

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 (HT) 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 HT 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 I error probability and the type II error-exponent of the HT is explored. Single-letter lower bounds on the optimal type II error-exponent are obtained by using two different coding schemes, a separate HT and channel coding scheme and a joint HT and channel coding scheme based on hybrid coding for the matched bandwidth case. Exact single-letter characterization of the same is established for the special case of testing against conditional independence, and it is shown to be achieved by the separate HT and channel coding scheme. An example is provided where the joint scheme achieves a strictly better performance than the separation based scheme.

I. INTRODUCTION

Given data samples, statistical hypothesis testing (HT) deals with the problem of ascertaining the true assumption, that is, the true hypothesis, about the data from among a set of hypotheses. In modern communication networks (like in sensor networks, cloud computing and Internet of things (IoT)), data is gathered at multiple remote nodes, referred to as *observers*, and transmitted over noisy links to another node for further processing. Often, there is some prior statistical knowledge available about the data, for example, that the joint probability distribution of the data belongs to a certain prescribed set. In such scenarios, it is of interest to identify the true underlying probability distribution, and this naturally leads to the problem of distributed HT over noisy channels. The simplest case of such a scenario is depicted in Fig. 1, where there is a single observer and two possibilities for the joint distribution of the data. The observer observes k independent and identically distributed (i.i.d) data samples U^k , and communicates its observation to the detector by n uses of the DMC, characterized by the conditional distribution $P_{Y|X}$. The detector performs a binary hypothesis test on the joint distribution of the data (U^k, V^k) to decide between them, based on

This work is supported in part by the European Research Council (ERC) through Starting Grant BEACON (agreement #677854). A part of this work was presented at the International Symposium on Information theory (ISIT), Aachen, 2017 [15].

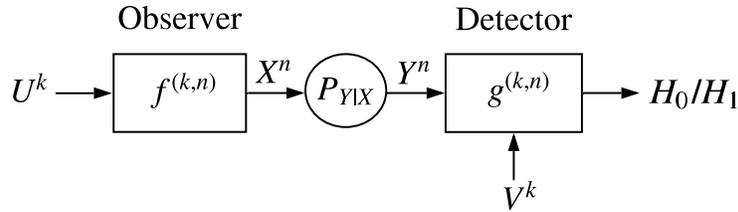


Fig. 1: Distributed HT over a DMC.

the channel outputs Y^n as well as its own observations V^k . The null and the alternate hypothesis of the hypothesis test are given by

$$H_0 : (U^k, V^k) \sim \prod_{i=1}^k P_{UV}, \quad (1a)$$

and

$$H_1 : (U^k, V^k) \sim \prod_{i=1}^k Q_{UV}, \quad (1b)$$

respectively. Our goal is to characterize the optimal exponential rate of decay of the type II error probability asymptotically, known as the *type II error-exponent* (henceforth, also referred to as *error-exponent*) for a prescribed constraint on the type I 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 error-exponent is characterized by the well-known lemma of Stein [1] (see also [2]). The study of distributed statistical inference under communication constraints was conceived by Berger in [3]. In [3], and in the follow up literature summarized below, communication from the observers to the detector are assumed to be over rate-limited error-free channel. Some of the fundamental results in this setting for the case of a single observer was established by Ahlswede and Csiszár in [4]. They obtained a tight single-letter characterization of the optimal error-exponent for a special case of HT known as *testing against independence* (TAI), in which, $Q_{UV} = P_U \times P_V$. Furthermore, the authors established a lower bound on the optimal error-exponent for the general HT case, and proved a *strong converse* result, which states that the optimal achievable error-exponent is independent of the constraint on the type I error probability. A tighter lower bound for the general HT problem is established by Han [5], which recovers the corresponding lower bound in [4]. Han also considered complete data compression in a related setting where either U , 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 error-exponent achieved in these three settings are equal. In contrast, a single-letter characterization of the optimal error-exponent for even the TAI with two-sided compression and general rate constraints remains open till date. Shalaby et al. [6] extended the complete data compression result of Han to show that the optimal error-exponent 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. [7] obtained a tighter lower bound on the optimal error-exponent for general HT by considering quantization and binning at the encoder along with a minimum empirical-entropy decoder. Rahman and Wagner [8] 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 error-exponent for *testing against conditional independence* (TACI), in which, $V = (E, Z)$ and $Q_{UEZ} = P_{UZ}P_{E|Z}$. Extensions of the distributed HT problem has also been considered in several other interesting scenarios involving multiple detectors [9], multiple observers [10], interactive HT [11], [12], collaborative HT [13], HT with lossy source reconstruction [14], HT over a multi-hop relay network [16], etc., in which, the authors obtain a single-letter characterization of the optimal error-exponent in some special cases.

While the works mentioned above have studied the unsymmetric case of focusing on the error-exponent for a constraint on the type I error probability, other works have analyzed the trade-off between the type I and type II error probabilities in the exponential sense. In this direction, the optimal trade-off between the type I and type II error-exponents in the centralized scenario is obtained in [17]. The distributed version of this problem is first studied in [18], 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 [19] and [20], 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 [21], 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 I and type II error-exponents for the general distributed HT problem remains open. Recently, an inner bound for this trade-off is obtained in [22], 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 error-exponent similar to the rate-constraint R in the noiseless setting. Another measure of the noisiness of the channel is the so-called *reliability function* $E(R, P_{Y|X})$ [23], 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 error-exponent for distributed HT over a noisy channel. Indeed, in Theorem 2 given below, we provide a lower bound on the optimal error-exponent that depends on the *expurgated exponent* at rate R , $E_x(R, P_{Y|X})$, which is a lower bound on $E(R, P_{Y|X})$ [24]. However, surprisingly, it will turn out that the reliability function does not play a role in the characterization of the error-exponent for TACI in the regime of vanishing type I error probability constraint.

The goal of this paper is to study the best attainable error-exponent for distributed HT over a DMC with a single observer 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. 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 [7], 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 [24] [23] of the channel along with the best channel coding error-exponent for a single special message. The channel coding scheme is based on the Borade-Nakiboğlu-Zheng unequal error-protection scheme [25]. As we show later, the SHTCC scheme achieves the optimal error-exponent for TACI.

Although the SHTCC scheme is attractive due to its modular design, *joint source channel coding* (JSCC) schemes are known to outperform separation based schemes in several different contexts, for example, the error exponent for reliable transmission of a source over a DMC [26], reliable transmission of correlated sources over a multiple-access channel [27], etc., to name a few. While in separation based schemes coding is usually performed by first quantizing the observed source sequence to an index, and transmitting the channel codeword corresponding to that index (independent of the source sequence), JSCC schemes allow the channel codeword to be dependent on the source sequence, in addition to the quantization index. Motivated by this, we propose a second scheme, referred to as the *joint HT and channel coding* (JHTCC) scheme, based on *hybrid coding* [28] for the communication between the observer and the detector.

Our main contributions can be summarized as follows.

- (i) We propose two different coding schemes (namely, SHTCC and JHTCC) for distributed HT over a DMC, and analyze the error-exponents achieved by these schemes.
- (ii) We obtain an exact single-letter characterization of the optimal error-exponent for the special case of TACI with a vanishing type I error probability constraint, and show that it is achievable by the SHTCC scheme.
- (iii) We provide an example where the JHTCC scheme achieves a strictly better error-exponent than the SHTCC scheme.

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. The achievable schemes are presented in Section III and the optimality results for special cases are discussed in Section IV. Finally, Section V 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 a finite set \mathcal{X} is denoted by $|\mathcal{X}|$. The set of all probability distributions on alphabet \mathcal{X} is denoted by $\mathcal{P}_{\mathcal{X}}$. Similar notations apply for set of conditional probability distributions, e.g., $\mathcal{P}_{\mathcal{Y}|\mathcal{X}}$. $X - Y - Z$ denotes that X , Y and Z form a Markov chain. For $m \in \mathbb{Z}^+$, X^m denotes the sequence X_1, \dots, X_m . Following the notation in [23], for a probability distribution P_X on r.v. X , $T_{P_X}^m$ and $T_{[P_X]_{\delta}}^m$ (or $T_{[X]_{\delta}}^m$) denote the set of sequences $x^m \in \mathcal{X}^m$ of type P_X and the set of P_X -typical sequences, respectively. The set of all possible types of sequences of length m with alphabet \mathcal{X} is denoted by $\mathcal{T}_{\mathcal{X}}^m$, and $\cup_{m \in \mathbb{Z}^+} \mathcal{T}_{\mathcal{X}}^m$ is denoted by $\mathcal{T}_{\mathcal{X}}$. Similar notations apply for pair's and other larger combinations of r.v.'s, e.g., $T_{P_{XY}}^m$, $T_{[P_{XY}]_{\delta}}^m$, $\mathcal{T}_{\mathcal{X}\mathcal{Y}}^m$, $\mathcal{T}_{\mathcal{X}\mathcal{Y}}$, etc.. The standard information theoretic quantities like 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} , are denoted by $D(P_X||Q_X)$, $H_{P_X}(X)$, $H_{P_{XY}}(X|Y)$ and $I_{P_{XY}}(X;Y)$, respectively. When the distribution of the r.v.'s involved are clear from the context, the last three quantities are denoted simply by $H(X)$, $H(X|Y)$ and $I(X;Y)$, respectively. Given realizations $X^m = x^m$ and $Y^m = y^m$, $H_e(x^m|y^m)$ denotes the conditional empirical entropy defined as

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

where $P_{\tilde{X}\tilde{Y}}$ denote the joint type of (x^m, y^m) , and $:=$ represents equality by definition (throughout this paper). For $a \in \mathbb{R}^+$, $[a]$ denotes the set of integers $\{1, 2, \dots, [a]\}$. All logarithms considered in this paper are with respect to the base e unless specified otherwise. For any set \mathcal{G} , \mathcal{G}^c denotes the set complement. $a_k \xrightarrow{(k)} b$ represents $\lim_{k \rightarrow \infty} a_k = b$. Similar notations are used for inequalities that hold asymptotically, e.g., $a_k \xrightarrow{(k)} \geq b_k$ denotes $\lim_{k \rightarrow \infty} a_k \geq b$. $\mathbb{P}(\mathcal{E})$ denotes the probability of event \mathcal{E} . For functions $f_1 : \mathcal{A} \rightarrow \mathcal{B}$ and $f_2 : \mathcal{B} \rightarrow \mathcal{C}$, $f_2 \circ f_1$ 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. The encoder (at the observer) observes U^k , and transmits codeword $X^n = f^{(k,n)}(U^k)$, where $f^{(k,n)} : \mathcal{U}^k \rightarrow \mathcal{X}^n$ represents the encoding function (possibly stochastic). Let $\tau := \frac{n}{k}$ denote the *bandwidth ratio*. The channel output Y^n is given by the probability law

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

i.e., the channels between the observers and the detector are independent of each other and memoryless. Depending on the received symbols Y^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 hypothesis test, where 0 and 1 denote H_0 and H_1 , respectively, and $\mathcal{A}_{(k,n)} \subseteq \mathcal{Y}^n \times \mathcal{V}^k$ denote the acceptance region for H_0 . Then, the decision rule $g^{(k,n)} : \mathcal{Y}^n \times \mathcal{V}^k \rightarrow \{0, 1\}$ is given by

$$g^{(k,n)}(y^n, v^k) = 1 - \mathbb{1}((y^n, v^k) \in \mathcal{A}_{(k,n)}).$$

Let

$$\begin{aligned} \alpha(k, n, f^{(k,n)}, g^{(k,n)}) &:= 1 - P_{Y^n V^k}(\mathcal{A}_{(k,n)}), \\ \text{and } \beta(k, n, f^{(k,n)}, g^{(k,n)}) &:= Q_{Y^n V^k}(\mathcal{A}_{(k,n)}), \end{aligned}$$

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

Definition 1. An error-exponent κ is (τ, ϵ) achievable if there exists a sequence of integers k , corresponding sequences of encoding function $f^{(k, n_k)}$ and decision rules $g^{(k, n_k)}$ such that $n_k \leq \tau k$, $\forall k$,

$$\liminf_{k \rightarrow \infty} \frac{-1}{k} \log \left(\beta \left(k, n_k, f^{(k, n_k)}, g^{(k, n_k)} \right) \right) \geq \kappa, \quad (4a)$$

$$\text{and } \limsup_{k \rightarrow \infty} \alpha \left(k, n_k, f^{(k, n_k)}, g^{(k, n_k)} \right) \leq \epsilon. \quad (4b)$$

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

$$\kappa(\tau, \epsilon) := \sup \{ \kappa' : \kappa' \text{ is } (\tau, \epsilon) \text{ achievable} \}. \quad (5)$$

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

It is well known that the Neyman-Pearson test [29] gives the optimal trade-off between the type I and type II error probabilities, and hence, also between the error-exponents in HT. It follows that the optimal error-exponent for distributed HT over a DMC is achieved when the channel-input X^n is generated correlated with U^k according to some optimal conditional distribution $P_{X^n|U^k}$, and the optimal Neyman-Pearson test is performed on the data available (both received and observed) at the detector. It can be shown, similarly to [4, Theorem 1], that the optimal error-exponent for vanishing type I error probability constraint is characterized by the multi-letter expression (see [30]) given by

$$\lim_{\epsilon \rightarrow 0} \kappa(\tau, \epsilon) = \sup_{\substack{P_{X^n|U^k} \in \mathcal{P}_{\mathcal{X}^n|U^k}, \\ k, n \in \mathbb{Z}^+, n \leq \tau k}} \frac{1}{k} D(P_{Y^n V^k} \| Q_{Y^n V^k}). \quad (6)$$

However, the above expression does not single-letterize in general, and hence, is intractable as it involves optimization over large dimensional probability simplexes when k and n are large. Moreover, the encoder and the detector of a scheme achieving the error-exponent given in (6) would be computationally complex to implement from a practical viewpoint. Consequently, we establish two computable single-letter lower bounds on $\kappa(\tau, \epsilon)$ in the next section by using the SHTCC and JHTCC schemes.

III. ACHIEVABLE SCHEMES

In [7], Shimokawa et al. obtained a lower bound on the optimal error-exponent 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 type¹ of the observed sequence $U^k = u^k$ is transmitted by the encoder to the detector, which is useful to improve the performance of the hypothesis test. In fact, in order to achieve the error-exponent proposed in [7], it is sufficient to send a message indicating whether U^k is typical or not, rather than sending the 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.

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

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 $f_s^{(k)} : \mathcal{U}^k \rightarrow \mathcal{M} = \{0, 1, \dots, \lceil e^{kR} \rceil\}$ generates an index $M = f_s^{(k)}(U^k)$ and the channel encoder $f_c^{(k,n)} : \mathcal{M} \rightarrow \tilde{\mathcal{C}} = \{X^n(j), j \in [0 : \lceil e^{kR} \rceil]\}$ 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 $g_c^{(k,n)} : \mathcal{Y}^n \rightarrow \mathcal{M}$ maps the channel-output Y^n into an index $\hat{M} = g_c^{(k,n)}(Y^n)$, and $g_s^{(k)} : \mathcal{M} \times \mathcal{V}^k \rightarrow \{0, 1\}$ 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 [25]. The red-alert exponent is defined as

$$E_m(R, P_{Y|X}) := \max_{\substack{P_{SX}: S=\mathcal{X}, \\ I(X;Y|S)=R, \\ S-X-Y}} \sum_{s \in \mathcal{S}} P_S(s) D(P_{Y|S=s} || P_{Y|X=s}). \quad (7)$$

Borade et al.'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

$$E_x(R, P_{Y|X}) := \max_{P_X} \max_{\rho \geq 1} \left\{ -\rho R - \rho \log \left(\sum_{x, \tilde{x}} P_X(x) P_X(\tilde{x}) \left(\sum_y \sqrt{P_{Y|X}(y|x) P_{Y|X}(y|\tilde{x})} \right)^{\frac{1}{\rho}} \right) \right\}, \quad (8)$$

is the *expurgated* exponent at rate R [24] [23]. Let

$$E_m(P_{SX}, P_{Y|X}) := \sum_{s \in \mathcal{S}} P_S(s) D(P_{Y|S=s} || P_{Y|X=s}), \quad (9)$$

where, $S = \mathcal{X}$ 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 et al.'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 $\tilde{\mathcal{C}}$, and encoder and

decoder as in Borade et al.'s scheme, such that the rate is $0 \leq R \leq I(X; Y|S)$ 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 $E_x(R, P_{SX}, P_{Y|X})$. Note that $E_m(P_{SX}, P_{Y|X})$ and $E_x(R, P_{SX}, P_{Y|X})$ denote Borade et al.'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 et al.'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

$$\begin{aligned} & \kappa_s(\tau) \\ & := \sup_{\substack{(P_{W|U}, P_{SX}, R) \\ \in \mathcal{B}(\tau, P_{Y|X})}} \min \{E_1(P_{W|U}), E_2(P_{W|U}, P_{SX}, \tau), E_3(P_{W|U}, P_{SX}, \tau), E_4(P_{W|U}, P_{SX}, \tau)\}, \end{aligned} \quad (10)$$

where

$$\mathcal{B}(\tau, P_{Y|X}) := \left\{ (P_{W|U}, P_{SX}, R) : \mathcal{S} = \mathcal{X}, P_{UVW SXY}(P_{W|U}, P_{SX}) := P_{UV} P_{W|U} P_{SX} P_{Y|X}, \right. \\ \left. I_P(U; W|V) \leq R < \tau I_P(X; Y|S) \right\}, \quad (11)$$

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

$$\begin{aligned} & E_2(P_{W|U}, P_{SX}, R) \\ & := \begin{cases} \min_{P_{\tilde{U}\tilde{V}\tilde{W}} \in \mathcal{T}_2(P_{UW}, P_V)} D(P_{\tilde{U}\tilde{V}\tilde{W}} \| Q_{UVW}) + R - I_P(U; W|V), & \text{if } I_P(U; W) > R, \\ \infty, & \text{otherwise,} \end{cases} \end{aligned} \quad (13)$$

$$\begin{aligned} & E_3(P_{W|U}, P_{SX}, R, \tau) \\ & := \begin{cases} \min_{P_{\tilde{U}\tilde{V}\tilde{W}} \in \mathcal{T}_3(P_{UW}, P_V)} D(P_{\tilde{U}\tilde{V}\tilde{W}} \| Q_{UVW}) + R - I_P(U; W|V) + \tau E_x\left(\frac{R}{\tau}, P_{SX}\right), & \text{if } I_P(U; W) > R, \\ \min_{P_{\tilde{U}\tilde{V}\tilde{W}} \in \mathcal{T}_3(P_{UW}, P_V)} D(P_{\tilde{U}\tilde{V}\tilde{W}} \| Q_{UVW}) + I_P(V; W) + \tau E_x\left(\frac{R}{\tau}, P_{SX}\right), & \text{otherwise,} \end{cases} \end{aligned} \quad (14)$$

$$E_4(P_{W|U}, P_{SX}, R, \tau) := \begin{cases} D(P_V \| Q_V) + R - I_P(U; W|V) + \tau E_m(P_{SX}), & \text{if } I_P(U; W) > R, \\ D(P_V \| Q_V) + I_P(V; W) + \tau E_m(P_{SX}), & \text{otherwise,} \end{cases} \quad (15)$$

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

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

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

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

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 $E_4(P_{W|U}, P_{SX}, R, \tau)$ can be understood to correspond to distinct events that can possibly lead to a type II error. Note that $E_1(P_{W|U})$ and $E_2(P_{W|U}, P_{SX}, R)$ are the same terms appearing in the error-exponent achieved by the Shimokawa et al.'s scheme [7] for the noiseless channel setting, while $E_3(P_{W|U}, P_{SX}, R, \tau)$ and $E_4(P_{W|U}, P_{SX}, R, \tau)$ are additional terms introduced due to the noisiness of the channel. $E_3(P_{W|U}, P_{SX}, R, \tau)$ corresponds to the event when $M \neq 0$, $\hat{M} \neq M$ and $g_s^{(k)}(\hat{M}, V^k) = 0$, whereas $E_4(P_{W|U}, P_{SX}, R, \tau)$ is due to the event when $M = 0$, $\hat{M} \neq M$ and $g_s^{(k)}(\hat{M}, V^k) = 0$. 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 $\tau E_m(P_{SX}) := 0$.

Remark 3. *In the SHTCC scheme, although we use Borade et al.'s scheme for channel coding, that is concerned specifically with the protection of a special message when the ordinary message rate is R , any other channel coding scheme with the same rate can be employed. For instance, the ordinary message can be transmitted with an error-exponent equal to the reliability function $E(R, P_{Y|X})$ [23] of the channel $P_{Y|X}$ at rate R , while the special message achieves the maximum reliability possible subject to this constraint. However, it should be noted that a computable characterization of neither $E(R, P_{Y|X})$ (for all values of R) nor the associated best reliability achievable for a single message is known in general.*

Remark 4. *Similarly to the zero-rate compression scenario considered in [5] for the case of a rate-limited noiseless channel, it is possible to achieve an error-exponent of $\kappa_0(\tau)$ in general by using a one-bit communication scheme (see [30]), where*

$$\kappa_0(\tau) := \begin{cases} D(P_V||Q_V) & , \text{ if } \tau = 0, \\ \min \{\beta_0, \tau E_c + D(P_V||Q_V)\} & , \text{ otherwise.} \end{cases} \quad (16)$$

Here,

$$\beta_0 := \beta_0(P_U, P_V, Q_{UV}) := \min_{\substack{P_{\tilde{U}\tilde{V}}: \\ P_{\tilde{U}}=P_U, P_{\tilde{V}}=P_V}} D(P_{\tilde{U}\tilde{V}}||Q_{UV}), \quad (17)$$

$$\text{and } E_c := E_c(P_{Y|X}) := D(P_{Y|X=a}||P_{Y|X=b}), \quad (18)$$

where a and b denote channel input symbols that satisfy

$$(a, b) = \arg \max_{(x, x') \in \mathcal{X} \times \mathcal{X}} D(P_{Y|X=x}||P_{Y|X=x'}). \quad (19)$$

Note that β_0 denotes the optimal error-exponent for distributed HT over a noiseless channel, when the communication rate-constraint is zero [5] [6].

In [30], it is shown that the one-bit communication scheme mentioned in Remark 4 achieves the optimal error-

exponent for HT over a DMC, i.e., when the detector has no side-information. Moreover, it is also proved that optimal error-exponent is not improved if the type I error probability constraint is relaxed; and hence, strong converse holds. In the limiting case of zero channel capacity, i.e., $C(P_{Y|X}) = 0$, it is intuitive to expect that communication from the observer to the detector does not improve the achievable error-exponent for distributed HT. In Appendix C below, we show that this is indeed the case in a strong converse sense, i.e., the optimal error-exponent depends only on the side-information V^k , and is given by $D(P_V||Q_V)$, for any constraint $\epsilon \in (0, 1)$ on the type I error probability. This is in contrast to the zero-rate compression case considered in [5], where one bit of communication between the observer and detector can achieve a strictly positive error-exponent, in general.

The SHTCC schemes introduced above performs 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.

B. JHTCC Scheme

Hybrid coding is a form of JSCC introduced in [28] 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 simplicity², we assume the *matched-bandwidth* scenario, i.e., $k = n$ ($\tau = 1$). 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{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 error-exponent 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{W}^n , the detector performs the hypothesis test using \hat{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.

²For the case $\tau \neq 1$, as mentioned in [28], we can consider hybrid coding over super symbols U^{k^*} and X^{n^*} , where k^* and n^* are some integers satisfying the constraint $n^* \leq \tau k^*$. This amounts to enlarging the source and side-information r.v.'s alphabets, and thus results in a harder optimization problem over the conditional probability distributions $P_{\bar{W}|U^{k^*}S}$ and $P_{X^{n^*}|U^{k^*}S\bar{W}}$ given in Theorem 5. However, we omit its description since the technique is standard and only adds notational clutter.

Theorem 5. $\kappa(1, \epsilon) \geq \kappa_h, \forall \epsilon \in (0, 1]$, where

$$\kappa_h := \sup_{\mathbf{b} \in \mathcal{B}_h} \min \left\{ E'_1(P_S, P_{\bar{W}|US}, P_{X|US\bar{W}}), E'_2(P_S, P_{\bar{W}|US}, P_{X|US\bar{W}}), \right. \\ \left. E'_3(P_S, P_{\bar{W}|US}, P_{X'|US}, P_{X|US\bar{W}}) \right\}, \quad (20)$$

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

$$E'_1(P_S, P_{\bar{W}|US}, P_{X|US\bar{W}}) := \min_{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} \in \mathcal{T}'_1(\hat{P}_{US\bar{W}}, \hat{P}_{VS\bar{W}Y})} D(P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} \| \hat{Q}_{UVS\bar{W}Y}), \quad (21)$$

$$E'_2(P_S, P_{\bar{W}|US}, P_{X|US\bar{W}}) := \min_{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} \in \mathcal{T}'_2(\hat{P}_{US\bar{W}}, \hat{P}_{VS\bar{W}Y})} D(P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} \| \hat{Q}_{UVS\bar{W}Y}) \\ + I_{\hat{P}}(\bar{W}; V, Y|S) - I_{\hat{P}}(U; \bar{W}|S), \quad (22)$$

$$E'_3(P_S, P_{\bar{W}|US}, P_{X'|US}, P_{X|US\bar{W}}) := D(\hat{P}_{VS\bar{W}Y} \| \check{Q}_{VS\bar{W}Y}) + I_{\hat{P}}(\bar{W}; V, Y|S) - I_{\hat{P}}(U; \bar{W}|S), \quad (23)$$

$$\hat{Q}_{UVS\bar{W}X'XY}(P_S, P_{\bar{W}|US}, P_{X'|US}, P_{X|US\bar{W}}) := Q_{UV}P_S P_{\bar{W}|US} P_{X'|US} P_{X|US\bar{W}} P_{Y|X}, \quad (24)$$

$$\check{Q}_{UVS\bar{W}X'XY}(P_S, P_{X'|US}) := Q_{UV}P_S P_{X'|US} \mathbb{1}(X = X') P_{Y|X}, \quad (25)$$

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

$$\mathcal{T}'_2(\hat{P}_{US\bar{W}}, \hat{P}_{VS\bar{W}Y}) := \{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} \in \mathcal{T}_{UVS\bar{W}Y} : P_{\tilde{U}\tilde{S}\tilde{W}} = \hat{P}_{US\bar{W}}, P_{\tilde{V}\tilde{S}\tilde{Y}} = \hat{P}_{VS\bar{W}Y},$$

$$H(\tilde{W}|\tilde{V}, \tilde{S}, \tilde{Y}) \geq H_{\hat{P}}(\bar{W}|V, S, Y)\}.$$

The proof of Theorem 5 is given in Appendix B. The different factors inside the minimum in (20) 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{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 $E'_3(P_S, P_{\bar{W}|US}, P_{X'|US}, P_{X|US\bar{W}})$ are related to the events that only the decoding error and both the encoding and decoding errors occur, respectively. From Theorem 2 and Theorem 5, we have the following corollary.

Corollary 6.

$$\kappa(1, \epsilon) \geq \max \{\kappa_h, \kappa_s(1)\}, \forall \epsilon \in (0, 1]. \quad (26)$$

It is well-known that in the context of JSCC, hybrid coding recovers separate source-channel coding as a special case [28]. It is also known that hybrid coding, of which uncoded transmission is a special case, strictly outperforms separation based schemes in certain multi-terminal settings [27]. Below, we provide an example where the error-exponent achieved by the JHTCC scheme is strictly better than that achieved by the SHTCC scheme, i.e., $\kappa_h > \kappa_s(1)$.

Example 1. Let $\mathcal{U} = \mathcal{V} = \mathcal{X} = \mathcal{Y} = \{0, 1\}$ and $P_U = Q_U = [0.5 \ 0.5]$. Let

$$P_{V|U} = \begin{bmatrix} 1-p_0 & p_0 \\ p_0 & 1-p_0 \end{bmatrix}, Q_{V|U} = \begin{bmatrix} 1-p_1 & p_1 \\ p_1 & 1-p_1 \end{bmatrix}, \text{ and } P_{Y|X} = \begin{bmatrix} 1-q & q \\ q & 1-q \end{bmatrix},$$

where $q = 0.2$, $p_0 = 0.8$ and $p_1 = 0.25$. For this example, we have $\kappa_h \geq 0.3244 > 0.161 \geq \kappa_s(1)$.

Proof: Note that $P_V = Q_V = [0.5 \ 0.5]$, and

$$H_Q(V|W) \geq H_P(\bar{V}|W) = H_P(V|W), \quad \bar{V} = V \oplus 1, \quad (27)$$

for any W that satisfies $V - U - W$, since

$$P_{\bar{V}|U} = \begin{bmatrix} 1-\bar{p}_0 & \bar{p}_0 \\ \bar{p}_0 & 1-\bar{p}_0 \end{bmatrix}$$

with $\bar{p}_0 = 0.2 < p_1$. Then, the lower bound $\kappa_s(1)$ simplifies as

$$\kappa_s(1) = \sup_{\substack{(P_{W|U}, P_{SX}, R) \\ \in \mathcal{B}(1, P_{Y|X})}} \min\{E_1(P_{W|U}), E_2(P_{W|U}, P_{SX}, R), E_3(P_{W|U}, P_{SX}, R, 1)\}. \quad (28)$$

To see this, consider an arbitrary $(P_{W|U}, P_{SX}, R) \in \mathcal{B}(1, P_{Y|X})$. We have

$$E_1(P_{W|U}) := \min_{P_{\bar{U}\bar{V}\bar{W}} \in \mathcal{T}_1(P_{UW}, P_{VW})} D(P_{\bar{U}\bar{V}\bar{W}} || Q_{UVW}), \quad (29)$$

$$E_2(P_{W|U}, P_{SX}, R) = \begin{cases} R - I_P(U; W|V), & \text{if } I_P(U; W) > R, \\ \infty, & \text{otherwise,} \end{cases} \quad (30)$$

$$E_3(P_{W|U}, P_{SX}, R, 1) := \begin{cases} R - I_P(U; W|V) + E_x(R, P_{SX}), & \text{if } I_P(U; W) > R, \\ I_P(V; W) + E_x(R, P_{SX}), & \text{otherwise,} \end{cases} \quad (31)$$

since $Q_{UVW} \in \mathcal{T}_2(P_{UW}, P_V) \cap \mathcal{T}_3(P_{UW}, P_V)$, which follows from (27), $P_{UW} = Q_{UW}$ and $P_V = Q_V$. This in turn implies that

$$P_{\bar{U}\bar{V}\bar{W}} \in \mathcal{T}_2(P_{UW}, P_V) \quad D(P_{\bar{U}\bar{V}\bar{W}} || Q_{UVW}) = P_{\bar{U}\bar{V}\bar{W}} \in \mathcal{T}_3(P_{UW}, P_V) \quad D(P_{\bar{U}\bar{V}\bar{W}} || Q_{UVW}) = 0. \quad (32)$$

Also, we have

$$E_4(P_{W|U}, P_{SX}, R, 1) := \begin{cases} R - I_P(U; W|V) + E_m(P_{SX}), & \text{if } I_P(U; W) > R, \\ I_P(V; W) + E_m(P_{SX}), & \text{otherwise,} \end{cases} \quad (33)$$

$$\geq E_3(P_{W|U}, P_{SX}, R, 1), \quad (34)$$

since $E_m(P_{SX}) \geq E_x(R, P_{SX})$ (the reliability of a special message in Borade et al.'s scheme is at least as good as that of an ordinary message), which implies (28). Given that (28) holds, $|\mathcal{S}|$ can be taken to be equal

to 1, and P_X can be chosen to be the capacity achieving channel input distribution ($P_X(0) = P_X(1) = 0.5$) which maximizes $E_x(R, P_{SX})$ (for any R) (see [24] and [23, Exercise 10.26]) without loss of generality. Hence, $I_P(X; Y) = C(P_{Y|X}) = 1 - h_b(q)$.

Let $r := h_b^{-1}(H_P(U|W)) = h_b^{-1}(H_Q(U|W))$, where $h_b^{-1} : [0, 1] \mapsto [0, 0.5]$ is the inverse of the binary entropy function given by $h_b(r) := -r \log_2(r) - (1-r) \log_2(1-r)$. First, consider

$$P_{W|U} \in \tilde{\mathcal{B}} := \{P_{W|U} : I_P(U; W) < I_P(X; Y) = 1 - h_b(q)\}. \quad (35)$$

Note that if $R \geq I_P(U; W)$, then $E_2(P_{W|U}, P_{SX}, R) = \infty$, and $E_3(P_{W|U}, P_{SX}, R, 1) = I_P(V; W) + E_x(R, P_{SX})$. Hence,

$$\begin{aligned} \min\{E_2(P_{W|U}, P_{SX}, R), E_3(P_{W|U}, P_{SX}, R, 1)\} &= I_P(V; W) + E_x(R, P_{SX}) \\ &\leq I_P(V; W) + E_x(I(U; W), P_{SX}), \end{aligned} \quad (36)$$

where (36) follows since $E_x(R, P_{SX})$ is a decreasing function of R . On the other hand, if $R < I_P(U; W)$, then $E_2(P_{W|U}, P_{SX}, R) = R - I_P(U; W|V)$ and $E_3(P_{W|U}, P_{SX}, R, 1) = R - I_P(U; W|V) + E_x(R, P_{SX})$ yielding that

$$\min\{E_2(P_{W|U}, P_{SX}, R), E_3(P_{W|U}, P_{SX}, R, 1)\} = R - I_P(U; W|V) \leq I_P(V; W). \quad (37)$$

Hence, from (36) and (37), we have

$$\sup_{\substack{(P_{W|U}, P_{SX}, R) \in \mathcal{B}(1, P_{Y|X}): \\ P_{W|U} \in \tilde{\mathcal{B}}}} \min\{E_2(P_{W|U}, P_{SX}, R), E_3(P_{W|U}, P_{SX}, R, 1)\} \leq I_P(V; W) + E_x(I(U; W), P_{SX}).$$

Also, note that (35) implies $h_b(r) \geq h_b(q)$; and hence, $r \in [q, 0.5]$. Thus, we can write

$$\begin{aligned} I_P(V; W) + E_x(I(U; W), P_{SX}) &= 1 - H_P(V|W) + E_x(I(U; W), P_{SX}) \\ &\leq 1 - h_b(h_b^{-1}(H(U|W)) * p_0) + E_x(I(U; W), P_{SX}) \end{aligned} \quad (38)$$

$$= 1 - h_b(r * p_0) + E_x(1 - h_b(r), P_{SX}) := f'(r), \quad (39)$$

where $p * q := (1-p)q + p(1-q)$, and (38) follows by an application of Mrs. Gerber's Lemma [31]. The plot of $f'(r)$ as a function of $r \in [q, 0.5]$ is shown in Fig. 2 below, which uses the expression for $E_x(R, P_{SX})$ given in [23, Exercise 10.26]. As is evident from the plot, the maximum value of $f'(r)$ is attained at $r = 0.5$, and equals $f'(0.5) = E_x(0) = -0.5 * 0.5 * \log_2(4q(1-q)) = 0.161$. It follows that

$$\sup_{\substack{(P_{W|U}, P_{SX}, R) \in \mathcal{B}(1, P_{Y|X}): \\ P_{W|U} \in \tilde{\mathcal{B}}}} \min\{E_2(P_{W|U}, P_{SX}, R), E_3(P_{W|U}, P_{SX}, R, 1)\} \leq 0.161. \quad (40)$$

Next, consider that

$$P_{W|U} \in \tilde{\mathcal{B}}^c := \{P_{W|U} : I_P(W; U) \geq 1 - h_b(q) \text{ and } I_P(U; W|V) \leq 1 - h_b(q)\}. \quad (41)$$

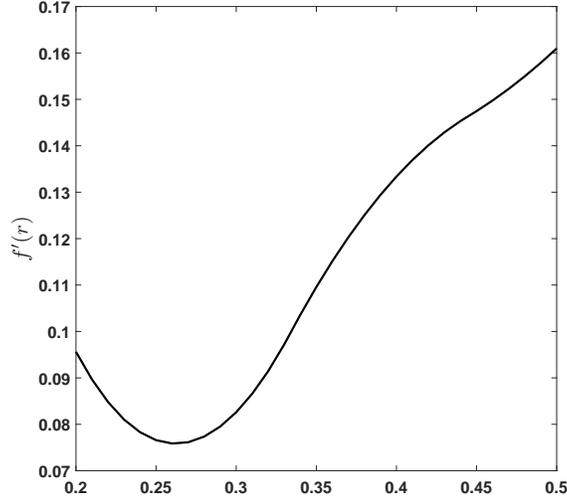


Fig. 2: Plot of $f'(r)$ in the range $r \in [0.2, 0.5]$.

Note that the first and second inequalities in (41) imply, respectively, that $r \in [0, q]$, and

$$1 - h_b(r) - (1 - h_b(r * p_0)) \leq 1 - h_b(q). \quad (42)$$

Also, since $R < 1 - h_b(q)$ holds for any $(P_{W|U}, P_{SX}, R) \in \mathcal{B}(1, P_{Y|X})$, we have $I_P(U; W) > R$, and hence,

$$\begin{aligned} & \sup_{\substack{(P_{W|U}, P_{SX}, R) \in \mathcal{B}(1, P_{Y|X}): \\ P_{W|U} \in \tilde{\mathcal{B}}^c}} \min\{E_2(P_{W|U}, P_{SX}, R), E_3(P_{W|U}, P_{SX}, R, 1)\} \\ & \leq \sup_{\substack{(P_{W|U}, P_{SX}, R) \in \mathcal{B}(1, P_{Y|X}): \\ P_{W|U} \in \tilde{\mathcal{B}}^c}} R - I_P(U; W|V) \\ & < 1 - h_b(q) - (h_b(r * p_0) - h_b(r)) \end{aligned} \quad (43)$$

$$\leq 1 - h_b(q * p_0) = 0.0956, \quad (44)$$

where (43) follows again from Mrs. Gerber's lemma, and (44) follows since the R.H.S. of (43) is an increasing function of r and hence the maximum is attained at $r = q$ in the range $[0, q]$. Thus, from (40) and (44), it follows that $\kappa_s(1) \leq 0.161$.

Finally, we show that the JHTCC scheme can achieve a strictly larger error-exponent, i.e., $\kappa_h > 0.161$. In fact, uncoded transmission which is a special case of the JHTCC scheme with $X = X' = U$, $W = S = \text{constant}$, achieves an error-exponent of

$$D(P_{VY}||Q_{VY}) = D_b(q * p_0||q * p_1) = D_b(0.68||0.35) = 0.3244, \quad (45)$$

where, D_b denotes the binary KL divergence defined as $D_b(p||q) = p \log_2 \left(\frac{p}{q}\right) + (1-p) \log_2 \left(\frac{1-p}{1-q}\right)$. Thus, we have shown that the error-exponent achieved by the JHTCC scheme is strictly greater than that achieved by the SHTCC scheme. \blacksquare

Thus far, we obtained lower bounds on the optimal error-exponent for distributed HT over a DMC, and showed via an example that the joint scheme strictly outperforms the separation based scheme in some cases. In order to get an exact characterization of the optimal error-exponent, a matching upper bound is required. However, obtaining a tight computable upper bound remains a challenging open problem in the general hypothesis testing case even when the channel is noiseless, and consequently, an exact computable characterization of the optimal error-exponent is unknown. However, as we show in the next section, the problem does admit single-letter characterization for TACI.

IV. OPTIMALITY RESULT FOR TACI

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

$$\kappa(\tau) = \lim_{\epsilon \rightarrow 0} \kappa(\tau, \epsilon). \quad (46)$$

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

$$\kappa(\tau) = \sup \left\{ \begin{array}{l} I(E; W|Z) : \exists W \text{ s.t. } I(U; W|Z) \leq \tau C(P_{Y|X}), \\ (Z, E) - U - W, \quad |\mathcal{W}| \leq |\mathcal{U}| + 1. \end{array} \right\}, \quad \tau \geq 0. \quad (47)$$

Proof: For the proof of achievability, we will show that $\kappa_s(\tau)$ when specialized to TACI recovers (47). Let $\mu > 0$ be a arbitrarily small positive number, and

$$\mathcal{B}'(\tau, P_{Y|X}) := \left\{ (P_{W|U}, P_{SX}, R_m) : \begin{array}{l} \mathcal{S} = \mathcal{X}, \quad P_{UEZWSXY}(P_{W|U}, P_{SX}) := P_{UEZ}P_{W|U}P_{SX}P_{Y|X}, \\ I(U; W|Z) \leq R_m := \tau I(X; Y|S) - \mu < \tau I(X; Y|S) \end{array} \right\}. \quad (48)$$

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

$$\begin{aligned} E_1(P_{W|U}) &= \min_{P_{\tilde{U}\tilde{E}\tilde{Z}\tilde{W}} \in \mathcal{T}_1(P_{UW}, P_{EZW})} D(P_{\tilde{U}\tilde{E}\tilde{Z}\tilde{W}} || P_Z P_{U|Z} P_{E|Z} P_{W|U}) \\ &\geq \min_{P_{\tilde{U}\tilde{E}\tilde{Z}\tilde{W}} \in \mathcal{T}_1(P_{UW}, P_{EZW})} D(P_{\tilde{U}\tilde{E}\tilde{Z}\tilde{W}} || P_Z P_{E|Z} P_{W|Z}) \\ &= I(E; W|Z), \end{aligned} \quad (49)$$

where (49) follows from the log-sum inequality [23]. Also,

$$E_2(P_{W|U}, P_{SX}, R_m) \geq R_m - I(U; W|E, Z) \geq I(U; W|Z) - I(U; W|E, Z) = I(E; W|Z),$$

$$\begin{aligned} &P_{\tilde{U}\tilde{E}\tilde{Z}\tilde{W}} \in \mathcal{T}_3(P_{UW}, P_{EZ}) \quad D(P_{\tilde{U}\tilde{E}\tilde{Z}\tilde{W}} || P_Z P_{U|Z} P_{E|Z} P_{W|U}) + R_m - I(U; W|E, Z) + \tau E_x \left(\frac{R_m}{\tau}, P_{SX} \right) \\ &\geq I(U; W|Z) - I(U; W|E, Z) = I(E; W|Z), \end{aligned} \quad (50)$$

$$\begin{aligned}
& P_{\tilde{U}\tilde{E}\tilde{Z}\tilde{W}} \in \min_{\mathcal{T}_3(P_{UW}, P_{EZ})} D(P_{\tilde{U}\tilde{E}\tilde{Z}\tilde{W}} || P_Z P_{U|Z} P_{E|Z} P_{W|U}) + I(E, Z; W) + \tau E_x \left(\frac{R_m}{\tau}, P_{SX} \right) \\
& \geq I(E; W|Z), \tag{51}
\end{aligned}$$

$$D(P_{EZ} || P_{EZ}) + R_m - I(U; W|E, Z) + \tau E_m(P_{SX}) \geq I(U; W|Z) - I(U; W|E, Z) = I(E; W|Z), \tag{52}$$

$$D(P_{EZ} || P_{EZ}) + I(E, Z; W) + \tau E_m(P_{SX}) \geq I(E; W|Z), \tag{53}$$

where in (50)-(53), we used the non-negativity of KL-divergence, $E_x(\cdot, \cdot)$ and $E_m(\cdot)$. Thus, from (50)-(53), it follows that

$$E_3(P_{W|U}, P_{SX}, R_m, \tau) \geq I(E; W|Z), \tag{54}$$

$$\text{and } E_4(P_{W|U}, P_{SX}, R_m, \tau) \geq I(E; W|Z). \tag{55}$$

Denoting $\mathcal{B}(\tau, P_{Y|X})$ and $\mathcal{B}'(\tau, P_{Y|X})$ by \mathcal{B} and \mathcal{B}' , respectively, we obtain

$$\begin{aligned}
& \kappa(\tau, \epsilon) \\
& \geq \sup_{(P_{W|U}, P_{SX}, R_m) \in \mathcal{B}} \min \left(E_1(P_{W|U}), E_2(P_{W|U}, P_{SX}, R_m), E_3(P_{W|U}, P_{SX}, R_m, \tau), E_4(P_{W|U}, P_{SX}, R_m, \tau) \right) \\
& \geq \sup_{(P_{W|U}, P_{SX}, R_m) \in \mathcal{B}} I(E; W|Z) \\
& \geq \sup_{(P_{W|U}, P_{SX}, R_m) \in \mathcal{B}'} I(E; W|Z) \tag{56}
\end{aligned}$$

$$= \sup_{P_{W|U}: I(W; U|Z) \leq \tau C(P_{Y|X}) - \mu} I(E; W|Z), \tag{57}$$

where (56) follows from the fact that $\mathcal{B}' \subseteq \mathcal{B}$; and (57) follows by maximizing over all P_{SX} and noting that $\sup_{P_{XS}} I(X; Y|S) = C(P_{Y|X})$. The proof of achievability is complete by noting that $\mu > 0$ is arbitrary and $I(E; W|Z)$ 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

$$\limsup_{k \rightarrow \infty} \alpha \left(k, n_k, f^{(k, n_k)}, g^{(k, n_k)} \right) = 0, \tag{58}$$

we have similar to [4, Theorem 1 (b)], that

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

$$= \limsup_{n \rightarrow \infty} \frac{1}{k} I(Y^{n_k}; E^k | Z^k) \tag{60}$$

$$= H(E|Z) - \liminf_{k \rightarrow \infty} \frac{1}{k} H(E^k | Y^{n_k}, Z^k), \tag{61}$$

where (60) follows since $Q_{Y^{n_k} E^k Z^k} = P_{Y^{n_k} Z^k} P_{E^k | Z^k}$. 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.

$$\begin{aligned}
H(E^k | Y^{n_k}, Z^k) &= \sum_{i=1}^k H(E_i | E^{i-1}, Y^{n_k}, Z^k) \\
&= \sum_{i=1}^k H(E_i | Z_i, W_i) \\
&= kH(E_T | Z_T, W_T, T) \\
&= kH(E | Z, W).
\end{aligned} \tag{62}$$

Substituting (62) in (61), we obtain

$$\limsup_{k \rightarrow \infty} \frac{-1}{k} \log \left(\beta \left(k, n_k, f_1^{(k, n_k)}, g^{(k, n_k)} \right) \right) \leq I(E; W | Z). \tag{63}$$

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

$$I(U^k; Y^{n_k}) - I(U^k; Z^k) \leq I(X^{n_k}; Y^{n_k}). \tag{64}$$

The R.H.S. of (64) can be upper bounded due to the memoryless nature of the channel as

$$I(X^{n_k}; Y^{n_k}) \leq n_k \max_{P_X} I(X; Y) = n_k C(P_{Y|X}), \tag{65}$$

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

$$I(U^k; Y^{n_k}) - I(U^k; Z^k) = I(U^k; Y^{n_k} | Z^k) \tag{66}$$

$$\begin{aligned}
&= \sum_{i=1}^k I(Y^{n_k}; U_i | U^{i-1}, Z^k) \\
&= \sum_{i=1}^k I(Y^{n_k}, U^{i-1}, Z^{i-1}, Z_{i+1}^k; U_i | Z_i)
\end{aligned} \tag{67}$$

$$= \sum_{i=1}^k I(Y^{n_k}, U^{i-1}, Z^{i-1}, Z_{i+1}^k, E^{i-1}; U_i | Z_i) \tag{68}$$

$$\begin{aligned}
&\geq \sum_{i=1}^k I(Y^{n_k}, Z^{i-1}, Z_{i+1}^k, E^{i-1}; U_i | Z_i) \\
&= \sum_{i=1}^k I(W_i; U_i | Z_i) = kI(W_T; U_T | Z_T, T) \\
&= kI(W_T, T; U_T | Z_T) \\
&= kI(W; U | Z).
\end{aligned} \tag{69}$$

Here, (66) follows due to $Z^k - U^k - Y^{n_k}$; (67) follows since the sequences (U^k, Z^k) are memoryless; (68) follows since $E^{i-1} - (Y^{n_k}, U^{i-1}, Z^{i-1}, Z_{i+1}^k) - U_i$; (69) 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 hence of the proposition. \blacksquare

As the above result shows, TACI is an instance of distributed HT over a DMC, in which, the optimal error-exponent is equal to that achieved over a noiseless channel of the same capacity. Hence, a noisy channel does not always degrade the achievable error-exponent. Also, notice that a separation based coding scheme that performs

independent HT and channel coding is sufficient to achieve the optimal error-exponent for TACI. The investigation of a single-letter characterization of the optimal error-exponent for TACI over a DMC is inspired from an analogous result for TACI over a noiseless channel. It would be interesting to explore whether the noisiness of the channel enables obtaining computable characterizations of the error-exponent for some other special cases of the problem.

V. CONCLUDING REMARKS

In this paper, we have studied the error-exponent achievable for distributed HT problem over a DMC with side information available at the detector. We obtained single-letter lower bounds on the optimal error-exponent for general HT, and exact single-letter characterization for TACI. 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 error-exponent for TACI, and only the channel capacity matters. We also showed via an example that the lower bound on the error-exponent obtained using our joint hypothesis testing and channel coding scheme is strictly better than that obtained using our separation based scheme. Although this does not imply that “separation does not hold” for distributed HT over a DMC, it points to the possibility that joint HT and channel coding schemes outperform separation based schemes, in general, and it is worthwhile investigating this aspect in greater detail. While a strong converse holds for distributed HT over a rate-limited noiseless channel [4], it remains an open question whether this property holds for noisy channels. As a first step, it is shown in [30] that this is indeed the case for HT over a DMC with no side-information. While we did not discuss the complexity of the schemes considered in this paper, it is an important factor that needs to be taken into account in any practical implementation of these schemes. In this regard, it is evident that the SHTCC and JHTCC schemes are in increasing 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 I and type II error probabilities achieved by this scheme, averaged over the ensemble of randomly generated codebooks. By the standard expurgation technique [24] (e.g., removing “worst” codebooks in the ensemble with the highest type I error probability such that the total probability of the removed codebooks lies in the interval $(0.5, 1)$), this guarantees the existence of at least one deterministic codebook that achieves type I and type II error probabilities of the same order, i.e., within a constant multiplicative factor. Since, in our scheme below, the type I error probability averaged over the random code ensemble vanishes asymptotically with the the number of samples k , the same holds for the codebook obtained after expurgation. Moreover, the error-exponent is not affected by a constant multiplicative factor on the type II error probability, and thus, this codebook asymptotically achieves the same type I error probability and error-exponent as the average.

For brevity, in the proof below, we denote the information theoretic quantities like $I_P(U; W)$, $T_{[P_{UW}]_\delta}^k$, etc., that are computed with respect to joint distribution P_{UVWSXY} given in (70) below by $I(U; W)$, $T_{[UW]_\delta}^k$, etc.

Codebook Generation: Let $k \in \mathbb{Z}^+$ and $n = \lfloor \tau k \rfloor$. Fix a finite alphabet \mathcal{W} , a positive number (small) $\delta > 0$, and distributions $P_{W|U}$ and $P_{S|X}$. Let $\delta' := \frac{\delta}{2}$, $\hat{\delta} := |\mathcal{U}|\delta$, $\tilde{\delta} := 2\delta$, $\bar{\delta} := \frac{\delta'}{|\mathcal{V}|}$, $\check{\delta} := |\mathcal{W}|\tilde{\delta}$ and

$$P_{UVWSXY}(P_{W|U}, P_{S|X}) := P_{UV}P_{W|U}P_{S|X}P_{Y|X}. \quad (70)$$

Let $\mu = O(\delta)$ (subject to constraints that will be specified below) and R be such that

$$I(U; W|V) + 2\mu \leq R \leq \tau I(X; Y|S) - \mu. \quad (71)$$

Denoting $M'_k := e^{k(I(U; W) + \mu)}$, the *source codebook* \mathcal{C} used by the source encoder $f_s^{(k)}$ is obtained by generating M'_k sequences $w^k(j)$, $j \in [M'_k]$, independently at random according to the distribution $\prod_{i=1}^k P_W(w_i)$, where

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

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 $\lceil e^{kR} \rceil$ ordinary codewords $x^n(m)$, $m \in [e^{kR}]$, 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 $P_{X|S=s_i}$. The sequence s^n is revealed to the detector.

Encoding: If $I(U; W) + \mu > R$, 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 $w^k(i)$, $i \in [M'_k]$ 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., $u^k \in T_{[U]_{\delta'}}^k$, the source encoder $f_s^{(k)}$ first looks for a sequence $w^k(j)$ in \mathcal{C} such that $(u^k, w^k(j)) \in T_{[UW]_{\delta}}^k$. 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 $I(U; W) + \mu > R$, or otherwise. If $u^k \notin T_{[U]_{\delta'}}^k$ 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 $T_{[P_{Y|S=s_i}]_{\delta}}^n$. 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 $\hat{w}^k = w^k(\hat{j}) \in T_{[W]_{\delta}}^k$, in codebook \mathcal{C} such that

$$\begin{aligned} \hat{j} &= \arg \min_{\substack{l: f_B(l) = \hat{m}, \\ w^k(l) \in T_{[W]_{\delta}}^k}} H_e(w^k(l)|v^k), \text{ if } I(U; W) + \mu > R, \\ \hat{j} &= \hat{m}, \text{ otherwise.} \end{aligned}$$

The decoder declares $\hat{H} = 0$ if $(\hat{w}^k, v^k) \in T_{[WV]_\delta}^k$. Else, $\hat{H} = 1$ is declared.

We next analyze the type I and type II error probabilities achieved by the above scheme.

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

$$\begin{aligned}\mathcal{E}_{TE} &= \left\{ (U^k, V^k) \notin T_{[UV]_\delta}^k \right\} \\ \mathcal{E}_{EE} &= \left\{ \nexists j \in [M'_k] : (U^k, W^k(j)) \in T_{[UW]_\delta}^k \right\} \\ \mathcal{E}_{ME} &= \left\{ (V^k, W^k(J)) \notin T_{[VW]_\delta}^k \right\} \\ \mathcal{E}_{DE} &= \left\{ \exists l \in [M'_k], l \neq J : f_B(l) = f_B(J), W^k(l) \in T_{[W]_\delta}^k, H_e(W^k(l)|V^k) \leq H_e(W^k(J)|V^k) \right\} \\ \mathcal{E}_{CD} &= \left\{ g_c^{(k,n)}(Y^n) \neq M \right\}\end{aligned}$$

$\mathbb{P}(\mathcal{E}_{TE}|H=0)$ tends to 0 asymptotically by the weak law of large numbers. Conditioned on \mathcal{E}_{TE}^c , $U^k \in T_{[U]_\delta}$, and by the covering lemma [23, Lemma 9.1], it is well known that for $\mu = O(\delta)$ chosen appropriately, $\mathbb{P}(\mathcal{E}_{EE}|\mathcal{E}_{TE}^c)$ tends to 0 doubly exponentially with k . Given $\mathcal{E}_{EE}^c \cap \mathcal{E}_{TE}^c$ holds, it follows from the Markov chain relation $V-U-W$ and the Markov lemma [31], that $\mathbb{P}(\mathcal{E}_{ME}|\mathcal{E}_{TE}^c \cap \mathcal{E}_{EE}^c)$ tends to zero as $k \rightarrow \infty$. Next, we consider $\mathbb{P}(\mathcal{E}_{DE})$. Given that $\mathcal{E}_{ME}^c \cap \mathcal{E}_{EE}^c \cap \mathcal{E}_{TE}^c$ holds, note that for k sufficiently large, $H_e(W^k(J)|V^k) \leq H(W|V) + O(\delta)$. Thus, we have (for sufficiently large k)

$$\begin{aligned}\mathbb{P}(\mathcal{E}_{DE} | V^k = v^k, W^k(J) = w^k, \mathcal{E}_{ME}^c \cap \mathcal{E}_{EE}^c \cap \mathcal{E}_{TE}^c) \\ \leq \sum_{\substack{l=1, \\ l \neq J}}^{M'_k} \sum_{\substack{\tilde{w}^k \in T_{[W]_\delta}^k : \\ H_e(\tilde{w}^k|v^k) \\ \leq H_e(w^k|v^k)}} \mathbb{P}\left(f_B(l) = f_B(J), W^k(l) = \tilde{w}^k | V^k = v^k, W^k(J) = w^k, \mathcal{E}_{ME}^c \cap \mathcal{E}_{EE}^c \cap \mathcal{E}_{TE}^c\right) \\ = \sum_{\substack{l=1, \\ l \neq J}}^{M'_k} \sum_{\substack{\tilde{w}^k \in T_{[W]_\delta}^k : \\ H_e(\tilde{w}^k|v^k) \leq H_e(w^k|v^k)}} \mathbb{P}(W^k(l) = \tilde{w}^k | V^k = v^k, W^k(J) = w^k, \mathcal{E}_{ME}^c \cap \mathcal{E}_{EE}^c \cap \mathcal{E}_{TE}^c) \frac{1}{e^{kR}} \\ \leq \sum_{\substack{l=1, \\ l \neq J}}^{M'_k} \sum_{\substack{\tilde{w}^k \in T_{[W]_\delta}^k : \\ H_e(\tilde{w}^k|v^k) \leq H_e(w^k|v^k)}} 2 \cdot e^{-kR} e^{-k(H(W)-O(\delta))}\end{aligned}\tag{72}$$

$$\leq \sum_{\substack{l=1, \\ l \neq J}}^{M'_k} (k+1)^{|\mathcal{V}||\mathcal{W}|} e^{k(H(W|V)+O(\delta))} \cdot 2 \cdot e^{-kR} e^{-k(H(W)-O(\delta))}\tag{73}$$

$$\leq e^{-k(R-I(U;W|V)-\delta_1^{(k)})},\tag{74}$$

where

$$\delta_1^{(k)} = \mu + O(\delta) + \frac{1}{k} |\mathcal{V}||\mathcal{W}| \log(k+1) + \frac{\log(2)}{k}.$$

To obtain (72), we used the fact that

$$\mathbb{P}(W^k(l) = \tilde{w}^k | \mathcal{E}_{ME}^c \cap \mathcal{E}_{EE}^c \cap \mathcal{E}_{TE}^c, W^k(J) = w^k, V^k = v^k) \leq 2 \cdot \mathbb{P}(W^k(l) = \tilde{w}^k). \quad (75)$$

This follows similarly to (96), which is discussed in the type II error analysis section below. In order to obtain the expression in (73), we first summed over the types $P_{\tilde{W}}$ of sequences within the typical set $T_{[W]_\delta}^k$ 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 $(k+1)^{|\mathcal{V}||\mathcal{W}|}$ [23]. Summing over all $(w^k, v^k) \in T_{[VW]_\delta}^k$, we obtain (for sufficiently large k) that

$$\begin{aligned} & \mathbb{P}(\mathcal{E}_{DE} | \mathcal{E}_{ME}^c \cap \mathcal{E}_{EE}^c \cap \mathcal{E}_{TE}^c) \\ & \leq \sum_{(w^k, v^k) \in T_{[VW]_\delta}^k} \mathbb{P}(W^k(J) = w^k, V^k = v^k | \mathcal{E}_{ME}^c \cap \mathcal{E}_{EE}^c \cap \mathcal{E}_{TE}^c) e^{-k(R-I(U;W|V)-\delta_1^{(k)})} \\ & \leq e^{-k(R-I(U;W|V)-\delta_1^{(k)})} \leq e^{-k\frac{\mu}{2}}, \end{aligned} \quad (76)$$

where, (76) follows from (71) by choosing $\mu = O(\delta)$ appropriately.

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 $T_{[P_{Y|S=s_i}]_\delta}^n$ for some $1 \leq i \leq |\mathcal{S}|$, it follows from the weak law of large numbers and the union bound, that

$$\mathbb{P}(\mathcal{E}_{CT} | \mathcal{E}_{EE}^c) \xrightarrow{(k)} 0. \quad (77)$$

Also, it follows from [23, Exercise 10.18, 10.24] that for sufficiently large n (depending on $\mu, \tau, |\mathcal{X}|$ and $|\mathcal{Y}|$),

$$\mathbb{P}(\mathcal{E}_{CD} | \mathcal{E}_{EE}^c \cap \mathcal{E}_{CT}^c) \leq e^{-nE_x(\frac{R}{\tau} + \frac{\mu}{2\tau}, P_{SX})}. \quad (78)$$

This implies that the probability that an error occurs at the channel decoder $g_c^{(k,n)}$ tends to 0 as $n \rightarrow \infty$ since $E_x(\frac{R}{\tau} + \frac{\mu}{2\tau}, P_{SX}) > 0$ for $R \leq \tau I(X; Y|S) - \mu$. Thus, since $I(U; W|V) + \mu \leq R \leq \tau I(X; Y|S) - \mu$, the probability of the events causing type I error tends to zero asymptotically.

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

$$\mathcal{E}_0 = \left\{ U^k \notin T_{[U]_{\delta'}}^k \right\}. \quad (79)$$

Then, the type II error probability can be written as

$$\begin{aligned} & \beta(k, n, f^{(k,n)}, g^{(k,n)}) \\ & = \sum_{(u^k, v^k) \in \mathcal{U}^k \times \mathcal{V}^k} \mathbb{P}(U^k = u^k, V^k = v^k | H = 1) \mathbb{P}(\mathcal{D}_0 | U^k = u^k, V^k = v^k). \end{aligned} \quad (80)$$

Let $\mathcal{E}_{NE} := \mathcal{E}_{EE}^c \cap \mathcal{E}_0^c$. The last term in (80) can be upper bounded as follows.

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

Thus, we have

$$\begin{aligned}
& \beta\left(k, n, f^{(k,n)}, g^{(k,n)}\right) \\
&\leq \sum_{\substack{(u^k, v^k) \\ \in \mathcal{U}^k \times \mathcal{V}^k}} \mathbb{P}(U^k = u^k, V^k = v^k | H = 1) \left[\mathbb{P}(\mathcal{D}_0|U^k = u^k, V^k = v^k, \mathcal{E}_{NE}) \right. \\
&\quad \left. + \mathbb{P}(\mathcal{D}_0|U^k = u^k, V^k = v^k, \mathcal{E}_{NE}^c) \right]. \tag{81}
\end{aligned}$$

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

$$\begin{aligned}
\mathbb{P}(\mathcal{D}_0|U^k = u^k, V^k = v^k, \mathcal{E}_{NE}) &= \sum_{j=1}^{M'_k} \sum_{m=1}^{e^{kR}} \mathbb{P}(J = j, f_B(J) = m | U^k = u^k, V^k = v^k, \mathcal{E}_{NE}) \\
&\quad \mathbb{P}(\mathcal{D}_0|U^k = u^k, V^k = v^k, J = j, f_B(J) = m, \mathcal{E}_{NE}). \tag{82}
\end{aligned}$$

By the symmetry of the codebook generation, encoding and decoding procedure, the term $\mathbb{P}(\mathcal{D}_0|U^k = u^k, V^k = v^k, J = j, f_B(J) = m, \mathcal{E}_{NE})$ in (82) 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

$$\begin{aligned}
& \mathbb{P}(\mathcal{D}_0|U^k = u^k, V^k = v^k, \mathcal{E}_{NE}) \\
&= \sum_{j=1}^{M'_k} \sum_{m=1}^{e^{kR}} \mathbb{P}(J = j, f_B(J) = m | U^k = u^k, V^k = v^k, \mathcal{E}_{NE}) \mathbb{P}(\mathcal{D}_0|U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, \mathcal{E}_{NE}) \\
&= \mathbb{P}(\mathcal{D}_0|U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, \mathcal{E}_{NE}) \\
&= \sum_{w^k \in \mathcal{W}^k} \mathbb{P}(W^k(1) = w^k | U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, \mathcal{E}_{NE}) \\
&\quad \mathbb{P}(\mathcal{D}_0|U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, W^k(1) = w^k, \mathcal{E}_{NE}). \tag{83}
\end{aligned}$$

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,

$$\mathcal{E}_{CE} = \{\mathcal{E}_{CT}^c \text{ and } \hat{M} = g_c^{(k,n)}(Y^n) \neq M\}, \quad (84)$$

$$\text{and } \mathcal{E}_{BE} = \left\{ \exists l \in [M'_k], l \neq J, f_B(l) = \hat{M}, W^k(l) \in T_{[W]_\delta}^k, (V^k, W^k(l)) \in T_{[VW]_\delta}^k \right\}. \quad (85)$$

Define the following events

$$\mathcal{F} = \{U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, W^k(1) = w^k, \mathcal{E}_{NE}\}, \quad (86)$$

$$\mathcal{F}_1 = \{U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, W^k(1) = w^k, \mathcal{E}_{NE}, \mathcal{E}_{CE}\}, \quad (87)$$

$$\mathcal{F}_2 = \{U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, W^k(1) = w^k, \mathcal{E}_{NE}, \mathcal{E}_{CE}^c\}, \quad (88)$$

$$\mathcal{F}_{21} = \{U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, W^k(1) = w^k, \mathcal{E}_{NE}, \mathcal{E}_{CE}, \mathcal{E}_{BE}\}, \quad (89)$$

$$\mathcal{F}_{22} = \{U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, W^k(1) = w^k, \mathcal{E}_{NE}, \mathcal{E}_{CE}^c, \mathcal{E}_{BE}^c\}. \quad (90)$$

The last term in (83) can be expressed as follows.

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

where

$$\mathbb{P}(\mathcal{D}_0|\mathcal{F}_2) = \mathbb{P}(\mathcal{E}_{BE}|\mathcal{F}_2) \mathbb{P}(\mathcal{D}_0|\mathcal{F}_{21}) + \mathbb{P}(\mathcal{E}_{BE}^c|\mathcal{F}_2) \mathbb{P}(\mathcal{D}_0|\mathcal{F}_{22}). \quad (91)$$

It follows from (78) that for sufficiently large k ,

$$\mathbb{P}(\mathcal{E}_{CE}|\mathcal{F}) \leq e^{-nE_x(\frac{R}{\tau} + \frac{\mu}{2\tau}, P_{SX})} = e^{-k\tau E_x(\frac{R}{\tau} + \frac{\mu}{2\tau}, P_{SX})}. \quad (92)$$

Next, consider the type II error event that happens when an error occurs at the channel decoder. We need to consider two separate cases: $I(U; W) + \mu > R$ and $I(U; W) + \mu \leq R$. Note that in the former case, binning is performed and type II 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 $(V^k, W^k(l)) \in T_{[VW]_\delta}^k$. As noted in [28], 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 [28], we analyze the probability of this event (averaged over codebooks \mathcal{C} and random binning) as follows. We first consider the case when $I(U; W) + \mu > R$.

$$\begin{aligned} \mathbb{P}(\mathcal{D}_0|\mathcal{F}_1) &\leq \mathbb{P}(\exists W^k(l) : f_B(l) = \hat{M} \neq 1, (W^k(l), v^k) \in T_{[WV]_\delta}^k | \mathcal{F}_1) \\ &\leq \sum_{l=2}^{M'_k} \sum_{\hat{m} \neq 1} \mathbb{P}(\hat{M} = \hat{m} | \mathcal{F}_1) \mathbb{P}((W^k(l), v^k) \in T_{[WV]_\delta}^k : f_B(l) = \hat{m} | \mathcal{F}_1) \\ &= \sum_{l=2}^{M'_k} \sum_{\hat{m} \neq 1} \mathbb{P}(\hat{M} = \hat{m} | \mathcal{F}_1) \sum_{\substack{\tilde{w}^k : \\ (\tilde{w}^k, v^k) \in T_{[WV]_\delta}^k}} \mathbb{P}(W^k(l) = \tilde{w}^k : f_B(l) = \hat{m} | \mathcal{F}_1) \end{aligned}$$

$$\begin{aligned}
&= \sum_{l=2}^{M'_k} \sum_{\hat{m} \neq 1} \mathbb{P}(\hat{M} = \hat{m} | \mathcal{F}_1) \sum_{\substack{\tilde{w}^k: \\ (\tilde{w}^k, v^k) \in T_{[WV]_\delta}^k}} \mathbb{P}(W^k(l) = \tilde{w}^k | \mathcal{F}_1) \frac{1}{e^{kR}} \\
&= \sum_{l=2}^{M'_k} \sum_{\substack{\tilde{w}^k: \\ (\tilde{w}^k, v^k) \in T_{[WV]_\delta}^k}} \mathbb{P}(W^k(l) = \tilde{w}^k | \mathcal{F}_1) \frac{1}{e^{kR}}. \tag{93}
\end{aligned}$$

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

$$\mathbb{P}(W^k(l) = \tilde{w}^k | \mathcal{F}_1) = \sum_{\mathcal{C}_{1,l}^- = c} \mathbb{P}(\mathcal{C}_{1,l}^- = c | \mathcal{F}_1) \mathbb{P}(W^k(l) = \tilde{w}^k | \mathcal{F}_1, \mathcal{C}_{1,l}^- = c). \tag{94}$$

The term in (94) can be upper bounded as follows:

$$\begin{aligned}
&\mathbb{P}(W^k(l) = \tilde{w}^k | \mathcal{F}_1, \mathcal{C}_{1,l}^- = c) \\
&= \mathbb{P}(W^k(l) = \tilde{w}^k | U^k = u^k, V^k = v^k, \mathcal{C}_{1,l}^- = c) \frac{\mathbb{P}(W^k(1) = w^k | W^k(l) = \tilde{w}^k, U^k = u^k, V^k = v^k, \mathcal{C}_{1,l}^- = c)}{\mathbb{P}(W^k(1) = w^k | U^k = u^k, V^k = v^k, \mathcal{C}_{1,l}^- = c)} \\
&\quad \frac{\mathbb{P}(J = 1 | W^k(1) = w^k, W^k(l) = \tilde{w}^k, U^k = u^k, V^k = v^k, \mathcal{C}_{1,l}^- = c)}{\mathbb{P}(J = 1 | W^k(1) = w^k, U^k = u^k, V^k = v^k, \mathcal{C}_{1,l}^- = c)} \\
&\quad \frac{\mathbb{P}(f_B(J) = 1 | J = 1, W^k(1) = w^k, W^k(l) = \tilde{w}^k, U^k = u^k, V^k = v^k, \mathcal{C}_{1,l}^- = c)}{\mathbb{P}(f_B(J) = 1 | J = 1, W^k(1) = w^k, U^k = u^k, V^k = v^k, \mathcal{C}_{1,l}^- = c)} \\
&\quad \frac{\mathbb{P}(\mathcal{E}_{NE}, \mathcal{E}_{CE} | f_B(J) = 1, J = 1, W^k(1) = w^k, W^k(l) = \tilde{w}^k, U^k = u^k, V^k = v^k, \mathcal{C}_{1,l}^- = c)}{\mathbb{P}(\mathcal{E}_{NE}, \mathcal{E}_{CE} | f_B(J) = 1, J = 1, W^k(1) = w^k, U^k = u^k, V^k = v^k, \mathcal{C}_{1,l}^- = c)}. \tag{95}
\end{aligned}$$

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

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

and

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

Also, note that

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

Next, consider the term in (95). 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

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

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

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

Hence, the term in (94) can be upper bounded as

$$\begin{aligned}
& \mathbb{P}(W^k(l) = \tilde{w}^k | \mathcal{F}_1) \\
&\leq \sum_{\mathcal{C}_{1,l}^- = c} \mathbb{P}(\mathcal{C}_{1,l}^- = c | \mathcal{F}_1) 2 \mathbb{P}(W^k(l) = \tilde{w}^k | U^k = u^k, V^k = v^k, \mathcal{C}_{1,l}^- = c) \\
&= 2 \mathbb{P}(W^k(l) = \tilde{w}^k | U^k = u^k, V^k = v^k) = 2 \mathbb{P}(W^k(l) = \tilde{w}^k). \tag{96}
\end{aligned}$$

Substituting (96) in (93), we obtain

$$\begin{aligned}
\mathbb{P}(\mathcal{D}_0 | \mathcal{F}_1) &\leq \sum_{l=1}^{M'_k} \sum_{\substack{\tilde{w}^k: \\ (\tilde{w}^k, v^k) \in T_{[WV]_\delta}^k}} 2 \mathbb{P}(W^k(l) = \tilde{w}^k) \frac{1}{e^{kR}} \\
&= \sum_{l=1}^{M'_k} \sum_{\substack{\tilde{w}^k: \\ (\tilde{w}^k, v^k) \in T_{[WV]_\delta}^k}} 2 \cdot e^{-k(H(W) - O(\delta))} \frac{1}{e^{kR}} \\
&= 2 M'_k e^{k(H(W|V) + \delta)} e^{-k(H(W) - O(\delta))} \frac{1}{e^{kR}} \\
&\leq e^{-k(R - I(U; W|V) - \delta_2^{(k)})}, \tag{97}
\end{aligned}$$

where $\delta_2^{(k)} := O(\delta) + \frac{\log(2)}{k}$. For the case $I(U; W) + \mu \leq R$ (when binning is not done), the terms can be bounded similarly using (96) as follows.

$$\mathbb{P}(\mathcal{D}_0 | \mathcal{F}_1) = \sum_{\hat{m} \neq 1} \mathbb{P}(\hat{M} = \hat{m} | \mathcal{F}_1) \mathbb{P}((W^k(\hat{m}), v^k) \in T_{[WV]_\delta}^k | \mathcal{F}_1)$$

$$\begin{aligned}
&\leq \sum_{\hat{m} \neq 1} \mathbb{P}(\hat{M} = \hat{m} | \mathcal{F}_1) \sum_{(\tilde{w}^k, v^k) \in T_{[WV]_\delta}^k} 2 \mathbb{P}(W^k(\hat{m}) = \tilde{w}^k) \\
&\leq e^{-k(I(V;W) - \delta_2^{(k)})}.
\end{aligned} \tag{98}$$

Next, consider the event when there are no encoding or channel errors, i.e., $\mathcal{E}_{NE} \cap \mathcal{E}_{CE}^c$. For the case $I(U;W) + \mu > R$, 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 $(W^k(l), V^k) \in T_{[WV]_\delta}^k$. Let $P_{\tilde{U}\tilde{V}\tilde{W}}$ denote the type of $P_{U^k V^k W^k(J)}$. Note that $P_{\tilde{U}\tilde{V}} \in \mathcal{T}_{[UW]_\delta}^k$ 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 is such that $(\hat{W}^k, V^k) \notin T_{[WV]_\delta}^k$ (asymptotically) since $(\hat{W}^k, V^k) \in T_{[WV]_\delta}^k$ necessarily implies that $H_e(\hat{W}^k|V^k) \geq H(W|V) - O(\delta)$ (for δ small enough). Consequently, a type II error can happen under the event \mathcal{E}_{BE} only when $H(\tilde{W}|\tilde{V}) \geq H(W|V) - O(\delta)$. The probability of the event \mathcal{E}_{BE} can be upper bounded under this condition as follows:

$$\begin{aligned}
&\mathbb{P}(\mathcal{E}_{BE} | \mathcal{F}_2) \\
&\leq \mathbb{P}\left(\exists l \neq 1, l \in [M'_k] : f_B(l) = 1 \text{ and } (W^k(l), v^k) \in T_{[WV]_\delta}^k | \mathcal{F}_2\right) \\
&\leq \sum_{l=2}^{M'_k} \mathbb{P}\left((W^k(l), v^k) \in T_{[WV]_\delta}^k | \mathcal{F}_2\right) \mathbb{P}\left(f_B(l) = 1 | \mathcal{F}_2, (W^k(l), v^k) \in T_{[WV]_\delta}^k\right) \\
&= \sum_{l=2}^{M'_k} \mathbb{P}\left((W^k(l), v^k) \in T_{[WV]_\delta}^k | \mathcal{F}_2\right) e^{-kR} \\
&\leq \sum_{l=2}^{M'_k} \sum_{(\tilde{w}^k, v^k) \in T_{[WV]_\delta}^k} 2 \mathbb{P}(W^k(l) = \tilde{w}^k) e^{-kR} \\
&= e^{-k(R - I(U;W|V) - \delta_2^{(k)})}.
\end{aligned} \tag{99}$$

$$= e^{-k(R - I(U;W|V) - \delta_2^{(k)})}. \tag{100}$$

In (99), we used the fact that

$$\mathbb{P}(W^k(l) = \tilde{w}^k | \mathcal{F}_2) \leq 2 \mathbb{P}(W^k(l) = \tilde{w}^k), \tag{101}$$

which follows in a similar way as (96). Also, note that, by definition, $\mathbb{P}(\mathcal{D}_0 | \mathcal{F}_{21}) = 1$.

We proceed to analyze the R.H.S of (81) which upper bounds the type II error probability. Towards this end, we first focus on the the case when \mathcal{E}_{NE} holds. From (83), it follows that

$$\sum_{(u^k, v^k) \in \mathcal{U}^k \times \mathcal{V}^k} \mathbb{P}(U^k = u^k, V^k = v^k | H = 1) \mathbb{P}(\mathcal{D}_0 | U^k = u^k, V^k = v^k, \mathcal{E}_{NE}) \tag{102}$$

$$= \sum_{(u^k, v^k) \in \mathcal{U}^k \times \mathcal{V}^k} \mathbb{P}(U^k = u^k, V^k = v^k | H = 1) \mathbb{P}(\mathcal{D}_0 | U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, \mathcal{E}_{NE}). \tag{103}$$

Rewriting the summation in (103) as the sum over the types and sequences within a type, we obtain

$$\begin{aligned}
& \mathbb{P}(\mathcal{D}_0 | \mathcal{E}_{NE}, H = 1) \\
&= \sum_{\substack{P_{\tilde{U}\tilde{V}\tilde{W}}(u^k, v^k, w^k) \\ \in \mathcal{T}_{\tilde{U}\tilde{V}\tilde{W}}^k}} \sum_{\substack{P_{\tilde{U}\tilde{V}\tilde{W}}(u^k, v^k, w^k) \\ \in \mathcal{T}_{\tilde{U}\tilde{V}\tilde{W}}^k}} \left[\mathbb{P}(U^k = u^k, V^k = v^k | H = 1) \mathbb{P}(\mathcal{D}_0 | \mathcal{F}) \right. \\
&\quad \left. \mathbb{P}(W^k(1) = w^k | U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, \mathcal{E}_{NE}) \right]. \tag{104}
\end{aligned}$$

We also have

$$\begin{aligned}
& \mathbb{P}(U^k = u^k, V^k = v^k | H = 1) \mathbb{P}(W^k(1) = w^k | U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, \mathcal{E}_{NE}) \\
&= \left[\prod_{i=1}^k Q_{UV}(u_i, v_i) \right] \mathbb{P}(W^k(1) = w^k | U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, \mathcal{E}_{NE}) \\
&\leq \left[\prod_{i=1}^k Q_{UV}(u_i, v_i) \right] \frac{1}{|T_{P_{\tilde{W}|\tilde{U}}}^k|} \leq e^{-k(H(\tilde{U}\tilde{V}) + D(P_{\tilde{U}\tilde{V}} || Q_{UV}) + H(\tilde{W}|\tilde{U}) - \frac{1}{k}|\mathcal{U}||\mathcal{W}|\log(k+1))}, \tag{105}
\end{aligned}$$

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

With (92), (97), (98), (100) and (105), we have the necessary machinery to analyze (104). First, consider that the event $\mathcal{E}_{NE} \cap \mathcal{E}_{CE}^c \cap \mathcal{E}_{BE}^c$ holds. In this case,

$$\begin{aligned}
\mathbb{P}(\mathcal{D}_0 | \mathcal{F}_{22}) &= \mathbb{P}(\mathcal{D}_0 | U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, W^k(1) = w^k, \mathcal{E}_{NE}, \mathcal{E}_{CE}^c, \mathcal{E}_{BE}^c) \\
&= \begin{cases} 1, & \text{if } P_{u^k w^k} \in T_{[UW]_\delta}^k \text{ and } P_{v^k w^k} \in T_{[VW]_\delta}^k, \\ 0, & \text{otherwise.} \end{cases} \tag{106}
\end{aligned}$$

Thus, the following terms in (104) can be simplified (for sufficiently large k) as follows:

$$\begin{aligned}
& \sum_{\substack{P_{\tilde{U}\tilde{V}\tilde{W}}(u^k, v^k, w^k) \\ \in \mathcal{T}_{\tilde{U}\tilde{V}\tilde{W}}^k}} \sum_{\substack{P_{\tilde{U}\tilde{V}\tilde{W}}(u^k, v^k, w^k) \\ \in \mathcal{T}_{\tilde{U}\tilde{V}\tilde{W}}^k}} \left[\mathbb{P}(U^k = u^k, V^k = v^k | H = 1) \mathbb{P}(\mathcal{E}_{CE}^c | \mathcal{F}) \mathbb{P}(\mathcal{E}_{BE}^c | \mathcal{F}_2) \mathbb{P}(\mathcal{D}_0 | \mathcal{F}_{22}) \right. \\
&\quad \left. \mathbb{P}(W^k(1) = w^k | U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, \mathcal{E}_{NE}) \right] \\
&\leq \sum_{\substack{P_{\tilde{U}\tilde{V}\tilde{W}}(u^k, v^k, w^k) \\ \in \mathcal{T}_{\tilde{U}\tilde{V}\tilde{W}}^k}} \sum_{\substack{P_{\tilde{U}\tilde{V}\tilde{W}}(u^k, v^k, w^k) \\ \in \mathcal{T}_{\tilde{U}\tilde{V}\tilde{W}}^k}} \left[\mathbb{P}(U^k = u^k, V^k = v^k | H = 1) \mathbb{P}(\mathcal{D}_0 | \mathcal{F}_{22}) \right. \\
&\quad \left. \mathbb{P}(W^k(1) = w^k | U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, \mathcal{E}_{NE}) \right] \\
&\leq (k+1)^{|\mathcal{U}||\mathcal{V}||\mathcal{W}|} \max_{\substack{P_{\tilde{U}\tilde{V}\tilde{W}} \in \\ \hat{\mathcal{T}}_1^{(k)}(P_{UW}, P_{VW})}} e^{kH(\tilde{U}\tilde{V}\tilde{W})} e^{-k(H(\tilde{U}\tilde{V}) + D(P_{\tilde{U}\tilde{V}} || Q_{UV}) + H(\tilde{W}|\tilde{U}) - \frac{1}{k}|\mathcal{U}||\mathcal{W}|\log(k+1))} \\
&= e^{-k\tilde{E}_{1k}}, \tag{107}
\end{aligned}$$

where,

$$\hat{\mathcal{T}}_1^{(k)}(P_{UW}, P_{VW}) := \{P_{\tilde{U}\tilde{V}\tilde{W}} : P_{\tilde{U}\tilde{W}} \in T_{[UW]_\delta}^k \text{ and } P_{\tilde{V}\tilde{W}} \in T_{[VW]_\delta}^k\}, \tag{108}$$

$$\text{and } \tilde{E}_{1k} := \min_{\substack{P_{\tilde{U}\tilde{V}\tilde{W}} \in \\ \hat{\mathcal{T}}_1^{(k)}(P_{UW}, P_{VW})}} H(\tilde{U}\tilde{V}) + D(P_{\tilde{U}\tilde{V}} \| Q_{UV}) + H(\tilde{W}|\tilde{U}) - H(\tilde{U}\tilde{V}\tilde{W}) - \frac{1}{k}|\mathcal{U}||\mathcal{V}||\mathcal{W}|\log(k+1) \\ - \frac{1}{k}|\mathcal{U}||\mathcal{W}|\log(k+1). \quad (109)$$

To obtain (107), we used (105) and (106). Note that for δ small enough,

$$\begin{aligned} \tilde{E}_{1k} &\stackrel{(k)}{\geq} \min_{\substack{P_{\tilde{U}\tilde{V}\tilde{W}} \in \\ \mathcal{T}_1(P_{UW}, P_{VW})}} \sum P_{\tilde{U}\tilde{V}\tilde{W}} \log \left(\frac{P_{\tilde{U}\tilde{V}}}{Q_{UV}} \frac{1}{P_{\tilde{U}\tilde{V}}} \frac{P_{\tilde{U}}}{P_{\tilde{U}\tilde{W}}} P_{\tilde{U}\tilde{V}\tilde{W}} \right) - O(\delta) \\ &= \min_{\substack{P_{\tilde{U}\tilde{V}\tilde{W}} \in \\ \mathcal{T}_1(P_{UW}, P_{VW})}} D(P_{\tilde{U}\tilde{V}\tilde{W}} \| Q_{UVW}) - O(\delta) = E_1(P_{W|U}) - O(\delta), \end{aligned} \quad (110)$$

Next, consider the terms corresponding to the event $\mathcal{E}_{NE} \cap \mathcal{E}_{CE}^c \cap \mathcal{E}_{BE}$ in (104). 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, $P_{u^k v^k} \in T_{[UW]_\delta}^k$. Also, \mathcal{D}_0 can happen only if $H_e(w^k|v^k) \geq H(W|V) - O(\delta)$, and $P_{v^k} \in T_{[V]_\delta}^k$. Using these facts to simplify the terms corresponding to the event $\mathcal{E}_{NE} \cap \mathcal{E}_{CE}^c \cap \mathcal{E}_{BE}$ in (104), we obtain

$$\begin{aligned} &\sum_{\substack{P_{\tilde{U}\tilde{V}\tilde{W}} \in \\ \mathcal{T}_{U^k V^k}^k}} \sum_{\substack{(u^k, v^k, w^k) \\ \in T_{P_{\tilde{U}\tilde{V}\tilde{W}}}^k}} \left[\mathbb{P}(U^k = u^k, V^k = v^k | H = 1) \mathbb{P}(\mathcal{E}_{CE}^c | \mathcal{F}) \mathbb{P}(\mathcal{E}_{BE} | \mathcal{F}_2) \mathbb{P}(\mathcal{D}_0 | \mathcal{F}_{21}) \right. \\ &\quad \left. \mathbb{P}(W^k(1) = w^k | U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, \mathcal{E}_{NE}) \right] \\ &\leq \sum_{\substack{P_{\tilde{U}\tilde{V}\tilde{W}} \in \\ \mathcal{T}_{U^k V^k}^k}} \sum_{\substack{(u^k, v^k, w^k) \\ \in T_{P_{\tilde{U}\tilde{V}\tilde{W}}}^k}} \left[\mathbb{P}(U^k = u^k, V^k = v^k | H = 1) \mathbb{P}(\mathcal{E}_{BE} | \mathcal{F}_2) \mathbb{P}(\mathcal{D}_0 | \mathcal{F}_{21}) \right. \\ &\quad \left. \mathbb{P}(W^k(1) = w^k | U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, \mathcal{E}_{NE}) \right] \\ &\leq \max_{\substack{P_{\tilde{U}\tilde{V}\tilde{W}} \in \\ \hat{\mathcal{T}}_2^{(k)}(P_{UW}, P_V)}} e^{kH(\tilde{U}\tilde{V}\tilde{W})} e^{-k(H(\tilde{U}\tilde{V}) + D(P_{\tilde{U}\tilde{V}} \| Q_{UV}) + H(\tilde{W}|\tilde{U}) + R - I(U; W|V) - O(\delta))} \\ &\quad e^{(|\mathcal{U}||\mathcal{V}||\mathcal{W}|\log(k+1) + |\mathcal{U}||\mathcal{W}|\log(k+1))} \\ &= e^{-k\tilde{E}_{2k}}, \end{aligned} \quad (111)$$

where,

$$\hat{\mathcal{T}}_2^{(k)}(P_{UW}, P_V) := \{P_{\tilde{U}\tilde{V}\tilde{W}} : P_{\tilde{U}\tilde{W}} \in T_{[UW]_\delta}^k, P_{\tilde{V}} \in T_{[V]_\delta}^k \text{ and } H(\tilde{W}|\tilde{V}) \geq H(W|V) - O(\delta)\}, \quad (112)$$

and

$$\begin{aligned} \tilde{E}_{2k} &:= \min_{\substack{P_{\tilde{U}\tilde{V}\tilde{W}} \in \\ \mathcal{T}_2(P_{UW}, P_V)}} H(\tilde{U}\tilde{V}) + D(P_{\tilde{U}\tilde{V}} \| Q_{UV}) + H(\tilde{W}|\tilde{U}) + R - I(U; W|V) - \frac{1}{k}|\mathcal{U}||\mathcal{V}||\mathcal{W}|\log(k+1) \\ &\quad - \frac{1}{k}|\mathcal{U}||\mathcal{W}|\log(k+1) - O(\delta) \\ &\stackrel{(k)}{\geq} E_2(P_{W|U}, P_{SX}, R) - O(\delta). \end{aligned} \quad (113)$$

Also, note that \mathcal{E}_{BE} occurs only when $I(U;W) + \mu > R$.

Next, consider that the event $\mathcal{E}_{NE} \cap \mathcal{E}_{CE}$ holds. As in the case above, note that given $\mathcal{F}_1 = \{U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, W^k(1) = w^k, \mathcal{E}_{NE}, \mathcal{E}_{CE}\}$, $P_{u^k w^k} \in T_{[UW]_\delta}^k$ and \mathcal{D}_0 occurs only if $P_{v^k} \in T_{[V]_\delta}^k$. Using these facts and eqns. (97), (98) and (92), it can be shown that the terms corresponding to this event in (104) results in the factor $E_3(P_{W|U}, P_{SX}, R, \tau) - O(\delta)$ in the error-exponent.

Finally, we analyze the case when the event \mathcal{E}_{NE}^c 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

$$\begin{aligned} \mathbb{P}(\mathcal{D}_0 | U^k = u^k, V^k = v^k, \mathcal{E}_{NE}^c) &= \mathbb{P}(\mathcal{E}_{CE} | U^k = u^k, V^k = v^k, \mathcal{E}_{NE}^c) \\ &= \mathbb{P}(\mathcal{D}_0 | U^k = u^k, V^k = v^k, \mathcal{E}_{NE}^c \cap \mathcal{E}_{CE}). \end{aligned} \quad (114)$$

It follows from Borade et al.'s coding scheme [25] that asymptotically,

$$\mathbb{P}(\mathcal{E}_{CE} | U^k = u^k, V^k = v^k, \mathcal{E}_{NE}^c) \leq e^{-n(E_m(P_{SX}) - O(\delta))} = e^{-k\tau(E_m(P_{SX}) - O(\delta))}. \quad (115)$$

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 $(\hat{W}^k, V^k) \in T_{[WV]_\delta}^k$. 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 $v^k \in T_{[V]_\delta}^k$,

$$\mathbb{P}(\mathcal{D}_0 | U^k = u^k, V^k = v^k, \mathcal{E}_{NE}^c \cap \mathcal{E}_{CE}) \leq e^{-k(R - I(U;W|V) - O(\delta))}. \quad (116)$$

Thus, from (114), (115) and (116), we obtain (asymptotically) that,

$$\begin{aligned} &\sum_{u^k, v^k} \mathbb{P}(U^k = u^k, V^k = v^k | H = 1) \mathbb{P}(\mathcal{D}_0 | U^k = u^k, V^k = v^k, \mathcal{E}_{NE}^c \cap \mathcal{E}_{CE}) \\ &\leq e^{-k(R - I(U;W|V) + D(P_V || Q_V) + \tau E_m(P_{SX}) - O(\delta))}. \end{aligned} \quad (117)$$

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

$$\begin{aligned} &\sum_{u^k, v^k} \mathbb{P}(U^k = u^k, V^k = v^k | H = 1) \mathbb{P}(\mathcal{D}_0 | U^k = u^k, V^k = v^k, \mathcal{E}_{NE}^c \cap \mathcal{E}_{CE}) \\ &\leq e^{-k(I(V;W) + D(P_V || Q_V) + \tau E_m(P_{SX}) - O(\delta))}. \end{aligned} \quad (118)$$

This results in the factor $E_4(P_{W|U}, P_{SX}, R, \tau) - O(\delta)$ in the error-exponent. Since the error-exponent is lower bounded by the minimal value of the exponent due to the various type II error events, the proof of the theorem is complete by noting that $\delta > 0$ is arbitrary.

APPENDIX B PROOF OF THEOREM 5

We only give a sketch of the proof as the intermediate steps follow similarly to those in the proof of Theorem 2. We will use the random coding method combined with the expurgation technique as explained in the proof

of Theorem 2, to guarantee the existence of at least one deterministic codebook that achieves the type I error probability and error-exponent claimed in Theorem 5. For brevity, we will denote information theoretic quantities like $I_{\hat{P}}(U, S; \bar{W})$, $T_{[\hat{P}_{US\bar{W}}]_\delta}^n$, etc., that are computed with respect to joint distribution $\hat{P}_{UVS\bar{W}X'XY}$ given below in (119) by $I(U, S; \bar{W})$, $T_{[US\bar{W}]_\delta}^n$, etc.

Fix distributions $(P_S, P_{\bar{W}|US}, P_{X'|US}, P_{X|US\bar{W}}) \in \mathcal{B}_h$ and a positive number $\delta > 0$. Let $\mu = O(\delta)$ subject to constraints that will be specified below. Let $\hat{\delta} := |\bar{W}|\delta$, $\delta' := \frac{\delta}{2}$, $\bar{\delta} := \frac{\delta'}{|\bar{W}|}$, $\tilde{\delta} := 2\delta$, and

$$\hat{P}_{UVS\bar{W}X'XY}(P_S, P_{\bar{W}|US}, P_{X'|US}, P_{X|US\bar{W}}) := P_{UV}P_S P_{\bar{W}|US} P_{X'|US} P_{X|US\bar{W}} P_{Y|X}. \quad (119)$$

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 $\mathcal{C} = \{\bar{w}^n(j), j \in [e^{n(I(U,S;\bar{W})+\mu)}]\}$, 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., $(u^n, s^n) \in T_{[US]_{\delta'}}^n$, the encoder first looks for a sequence $\bar{w}^n(j)$ such that $(u^n, s^n, \bar{w}^n(j)) \in T_{[US\bar{W}]_\delta}^n$. 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 $\prod_{i=1}^n P_{X|US\bar{W}}(x_i|u_i, s_i, \bar{w}_i(j))$. If $(u^n, s^n) \notin T_{[US]_{\delta'}}^n$ or such an index j does not exist, the encoder generates the channel input $X'^n = x'^n$ randomly according to $\prod_{i=1}^n P_{X'|US}(x'_i|u_i, s_i)$.

Decoding: Given the side information sequence $V^n = v^n$, received sequence $Y^n = y^n$ and s^n , the detector first checks if $(v^n, s^n, y^n) \in T_{[VSY]_\delta}^n$, $\tilde{\delta} > \delta$. If the check is unsuccessful, $\hat{H} = 1$. Else, it searches for a typical sequence $\hat{w}^n = \bar{w}^n(\hat{j}) \in T_{[\bar{W}]_\delta}^n$, in the codebook such that

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

If $(v^n, s^n, y^n, \hat{w}^n) \in T_{[VSY\bar{W}]_\delta}^n$, $\hat{H} = 0$. Else, $\hat{H} = 1$.

Analysis of Type I error:

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

$$\begin{aligned} \tilde{\mathcal{E}}_{TE} &= \left\{ (U^n, V^n, S^n) \notin T_{[UVS]_\delta}^n \right\} \\ \tilde{\mathcal{E}}_{EE} &= \left\{ \nexists j \in [e^{n(I(U,S;\bar{W})+\mu)}] : (U^n, S^n, \bar{W}^n(j)) \in T_{[US\bar{W}]_\delta}^n \right\} \\ \tilde{\mathcal{E}}_{ME} &= \left\{ (V^n, S^n, \bar{W}^n(J)) \notin T_{[VSW]_\delta}^n \right\} \\ \tilde{\mathcal{E}}_{CE} &= \left\{ (V^n, S^n, \bar{W}^n(J), Y^n) \notin T_{[VSWY]_\delta}^n \right\} \\ \tilde{\mathcal{E}}_{DE} &= \left\{ \exists l \in [e^{n(I(U,S;\bar{W})+\mu)}], l \neq J, \bar{W}^n(l) \in T_{[\bar{W}]_\delta}^n, H_e(\bar{W}^n(l)|V^n, S^n, Y^n) \leq H_e(\bar{W}^n(J)|V^n, S^n, Y^n) \right\} \end{aligned}$$

By the weak law of large numbers, $\tilde{\mathcal{E}}_{TE}$ tends to 0 asymptotically with n . The covering lemma guarantees that $\tilde{\mathcal{E}}_{EE} \cap \tilde{\mathcal{E}}_{TE}^c$ tends to 0 doubly exponentially if $\mu = O(\delta)$ is chosen appropriately. Given $\tilde{\mathcal{E}}_{EE}^c \cap \tilde{\mathcal{E}}_{TE}^c$ 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. Next, we consider the probability of the event $\tilde{\mathcal{E}}_{DE}$. Given that $\tilde{\mathcal{E}}_{CE}^c \cap \tilde{\mathcal{E}}_{ME}^c \cap \tilde{\mathcal{E}}_{EE}^c \cap \tilde{\mathcal{E}}_{TE}^c$ holds, note that $H_e(\bar{W}^n(J)|V^n, S^n, Y^n) \stackrel{(n)}{\geq} H(\bar{W}|V, S, Y) - O(\delta)$. Hence, similarly to (74) in Appendix A, it can be shown that

$$\mathbb{P}(\tilde{\mathcal{E}}_{DE}|\tilde{\mathcal{E}}_{CE}^c \cap \tilde{\mathcal{E}}_{ME}^c \cap \tilde{\mathcal{E}}_{EE}^c \cap \tilde{\mathcal{E}}_{TE}^c) \leq e^{-n(I_{\bar{P}}(\bar{W};V,S,Y) - I_{\bar{P}}(U,S;\bar{W}) - \delta_3^{(n)})}.$$

where $\delta_3^{(n)} \xrightarrow{(n)} O(\delta)$. Hence, for $\delta > 0$ small enough, the probability of the events causing type I error tends to zero asymptotically since $I(U; \bar{W}|S) < I(\bar{W}; Y, V|S)$.

Analysis of Type II error: The analysis of the error-exponent 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

$$\bar{\mathcal{E}}_0 := \{(U^n, S^n) \notin T_{[US]_\delta}^n\}. \quad (120)$$

Then, the type 2 error probability can be written as

$$\begin{aligned} & \beta(n, n, f^{(n,n)}, g^{(n,n)}) \\ & \leq \sum_{(u^n, v^n) \in \mathcal{U}^n \times \mathcal{V}^n} \mathbb{P}(U^n = u^n, V^n = v^n | H = 1) \left[\mathbb{P}(\tilde{\mathcal{E}}_{EE} \cap \bar{\mathcal{E}}_0^c | U^n = u^n, V^n = v^n) \right. \\ & \quad \left. + \mathbb{P}(\mathcal{D}_0 | U^n = u^n, V^n = v^n, \tilde{\mathcal{E}}_{NE}) + \mathbb{P}(\mathcal{D}_0 | U^n = u^n, V^n = v^n, \bar{\mathcal{E}}_0) \right], \end{aligned} \quad (121)$$

where, $\tilde{\mathcal{E}}_{NE} := \tilde{\mathcal{E}}_{EE}^c \cap \bar{\mathcal{E}}_0^c$. 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

$$\tilde{\mathcal{E}}_{T2} = \left\{ \exists l \in \left[e^{n(I(U,S;\bar{W})+\mu)} \right], l \neq J, \bar{W}^n(l) \in T_{[\bar{W}]_\delta}^n, (V^n, \bar{W}^n(l), S^n, Y^n) \in T_{[VS\bar{W}Y]_\delta}^n \right\}, \quad (122)$$

$$\tilde{\mathcal{F}} = \{U^n = u^n, V^n = v^n, J = 1, \bar{W}^n(1) = \bar{w}^n, S^n = s^n, Y^n = y^n, \tilde{\mathcal{E}}_{NE}\}, \quad (123)$$

$$\tilde{\mathcal{F}}_1 = \{U^n = u^n, V^n = v^n, J = 1, \bar{W}^n(1) = \bar{w}^n, S^n = s^n, Y^n = y^n, \tilde{\mathcal{E}}_{NE}, \tilde{\mathcal{E}}_{T2}^c\}, \quad (124)$$

$$\tilde{\mathcal{F}}_2 = \{U^n = u^n, V^n = v^n, J = 1, \bar{W}^n(1) = \bar{w}^n, S^n = s^n, Y^n = y^n, \tilde{\mathcal{E}}_{NE}, \tilde{\mathcal{E}}_{T2}\}. \quad (125)$$

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

$$\begin{aligned} & \mathbb{P}(\mathcal{D}_0 | U^n = u^n, V^n = v^n, \tilde{\mathcal{E}}_{NE}) \\ & = \sum_{j=1}^{e^{n(I(U,S;\bar{W})+\mu)}} \mathbb{P}(J = j | U^n = u^n, V^n = v^n, \tilde{\mathcal{E}}_{NE}) \mathbb{P}(\mathcal{D}_0 | U^n = u^n, V^n = v^n, J = 1, \tilde{\mathcal{E}}_{NE}) \\ & = \mathbb{P}(\mathcal{D}_0 | U^n = u^n, V^n = v^n, J = 1, \tilde{\mathcal{E}}_{NE}) \end{aligned}$$

$$\begin{aligned}
&= \sum_{\substack{(\bar{w}^n, s^n, y^n) \\ \in \bar{\mathcal{W}}^n \times \mathcal{S}^n \times \mathcal{Y}^n}} \mathbb{P}(\bar{W}^n(1) = \bar{w}^n, S^n = s^n, Y^n = y^n | U^n = u^n, V^n = v^n, J = 1, \tilde{\mathcal{E}}_{NE}) \\
&\quad \mathbb{P}(\mathcal{D}_0 | U^n = u^n, V^n = v^n, J = 1, \bar{W}^n(1) = \bar{w}^n, S^n = s^n, Y^n = y^n, \tilde{\mathcal{E}}_{NE}) \\
&= \sum_{\substack{(\bar{w}^n, s^n, y^n) \\ \in \bar{\mathcal{W}}^n \times \mathcal{S}^n \times \mathcal{Y}^n}} \mathbb{P}(\bar{W}^n(1) = \bar{w}^n, S^n = s^n, Y^n = y^n | U^n = u^n, V^n = v^n, J = 1, \tilde{\mathcal{E}}_{NE}) \mathbb{P}(\mathcal{D}_0 | \tilde{\mathcal{F}}). \tag{126}
\end{aligned}$$

The last term in (126) can be upper bounded using the events in (123)-(125) as follows.

$$\mathbb{P}(\mathcal{D}_0 | \tilde{\mathcal{F}}) \leq \mathbb{P}(\mathcal{D}_0 | \tilde{\mathcal{F}}_1) + \mathbb{P}(\tilde{\mathcal{E}}_{T_2} | \tilde{\mathcal{F}}) \mathbb{P}(\mathcal{D}_0 | \tilde{\mathcal{F}}_2).$$

We next analyze the R.H.S of (121), which upper bounds the type 2 error probability. We can write,

$$\mathbb{P}(\mathcal{D}_0 | \tilde{\mathcal{F}}_1) = \begin{cases} 1, & \text{if } P_{u^n s^n \bar{w}^n} \in T_{[US\bar{W}]_\delta}^n \text{ and } P_{v^n \bar{w}^n s^n y^n} \in T_{[VSWY]_\delta}^k, \\ 0, & \text{otherwise.} \end{cases} \tag{127}$$

Hence, the terms corresponding to the event $\tilde{\mathcal{F}}_1$ in (121) can be upper bounded (in the limit $\delta, \tilde{\delta} \rightarrow 0$) as

$$\begin{aligned}
&\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}} \left[\mathbb{P}(U^n = u^n, V^n = v^n | H = 1) \mathbb{P}(\mathcal{D}_0 | \tilde{\mathcal{F}}_1) \right. \\
&\quad \left. \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}) \right] \\
&\leq \sum_{\substack{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} \\ \in \mathcal{T}_{\mathcal{U}\mathcal{V}\mathcal{W}\mathcal{S}\mathcal{Y}}^n}} \sum_{\substack{(u^n, v^n, \bar{w}^n, s^n, y^n) \\ \in \mathcal{T}_{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}}}}} \left[\mathbb{P}(U^n = u^n, V^n = v^n | H = 1) \mathbb{P}(\mathcal{D}_0 | \tilde{\mathcal{F}}_1) \right. \\
&\quad \mathbb{P}(S^n = s^n, \bar{W}^n(1) = \bar{w}^n | U^n = u^n, J = 1, \tilde{\mathcal{E}}_{NE}) \\
&\quad \left. \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}) \right] \\
&\leq \sum_{\substack{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} \\ \in \mathcal{T}_{\mathcal{U}\mathcal{V}\mathcal{W}\mathcal{S}\mathcal{Y}}^n}} \sum_{\substack{(u^n, v^n, \bar{w}^n, s^n, y^n) \\ \in \mathcal{T}_{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}}}}} \left[\mathbb{P}(\mathcal{D}_0 | \tilde{\mathcal{F}}_1) e^{-n(H(\tilde{U}\tilde{V}) + D(P_{\tilde{U}\tilde{V}} || Q_{UV}))} \right. \\
&\quad \left. e^{-n(H(\tilde{S}\tilde{W}|\tilde{U}) - \frac{1}{n}|\mathcal{U}||\tilde{\mathcal{W}}||\mathcal{S}|\log(n+1))} e^{-n(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] \\
&\leq \max_{\substack{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} \\ \in \mathcal{T}_1^{(n)}(\hat{P}_{US\bar{W}}, \hat{P}_{VSWY})}} \left[e^{-n(H(\tilde{U}\tilde{V}) + D(P_{\tilde{U}\tilde{V}} || Q_{UV}))} e^{-n(H(\tilde{S}\tilde{W}|\tilde{U}) - \frac{1}{n}|\mathcal{U}||\tilde{\mathcal{W}}||\mathcal{S}|\log(n+1))} \right. \\
&\quad \left. e^{-n(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}}))} e^{n(H(\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}) - \frac{1}{n}|\mathcal{U}||\mathcal{V}||\tilde{\mathcal{W}}||\mathcal{S}||\mathcal{Y}|\log(n+1))} \right] \\
&= e^{-nE_{1n}^*}, \tag{128}
\end{aligned}$$

where

$$\mathcal{T}_1^{(n)}(\hat{P}_{US\bar{W}}, \hat{P}_{VSWY}) := \{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} \in \mathcal{T}_{\mathcal{U}\mathcal{V}\mathcal{W}\mathcal{S}\mathcal{Y}} : P_{\tilde{U}\tilde{S}\tilde{W}} \in T_{[US\bar{W}]_\delta}^n, P_{\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} \in T_{[VSWY]_\delta}^n\},$$

and

$$\begin{aligned}
E_{1n}^* &:= \min_{\substack{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} \in \\ \mathcal{T}'_1(\hat{P}_{US\tilde{W}}, \hat{P}_{VSWY})}} \left[H(\tilde{U}\tilde{V}) + D(P_{\tilde{U}\tilde{V}} \| Q_{UV}) + H(\tilde{S}\tilde{W}|\tilde{U}) + H(\tilde{Y}|\tilde{U}\tilde{S}\tilde{W}) - H(\tilde{U}\tilde{V}\tilde{W}\tilde{S}\tilde{Y}) \right. \\
&\quad \left. + D(P_{\tilde{Y}|\tilde{U}\tilde{S}\tilde{W}} \| \hat{P}_{Y|US\tilde{W}} | P_{\tilde{U}\tilde{S}\tilde{W}}) - \frac{1}{n} (|\mathcal{U}||\mathcal{W}| + |\mathcal{U}||\mathcal{V}||\mathcal{W}||\mathcal{S}||\mathcal{Y}|) \log(n+1) \right] \\
&\stackrel{(n)}{\geq} \min_{\substack{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} \in \\ \mathcal{T}'_1(\hat{P}_{US\tilde{W}}, \hat{P}_{VSWY})}} \left[\sum_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} \log \left(\frac{1}{P_{\tilde{U}\tilde{V}}} \frac{P_{\tilde{U}}}{Q_{UV}} \frac{1}{P_{\tilde{U}\tilde{S}\tilde{W}}} \frac{P_{\tilde{Y}|\tilde{U}\tilde{S}\tilde{W}}}{\hat{P}_{Y|US\tilde{W}}} P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} \right) - O(\delta) \right] \\
&= \min_{\substack{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} \in \\ \mathcal{T}'_1(\hat{P}_{US\tilde{W}}, \hat{P}_{VSWY})}} \left[D(P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} | Q_{UV} P_{\tilde{S}\tilde{W}|\tilde{U}} \hat{P}_{Y|US\tilde{W}}) - O(\delta) \right] \\
&= E'_1(P_S, P_{\tilde{W}|US}, P_{X|US\tilde{W}}) - O(\delta). \tag{129}
\end{aligned}$$

Here, (129) follows from the fact that $P_{\tilde{S}\tilde{W}|\tilde{U}} \rightarrow P_{S\tilde{W}|U}$ given $\tilde{\mathcal{E}}_{NE}$, as $\delta \rightarrow 0$.

Next, consider the terms corresponding to the event $\tilde{\mathcal{F}}_2$ in (121). Given $\tilde{\mathcal{F}}_2$, $P_{\tilde{U}\tilde{S}\tilde{W}} \in T^n_{[US\tilde{W}]_\delta}$ and \mathcal{D}_0 occurs only if $(V^n, S^n, Y^n) \in T^n_{[VSY]_{\delta''}}$, $\delta'' = |\tilde{\mathcal{W}}|\delta$, and $H(\tilde{W}|\tilde{V}, \tilde{S}, \tilde{Y}) \geq H(\tilde{W}|V, S, Y) - O(\delta)$. Thus, we have,

$$\begin{aligned}
&\sum_{\substack{(u^n, v^n, \bar{w}^n, s^n, y^n) \\ \in \mathcal{U}^n \times \mathcal{V}^n \times \mathcal{W}^n \times \mathcal{S}^n \times \mathcal{Y}^n}} \left[\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}}_{T_2} | \tilde{\mathcal{F}}) \right. \\
&\quad \left. \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}) \right] \\
&\leq \sum_{\substack{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} \in \\ \mathcal{T}^n(\mathcal{U} \times \mathcal{V} \times \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}}}^n}} \left[\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}}_{T_2} | \tilde{\mathcal{F}}) \right. \\
&\quad \left. \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}) \right] \\
&\leq \sum_{\substack{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} \in \\ \mathcal{T}^n(\mathcal{U} \times \mathcal{V} \times \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}}}^n}} \left[e^{-n(H(\tilde{U}\tilde{V}) + D(P_{\tilde{U}\tilde{V}} \| Q_{UV}))} \mathbb{P}(\mathcal{D}_0 | \tilde{\mathcal{F}}_2) \cdot 2 \cdot e^{-n(I(\tilde{W}; V, S, Y) - I(U, S; \tilde{W}) - O(\delta))} \right. \\
&\quad \left. e^{-n(H(\tilde{S}\tilde{W}|\tilde{U}) - \frac{1}{n}|\mathcal{U}||\mathcal{W}||\mathcal{S}| \log(n+1))} e^{-n(H(\tilde{Y}|\tilde{U}\tilde{S}\tilde{W}) + D(P_{\tilde{Y}|\tilde{U}\tilde{S}\tilde{W}} \| \hat{P}_{Y|US\tilde{W}} | P_{\tilde{U}\tilde{S}\tilde{W}}))} \right] \tag{130} \\
&\leq \max_{\substack{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} \in \\ \mathcal{T}_2^{(n)}(\hat{P}_{UW}, \hat{P}_{VSWY})}} \left[e^{-n(H(\tilde{U}\tilde{V}) + D(P_{\tilde{U}\tilde{V}} \| Q_{UV}))} e^{-n(H(\tilde{S}\tilde{W}|\tilde{U}) - \frac{1}{n}|\mathcal{U}||\mathcal{W}||\mathcal{S}| \log(n+1))} \right. \\
&\quad e^{-n(I(\tilde{W}; V, S, Y) - I(U, S; \tilde{W}) - O(\delta) - \frac{1}{n})} e^{-n(H(\tilde{Y}|\tilde{U}\tilde{S}\tilde{W}) + D(P_{\tilde{Y}|\tilde{U}\tilde{S}\tilde{W}} \| \hat{P}_{Y|US\tilde{W}} | P_{\tilde{U}\tilde{S}\tilde{W}}))} \\
&\quad \left. e^{n(H(\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}) - \frac{1}{n}(|\mathcal{U}||\mathcal{V}||\mathcal{W}||\mathcal{S}||\mathcal{Y}| \log(n+1)))} \right] \\
&= e^{-nE_{2n}^*}, \tag{131}
\end{aligned}$$

where,

$$\begin{aligned}
\mathcal{T}_2^{(n)}(\hat{P}_{US\tilde{W}}, \hat{P}_{VSWY}) &:= \{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} \in \mathcal{T}_{UVSWY} : P_{\tilde{U}\tilde{S}\tilde{W}} \in T^n_{[US\tilde{W}]_\delta}, P_{\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} \in T^n_{[VSWY]_\delta} \\
&\quad \text{and } H(\tilde{W}|\tilde{V}, \tilde{S}, \tilde{Y}) \geq H(\tilde{W}|V, S, Y) - O(\delta)\},
\end{aligned}$$

and

$$\begin{aligned}
E_{2n}^* &\stackrel{(n)}{\geq} \min_{\substack{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} \in \\ \mathcal{T}'_2(P_{US\bar{W}}, P_{VSWY})}} \left[D(P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} | Q_{UV} P_{\tilde{S}\tilde{W}|\tilde{U}} \hat{P}_{Y|US\bar{W}}) + I(\bar{W}; V, Y | S) - I(U; \bar{W} | S) - O(\delta) \right] \\
&= E'_2(P_S, P_{\bar{W}|US}, P_{X|US\bar{W}}) - O(\delta).
\end{aligned} \tag{132}$$

In (130), we used the fact that

$$\mathbb{P}(\tilde{\mathcal{E}}_{T_2} | \tilde{\mathcal{F}}) \leq 2 \cdot e^{-n(I(\bar{W}; V, Y | S) - I(U; \bar{W} | S) - O(\delta))},$$

which follows from

$$\mathbb{P}(\bar{W}^n(l) = \tilde{w}^n | \tilde{\mathcal{F}}) \leq 2 \mathbb{P}(\bar{W}^n(l) = \tilde{w}^n). \tag{133}$$

Eqn. (133) can be proved similarly to (96).

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

$$\begin{aligned}
&\sum_{u^n, v^n} \mathbb{P}(U^n = u^n, V^n = v^n | H = 1) \mathbb{P}(\mathcal{D}_0 | U^n = u^n, V^n = v^n, \bar{\mathcal{E}}_0) \\
&= \sum_{u^n, v^n} \mathbb{P}(U^n = u^n, V^n = v^n | H = 1) \sum_{s^n, y^n} \mathbb{P}(S^n = s^n, Y^n = y^n, \mathcal{D}_0 | U^n = u^n, V^n = v^n, \bar{\mathcal{E}}_0) \\
&= \sum_{u^n, v^n} \mathbb{P}(U^n = u^n, V^n = v^n | H = 1) \left[\sum_{s^n, y^n} \mathbb{P}(S^n = s^n, Y^n = y^n | U^n = u^n, V^n = v^n, \bar{\mathcal{E}}_0) \right. \\
&\quad \left. \mathbb{P}(\mathcal{D}_0 | U^n = u^n, V^n = v^n, S^n = s^n, Y^n = y^n, \bar{\mathcal{E}}_0) \right] \\
&= \sum_{u^n, v^n} \mathbb{P}(U^n = u^n, V^n = v^n | H = 1) \left[\sum_{s^n, y^n} \mathbb{P}(S^n = s^n, Y^n = y^n | U^n = u^n, \bar{\mathcal{E}}_0) \right. \\
&\quad \left. \mathbb{P}(\mathcal{D}_0 | U^n = u^n, V^n = v^n, S^n = s^n, Y^n = y^n, \bar{\mathcal{E}}_0) \right]
\end{aligned} \tag{134}$$

The event \mathcal{D}_0 occurs only if there exists a sequence $(\bar{W}^n(l), V^n, S^n, Y^n) \in T_{[\bar{W}VSY]_{\delta}}^n$ for some $l \in [e^{n(I(U, S; \bar{W}) + \mu)}]$. 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

$$\mathbb{P}(\mathcal{D}_0 | U^n = u^n, V^n = v^n, S^n = s^n, Y^n = y^n, \bar{\mathcal{E}}_0) \leq e^{-n(I(\bar{W}; V, Y | S) - I(U; \bar{W} | S) - O(\delta))}. \tag{135}$$

Also,

$$\mathbb{P}(S^n = s^n, Y^n = y^n | U^n = u^n, \bar{\mathcal{E}}_0) \leq e^{-n(H(\tilde{S}\tilde{Y}|\tilde{U}) + D(P_{\tilde{S}\tilde{Y}|\tilde{U}} || \tilde{Q}_{SY|U} | P_{\tilde{U}}))}. \tag{136}$$

Hence, using (135) and (136) in (134), we obtain

$$\sum_{u^n, v^n} \mathbb{P}(U^n = u^n, V^n = v^n | H = 1) \mathbb{P}(\mathcal{D}_0 | U^n = u^n, V^n = v^n, \bar{\mathcal{E}}_0)$$

$$\begin{aligned}
&\leq (n+1)^{|\mathcal{U}||\mathcal{V}||\mathcal{S}||\mathcal{Y}|} \max_{\substack{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{Y}}: \\ P_{\tilde{V}\tilde{S}\tilde{Y}} = \hat{P}_{VSY}}} e^{nH(\tilde{U}\tilde{V}\tilde{S}\tilde{Y})} e^{-n(H(\tilde{U}\tilde{V}) + D(P_{\tilde{U}\tilde{V}} \| Q_{UV}))} e^{-n(H(\tilde{S}\tilde{Y}|\tilde{U}) + D(P_{\tilde{S}\tilde{Y}|\tilde{U}} \| \check{Q}_{SY|U}|P_{\tilde{U}}))} \\
&\quad e^{-n(I(\bar{W}; V, Y|S) - I(U; \bar{W}|S) - O(\delta))} \\
&= e^{-nE_{3n}^*},
\end{aligned}$$

where,

$$\begin{aligned}
E_{3n}^* &= \min_{P_{\tilde{V}\tilde{S}\tilde{Y}} = \hat{P}_{VSY}} D(P_{\tilde{V}\tilde{S}\tilde{Y}} \| \check{Q}_{VSY}) + I(\bar{W}; V, Y|S) - I(U; \bar{W}|S) - |\mathcal{U}||\mathcal{V}||\mathcal{S}||\mathcal{Y}| \log(n+1) - O(\delta) \\
&\stackrel{(n)}{\rightarrow} E'_3(P_S, P_{\bar{W}|US}, P_{X'|US}, P_{X|US\bar{W}}) - O(\delta).
\end{aligned}$$

Since the error-exponent 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

OPTIMAL SINGLE-LETTER CHARACTERIZATION OF ERROR-EXPONENT WHEN $C(P_{Y|X}) = 0$

The proof of achievability follows from the one-bit scheme mentioned in Remark 4 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 [23]) that $C(P_{Y|X}) = 0$ only if

$$P_Y^* := P_{Y|X=x} = P_{Y|X=x'}, \quad \forall x, x' \in \mathcal{X}. \quad (137)$$

From (137), it follows that $E_c(P_{Y|X}) = 0$. Also,

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

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 (46). 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 (58), it follows similarly to (59), that

$$\limsup_{k \rightarrow \infty} \frac{-1}{k} \log \left(\beta \left(k, n_k, f^{(k, n_k)}, g^{(k, n_k)} \right) \right) \leq \limsup_{k \rightarrow \infty} \frac{1}{k} D(P_{Y^{n_k} V^k} \| Q_{Y^{n_k} V^k}). \quad (138)$$

The terms in the R.H.S. of (138) can be expanded as

$$\begin{aligned}
&\frac{1}{k} D(P_{Y^{n_k} V^k} \| Q_{Y^{n_k} V^k}) \\
&= D(P_V \| Q_V) + \frac{1}{k} \sum_{\substack{(v^k, y^{n_k}) \\ \in \mathcal{V}^k \times \mathcal{Y}^{n_k}}} P_{V^k Y^{n_k}}(v^k, y^{n_k}) \log \left(\frac{P_{Y^{n_k} | V^k}(y^{n_k} | v^k)}{Q_{Y^{n_k} | V^k}(y^{n_k} | v^k)} \right). \quad (139)
\end{aligned}$$

Next, note that

$$\begin{aligned}
P_{Y^{n_k}|V^k}(y^{n_k}|v^k) &= \sum_{\substack{(u^k, x^{n_k}) \\ \in \mathcal{U}^k \times \mathcal{X}^{n_k}}} P_{U^k|V^k}(u^k|v^k) P_{X^{n_k}|U^k}(x^{n_k}|u^k) P_{Y^{n_k}|X^{n_k}}(y^{n_k}|x^{n_k}) \\
&= \left(\prod_{i=1}^{n_k} P_Y^*(y_i) \right) \sum_{\substack{(u^k, x^{n_k}) \\ \in \mathcal{U}^k \times \mathcal{X}^{n_k}}} P_{U^k|V^k}(u^k|v^k) P_{X^{n_k}|U^k}(x^{n_k}|u^k) \tag{140}
\end{aligned}$$

$$= \prod_{i=1}^{n_k} P_Y^*(y_i), \tag{141}$$

where, (140) follows from (3) and (137). Similarly, it follows that

$$Q_{Y^{n_k}|V^k}(y^{n_k}|v^k) = \prod_{i=1}^{n_k} P_Y^*(y_i). \tag{142}$$

From (138), (139), (141) and (142), we obtain that

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

This completes the proof of the weak converse.

Next, we proceed to show that $D(P_V || Q_V)$ is the optimal error-exponent 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

$$\limsup_{k \rightarrow \infty} \alpha \left(k, n_k, f^{(k, n_k)}, g^{(k, n_k)} \right) \leq \epsilon. \tag{143}$$

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

$$P_{V^k Y^{n_k}}(v^k, y^{n_k}) = \left(\prod_{i=1}^k P_V(v_i) \right) \left(\prod_{j=1}^{n_k} P_Y^*(y_j) \right), \tag{144}$$

$$\text{and } Q_{V^k Y^{n_k}}(v^k, y^{n_k}) = \left(\prod_{i=1}^k Q_V(v_i) \right) \left(\prod_{j=1}^{n_k} P_Y^*(y_j) \right), \tag{145}$$

respectively. By the weak law of large numbers, for any $\delta > 0$, (144) implies that

$$\lim_{k \rightarrow \infty} P_{V^k Y^{n_k}} \left(T_{[P_V]_\delta}^k \times T_{[P_Y^*]_\delta}^{n_k} \right) = 1. \tag{146}$$

Also, from (143), we have

$$\liminf_{k \rightarrow \infty} P_{V^k Y^{n_k}} \left(\mathcal{A}_{(k, n_k)} \right) \geq (1 - \epsilon). \tag{147}$$

From (146) and (147), it follows that

$$P_{V^k Y^{n_k}} \left(\mathcal{A}_{(k, n_k)} \cap T_{[P_V]_\delta}^k \times T_{[P_Y^*]_\delta}^{n_k} \right) \geq 1 - \epsilon', \tag{148}$$

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

$$\mathcal{A}(v^k, \delta) := \left\{ y^{n_k} : (v^k, y^{n_k}) \in \mathcal{A}_{(k, n_k)} \cap T_{[P_V]_\delta}^k \times T_{[P_Y^*]_\delta}^{n_k} \right\}, \quad (149)$$

$$\text{and } \mathcal{D}(\eta, \delta) := \left\{ v^k \in T_{[P_V]_\delta}^k : P_{Y^{n_k}}(\mathcal{A}(v^k, \delta)) \geq \eta \right\}. \quad (150)$$

Fix $0 < \eta' < 1 - \epsilon'$. Then, we have from (148) that for any $\delta > 0$ and sufficiently large k ,

$$P_{V^k}(\mathcal{D}(\eta', \delta)) \geq \frac{1 - \epsilon' - \eta'}{1 - \eta'}. \quad (151)$$

From [23, Lemma 2.14], (151) implies that $\mathcal{D}(\eta', \delta)$ should contain at least $\frac{1 - \epsilon' - \eta'}{1 - \eta'}$ fraction (approx.) of sequences in $T_{[P_V]_\delta}^k$ and for each $v^k \in \mathcal{D}(\eta', \delta)$, (150) implies that $\mathcal{A}(v^k, \delta)$ should contain at least η' fraction (approx.) of sequences in $T_{[P_Y^*]_\delta}^{n_k}$, asymptotically. Hence, for sufficiently large k , we have

$$Q_{V^k Y^{n_k}}(\mathcal{A}_{(k, n_k)}) \geq \sum_{v^k \in \mathcal{D}(\eta', \delta)} Q_{V^k}(v^k) \sum_{y^{n_k} \in \mathcal{A}(v^k, \delta)} P_{Y^{n_k}}(y^{n_k}) \quad (152)$$

$$\geq e^{-k \left(D(P_V \| Q_V) - \frac{\log\left(\frac{1 - \epsilon' - \eta'}{1 - \eta'}\right)}{k} - \frac{\log(\eta')}{k} - O(\delta) \right)}. \quad (153)$$

Here, (153) follows from [23, Lemma 2.6].

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

$$P_{V^k Y^{n_k}}(\mathcal{A}'_{(k, n_k)}) \xrightarrow{(k)} 1, \quad (154)$$

$$\text{and } Q_{V^k Y^{n_k}}(\mathcal{A}'_{(k, n_k)}) \leq e^{-k(D(P_V \| Q_V) - O(\delta))}, \quad (155)$$

where, (154) and (155) follows from weak law of large numbers and [23, Lemma 2.6], respectively. Together (153), (154) and (155) implies that

$$|\kappa(\tau, \epsilon) - \kappa(\tau)| \leq O(\delta),$$

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

REFERENCES

- [1] 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.
- [2] W. Hoeffding, "Asymptotically optimal tests for multinomial distributions," *Ann. Math. Stat.*, vol. 36, no. 2, pp. 369–400, 1965.
- [3] T. Berger, "Decentralized estimation and decision theory," in *IEEE 7th. Spring Workshop on Inf. Theory*, Mt. Kisco, NY, Sep. 1979.
- [4] R. Ahlswede and I. Csiszár, "Hypothesis testing with communication constraints," *IEEE Trans. Inf. Theory*, vol. 32, no. 4, pp. 533–542, Jul. 1986.
- [5] T. S. Han, "Hypothesis testing with multiterminal data compression," *IEEE Trans. Inf. Theory*, vol. 33, no. 6, pp. 759–772, Nov. 1987.
- [6] 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.
- [7] 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.

- [8] 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.
- [9] M. Wigger and R. Timo, "Testing against independence with multiple decision centers," in *Int. Conf. on Signal Processing and Communication*, Bengaluru, India, Jun. 2016.
- [10] 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.
- [11] Y. Xiang and Y. H. Kim, "Interactive hypothesis testing against independence," in *Proc. IEEE Int. Symp. Inf. Theory*, Istanbul, Turkey, Nov. 2013.
- [12] —, "Interactive hypothesis testing with communication constraints," in *50th Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, USA, Oct. 2012.
- [13] G. Katz, P. Piantanida, and M. Debbah, "Collaborative distributed hypothesis testing," *arXiv:1604.01292 [cs.IT]*, Apr. 2016.
- [14] —, "Distributed binary detection with lossy data compression," *IEEE Trans. Inf. Theory*, vol. 63, no. 8, pp. 5207–5227, Mar. 2017.
- [15] S. Sreekumar and D. Gündüz, "Distributed hypothesis testing over noisy channels," in *Proc. IEEE Int. Symp. Inf. Theory*, Aachen, Germany, Jun. 2017.
- [16] S. Salehkalaibar, M. Wigger, and L. Wang, "Hypothesis testing over the two-hop relay network," *IEEE Trans. Inf. Theory*, vol. 65, no. 7, pp. 4411–4433, Jul. 2019.
- [17] R. E. Blahut, "Hypothesis testing and information theory," *IEEE Trans. Inf. Theory*, vol. 20, no. 4, pp. 405–417, Jul. 1974.
- [18] 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.
- [19] 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.
- [20] 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.
- [21] S. Watanabe, "Neyman-pearson test for zero-rate multiterminal hypothesis testing," *IEEE Trans. Inf. Theory*, vol. 64, no. 7, Jul. 2018.
- [22] N. Weinberger and Y. Kochman, "On the reliability function of distributed hypothesis testing under optimal detection," *IEEE Trans. Inf. Theory*, vol. 65, no. 8, pp. 4940–4965, Aug. 2019.
- [23] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge University Press, 2011.
- [24] R. G. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Trans. Inf. Theory*, vol. 11, pp. 3–18, Jan. 1965.
- [25] S. Borade, B. Nakiboğlu, and L. Zheng, "Unequal error protection: An information-theoretic perspective," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5511–5539, Dec. 2009.
- [26] I. Csiszár, "Joint source-channel error exponent," *Prob. of Control and Inf. Theory*, vol. 9, no. 5, pp. 315–328, 1980.
- [27] T. Cover, A. E. Gamal, and M. Salehi, "Multiple access channels with arbitrarily correlated sources," *IEEE Trans. Inf. Theory*, vol. 26, no. 6, pp. 648–657, Nov. 1980.
- [28] 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.
- [29] 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.
- [30] S. Sreekumar and D. Gündüz, "Hypothesis testing over a noisy channel," in *Proc. IEEE Int. Symp. Inf. Theory*, Paris, France, Jul. 2019.
- [31] A. E. Gamal and Y.-H. Kim, *Network Information theory*. Cambridge University Press, 2011.