M\}, \text{ and} \tag{134}$$

$$\mathcal{E}_{BE} = \Big\{ \exists\ l \in \Big[e^{k(I(U;W)+\delta')} \Big],\ l \neq J,\ f_B(l) = \hat{M},\ W^k(l)) \in T^k_{[W]_{\hat{\delta}}}, \\ (V^k, W^k(l)) \in T^k_{[VW]_{\tilde{\delta}}} \Big\}. \tag{135}$$

Define the following events

$$\mathcal{F} = \{U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, W^k(1) = w^k,\ \mathcal{E}_{NE}\}, \tag{136}$$

$$\mathcal{F}_1 = \{U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, W^k(1) = w^k,\ \mathcal{E}_{NE},\ \mathcal{E}_{CE}\}, \tag{137}$$

$$\mathcal{F}_2 = \{U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, W^k(1) = w^k,\ \mathcal{E}_{NE},\ \mathcal{E}^c_{CE}\}, \tag{138}$$

$$\mathcal{F}_{21} = \{U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, W^k(1) = w^k,\ \mathcal{E}_{NE},\ \mathcal{E}^c_{CE},\ \mathcal{E}_{BE}\}, \tag{139}$$

$$\mathcal{F}_{22} = \{U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, W^k(1) = w^k,\ \mathcal{E}_{NE},\ \mathcal{E}^c_{CE},\ \mathcal{E}^c_{BE}\}. \tag{140}$$

The last term in (133) can be expressed as follows.

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

where

$$\mathbb{P}(\mathcal{D}_0|\mathcal{F}_2) = \mathbb{P}(\mathcal{E}_{BE}|\mathcal{F}_2)\ \mathbb{P}(\mathcal{D}_0|\mathcal{F}_{21}) + \mathbb{P}(\mathcal{E}^c_{BE}|\mathcal{F}_2)\ \mathbb{P}(\mathcal{D}_0|\mathcal{F}_{22}). \tag{141}$$

It follows from (128) that

$$\mathbb{P}(\mathcal{E}_{CE}|\mathcal{F}) \leq e^{-nE_x(I(X;Y|S)-\mu, P_{SX}, P_{Y|X})} = e^{-k\tau E_x(I(X;Y|S)-\mu, P_{SX}, P_{Y|X})}. \tag{142}$$

Next, consider the type 2 error event that happens when an error occurs at the channel decoder. We need to consider two separate cases: $I(U;W) > R$ 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 $(V^k, W^k(l)) \in T^k_{[VW]_{\tilde{\delta}}}$. As noted in [32], the calculation of the probability of this event does not follow from the standard random coding argument usually encountered in achievability proofs due to the fact that the chosen codeword $W^k(J)$ depends on the entire codebook. Following steps similar to those in [32], we analyze the probability of this event (averaged over codebooks $\mathcal{C}$ and random binning) as follows. We first consider the case when $I(U;W) > R$.

$$\mathbb{P}(\mathcal{D}_0|\mathcal{F}_1) \leq \mathbb{P}(\ \exists\ W^k(l):\ \ f_B(l) = \hat{M} \neq 1,\ (W^k(l), v^k) \in T^k_{[WV]_{\tilde{\delta}}}|\mathcal{F}_1)$$

$$\leq \sum_{l=2}^{e^{k(I(U;W)+\delta')}} \sum_{\hat{m}\neq 1} \mathbb{P}(\hat{M}=\hat{m}|\mathcal{F}_1)\ \mathbb{P}((W^k(l),v^k)\in T^k_{[WV]_{\tilde{\delta}}}:\ f_B(l)=\hat{m}|\mathcal{F}_1)$$

$$= \sum_{l=2}^{e^{k(I(U;W)+\delta')}} \sum_{\hat{m}\neq 1} \mathbb{P}(\hat{M}=\hat{m}|\mathcal{F}_1) \sum_{\substack{\tilde{w}^k:\\(\tilde{w}^k,v^k)\in T^k_{[WV]_{\tilde{\delta}}}}} \mathbb{P}(W^k(l)=\tilde{w}^k:\ f_B(l)=\hat{m}|\mathcal{F}_1)$$

$$= \sum_{l=2}^{e^{k(I(U;W)+\delta')}} \sum_{\hat{m}\neq 1} \mathbb{P}(\hat{M}=\hat{m}|\mathcal{F}_1) \sum_{\substack{\tilde{w}^k:\\(\tilde{w}^k,v^k)\in T^k_{[WV]_{\tilde{\delta}}}}} \mathbb{P}(W^k(l)=\tilde{w}^k|\mathcal{F}_1)\ \frac{1}{e^{kR}} \tag{143}$$

$$= \sum_{l=2}^{e^{k(I(U;W)+\delta')}} \sum_{\substack{\tilde{w}^k:\\(\tilde{w}^k,v^k)\in T^k_{[WV]_{\tilde{\delta}}}}} \mathbb{P}(W^k(l)=\tilde{w}^k|\mathcal{F}_1)\ \frac{1}{e^{kR}}. \tag{144}$$

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

$$\mathbb{P}(W^k(l)=\tilde{w}^k|\mathcal{F}_1)=\sum_{\mathcal{C}^-_{1,l}=c}\mathbb{P}(\mathcal{C}^-_{1,l}=c|\mathcal{F}_1)\mathbb{P}(W^k(l)=\tilde{w}^k|\mathcal{F}_1,\mathcal{C}^-_{1,l}=c). \tag{145}$$

The term in (145) can be upper bounded as follows:

$$\begin{aligned}
&\mathbb{P}(W^k(l)=\tilde{w}^k|\mathcal{F}_1,\ \mathcal{C}^-_{1,l}=c)\\
&=\mathbb{P}(W^k(l)=\tilde{w}^k|U^k=u^k,V^k=v^k,\ \mathcal{C}^-_{1,l}=c)\\
&\quad\frac{\mathbb{P}(W^k(1)=w^k|W^k(l)=\tilde{w}^k,U^k=u^k,V^k=v^k,\mathcal{C}^-_{1,l}=c)}{\mathbb{P}(W^k(1)=w^k|U^k=u^k,V^k=v^k,\mathcal{C}^-_{1,l}=c)}\\
&\quad\frac{\mathbb{P}(J=1|W^k(1)=w^k,W^k(l)=\tilde{w}^k,U^k=u^k,V^k=v^k,\mathcal{C}^-_{1,l}=c)}{\mathbb{P}(J=1|W^k(1)=w^k,U^k=u^k,V^k=v^k,\mathcal{C}^-_{1,l}=c)}\\
&\quad\frac{\mathbb{P}(f_B(J)=1|J=1,W^k(1)=w^k,W^k(l)=\tilde{w}^k,U^k=u^k,V^k=v^k,\ \mathcal{C}^-_{1,l}=c)}{\mathbb{P}(f_B(J)=1|J=1,W^k(1)=w^k,U^k=u^k,V^k=v^k,\mathcal{C}^-_{1,l}=c)}\\
&\quad\frac{\mathbb{P}(\mathcal{E}_{NE},\ \mathcal{E}_{CE}|f_B(J)=1,J=1,W^k(1)=w^k,W^k(l)=\tilde{w}^k,U^k=u^k,V^k=v^k,\ \mathcal{C}^-_{1,l}=c)}{\mathbb{P}(\mathcal{E}_{NE},\ \mathcal{E}_{CE}|f_B(J)=1,J=1,W^k(1)=w^k,U^k=u^k,V^k=v^k,\mathcal{C}^-_{1,l}=c)}.
\end{aligned} \tag{146}$$

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

$$\begin{aligned}
&\mathbb{P}(W^k(1)=w^k|W^k(l)=\tilde{w}^k,U^k=u^k,V^k=v^k,\mathcal{C}^-_{1,l}=c)\\
&\qquad\qquad=\mathbb{P}(W^k(1)=w^k|U^k=u^k,V^k=v^k,\mathcal{C}^-_{1,l}=c),
\end{aligned}$$

and

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

Also, note that

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

Next, consider the term in (146). Let $N(u^k, \mathcal{C}_{1,l}^-)=|\{w^k(l')\in \mathcal{C}_{1,l}^- : l'\neq 1,\ l'\neq l,\ (w^k(l'), u^k)\in T_{[WU]_\delta}^k\}|$. Recall that if there are multiple sequences in codebook $\mathcal{C}$ that are jointly typical with the observed sequence U^k, then the encoder selects one of them uniformly at random. Also, note that given $\mathcal{F}_1$, $(w^k, u^k)\in T_{[WU]_\delta}^k$. Thus, if $(\tilde{w}^k, u^k)\in T_{[WU]_\delta}^k$, then

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

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

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

Hence, the term in (145) can be upper bounded as

$$\mathbb{P}(W^k(l)=\tilde{w}^k|\mathcal{F}_1)$$
$$\leq \sum_{\mathcal{C}_{1,l}^-=c}\mathbb{P}(\mathcal{C}_{1,l}^-=c|\mathcal{F}_1)\ 2\ \mathbb{P}(W^k(l)=\tilde{w}^k|U^k=u^k, V^k=v^k,\ \mathcal{C}_{1,l}^-=c)$$

$$= 2\ \mathbb{P}(W^k(l) = \tilde{w}^k | U^k = u^k, V^k = v^k) = 2\ \mathbb{P}(W^k(l) = \tilde{w}^k). \tag{147}$$

Substituting (147) in (144), we obtain

$$\begin{aligned}
\mathbb{P}(\mathcal{D}_0|\mathcal{F}_1) &\leq \sum_{l=1}^{e^{k(I(U:W)+\delta')}} \sum_{\substack{\tilde{w}^k: \\ (\tilde{w}^k, v^k) \in T^k_{[WV]_{\tilde{\delta}}}}} 2\ \mathbb{P}(W^k(l) = \tilde{w}^k)\ \frac{1}{e^{kR}} \\
&= \sum_{l=1}^{e^{k(I(U:W)+\delta')}} \sum_{\substack{\tilde{w}^k: \\ (\tilde{w}^k, v^k) \in T^k_{[WV]_{\tilde{\delta}}}}} 2 \cdot e^{-k(H(W)-\delta_1)}\ \frac{1}{e^{kR}} \\
&= 2 \cdot e^{k(I(U:W)+\delta')}\ e^{k(H(W|V)+\delta_3)}\ e^{-k(H(W)-\delta_1)}\ \frac{1}{e^{kR}} \\
&\leq e^{-k(R - I(U;W|V) - \delta_4^{(k)})},
\end{aligned} \tag{148}$$

where $\delta_4^{(k)} := \delta' + \delta_1 + \delta_3 + \frac{\log(2)}{k} \xrightarrow{k} 0$ 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 (147) as follows.

$$\begin{aligned}
\mathbb{P}(\mathcal{D}_0|\mathcal{F}_1) &= \sum_{\hat{m} \neq 1} \mathbb{P}(\hat{M} = \hat{m}|\mathcal{F}_1)\ \mathbb{P}((W^k(\hat{m}), v^k) \in T^k_{[WV]_{\tilde{\delta}}}|\mathcal{F}_1) \\
&\leq \sum_{\hat{m} \neq 1} \mathbb{P}(\hat{M} = \hat{m}|\mathcal{F}_1) \sum_{\substack{\tilde{w}^k: \\ (\tilde{w}^k, v^k) \in T^k_{[WV]_{\tilde{\delta}}}}} 2\ \mathbb{P}(W^k(\hat{m}) = \tilde{w}^k) \\
&\leq \sum_{\hat{m} \neq 1} \mathbb{P}(\hat{M} = \hat{m}|\mathcal{F}_1)\ e^{-k(I(V;W) - (\delta_1 + \delta_3 + \frac{1}{k}))} \leq e^{-k(I(V;W) - \delta_4^{(k)})}.
\end{aligned} \tag{149}$$

Next, consider the event when there are no encoding or channel errors, i.e., $\mathcal{E}_{NE} \cap \mathcal{E}^c_{CE}$. 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 $P_{\tilde{U}\tilde{V}\tilde{W}}$ denote the type of $P_{U^kV^kW^k(J)}$. Note that $P_{\tilde{U}\tilde{W}} \in \mathcal{T}^k_{[UW]_\delta}$ when $\mathcal{E}_{NE}$ holds. If $H(\tilde{W}|\tilde{V}) < H(W|V)$, then in the bin with index M, there exists a codeword with empirical entropy strictly less than $H(W|V)$. Hence, the decoded codeword $\hat{W}^k \notin T^k_{[WV]_{\tilde{\delta}}}$(asymptotically) since $(\hat{W}^k,\ V^k) \in T^k_{[WV]_{\tilde{\delta}}}$ necessarily implies that $H_e(\hat{W}^k|V^k) := H(P_{\hat{W}^k}|P_{V^k}) \to H(W|V)$ 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:

$$
\begin{aligned}
&\mathbb{P}(\mathcal{E}_{BE}|\mathcal{F}_2)\\
&\leq \mathbb{P}\left(\exists\ l \neq 1,\ l \in [e^{k(I(U:W)+\delta')}]:\ f_B(l)=1 \text{ and } (W^k(l),v^k) \in T^k_{[WV]_{\tilde{\delta}}}|\mathcal{F}_2\right)\\
&\leq \sum_{l=2}^{e^{k(I(U;W)+\delta')}} \mathbb{P}\left((W^k(l),v^k) \in T^k_{[WV]_{\tilde{\delta}}}|\mathcal{F}_2\right)\ \mathbb{P}\left(f_B(l)=1|\mathcal{F}_2,(W^k(l),v^k) \in T^k_{[WV]_{\tilde{\delta}}}\right)\\
&= \sum_{l=2}^{e^{k(I(U;W)+\delta')}} \mathbb{P}\left((W^k(l),v^k) \in T^k_{[WV]_{\tilde{\delta}}}|\mathcal{F}_2\right)\ e^{-kR}\\
&\leq \sum_{l=2}^{e^{k(I(U;W)+\delta')}} \sum_{\substack{\tilde{w}^k:\\ (\tilde{w}^k,v^k)\in T^k_{[WV]_{\tilde{\delta}}}}} 2\ \mathbb{P}(W^k(l)=\tilde{w}^k)\ e^{-kR} \qquad (150)\\
&= e^{-k(R-I(U;W|V)-\delta_4^{(k)})}. \qquad (151)
\end{aligned}
$$

In (150), we used the fact that

$$
\mathbb{P}\left(W^k(l)=\tilde{w}^k|\mathcal{F}_2\right) \leq 2\ \mathbb{P}(W^k(l)=\tilde{w}^k), \qquad (152)
$$

which follows in a similar way as (147). Also, note that, by definition, $\mathbb{P}(\mathcal{D}_0|\mathcal{F}_{21})=1$.

We proceed to analyze the R.H.S of (131) which upper bounds the type 2 error probability, in the limit $k \to \infty$ and $\delta,\ \delta',\tilde{\delta} \to 0$. Towards this end, we first focus on the the case when $\mathcal{E}_{NE}$ holds. From (133), it follows that

$$
\begin{aligned}
&\lim_{k\to\infty}\ \lim_{\delta,\tilde{\delta},\delta'\to 0} \sum_{(u^k,v^k)\in\mathcal{U}^k\times\mathcal{V}^k} \mathbb{P}(U^k=u^k,V^k=v^k|H=1)\ \mathbb{P}(\mathcal{D}_0|U^k=u^k,V^k=v^k,\mathcal{E}_{NE}) \qquad (153)\\
&= \lim_{k\to\infty}\ \lim_{\delta,\tilde{\delta},\delta'\to 0} \sum_{(u^k,v^k)\in\mathcal{U}^k\times\mathcal{V}^k} \mathbb{P}(U^k=u^k,V^k=v^k|H=1)\\
&\qquad\qquad \mathbb{P}(\mathcal{D}_0|U^k=u^k,V^k=v^k,J=1,f_B(J)=1,\mathcal{E}_{NE}). \qquad (154)
\end{aligned}
$$

Rewriting the summation in (154) as the sum over the types and sequences within a type, we obtain

$$\mathbb{P}(\mathcal{D}_0|\ \mathcal{E}_{NE}, H=1)$$
$$= \sum_{\substack{P_{\tilde{U}\tilde{V}\tilde{W}} \\ \in \mathcal{T}^k_{\mathcal{U}\mathcal{V}\mathcal{W}}}} \sum_{\substack{(u^k,v^k,w^k) \\ \in T_{P_{\tilde{U}\tilde{V}\tilde{W}}}}} \Big[\mathbb{P}(U^k=u^k, V^k=v^k|H=1)\ \mathbb{P}(\mathcal{D}_0|\mathcal{F})$$
$$\mathbb{P}(W^k(1)=w^k|U^k=u^k, V^k=v^k, J=1, f_B(J)=1, \mathcal{E}_{NE})\Big]. \quad (155)$$

We also have

$$\mathbb{P}(U^k=u^k, V^k=v^k|H=1)\ \mathbb{P}(W^k(1)=w^k|U^k=u^k, V^k=v^k, J=1, f_B(J)=1, \mathcal{E}_{NE})$$
$$= \left[\prod_{i=1}^{k} Q_{UV}(u_i,v_i)\right] \mathbb{P}(W^k(1)=w^k|U^k=u^k, V^k=v^k, J=1, f_B(J)=1, \mathcal{E}_{NE})$$
$$\leq \left[\prod_{i=1}^{k} Q_{UV}(u_i,v_i)\right] \frac{1}{|T_{P_{\tilde{W}|\tilde{U}}}|} \leq e^{-k(H(\tilde{U}\tilde{V})+D(P_{\tilde{U}\tilde{V}}||Q_{UV})+H(\tilde{W}|\tilde{U})-\frac{1}{k}|\mathcal{U}||\mathcal{W}|\log(k+1))}, \quad (156)$$

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

With (142), (148), (149), (151) and (156), we have the necessary machinery to analyze (155). First, consider that the event $\mathcal{E}_{NE} \cap \mathcal{E}^c_{CE} \cap \mathcal{E}^c_{BE}$ holds. In this case,

$$\mathbb{P}(\mathcal{D}_0|\mathcal{F}_{22}) = \mathbb{P}(\mathcal{D}_0|U^k=u^k, V^k=v^k, J=1, f_B(J)=1, W^k(1)=w^k, \mathcal{E}_{NE}, \mathcal{E}^c_{CE}, \mathcal{E}^c_{BE})$$
$$= \begin{cases} 1, & \text{if } P_{u^k w^k} \in T^k_{[UW]_\delta} \\ & \quad \text{and } P_{v^k w^k} \in T^k_{[VW]_{\tilde{\delta}}}, \\ 0, & \text{otherwise.} \end{cases} \quad (157)$$

Thus, the following terms in (155) can be simplified (in the limit $\delta, \tilde{\delta} \to 0$) as

$$\lim_{k\to\infty} \lim_{\delta,\tilde{\delta},\delta'\to 0} \sum_{\substack{P_{\tilde{U}\tilde{V}\tilde{W}} \\ \in \mathcal{T}^k_{\mathcal{U}\mathcal{V}\mathcal{W}}}} \sum_{\substack{(u^k,v^k,w^k) \\ \in T_{P_{\tilde{U}\tilde{V}\tilde{W}}}}} \Big[\mathbb{P}(U^k=u^k, V^k=v^k|H=1)\ \mathbb{P}(\mathcal{E}^c_{CE}|\mathcal{F})\ \mathbb{P}(\mathcal{E}^c_{BE}|\mathcal{F}_2)\ \mathbb{P}(\mathcal{D}_0|\mathcal{F}_{22})$$
$$\mathbb{P}(W^k(1)=w^k|U^k=u^k, V^k=v^k, J=1, f_B(J)=1, \mathcal{E}_{NE})\Big]$$
$$\leq \lim_{k\to\infty} \lim_{\delta,\tilde{\delta},\delta'\to 0} \sum_{\substack{P_{\tilde{U}\tilde{V}\tilde{W}} \\ \in \mathcal{T}^k_{\mathcal{U}\mathcal{V}\mathcal{W}}}} \sum_{\substack{(u^k,v^k,w^k) \\ \in T_{P_{\tilde{U}\tilde{V}\tilde{W}}}}} \Big[\mathbb{P}(U^k=u^k, V^k=v^k|H=1)\ \mathbb{P}(\mathcal{D}_0|\mathcal{F}_{22})$$
$$\mathbb{P}(W^k(1)=w^k|U^k=u^k, V^k=v^k, J=1, f_B(J)=1, \mathcal{E}_{NE})\Big]$$

$$\leq \lim_{k\to\infty} (k+1)^{|\mathcal{U}||\mathcal{V}||\mathcal{W}|} \max_{\substack{P_{\tilde{U}\tilde{V}\tilde{W}}\in \\ \mathcal{T}_1(P_{UW},P_{VW})}} e^{kH(\tilde{U}\tilde{V}\tilde{W})} e^{-k(H(\tilde{U}\tilde{V})+D(P_{\tilde{U}\tilde{V}}||Q_{UV})+H(\tilde{W}|\tilde{U})-\frac{1}{k}|\mathcal{U}||\mathcal{W}|\log(k+1))}$$

$$= \lim_{k\to\infty} e^{-k\tilde{E}_{1k}}. \tag{158}$$

Here,

$$\begin{aligned}\tilde{E}_{1k} &:= \min_{\substack{P_{\tilde{U}\tilde{V}\tilde{W}}\in \\ \mathcal{T}_1(P_{UW},P_{VW})}} H(\tilde{U}\tilde{V}) + D(P_{\tilde{U}\tilde{V}}||Q_{UV}) + H(\tilde{W}|\tilde{U}) - H(\tilde{U}\tilde{V}\tilde{W}) \\ &\qquad - \frac{1}{k}|\mathcal{U}||\mathcal{V}||\mathcal{W}|\log(k+1) - \frac{1}{k}|\mathcal{U}||\mathcal{W}|\log(k+1) \\ &= \min_{\substack{P_{\tilde{U}\tilde{V}\tilde{W}}\in \\ \mathcal{T}_1(P_{UW},P_{VW})}} \sum P_{\tilde{U}\tilde{V}\tilde{W}} \log\left(\frac{P_{\tilde{U}\tilde{V}}}{Q_{UV}}\frac{1}{P_{\tilde{U}\tilde{V}}}\frac{P_{\tilde{U}}}{P_{\tilde{U}\tilde{W}}}P_{\tilde{U}\tilde{V}\tilde{W}}\right) - o(1) \\ &= \min_{\substack{P_{\tilde{U}\tilde{V}\tilde{W}}\in \\ \mathcal{T}_1(P_{UW},P_{VW})}} D(P_{\tilde{U}\tilde{V}\tilde{W}}||Q_{UVW}) - o(1) \xrightarrow{(k)} E_1(P_{W|U}),\end{aligned} \tag{159}$$

and $Q_{UVW} := Q_{UV}P_{W|U}$. To obtain (158), we used (156) and (157). This results in the term $E_1(P_{W|U})$ in (12).

Next, consider the terms corresponding to the event $\mathcal{E}_{NE}\cap\mathcal{E}_{CE}^c\cap\mathcal{E}_{BE}$ in (155). Note that given the event $\mathcal{F}_{21} = \{U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, W^k(1) = w^k,\ \mathcal{E}_{NE},\ \mathcal{E}_{CE}^c,\ \mathcal{E}_{BE}\}$ occurs, $P_{u^k w^k} \in T^k_{[UW]_\delta}$. Also, $\mathcal{D}_0$ can happen only if $H_e(w^k|v^k) \geq H(W|V) - \gamma_2(\tilde{\delta})$ for some positive function $\gamma_2(\tilde{\delta}) \in O(\tilde{\delta})$ and $P_{v^k} \in T^k_{[V]_{\delta'''}}$. Using these facts to simplify the terms corresponding to the event $\mathcal{E}_{NE}\cap\mathcal{E}_{CE}^c\cap\mathcal{E}_{BE}$ in (155), we obtain

$$\begin{aligned}&\lim_{k\to\infty} \lim_{\delta',\delta,\tilde{\delta}\to 0} \sum_{\substack{P_{\tilde{U}\tilde{V}\tilde{W}} \\ \in\mathcal{T}^k_{\mathcal{U}\mathcal{V}\mathcal{W}}}} \sum_{\substack{(u^k,v^k,w^k) \\ \in T_{P_{\tilde{U}\tilde{V}\tilde{W}}}}} \Big[\mathbb{P}(U^k = u^k, V^k = v^k|H=1)\ \mathbb{P}(\mathcal{E}_{CE}^c|\mathcal{F})\ \mathbb{P}(\mathcal{E}_{BE}|\mathcal{F}_2)\ \mathbb{P}(\mathcal{D}_0|\mathcal{F}_{21}) \\ &\qquad\qquad \mathbb{P}(W^k(1) = w^k|U^k = u^k, V^k = v^k, J=1, f_B(J)=1, \mathcal{E}_{NE})\Big] \\ &\leq \lim_{k\to\infty} \lim_{\delta',\delta,\tilde{\delta}\to 0} \sum_{\substack{P_{\tilde{U}\tilde{V}\tilde{W}} \\ \in\mathcal{T}^k_{\mathcal{U}\mathcal{V}\mathcal{W}}}} \sum_{\substack{(u^k,v^k,w^k) \\ \in T_{P_{\tilde{U}\tilde{V}\tilde{W}}}}} \Big[\mathbb{P}(U^k = u^k, V^k = v^k|H=1)\ \mathbb{P}(\mathcal{E}_{BE}|\mathcal{F}_2)\ \mathbb{P}(\mathcal{D}_0|\mathcal{F}_{21}) \\ &\qquad\qquad \mathbb{P}(W^k(1) = w^k|U^k = u^k, V^k = v^k, J=1, f_B(J)=1, \mathcal{E}_{NE})\Big] \\ &\leq \lim_{k\to\infty} \max_{\substack{P_{\tilde{U}\tilde{V}\tilde{W}}\in \\ \mathcal{T}_2(P_{UW},P_V)}} e^{kH(\tilde{U}\tilde{V}\tilde{W})} e^{-k\left(H(\tilde{U}\tilde{V})+D(P_{\tilde{U}\tilde{V}}||Q_{UV})+H(\tilde{W}|\tilde{U})+R-I(U;W|V)\right)} \\ &\qquad e^{(|\mathcal{U}||\mathcal{V}||\mathcal{W}|\log(k+1)+|\mathcal{U}||\mathcal{W}|\log(k+1))}\end{aligned}$$

$$= \lim_{k \to \infty} e^{-k\tilde{E}_{2k}}, \tag{160}$$

where,

$$\begin{aligned}\tilde{E}_{2k} := \min_{\substack{P_{\tilde{U}\tilde{V}\tilde{W}} \in \\ \mathcal{T}_2(P_{UW}, P_V)}} & H(\tilde{U}\tilde{V}) + D(P_{\tilde{U}\tilde{V}}||Q_{UV}) + H(\tilde{W}|\tilde{U}) + R - I(U;W|V) \\ & - \frac{1}{k}|\mathcal{U}||\mathcal{V}||\mathcal{W}| \log(k+1) - \frac{1}{k}|\mathcal{U}||\mathcal{W}| \log(k+1) \\ \xrightarrow{(k)} & E_2(P_{W|U}, P_{SX}, \tau). \end{aligned} \tag{161}$$

Note that $\mathcal{E}_{BE}$ occurs only when $I(U;W) > R$.

Next, consider 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. (148), (149) and (142), it can be shown that the terms corresponding to this event in (155) result in the factor $E_3(P_{W|U}, P_{SX}, \tau)$ given in (14).

Finally, we analyze the case when the event $\mathcal{E}^c_{NE}$ 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

$$\begin{aligned}\mathbb{P}(\mathcal{D}_0 |\ U^k = u^k, V^k = v^k,\ \mathcal{E}^c_{NE}) =& \mathbb{P}(\mathcal{E}_{CE} |\ U^k = u^k, V^k = v^k,\ \mathcal{E}^c_{NE}) \\ & \mathbb{P}(\mathcal{D}_0 |\ U^k = u^k, V^k = v^k,\ \mathcal{E}^c_{NE} \cap\ \mathcal{E}_{CE}). \end{aligned} \tag{162}$$

It follows from Borade's coding scheme [31] that

$$\mathbb{P}(\mathcal{E}_{CE} |\ U^k = u^k, V^k = v^k,\ \mathcal{E}^c_{NE}) \leq e^{-nE_m\left(P_{SX}, P_{Y|X}\right)} = e^{-k\tau E_m\left(P_{SX}, P_{Y|X}\right)}. \tag{163}$$

When binning is performed at the encoder, $\mathcal{D}_0$ occurs only if there exists a sequence $\hat{W}^k$ in the bin $\hat{M} \neq 0$ such that $(\hat{W}^k, V^k) \in T^k_{[WV]_{\tilde{\delta}}}$. Also, recalling that the encoder sends the error message $M = 0$ independent of the source codebook $\mathcal{C}$, it can be shown using standard arguments that for such $v^k \in T^k_{[V]_{\delta'''}}$,

$$\mathbb{P}(\mathcal{D}_0 |\ U^k = u^k, V^k = v^k,\ \mathcal{E}^c_{NE} \cap\ \mathcal{E}_{CE}) \leq e^{-k(R - I(U;W|V) - \delta_5)}, \tag{164}$$

where $\delta_5 = \delta_1 + \delta_3 + \delta'$. Thus, from (162), (163) and (164), we obtain

$$\lim_{\delta,\delta',\tilde{\delta}\to 0} \sum_{u^k,v^k} \mathbb{P}(U^k=u^k, V^k=v^k|H=1)\ \mathbb{P}(\mathcal{D}_0|\ U^k=u^k, V^k=v^k,\ \mathcal{E}_{NE}^c \cap\ \mathcal{E}_{CE})$$
$$\leq e^{-kD(P_V||Q_V)}\ e^{-nE_m\left(P_{SX},P_{Y|X}\right)}\ e^{-k(R-I(U;W|V)))}. \tag{165}$$

On the other hand, when binning is not performed, $\mathcal{D}_0$ occurs only if $(W^k(\hat{M}), V^k) \in T^k_{[WV]_{\tilde{\delta}}}$ and in this case, we obtain

$$\lim_{\delta,\delta',\tilde{\delta}\to 0} \sum_{u^k,v^k} \mathbb{P}(U^k=u^k, V^k=v^k|H=1)\ \mathbb{P}(\mathcal{D}_0|\ U^k=u^k, V^k=v^k,\ \mathcal{E}_{NE}^c \cap\ \mathcal{E}_{CE})$$
$$\leq e^{-kD(P_V||Q_V)}\ e^{-nE_m\left(P_{SX},P_{Y|X}\right)}\ e^{-kI(V;W)}. \tag{166}$$

This results in the factor $E_4(P_{W|U}, P_{SX}, \tau)$ in (15). Since the T2EE is lower bounded by the minimal value of the exponent due to the various type 2 error events, this completes the proof of the theorem.

Appendix B
Proof of Theorem 6

We only give a sketch of the proof as the intermediate steps follow similarly to those in the proof of Theorem 2. For brevity, in the proof below, we denote the information theoretic quantities like $I_{\hat{P}}(U,S;\bar{W})$, $T^n_{[\hat{P}_{US\bar{W}}]_\delta}$, etc., that are computed with respect to joint distribution $\hat{P}_{UVS\bar{W}X'XY}(P_S, P_{\bar{W}|US}, P_{X'|S}, P_{X|US\bar{W}}) := P_{UV}P_S P_{\bar{W}|US}P_{X'|S}P_{X|US\bar{W}}P_{Y|X}$ by $I(U,S;\bar{W})$, $T^n_{[US\bar{W}]_\delta}$, etc. As in the proof of Theorem 2, δ, δ', δ'' and $\tilde{\delta}$ appearing in the proof below denote arbitrarily small positive numbers subject to delta-convention [24] and certain other constraints that will be specified in the course of the proof.

Codebook Generation:

Fix distributions $(P_S, P_{\bar{W}|US}, P_{X'|S}, P_{X|US\bar{W}}) \in \mathcal{B}_h$ and let

$$\hat{P}_{UVS\bar{W}X'XY}(P_S, P_{\bar{W}|US}, P_{X'|S}, P_{X|US\bar{W}}) = P_{UV}P_S P_{\bar{W}|US}P_{X'|S}\ P_{X|US\bar{W}}\ P_{Y|X}.$$

Generate a sequence S^n i.i.d. according to $\prod_{i=1}^n P_S(s_i)$. The realization $S^n=s^n$ is revealed to both the encoder and detector. Generate the quantization codebook $\mathcal{C} = \{\bar{w}^n(j),\ j \in [e^{n(I(U,S;\bar{W})+\delta')}]\}$, where each codeword $\bar{w}^n(j)$ is generated independently according to the distribution $\prod_{i=1}^n \hat{P}_{\bar{W}}$, where $\hat{P}_{\bar{W}} = \sum_{(u,s)\in\mathcal{U}\times\mathcal{S}} P_U(u)P_S(s)P_{\bar{W}|US}(\bar{w}|u,s)$.

Encoding: If (u^n, s^n) is typical, i.e., $(u^n, s^n) \in T^n_{[US]_{\delta''}}$, the encoder first looks for a sequence $\bar{w}^n(j)$ such that $(u^n, s^n, \bar{w}^n(j)) \in T^n_{[USW]_\delta}$, $\delta > \delta''$. If there exists multiple such codewords, it chooses one among them uniformly at random. The encoder transmits $X^n = x^n$ over the channel, where X^n is generated according to the distribution $\prod_{i=1}^n P_{X|US\bar{W}}(x_i|u_i, s_i, \bar{w}_i)$. If $(u^n, s^n) \notin T^k_{[US]_{\delta''}}$ or such an index j does not exist, the encoder generates the channel input $X'^n = x'^n$ randomly according to $\prod_{i=1}^n P_{X'|S}(x'_i|s_i)$.

Decoding: Given the side information sequence $V^n = v^n$, received sequence $Y^n = y^n$ and s^n, the detector first checks if $(v^n, s^n, y^n) \in T^n_{[VSY]_{\tilde{\delta}}}$, $\tilde{\delta} > \delta$. If the check is unsuccessful, H_1 is declared. Else, searches for a typical sequence $\hat{\bar{w}}^n = \bar{w}^n(\hat{j}) \in T^k_{[\bar{W}]_{\hat{\delta}}}$, $\hat{\delta} = |\bar{\mathcal{W}}|\delta$ in the codebook such that

$$\hat{j} = \underset{l: \bar{w}^n(l) \in T^n_{[\bar{W}]_{\hat{\delta}}}}{\arg\min} \; H_e(\bar{w}^n(l)|v^n, s^n, y^n).$$

If $(v^n, s^n, y^n, \hat{\bar{w}}^n) \in T^n_{[VSY\bar{W}]_{\tilde{\delta}}}$, H_0 is declared, else H_1 is declared.

Analysis of Type 1 error:

A type 1 error occurs only if one of the following events happen.

$$\tilde{\mathcal{E}}_{TE} = \left\{ (U^n, V^n, S^n) \notin T^n_{[UVS]_{\bar{\delta}}}, \; \bar{\delta} = \frac{\delta''}{|\mathcal{V}|} \right\}$$

$$\tilde{\mathcal{E}}_{EE} = \left\{ \nexists \; j \in \left[e^{n(I(U,S;\bar{W}) + \delta')} \right] : (U^n, S^n, \bar{W}^n(j)) \in T^n_{[US\bar{W}]_\delta} \right\}$$

$$\tilde{\mathcal{E}}_{ME} = \left\{ (V^n, S^n, \bar{W}^n(J)) \notin T^n_{[VS\bar{W}]_{\tilde{\delta}}} \right\}$$

$$\tilde{\mathcal{E}}_{CE} = \left\{ (V^n, S^n, \bar{W}^n(J), Y^n) \notin T^n_{[VS\bar{W}Y]_{\tilde{\delta}}} \right\}$$

$$\tilde{\mathcal{E}}_{DE} = \Bigg\{ \exists \; l \in \left[e^{n(I(U,S;\bar{W}) + \delta')} \right], \; l \neq J, \; \bar{W}^n(l)) \in T^n_{[\bar{W}]_{\hat{\delta}}}, \\ H_e(\bar{W}^n(l)|V^n, S^n, Y^n) \leq H_e(\bar{W}^n(J)|V^n, S^n, Y^n) \Bigg\}$$

By the weak law of large numbers, $\tilde{\mathcal{E}}_{TE}$ tends to 0 asymptotically with n for any $\bar{\delta} > 0$. The covering lemma guarantees that $\tilde{\mathcal{E}}_{EE} \cap \tilde{\mathcal{E}}^c_{TE}$ tends to 0 doubly exponentially for $\bar{\delta} < \delta$ and δ' appropriately chosen. Given $\tilde{\mathcal{E}}^c_{EE} \cap \tilde{\mathcal{E}}^c_{TE}$ holds, it follows from the Markov lemma and the weak law of large numbers, respectively, that $\mathbb{P}(\tilde{\mathcal{E}}_{ME})$ and $\mathbb{P}(\tilde{\mathcal{E}}_{CE})$ tends to zero asymptotically for $\tilde{\delta} > \delta$ (appropriately chosen). Next, we consider the probability of the event $\tilde{\mathcal{E}}_{DE}$. Given that $\tilde{\mathcal{E}}^c_{CE} \cap \tilde{\mathcal{E}}^c_{ME} \cap \tilde{\mathcal{E}}^c_{EE} \cap \tilde{\mathcal{E}}^c_{TE}$ holds, note that $\lim_{n \to \infty} H_e(\bar{W}^n(J)|V^n, S^n, Y^n) \to H(\bar{W}|V, S, Y)$ as

$\tilde{\delta} \rightarrow 0$. Hence, similarly to that shown in Appendix A, it can be shown that

$$\mathbb{P}(\tilde{\mathcal{E}}_{DE}|\tilde{\mathcal{E}}_{CE}^c \cap \tilde{\mathcal{E}}_{ME}^c \cap \tilde{\mathcal{E}}_{EE}^c \cap \tilde{\mathcal{E}}_{TE}^c) \leq e^{-n(I_{\hat{P}}(\bar{W};V,S,Y) - I_{\hat{P}}(U,S;\bar{W}) - \delta_6^{(n)})}.$$

where $\delta_6^{(n)} \xrightarrow{(n)} 0$ as $\tilde{\delta}, \delta' \rightarrow 0$. Hence, if $I(U;\bar{W}|S) < I(\bar{W};Y,V|S)$, the probability of the events causing Type 1 error tends to zero asymptotically.

Analysis of Type 2 error: The analysis of the T2EE is very similar to that of the SHTCC scheme given in Appendix A. Hence, only a sketch of the proof is provided, with the differences from the proof of the SHTCC scheme highlighted.

Let

$$\bar{\mathcal{E}}_0 := \{(U^n, S^n) \notin T^n_{[US]_{\delta''}}\}. \tag{167}$$

Then, as in Appendix A, the type 2 error probability can be written as

$$\begin{aligned}
&\beta\left(n, n, f^{(n,n)}, g^{(n,n)}\right) \\
&\leq \sum_{(u^n,v^n) \in \mathcal{U}^n \times \mathcal{V}^n} \mathbb{P}(U^n = u^n, V^n = v^n | H = 1)\Big[\mathbb{P}(\tilde{\mathcal{E}}_{EE} \cap \bar{\mathcal{E}}_0^c | U^n = u^n, V^n = v^n) \\
&\qquad + \mathbb{P}(\mathcal{D}_0 | U^n = u^n, V^n = v^n, \tilde{\mathcal{E}}_{NE}) + \mathbb{P}(\mathcal{D}_0 | U^n = u^n, V^n = v^n, \bar{\mathcal{E}}_0)\Big], \qquad (168)
\end{aligned}$$

where, $\tilde{\mathcal{E}}_{NE} := \tilde{\mathcal{E}}_{EE}^c \cap \bar{\mathcal{E}}_0^c$. As before, it is sufficient to restrict the analysis to the events $\tilde{\mathcal{E}}_{NE}$ and $\bar{\mathcal{E}}_0$ that dominate the type 2 error. Define the events

$$\begin{aligned}
\tilde{\mathcal{E}}_{T2} = \Big\{&\exists\ l \in \left[e^{n(I(U,S;\bar{W})+\delta')}\right],\ l \neq J,\ \bar{W}^n(l) \in T^n_{[\bar{W}]_{\hat{\delta}}}, \\
&(V^n, \bar{W}^n(l), S^n, Y^n) \in T^n_{[VS\bar{W}Y]_{\tilde{\delta}}}\Big\}, \qquad (169)
\end{aligned}$$

$$\tilde{\mathcal{F}} = \{U^n = u^n, V^n = v^n, J = 1, \bar{W}^n(1) = \bar{w}^n, S^n = s^n, Y^n = y^n, \tilde{\mathcal{E}}_{NE}\}, \tag{170}$$

$$\tilde{\mathcal{F}}_1 = \{U^n = u^n, V^n = v^n, J = 1, \bar{W}^n(1) = \bar{w}^n, S^n = s^n, Y^n = y^n, \tilde{\mathcal{E}}_{NE}, \tilde{\mathcal{E}}_{T2}^c\}, \tag{171}$$

$$\tilde{\mathcal{F}}_2 = \{U^n = u^n, V^n = v^n, J = 1, \bar{W}^n(1) = \bar{w}^n, S^n = s^n, Y^n = y^n, \tilde{\mathcal{E}}_{NE}, \tilde{\mathcal{E}}_{T2}\}. \tag{172}$$

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

$$
\begin{aligned}
&\mathbb{P}(\mathcal{D}_0|\ U^n=u^n, V^n=v^n, \tilde{\mathcal{E}}_{NE})\\
&=\sum_{j=1}^{e^{n(I(U,S;\bar{W})+\delta')}} \mathbb{P}(J=j|\ U^n=u^n, V^n=v^n, \tilde{\mathcal{E}}_{NE})\ \mathbb{P}(\mathcal{D}_0|U^n=u^n, V^n=v^n, J=1,\ \tilde{\mathcal{E}}_{NE})\\
&=\mathbb{P}(\mathcal{D}_0|U^n=u^n, V^n=v^n, J=1,\ \tilde{\mathcal{E}}_{NE})\\
&=\sum_{\substack{(\bar{w}^n,s^n,y^n)\\ \in\ \bar{\mathcal{W}}^n\times\mathcal{S}^n\times\mathcal{Y}^n}} \mathbb{P}(\bar{W}^n(1)=\bar{w}^n, S^n=s^n, Y^n=y^n|U^n=u^n, V^n=v^n, J=1,\ \tilde{\mathcal{E}}_{NE})\\
&\qquad \mathbb{P}(\mathcal{D}_0|U^n=u^n, V^n=v^n, J=1, \bar{W}^n(1)=\bar{w}^n, S^n=s^n, Y^n=y^n,\ \tilde{\mathcal{E}}_{NE})\\
&=\sum_{\substack{(\bar{w}^n,s^n,y^n)\\ \in\ \bar{\mathcal{W}}^n\times\mathcal{S}^n\times\mathcal{Y}^n}} \mathbb{P}(\bar{W}^n(1)=\bar{w}^n, S^n=s^n, Y^n=y^n|U^n=u^n, V^n=v^n, J=1,\ \tilde{\mathcal{E}}_{NE})\\
&\qquad \mathbb{P}(\mathcal{D}_0|\ \tilde{\mathcal{F}}).
\end{aligned} \tag{173}
$$

The last term in (173) can be upper bounded using the events in (170)-(172) as follows.

$$
\mathbb{P}(\mathcal{D}_0|\ \tilde{\mathcal{F}}) \leq \mathbb{P}(\mathcal{D}_0|\ \tilde{\mathcal{F}}_1) + \mathbb{P}(\tilde{\mathcal{E}}_{T2}|\ \tilde{\mathcal{F}})\ \mathbb{P}(\mathcal{D}_0|\ \tilde{\mathcal{F}}_2).
$$

We next analyze the R.H.S of (168), which upper bounds the type 2 error probability, in the limit $n \to \infty$ and $\delta, \delta', \tilde{\delta} \to 0$. We have,

$$
\mathbb{P}(\mathcal{D}_0|\tilde{\mathcal{F}}_1) = \begin{cases} 1, & \text{if } P_{u^n\bar{w}^n} \in T^n_{[U\bar{W}]_\delta} \\ & \text{and } P_{v^n\bar{w}^n s^n y^n} \in T^k_{[VS\bar{W}Y]_{\tilde{\delta}}}, \\ 0, & \text{otherwise.} \end{cases} \tag{174}
$$

Hence, the terms corresponding to the event $\tilde{\mathcal{F}}_1$ in (168) can be upper bounded (in the limit $\delta, \tilde{\delta} \to 0$) as

$$
\begin{aligned}
&\lim_{n\to\infty}\ \lim_{\delta',\delta,\tilde{\delta}\to 0} \sum_{\substack{(u^n,v^n,\bar{w}^n,s^n,y^n)\\ \in\ \mathcal{U}^n\times\mathcal{V}^n\times\bar{\mathcal{W}}^n\times\mathcal{S}^n\times\mathcal{Y}^n}} \Big[\mathbb{P}(U^n=u^n, V^n=v^n|H=1)\ \mathbb{P}(\mathcal{D}_0|\tilde{\mathcal{F}}_1)\\
&\qquad \mathbb{P}(\bar{W}^n(1)=\bar{w}^n, S^n=s^n, Y^n=y^n|U^n=u^n, V^n=v^n, J=1,\ \tilde{\mathcal{E}}_{NE})\Big]
\end{aligned}
$$

$$\leq \lim_{n\to\infty} \lim_{\delta',\delta,\tilde{\delta}\to 0} \sum_{\substack{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} \\ \in \mathcal{T}^n_{\mathcal{U}\mathcal{V}\bar{\mathcal{W}}\mathcal{S}\mathcal{Y}}}} \sum_{\substack{(u^n,v^n,\bar{w}^n,s^n,y^n) \\ \in T_{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}}}}} \Big[\mathbb{P}(U^n = u^n, V^n = v^n|H=1)\ \mathbb{P}(\mathcal{D}_0|\tilde{\mathcal{F}}_1)$$

$$\mathbb{P}(S^n = s^n, \bar{W}^n(1) = \bar{w}^n|U^n = u^n, J=1,\ \tilde{\mathcal{E}}_{NE})$$

$$\mathbb{P}(Y^n = y^n|U^n = u^n, S^n = s^n, J=1, \bar{W}^n(1) = \bar{w}^n,\ \tilde{\mathcal{E}}_{NE})\Big]$$

$$\leq \lim_{n\to\infty} \lim_{\delta',\delta,\tilde{\delta}\to 0} \sum_{\substack{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} \\ \in \mathcal{T}^n_{\mathcal{U}\mathcal{V}\bar{\mathcal{W}}\mathcal{S}\mathcal{Y}}}} \sum_{\substack{(u^n,v^n,\bar{w}^n,s^n,y^n) \\ \in T_{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}}}}} \Big[\mathbb{P}(\mathcal{D}_0|\tilde{\mathcal{F}}_1)\ e^{-n\left(H(\tilde{U}\tilde{V}) + D(P_{\tilde{U}\tilde{V}}||Q_{UV})\right)}$$

$$e^{-n\left(H(\tilde{S}\tilde{W}|\tilde{U}) - \frac{1}{n}|\mathcal{U}||\bar{\mathcal{W}}||\mathcal{S}|\log(n+1)\right)}\ e^{-n\left(H(\tilde{Y}|\tilde{U}\tilde{S}\tilde{W}) + D(P_{\tilde{Y}|\tilde{U}\tilde{S}\tilde{W}}||\hat{P}_{Y|US\bar{W}}|P_{\tilde{U}\tilde{S}\tilde{W}})\right)}\Big]$$

$$\leq \lim_{n\to\infty} \max_{\substack{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} \in \\ \mathcal{T}'_1(\hat{P}_{US\bar{W}}, \hat{P}_{VS\bar{W}Y})}} \Big[e^{-n\left(H(\tilde{U}\tilde{V}) + D(P_{\tilde{U}\tilde{V}}||Q_{UV})\right)}\ e^{-n\left(H(\tilde{S}\tilde{W}|\tilde{U}) - \frac{1}{n}|\mathcal{U}||\bar{\mathcal{W}}||\mathcal{S}|\log(n+1)\right)}$$

$$e^{-n\left(H(\tilde{Y}|\tilde{U}\tilde{S}\tilde{W}) + D(P_{\tilde{Y}|\tilde{U}\tilde{S}\tilde{W}}||\hat{P}_{Y|US\bar{W}}|P_{\tilde{U}\tilde{S}\tilde{W}})\right)} e^{n\left(H(\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}) - \frac{1}{n}||\mathcal{U}||\mathcal{V}||\bar{\mathcal{W}}||\mathcal{S}||\mathcal{Y}|\log(n+1)\right)}\Big]$$

$$= \lim_{n\to\infty} e^{-nE^*_{1n}}, \tag{175}$$

where,

$$E^*_{1n} = \min_{\substack{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} \in \\ \mathcal{T}'_1(\hat{P}_{US\bar{W}}, \hat{P}_{VS\bar{W}Y})}} \Bigg[H(\tilde{U}\tilde{V}) + D(P_{\tilde{U}\tilde{V}}||Q_{UV}) + H(\tilde{S}\tilde{W}|\tilde{U}) + H(\tilde{Y}|\tilde{U}\tilde{S}\tilde{W})$$

$$+ D(P_{\tilde{Y}|\tilde{U}\tilde{S}\tilde{W}}||\hat{P}_{Y|US\bar{W}}|P_{\tilde{U}\tilde{S}\tilde{W}}) - H(\tilde{U}\tilde{V}\tilde{W}\tilde{S}\tilde{Y}) - \frac{1}{n}(|\mathcal{U}||\bar{\mathcal{W}}| + |\mathcal{U}||\mathcal{V}||\bar{\mathcal{W}}||\mathcal{S}||\mathcal{Y}|)\log(n+1)\Bigg]$$

$$= \min_{\substack{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} \in \\ \mathcal{T}'_1(\hat{P}_{US\bar{W}}, \hat{P}_{VS\bar{W}Y})}} \left[\sum_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} \log\left(\frac{1}{P_{\tilde{U}\tilde{V}}} \frac{P_{\tilde{U}\tilde{V}}}{Q_{UV}} \frac{P_{\tilde{U}}}{P_{\tilde{U}\tilde{S}\tilde{W}}} \frac{1}{P_{\tilde{Y}|\tilde{U}\tilde{S}\tilde{W}}} \frac{P_{\tilde{Y}|\tilde{U}\tilde{S}\tilde{W}}}{\hat{P}_{Y|US\bar{W}}} P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}}\right) - o(1)\right]$$

$$= \min_{\substack{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} \in \\ \mathcal{T}'_1(\hat{P}_{US\bar{W}}, \hat{P}_{VS\bar{W}Y})}} \left[D(P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}}|Q_{UV}P_{\tilde{S}\tilde{W}|\tilde{U}}\hat{P}_{Y|US\bar{W}}) - o(1)\right]$$

$$\xrightarrow{(n)} E'_1(P_S, P_{\bar{W}|US}, P_{X|US\bar{W}}). \tag{176}$$

Here, (176) follows from the fact that $P_{\tilde{S}\tilde{W}|\tilde{U}} \to P_{S\bar{W}|U}$ given $\tilde{\mathcal{E}}_{NE}$, as $\delta \to 0$.

Next, consider the terms corresponding to the event $\tilde{\mathcal{F}}_2$ in (168). Given $\tilde{\mathcal{F}}_2$, $P_{\tilde{U}\tilde{W}} \in T^n_{[U\bar{W}]_\delta}$ and $\mathcal{D}_0$ occurs only if $(V^n, S^n, Y^n) \in T^n_{[VSY]_{\delta'''}}$, $\delta''' = |\bar{\mathcal{W}}|\tilde{\delta}$, and $H(\tilde{W}|\tilde{V},\tilde{S},\tilde{Y}) \geq H(\bar{W}|V,S,Y) -$

$\gamma_2(\tilde{\delta})$, for some $\gamma_2(\tilde{\delta}) \in O(\tilde{\delta})$. Thus, we have,

$$\lim_{n\to\infty} \lim_{\delta',\delta,\tilde{\delta}\to 0} \sum_{\substack{(u^n,v^n,\bar{w}^n,s^n,y^n)\\ \in\ \mathcal{U}^n\times\mathcal{V}^n\times\bar{\mathcal{W}}^n\times\mathcal{S}^n\times\mathcal{Y}^n}} \Big[\mathbb{P}(U^n=u^n,V^n=v^n|H=1)\ \mathbb{P}(\mathcal{D}_0|\tilde{\mathcal{F}}_2)\ \mathbb{P}(\tilde{\mathcal{E}}_{T2}|\tilde{\mathcal{F}})$$
$$\mathbb{P}(\bar{W}^n(1)=\bar{w}^n,S^n=s^n,Y^n=y^n|U^n=u^n,V^n=v^n,J=1,\ \tilde{\mathcal{E}}_{NE})\Big]$$
$$\leq \lim_{n\to\infty} \lim_{\delta',\delta,\tilde{\delta}\to 0} \sum_{\substack{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}}\in\\ \mathcal{T}^n(\mathcal{U}\times\mathcal{V}\times\bar{\mathcal{W}}\times\mathcal{S}\times\mathcal{Y})}} \sum_{\substack{(u^n,v^n,\bar{w}^n,s^n,y^n)\\ \in T_{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}}}}} \Big[\mathbb{P}(U^n=u^n,V^n=v^n|H=1)$$
$$\mathbb{P}(\mathcal{D}_0|\tilde{\mathcal{F}}_2)\ \mathbb{P}(\tilde{\mathcal{E}}_{T2}|\tilde{\mathcal{F}})\ \mathbb{P}(S^n=s^n,\bar{W}^n(1)=\bar{w}^n|U^n=u^n,J=1,\ \tilde{\mathcal{E}}_{NE})$$
$$\mathbb{P}(Y^n=y^n|U^n=u^n,S^n=s^n,J=1,\bar{W}^n(1)=\bar{w}^n,\ \tilde{\mathcal{E}}_{NE})\Big]$$
$$\leq \lim_{n\to\infty} \lim_{\delta',\delta,\tilde{\delta}\to 0} \sum_{\substack{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}}\in\\ \mathcal{T}^n(\mathcal{U}\times\mathcal{V}\times\bar{\mathcal{W}}\times\mathcal{S}\times\mathcal{Y})}} \sum_{\substack{(u^n,v^n,\bar{w}^n,s^n,y^n)\\ \in T_{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}}}}} \Big[e^{-n\left(H(\tilde{U}\tilde{V})+D(P_{\tilde{U}\tilde{V}}||Q_{UV})\right)}\mathbb{P}(\mathcal{D}_0|\tilde{\mathcal{F}}_2)$$
$$2\cdot e^{-n\left(I(\bar{W};V,S,Y)-I(U,S;\bar{W})-\delta_7\right)}\ e^{-n\left(H(\tilde{S}\tilde{W}|\tilde{U})-\frac{1}{n}|\mathcal{U}||\bar{\mathcal{W}}||\mathcal{S}|\log(n+1)\right)}$$
$$e^{-n\left(H(\tilde{Y}|\tilde{U}\tilde{S}\tilde{W})+D(P_{\tilde{Y}|\tilde{U}\tilde{S}\tilde{W}}||\hat{P}_{Y|US\bar{W}}|P_{\tilde{U}\tilde{S}\tilde{W}})\right)}\Big] \tag{177}$$

$$\leq \lim_{n\to\infty} \max_{\substack{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}}\in\\ \mathcal{T}_2'(\hat{P}_{UW},\hat{P}_{VSWY})}} \Big[e^{-n\left(H(\tilde{U}\tilde{V})+D(P_{\tilde{U}\tilde{V}}||Q_{UV})\right)}\ e^{-n\left(H(\tilde{S}\tilde{W}|\tilde{U})-\frac{1}{n}|\mathcal{U}||\bar{\mathcal{W}}||\mathcal{S}|\log(n+1)\right)}$$
$$e^{-n\left(I(\bar{W};V,S,Y)-I(U,S;\bar{W})-\delta_7-\frac{1}{n}\right)}$$
$$e^{-n\left(H(\tilde{Y}|\tilde{U}\tilde{S}\tilde{W})+D(P_{\tilde{Y}|\tilde{U}\tilde{S}\tilde{W}}||\hat{P}_{Y|US\bar{W}}|P_{\tilde{U}\tilde{S}\tilde{W}})\right)}\ e^{n\left(H(\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y})-\frac{1}{n}||\mathcal{U}||\mathcal{V}||\bar{\mathcal{W}}||\mathcal{S}||\mathcal{Y}|\log(n+1)\right)}\Big]$$
$$= \lim_{n\to\infty}\ e^{-nE_{2n}^*}, \tag{178}$$

where,

$$E_{2n}^* = \min_{\substack{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}}\in\\ \mathcal{T}_2'(\hat{P}_{US\bar{W}},\hat{P}_{VS\bar{W}Y})}} \left[D(P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}}|Q_{UV}P_{\tilde{S}\tilde{W}|\tilde{U}}\hat{P}_{Y|US\bar{W}}) + I(\bar{W};V,Y|S) - I(U;\bar{W}|S) - o(1)\right]$$
$$\xrightarrow{(n)} E_2'(P_S,P_{\bar{W}|US},P_{X|US\bar{W}}). \tag{179}$$

In (177), we used the fact that

$$\mathbb{P}(\tilde{\mathcal{E}}_{T2}|\tilde{\mathcal{F}}) \leq 2\cdot e^{-n\left(I(\bar{W};V,Y|S)-I(U;\bar{W}|S)-\delta_7\right)},$$

which follows from

$$\mathbb{P}\left(\bar{W}^n(l)=\tilde{w}^n|\tilde{\mathcal{F}}\right) \leq 2\ \mathbb{P}(\bar{W}^n(l)=\tilde{w}^n). \tag{180}$$

Eqn. (180) can be proved similarly to (147).

Finally, we consider the case when $\bar{\mathcal{E}}_0$ holds.

$$\begin{aligned}
&\lim_{n\to\infty}\lim_{\delta,\delta',\tilde{\delta}\to 0}\sum_{u^n,v^n}\mathbb{P}(U^n=u^n,V^n=v^n|H=1)\ \mathbb{P}(\mathcal{D}_0|\ U^n=u^n,V^n=v^n,\ \bar{\mathcal{E}}_0)\\
&=\lim_{n\to\infty}\lim_{\delta,\delta',\tilde{\delta}\to 0}\sum_{u^n,v^n}\mathbb{P}(U^n=u^n,V^n=v^n|H=1)\\
&\qquad\qquad\sum_{s^n,y^n}\mathbb{P}(S^n=s^n,Y^n=y^n,\mathcal{D}_0|\ U^n=u^n,V^n=v^n,\ \bar{\mathcal{E}}_0)\\
&=\lim_{n\to\infty}\lim_{\delta,\delta',\tilde{\delta}\to 0}\sum_{u^n,v^n}\mathbb{P}(U^n=u^n,V^n=v^n|H=1)\\
&\qquad\qquad\Big[\sum_{s^n,y^n}\mathbb{P}(S^n=s^n,Y^n=y^n|\ U^n=u^n,V^n=v^n,\ \bar{\mathcal{E}}_0)\\
&\qquad\qquad\mathbb{P}(\mathcal{D}_0|\ U^n=u^n,V^n=v^n,S^n=s^n,Y^n=y^n,\ \bar{\mathcal{E}}_0)\Big]\\
&=\lim_{n\to\infty}\lim_{\delta,\delta',\tilde{\delta}\to 0}\sum_{u^n,v^n}\mathbb{P}(U^n=u^n,V^n=v^n|H=1)\Big[\sum_{s^n,y^n}\mathbb{P}(S^n=s^n,Y^n=y^n|\ \bar{\mathcal{E}}_0)\\
&\qquad\qquad\mathbb{P}(\mathcal{D}_0|\ V^n=v^n,S^n=s^n,Y^n=y^n,\ \bar{\mathcal{E}}_0)\Big]\\
&=\lim_{n\to\infty}\lim_{\delta,\delta',\tilde{\delta}\to 0}\sum_{v^n,s^n,y^n}\mathbb{P}(V^n=v^n|H=1)\ \mathbb{P}(S^n=s^n,Y^n=y^n|\ \bar{\mathcal{E}}_0)\\
&\qquad\qquad\mathbb{P}(\mathcal{D}_0|\ V^n=v^n,S^n=s^n,Y^n=y^n,\ \bar{\mathcal{E}}_0).
\end{aligned} \tag{181}$$

The event $\mathcal{D}_0$ occurs only if there exists a sequence $(\bar{W}^n(l),V^n,S^n,Y^n)\in T^n_{[\bar{W}VSY]_{\tilde{\delta}}}$ for some $l\in[e^{n(I(U,S;\bar{W})+\delta')}]$. Noting that the quantization codebook is independent of the (V^n,S^n,Y^n) given that $\bar{\mathcal{E}}_0$ holds, it can be shown using standard arguments that

$$\mathbb{P}(\mathcal{D}_0|\ V^n=v^n,S^n=s^n,Y^n=y^n,\ \bar{\mathcal{E}}_0)\leq e^{-n(I(\bar{W};V,Y|S)-I(U;\bar{W}|S)-\delta_7)}. \tag{182}$$

Also,

$$\mathbb{P}(S^n=s^n,Y^n=y^n|\ \bar{\mathcal{E}}_0)\leq e^{-n(H(\tilde{S}\tilde{Y})+D(P_{\tilde{S}\tilde{Y}}||\check{Q}_{SY}))}. \tag{183}$$

Hence, using (182) and (183) in (181), we obtain

$$\lim_{n\to\infty} \lim_{\delta,\delta',\tilde{\delta}\to 0} \sum_{v^n,s^n,y^n} \mathbb{P}(V^n = v^n|H=1)\ \mathbb{P}(S^n = s^n, Y^n = y^n|\ \bar{\mathcal{E}}_0)$$

$$\mathbb{P}(\mathcal{D}_0|\ V^n = v^n, S^n = s^n, Y^n = y^n,\ \bar{\mathcal{E}}_0)$$

$$\leq \lim_{n\to\infty} \lim_{\delta,\delta',\tilde{\delta}\to 0} \sum_{v^n,s^n,y^n} e^{-n(H(\tilde{V})+D(P_{\tilde{V}}||Q_V))}\ e^{-n(H(\tilde{S}\tilde{Y})+D(P_{\tilde{S}\tilde{Y}}||\check{Q}_{SY}))}\ e^{-n(I(\bar{W};V,Y|S)-I(U;\bar{W}|S)-\delta_7)}$$

$$\leq \lim_{n\to\infty} (n+1)^{|\mathcal{V}||\mathcal{S}||\mathcal{Y}|} \max_{P_{\tilde{V}\tilde{S}\tilde{Y}}=\hat{P}_{VSY}} e^{nH(\tilde{V}\tilde{S}\tilde{Y})}\ e^{-n(H(\tilde{V})+D(P_{\tilde{V}}||Q_V))}\ e^{-n(H(\tilde{S}\tilde{Y})+D(P_{\tilde{S}\tilde{Y}}||\check{Q}_{SY}))}$$

$$e^{-n(I(\bar{W};V,Y|S)-I(U;\bar{W}|S))}$$

$$= \lim_{n\to\infty} e^{-nE^*_{3n}},$$

where,

$$E^*_{3n} = \min_{P_{\tilde{V}\tilde{S}\tilde{Y}}=\hat{P}_{VSY}} D(P_{\tilde{V}\tilde{S}\tilde{Y}}||\check{Q}_{VSY}) + I(\bar{W};V,Y|S) - I(U;\bar{W}|S) - |\mathcal{V}||\mathcal{S}||\mathcal{Y}|\log(n+1)$$

$$\xrightarrow{(n)} E'_3\left(P_S, P_{\bar{W}|US}, P_{X'|S}, P_{X|US\bar{W}}\right).$$

Since the T2EE is lower bounded by the minimal value of the exponent due to the various type 2 error events, this completes the proof of the theorem.

Appendix C
Proof of Theorem 8

We can write

$$\begin{aligned} I_{\hat{P}^*}(\bar{W};V,Y|S) - I_{\hat{P}^*}(U;\bar{W}|S) &= I_{\hat{P}^*}(W,X;V,Y|S) - I_{\hat{P}^*}(U;W,X|S) \\ &= I_{\hat{P}^*}(X;Y|S) - I_{\hat{P}^*}(U;W) + I_{\hat{P}^*}(V;W) \\ &= I_{\hat{P}^*}(X;Y|S) - I_{\hat{P}^*}(U;W|V) > 0. \end{aligned} \tag{184}$$

This implies that $(P^*_S, P^*_{W|U}, P_{X'|S}, P^*_{X|S}) \in \mathcal{B}_h$. Now,

$$E'_1(P^*_S, P^*_{W|U}, P^*_{X|S})$$

$$= \min_{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{X}\tilde{Y}}\in\mathcal{T}'_1(\hat{P}^*_{USWX},\hat{P}^*_{VSWXY})} D(P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{X}\tilde{Y}}||\hat{Q}_{UVSWXY})$$

$$\geq \min_{P_{\tilde{U}\tilde{V}\tilde{W}} \in \mathcal{T}_1(\hat{P}^*_{UW}, \hat{P}^*_{VW})} D(P_{\tilde{U}\tilde{V}\tilde{W}}||\hat{Q}_{UVW})$$

$$+ \min_{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{X}\tilde{Y}} \in \mathcal{T}_1'(\hat{P}^*_{USWX}, \hat{P}^*_{VSWXY})} D(P_{\tilde{S}\tilde{X}\tilde{Y}|\tilde{U}\tilde{V}\tilde{W}}||\hat{Q}_{SXY}|P_{\tilde{U}\tilde{V}\tilde{W}})$$

$$= E_1(P^*_{W|U}) + \min_{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{X}\tilde{Y}} \in \mathcal{T}_1'(\hat{P}^*_{USWX}, \hat{P}^*_{VSWXY})} D(P_{\tilde{S}\tilde{X}\tilde{Y}|\tilde{U}\tilde{V}\tilde{W}}||\hat{Q}_{SXY}|P_{\tilde{U}\tilde{V}\tilde{W}}). \tag{185}$$

Also, by choosing $X' = S$, it follows from (184) that

$$E_3'\left(P^*_S, P^*_{W|U}, P_{X'|S}, P^*_{X|S}\right) - E_4(P^*_{W|U}, P^*_{SX}, 1)$$

$$\geq D(\hat{P}^*_{VSY}||\check{Q}^*_{VSY}) - E_m\left(P^*_{SX}, P_{Y|X}\right) - D(P_V||Q_V) \tag{186}$$

$$= D(\hat{P}^*_{SY}||\check{Q}^*_{SY}) - E_m\left(P^*_{SX}, P_{Y|X}\right) = 0. \tag{187}$$

Here, (187) follows from the fact that since $\mathcal{S} = \mathcal{X}$ and $X' = S$, $D(\hat{P}^*_{SY}||\check{Q}^*_{SY})$ is equal to $E_m\left(P^*_{SX}, P_{Y|X}\right)$.

Again, it follows from (184) that

$$E_2'\left(P^*_S, P^*_{W|U}, P^*_{X|S}\right) - E_3(P^*_{W|U}, P^*_{SX}, 1)$$

$$\geq \min_{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{X}\tilde{Y}} \in \mathcal{T}_2'\left(\hat{P}^*_{USWX}, \hat{P}^*_{VSWXY}\right)} D\left(P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{X}\tilde{Y}}||\hat{Q}^*_{UVSWXY}\right)$$

$$- \min_{P_{\tilde{U}\tilde{V}\tilde{W}} \in \mathcal{T}_3(P_{UW}, P_V)} D(P_{\tilde{U}\tilde{V}\tilde{W}}||\hat{Q}^*_{UVW}) - E_x\left(I_{\hat{P}^*}(X;Y|S), P^*_{SX}, P_{Y|X}\right)$$

$$= E_h\left(P^*_S, P^*_{W|U}, P_{X'|S}, P^*_{X|S}\right) - E_s\left(P^*_S, P^*_{W|U}, P_{X'|S}, P^*_{X|S}\right)$$

$$\geq 0, \tag{188}$$

where, (188) follows from the condition given in the theorem. This shows that for $\mathbf{b} = (P^*_S, P^*_{W|U}, P_{X'|S}, P^*_{X|S}) \in \mathcal{B}_h$, each of the argument inside the minimum in (23) is greater than or equal to $\kappa_s(1)$, thus implying that $\kappa_h \geq \kappa_s(1)$. This completes the proof.

Appendix D

JHTCC scheme achieves optimal T2EE for TACI

Let $\tau = 1$ and recall that for TACI, $V = (E, Z)$ and $Q_{UEZ} = P_{UZ}P_{E|Z}$. To show the above claim, note that (185) and (187) holds for any $(P_S, P_{W|U}, P_{X'|S}, P_{X|S}) \in \mathcal{B}_h$ such that $(P_{W|U}, P_{SX}) \in \mathcal{B}$ (in place of $(P^*_S, P^*_{W|U}, P_{X'|S}, P^*_{X|S}) \in \mathcal{B}_h$ such that $(P^*_{W|U}, P^*_{SX}) \in \mathcal{B}$). From this and the achievability proof of Theorem 9 where it is shown that $\kappa_s(1) \geq I(E;W|Z)$, it follows that proving

$$E_2'\left(P_S, P_{W|U}, P_{X|S}\right) \geq I(E;W|Z), \tag{189}$$

for any $(P_{W|U}, P_{SX}) \in \mathcal{B}'(1, C)$ suffices, where $\mathcal{B}'(1, C)$ is defined in (38). This can be done as follows.

$$\begin{aligned}
E_2'\left(P_S, P_{W|U}, P_{X|S}\right) &= \min_{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} \in \mathcal{T}_2'\left(\hat{P}_{US\bar{W}}, \hat{P}_{VS\bar{W}Y}\right)} D\left(P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} || \hat{Q}_{UVS\bar{W}Y}\right) \\
&\qquad + I_{\hat{P}}(W, X; E, Z, Y|S) - I_{\hat{P}}(U; W, X|S) \\
&\geq I_{\hat{P}}(X;Y|S) + I_{\hat{P}}(W;E,Z) - I_{\hat{P}}(U;W) \\
&= I_{\hat{P}}(X;Y|S) - I_{\hat{P}}(U;W|E,Z) \\
&\geq I_{\hat{P}}(U;W|Z) - I_{\hat{P}}(U;W|E,Z) = I_{\hat{P}}(E;W|Z) \qquad (190) \\
&= I(E;W|Z),
\end{aligned}$$

where, (190) follows from the assumption that $(P_{W|U}, P_{SX}) \in \mathcal{B}'(1, C)$. Thus, the JHTCC scheme achieves the optimal T2EE for TACI over a DMC.

APPENDIX E
PROOF OF LEMMA 13

Note that for $\tau = 0$, $n = 0$, which implies that the observer does not transmit anything. Then, from Stein's lemma [5] for ordinary hypothesis testing, (i) and (ii) follows, where $\theta(0) := D(P_V||Q_V)$. When $\tau > 0$, the proof is similar to that of Theorem 1 in [5]. 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 [5] and is omitted here.

For given encoding functions $f_1^{(k,n)}, \ldots, f_L^{(k,n)}$, define

$$\beta'\left(k, n, f_1^{(k,n)}, \ldots, f_L^{(k,n)}, \epsilon\right) := \inf_{g^{(k,n)}} \beta\left(k, n, f_1^{(k,n)}, \ldots, f_L^{(k,n)},\ g^{(k,n)}\right), \tag{191}$$

such that

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

and

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

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

$$\bar{\beta}(k,\tau,\epsilon) := \inf_{\substack{f_1^{(k,n)},\ldots,f_L^{(k,n)} \\ n \leq \tau k}} \beta'\left(k,n,f_1^{(k,n)},\ldots,f_L^{(k,n)},\epsilon\right).$$

Now, let $k \in \mathbb{Z}^+$, $\tilde{\delta} > 0$ 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 blocklength, encoding functions and channel outputs respectively, such that $k\theta(k,\tau) - D\big(P_{Y_{\mathcal{L}}^{\tilde{n}_k}V^k} || Q_{Y_{\mathcal{L}}^{\tilde{n}_k}V^k}\big) < k\tilde{\delta}$. 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 [5] to the sequences $\left\{\tilde{Y}_{\mathcal{L}}^{\tilde{n}_k}(j), V^k(j)\right\}_{j \in \mathbb{Z}^+}$, we have

$$\begin{aligned}\limsup_{j \to \infty} \frac{\log\left(\bar{\beta}(kj,\tau,\epsilon)\right)}{kj} &\leq \frac{-D\big(P_{Y_{\mathcal{L}}^{\tilde{n}_k}V^k} || Q_{Y_{\mathcal{L}}^{\tilde{n}_k}V^k}\big)}{k}, \\ &\leq -(\theta(k,\tau) - \tilde{\delta}). \end{aligned} \tag{193}$$

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

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

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

APPENDIX F
PROOF OF THEOREM 15

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 $\prod_{j'=1}^{j} P_{X_l^n|U_l^k = u_l^k(j')}$. The main encoder performs uniform random binning on E^k, i.e., $f_s^{kj}(E^{kj}) = M$, where M is selected uniformly at random from the set $\mathcal{M} := \{1, 2, \cdots, e^{kjR}\}$.

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

Analysis of the probability of error: The events that can possibly lead to an error under the above encoding and decoding rules are given below:

$$\mathcal{E}_1 = \left\{(E^{kj}, Y_{\mathcal{L}}^{nj}, Z^{kj}) \notin T^j_{[E^k Y_{\mathcal{L}}^n, Z^k]_\delta}\right\}$$

$$\mathcal{E}_2 = \begin{Bmatrix} \exists\ \tilde{E}^{kj} \neq E^{kj},\ f_s^{kj}(\tilde{E}^{kj}) = f_s^{kj}(E^{kj}) \\ (\tilde{E}^{kj}, Y_{\mathcal{L}}^{nj}, Z^{kj}) \in T^j_{[E^k Y^n_{\mathcal{L}} Z^k]_\delta} \end{Bmatrix}.$$

By the joint typicality lemma [33], $\mathbb{P}(\mathcal{E}_1) \to 0$ as $j \to \infty$. Also,

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

Hence, $\mathbb{P}(\mathcal{E}_2) \to 0$ as $j \to \infty$ if $R > H(E^k|Y^n_{\mathcal{L}}, Z^k) + \delta$, $(Z^k, E^k) - U_l^k - X_l^n - Y_l^n,\ l \in \mathcal{L}$. Since $\delta > 0$ is arbitrary, this proves that $R > \frac{H(E^k|Y^n_{\mathcal{L}}, Z^k)}{k}$ is an achievable rate.

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

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

This proves the converse by noting that $(Z^k, E^k) - U_l^k - X_l^n - Y_l^n,\ l \in \mathcal{L}$ holds for any communication scheme.

Appendix G
Proof of Theorem 16

From the source-channel separation theorem, an upper bound on $R(\tau)$ can be obtained by the intersection of the BT inner bound [33, Th. 12.1] with the capacity region $(C_1, \ldots, C_L, C_s)$, where C_s is the rate available over the noiseless link from the encoder of source E to the decoder.

Writing the BT inner bound explicitly, we obtain that for all $\mathcal{G} \subseteq \mathcal{L}$ (including the null-set),

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

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

where the auxiliary r.v.'s $W_{\mathcal{L}}$ satisfy (115) and $|\mathcal{W}_l| \leq |\mathcal{U}_l| + 4$. Taking the infimum of C_s over all such $W_{\mathcal{L}}$ and denoting it by $R^i(\tau)$, we obtain the second inequality in (118). The other direction in (118) is obtained similarly by using the BT outer bound [33, Th. 12.2]. Since $R(\tau)$ is equal to the infimum in (110), substituting (118) in (110) proves (119).

APPENDIX H
ALTERNATE PROOF OF PROPOSITION 9

For $L = 1$, note that the Markov chain conditions in (115) and (117) are identical. Hence,

$$R^i(\tau) = R^o(\tau) = R(\tau). \tag{195}$$

Using the BT inner bound in [33, Ch.12], we obtain $R(\tau)$ as the infimum of R' such that

$$H(E|Z, W) \leq R', \tag{196}$$

$$I(U; W|E, Z) \leq \tau C, \tag{197}$$

$$H(E|Z, W) + I(U; W|Z) \leq \tau C + R', \tag{198}$$

for some auxiliary r.v. satisfying $(E, Z) - U - W$. Hence,

$$\begin{aligned} R(\tau) = \inf_W \max \big(&H(E|W, Z),\ H(E|W, Z) \\ &+ I(U; W|Z) - \tau C \big), \end{aligned} \tag{199}$$

such that $(E, Z) - U - W$ and (197) hold. We next prove that (199) can be simplified as

$$R(\tau) = \inf_W H(E|Z, W), \tag{200}$$

such that $I(U;W|Z) \leq \tau C$ and $(E,Z) - U - W$ 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. W' such that $(E,Z) - U - W'$,

$$I(U;W'|Z) = \tau C, \tag{201}$$

$$H(E|W',Z) \leq H(E|W,Z) + I(U;W|Z) - \tau C, \tag{202}$$

and (197) is satisfied with W replaced by W'. Setting

$$W' = \begin{cases} W, & \text{with probability 1-p,} \\ \text{constant,} & \text{with probability p,} \end{cases} \tag{203}$$

suffices, where p is chosen such that $I(U;W'|Z) = \tau C$. To see this, first note that $H(E|W',Z)$ is an increasing function of p, while $I(U;W'|Z)$ and $I(U;W'|E,Z)$ are decreasing functions of p. Hence, it is possible to choose p such that (201) and (197) are satisfied with W' in place of W. It is clear that such a choice of W' also satisfies $(E,Z) - U - W'$. To complete the proof of (200), it remains to be shown that for such a W', (202) holds. We can write,

$$H(E|W',Z) = (1-p)H(E|W,Z) + pH(E|Z). \tag{204}$$

Taking derivative with respect to p, we obtain

$$\frac{d}{dp}H(E|W',Z) = I(E;W|Z). \tag{205}$$

Similarly,

$$\frac{d}{dp}H(U|W',Z) = I(U;W|Z). \tag{206}$$

By the data processing inequality [24] applied to $(E,Z) - U - W$, we have that $I(E;W|Z) \leq I(U;W|Z)$. Hence,

$$\frac{d}{dp}H(E|W',Z) \leq \frac{d}{dp}H(U|W',Z). \tag{207}$$

From (207), it follows that

$$F(p) := H(E|W',Z) + I(U;W'|Z) - \tau C \tag{208}$$

is a decreasing function of p. Together with the fact that $H(E|W', Z)$ is increasing with p, it then follows that (202) is satisfied for W' chosen in (203). Having shown (200), the proof is now complete from (119) and (195).

References

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

[2] H. Chernoff, "A measure of asymptotic efficiency for tests of a hypothesis based on a sum of observations," *Ann. Math. Statist.*, vol. 23, no. 4, pp. 493–507, 1952.

[3] W. Hoeffding, "Asymptotically optimal tests for multinominal distributions," *Ann. Math. Stat.*, vol. 36, no. 2, pp. 369–400, 1965.

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

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

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

[7] 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.

[8] 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.

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

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

[11] 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.

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

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

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

[15] ——, "Distributed binary detection with lossy data compression," *IEEE Trans. Inf. Theory*, vol. 63, no. 8, pp. 5207–5227, Mar. 2017.

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

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

[18] R. E. Blahut, "Hypothesis testing and information theory," *IEEE Trans. Inf. Theory*, vol. 20, no. 4, pp. 405–417, Jul. 1974.

[19] T. S. Han and K. Kobayashi, "Exponential-type error probabilities for multiterminal hypothesis testing," *IEEE Trans. Inf. Theory*, vol. 35, no. 1, pp. 2–14, Jan. 1989.

[20] S. Amari and T. S. Han, “Statistical inference under multiterminal rate restrictions: A differential geometric approach,” *IEEE Trans. Inf. Theory*, vol. 35, no. 2, pp. 217–227, Mar. 1989.

[21] T. S. Han and S. Amari, “Statistical inference under multiterminal data compression,” *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2300–2324, Oct. 1998.

[22] S. Watanabe, “Neyman-pearson test for zero-rate multiterminal hypothesis testing,” *arXiv:1611.08175*.

[23] N. Weinberger and Y. Kochman, “On the reliability function of distributed hypothesis testing under optimal detection,” *arXiv:1801.03687*.

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

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

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

[27] 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.

[28] T. Berger, *Multiterminal source coding. In G. Longo (ed.) The Information Theory Approach to Communications*. Springer-Verlag, NewYork, 1978.

[29] S. Y. Tung, “Multiterminal source coding,” *Ph.D. thesis, Cornell University, Ithaca, NY*.

[30] 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.

[31] S. P. Borade, “When all information is not created equal,” *Ph.D. Thesis, MIT*, 2008.

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

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

[34] J. Korner and K. Marton, “How to encode the modulo-two sum of binary sources,” *IEEE Trans. Inf. Theory*, vol. 25, no. 2, pp. 219–221, Mar. 1979.

[35] A. B. Wagner, B. G. Kelly, and Y. Altuğ, “Distributed rate-distortion with common components,” *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4035–4057, Jul. 2011.

[36] A. B. Wagner and V. Anantharam, “An improved outer bound for multiterminal source coding,” *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 1919–1937, May. 2008.

[37] W. Kang and S. Ulukus, “A new data processing inequality and its applications in distributed source and channel coding,” *IEEE Trans. Inf. Theory*, vol. 57, no. 1, pp. 56–69, Jan. 2011.