Article

Distributed Hypothesis Testing over a Noisy Channel: Error-Exponents Trade-Off

Sreejith Sreekumar [1,*] and **Deniz Gündüz** [2]

1. Department of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14850, USA
2. Department of Electrical and Electronic Engineering, Imperial College London, London SW72AZ, UK
* Correspondence: sreejithsreekumar@cornell.edu

Abstract: A two-terminal distributed binary hypothesis testing problem over a noisy channel is studied. The two terminals, called the observer and the decision maker, each has access to n independent and identically distributed samples, denoted by $\mathbf{U}$ and $\mathbf{V}$, respectively. The observer communicates to the decision maker over a discrete memoryless channel, and the decision maker performs a binary hypothesis test on the joint probability distribution of $(\mathbf{U}, \mathbf{V})$ based on $\mathbf{V}$ and the noisy information received from the observer. The trade-off between the exponents of the type I and type II error probabilities is investigated. Two inner bounds are obtained, one using a separation-based scheme that involves type-based compression and unequal error-protection channel coding, and the other using a joint scheme that incorporates type-based hybrid coding. The separation-based scheme is shown to recover the inner bound obtained by Han and Kobayashi for the special case of a rate-limited noiseless channel, and also the one obtained by the authors previously for a corner point of the trade-off. Finally, we show via an example that the joint scheme achieves a strictly tighter bound than the separation-based scheme for some points of the error-exponents trade-off.

Keywords: distributed hypothesis testing; noisy channel; error-exponents; source-channel separation; joint source-channel coding; hybrid coding

1. Introduction

Hypothesis testing (HT), which refers to the problem of choosing between one or more alternatives based on available data, plays a central role in statistics and information theory. Distributed HT (DHT) problems arise in situations where the test data are scattered across multiple terminals, and need to be communicated to a central terminal, called the *decision maker*, which performs the hypothesis test. The need to jointly optimize the communication scheme and the hypothesis test makes DHT problems much more challenging than their centralized counterparts. Indeed, while an efficient characterization of the optimal hypothesis test and its asymptotic performance is well known in the centralized setting, thanks to [1–5], the same problem in even the simplest distributed setting remains open, except for some special cases (see [6–11]).

In this work, we consider a DHT problem with two parties, an *observer* and a decision maker, such that the former communicates to the latter over a noisy channel. The observer and the decision maker each has access to independent and identically distributed samples, denoted by $\mathbf{U}$ and $\mathbf{V}$, respectively. Based on the information received from the observer and its own observations $\mathbf{V}$, the decision maker performs a binary hypothesis test on the joint distribution of $(\mathbf{U}, \mathbf{V})$. Our goal is to characterize the trade-off between the best achievable rate of decay (or exponent) of the type I and type II error probabilities with respect to the sample size. We will refer to this problem as *DHT over a noisy channel*, and its special instance with the noisy channel replaced by a rate-limited noiseless channel as *DHT over a noiseless channel*.

1.1. Background

Distributed statistical inference problems were first conceived in [12] and the information-theoretic study of DHT over a noiseless channel was first investigated in [6], where the objective is to characterize Stein's exponent $\kappa_{\text{se}}(\epsilon)$, i.e., the optimal *type II error-exponent* subject to the type I error probability constrained to be at most $\epsilon \in (0, 1)$. The authors therein established a multi-letter characterization of this quantity including a strong converse, which shows that $\kappa_{\text{se}}(\epsilon)$ is independent of ϵ. Furthermore, a single-letter characterization of $\kappa_{\text{se}}(\epsilon)$ is obtained for a special case of HT known as *testing against independence* (TAI), in which the joint distribution factors as a product of the marginal distributions under the alternative hypothesis. Improved lower bounds on $\kappa_{\text{se}}(\epsilon)$ were subsequently obtained in [7,8], respectively, and the strong converse was extended to zero-rate settings [13]. While all the aforementioned works focus on $\kappa_{\text{se}}(\epsilon)$, the trade-off between the exponents of both the type I and type II error probabilities in the same setting was first explored in [14].

In the recent years, there has been a renewed interest in distributed statistical inference problems motivated by emerging machine learning applications to be served at the wireless edge, particularly in the context of semantic communications in 5G/6G communication systems [15,16]. Several extensions of the DHT over a noiseless channel problem have been studied, such as generalizations to multi-terminal settings [9,17–21], DHT under security or privacy constraints [22–25], DHT with lossy compression [26], interactive settings [27,28], successive refinement models [29], and more. Improved bounds have been obtained on the type I and type II error-exponents region [30,31], and on $\kappa_{\text{se}}(\epsilon)$ for testing correlation between bivariate standard normal distributions [32]. In the simpler zero-rate communication setting, there has been some progress in terms of second-order optimal schemes [33], geometric interpretation of type I and type II error-exponent region [34], and characterization of $\kappa_{\text{se}}(\epsilon)$ for sequential HT [35]. DHT over noisy communication channels with the goal of characterizing $\kappa_{\text{se}}(\epsilon)$ has been considered in [10,11,36,37].

1.2. Contributions

In this work, our objective is to explore the trade-off between the type I and type II error-exponents for DHT over a noisy channel. This problem is a generalization of [14] from noiseless rate-limited channels to noisy channels, and also of [10,11] from a type I error probability constraint to a positive type I error-exponent constraint.

Our main contributions can be summarized as follows:

(i)   We obtain an inner bound (Theorem 1) on the error-exponents trade-off by using a *separate HT and channel coding* scheme (SHTCC) that is a combination of a type-based (type here refers to the empirical probability distribution of a sequence, see [38]) quantize-bin strategy and unequal error-protection scheme of [39]. This result is shown to recover the bounds established in [10,14]. Furthermore, we evaluate Theorem 1 for two important instances of DHT, namely TAI and its opposite, i.e., *testing against dependence* (TAD) in which the joint distribution under the null hypothesis factors as a product of marginal distributions.

(ii)   We also obtain a second inner bound (Theorem 2) on the error-exponents trade-off by using a *joint HT and channel coding scheme* (JHTCC) based on *hybrid coding* [40]. Subsequently, we show via an example that the JHTCC scheme strictly outperforms the SHTCC scheme for some points on the error-exponent trade-off.

While the above schemes are inspired from those in [10], which have been proposed with the goal of maximizing the type II error-exponent, novel modifications in its design and analysis are required when considering both of the error-exponents. More specifically, the schemes presented here perform separate quantization-binning or hybrid coding on each individual source sequence type at the observer/encoder (as opposed to a typical ball in [10]) with the corresponding reverse operation implemented at the decision-maker/decoder. This necessitates a different analysis to compute the probabilities of the various error events contributing to the overall error-exponents. We finally mention that the DHT problem considered here was recently investigated in [41], where an inner bound on the error-exponents

trade-off (Theorem 2 in [41]) is obtained using a combination of a type-based quantization scheme and unequal error protection scheme of [42] with two special messages. A qualitative comparison between Theorem 2 and Theorem 2 in [41] seems to suggest that the JHTCC scheme here uses a stronger decoding rule depending jointly on the source-channel statistics. In comparison, the metric used at the decoder for the scheme in [41] factors as the sum of two metrics, one which depends only on the source statistics, and the other which depends only on the channel statistics. Importantly, this hints that the inner bound achieved by JHTCC scheme is not subsumed by that in [41]. That said, a direct computational comparison appears difficult, as evaluating the latter requires optimization over several parameters as mentioned in the last paragraph of [41].

The remainder of the paper is organized as follows. Section 2 formulates the operational problem along with the required definitions. The main results are presented in Section 3. The proofs are furnished in Section 4. Finally, concluding remarks are given in Section 5.

2. Preliminaries

2.1. Notation

We use the following notation. All logarithms are with respect to the natural base e. $\mathbb{N}$, $\mathbb{R}$, $\mathbb{R}_{\geq 0}$, and $\bar{\mathbb{R}}$ denotes the set of natural, real, non-negative real and extended real numbers, respectively. For $a, b \in \mathbb{R}_{\geq 0}$, $[a : b] := \{n \in \mathbb{N} : a \leq n \leq b\}$ and $[b] := [1 : b]$. Calligraphic letters, e.g., $\mathcal{X}$, denote sets, while $\mathcal{X}^c$ and $|\mathcal{X}|$ stands for its complement and cardinality, respectively. For $n \in \mathbb{N}$, $\mathcal{X}^n$ denotes the n-fold Cartesian product of $\mathcal{X}$, and $x^n = (x_1, \cdots, x_n)$ denotes an element of $\mathcal{X}^n$. Bold-face letters denote vectors or sequences, e.g., $\mathbf{x}$ for x^n; its length n will be clear from the context. For $i, j \in \mathbb{N}$ such that $i \leq j$, $x_i^j := (x_i, x_{i+1}, \cdots, x_j)$, the subscript is omitted when $i = 1$. $\mathbb{1}_{\mathcal{A}}$ denotes the indicator of set $\mathcal{A}$. For a real sequence $\{a_n\}_{n \in \mathbb{N}}$, $a_n \xrightarrow{(n)} b$ stands for $\lim_{n \to \infty} a_n = b$, while $a_n \gtrsim b$ denotes $\lim_{n \to \infty} a_n \geq b$. Similar notations apply for other inequalities. $O(\cdot)$, $\Omega(\cdot)$ and $o(\cdot)$ denote standard asymptotic notations.

Random variables and their realizations are denoted by uppercase and lowercase letters, respectively, e.g., X and x. Similar conventions apply for random vectors and their realizations. The set of all probability mass functions (PMFs) on a finite set $\mathcal{X}$ is denoted by $\mathcal{P}(\mathcal{X})$. The joint PMF of two discrete random variables X and Y is denoted by P_{XY}; the corresponding marginals are P_X and P_Y. The conditional PMF of X given Y is represented by $P_{X|Y}$. Expressions such as $P_{XY} = P_X P_{Y|X}$ are to be understood as pointwise equality, i.e., $P_{XY}(x, y) = P_X(x) P_{Y|X}(y|x)$, for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$. When the joint distribution of a triple (X, Y, Z) factors as $P_{XYZ} = P_{XY} P_{Z|X}$, these variables form a Markov chain $X - Y - Z$. When X and Y are statistically independent, we write $X \perp\!\!\!\perp Y$. If the entries of X^n are drawn in an independent and identically distributed manner, i.e., if $P_{X^n}(\mathbf{x}) = \prod_{i=1}^n P_X(x_i)$, $\forall \mathbf{x} \in \mathcal{X}^n$, then the PMF P_{X^n} is denoted by $P_X^{\otimes n}$. Similarly, if $P_{Y^n|X^n}(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^n P_{Y|X}(y_i|x_i)$ for all $(\mathbf{x}, \mathbf{y}) \in \mathcal{X}^n \times \mathcal{Y}^n$, then we write $P_{Y|X}^{\otimes n}$ for $P_{Y^n|X^n}$. The conditional product PMF given a fixed $\mathbf{x} \in \mathcal{X}^n$ is designated by $P_{Y|X}^{\otimes n}(\cdot|\mathbf{x})$. The probability measure induced by a PMF P is denoted by $\mathbb{P}_P$. The corresponding expectation is designated by $\mathbb{E}_P$.

The *type* or empirical PMF of a sequence $\mathbf{x} \in \mathcal{X}^n$ is designated by $P_{\mathbf{x}}$, i.e., $P_{\mathbf{x}}(x) := \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{x_i = x\}}$. The set of n-length sequences $\mathbf{x} \in \mathcal{X}^n$ of type P_X is $\mathcal{T}_n(P_X, \mathcal{X}^n) := \{\mathbf{x} \in \mathcal{X}^n : P_{\mathbf{x}} = P_X\}$. Whenever the underlying alphabet $\mathcal{X}^n$ is clear from the context, $\mathcal{T}_n(P_X, \mathcal{X}^n)$ is simplified to $\mathcal{T}_n(P_X)$. The set of all possible types of n-length sequences $\mathbf{x} \in \mathcal{X}^n$ is $\mathcal{T}(\mathcal{X}^n) := \{P_X \in \mathcal{P}(\mathcal{X}) : |\mathcal{T}_n(P_X, \mathcal{X}^n)| \geq 1\}$. Similar notations are used for larger combinations, e.g., $P_{\mathbf{xy}}$, $\mathcal{T}_n(P_{XY}, \mathcal{X} \times \mathcal{Y})$ and $\mathcal{T}(\mathcal{X}^n \times \mathcal{Y}^n)$. For a given $\mathbf{x} \in \mathcal{T}_n(P_X, \mathcal{X}^n)$ and a conditional PMF $P_{Y|X}$, $\mathcal{T}_n(P_{Y|X}, \mathbf{x}) := \{\mathbf{y} \in \mathcal{Y}^n : (\mathbf{x}, \mathbf{y}) \in \mathcal{T}_n(P_{XY}, \mathcal{X}^n \times \mathcal{Y}^n)\}$ stands for the $P_{Y|X}$-conditional type class of $\mathbf{x}$.

For PMFs $P, Q \in \mathcal{P}(\mathcal{X})$, the Kullback–Leibler (KL) divergence between P and Q is $D(P||Q) := \sum_{x \in \mathcal{X}} P(x) \log (P(x)/Q(x))$. The conditional KL divergence between $P_{Y|X}$ and $Q_{Y|X}$ given P_X is $D(P_{Y|X}||Q_{Y|X}|P_X) := \sum_{x \in \mathcal{X}} P_X(x) D(P_{Y|X}(\cdot|x)||Q_{Y|X}(\cdot|x))$. The mutual

information and entropy terms are denoted by $I_P(\cdot)$ and $H_P(\cdot)$, respectively, where P denotes the PMF of the relevant random variables. When the PMF is clear from the context, the subscript is omitted. For $(\mathbf{x}, \mathbf{y}) \in \mathcal{X}^n \times \mathcal{Y}^n$, the empirical conditional entropy of $\mathbf{y}$ given $\mathbf{x}$ is $H_e(\mathbf{y}|\mathbf{x}) := H_P(\tilde{Y}|\tilde{X})$, where $P_{\tilde{X}\tilde{Y}} = P_{\mathbf{xy}}$. For a given function $f : \mathcal{Z} \to \mathbb{R}$ and a random variable $Z \sim P_Z$, the log-moment generating function of Z with respect to f is $\psi_{P_Z,f}(\lambda) := \log \mathbb{E}_{P_Z}[e^{\lambda f(Z)}]$ whenever the expectation exists. Finally, let

$$\psi^*_{P_Z,f}(\theta) := \sup_{\lambda \in \mathbb{R}} \theta\lambda - \psi_{P_Z,f}(\lambda), \tag{1}$$

denote the rate function (see, e.g., Definition 15.5 in [43]).

2.2. Problem Formulation

Let $\mathcal{U}$, $\mathcal{V}$, $\mathcal{X}$ and $\mathcal{Y}$ be finite sets, and $n \in \mathbb{N}$. The DHT over a noisy channel setting is depicted in Figure 1. Herein, the observer and the decision maker observe n independent and identically distributed samples, denoted by $\mathbf{u}$ and $\mathbf{v}$, respectively. Based on its observations $\mathbf{u}$, the observer outputs a sequence $\mathbf{x} \in \mathcal{X}^n$ as the channel input sequence (note that the ratio of the number of channel uses to the number of data samples, termed the bandwidth ratio, is taken to be 1 for simplicity; however, our results easily generalize to arbitrary bandwidth ratios). The discrete memoryless channel (DMC) with transition kernel $P_{Y|X}$ produces a sequence $\mathbf{y} \in \mathcal{Y}^n$ according to the probability law $P^{\otimes n}_{Y|X}(\cdot|\mathbf{x})$ as its output. We will assume that $P_{Y|X}(\cdot|x) \ll P_{Y|X}(\cdot|x')$, $\forall (x, x') \in \mathcal{X}^2$, where $P \ll Q$ indicates the absolute continuity of P with respect to Q. Based on its observations, $\mathbf{y}$ and $\mathbf{v}$, the decision maker performs binary HT on the joint probability distribution of $(\mathbf{U}, \mathbf{V})$ with the null (H_0) and alternative (H_1) hypotheses given by

$$H_0 : (\mathbf{U}, \mathbf{V}) \sim P^{\otimes n}_{UV},$$
$$H_1 : (\mathbf{U}, \mathbf{V}) \sim Q^{\otimes n}_{UV}.$$

The decision maker outputs $\hat{h} \in \hat{\mathcal{H}} := \{0, 1\}$ as the decision of the hypothesis test, where 0 and 1 denote H_0 and H_1, respectively.
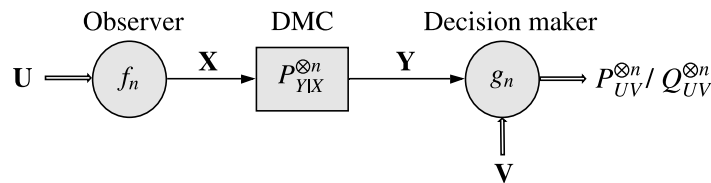

Figure 1. DHT over a noisy channel. The observer observes an n-length independent and identically distributed sequence $\mathbf{U}$, and transmits $\mathbf{X}$ over the DMC $P^{\otimes n}_{Y|X}$. Based on the channel output $\mathbf{Y}$ and the n-length independent and identically distributed sequence $\mathbf{V}$, the decision maker performs a binary HT to determine whether $(\mathbf{U}, \mathbf{V}) \sim P^{\otimes n}_{UV}$ or $(\mathbf{U}, \mathbf{V}) \sim Q^{\otimes n}_{UV}$.

A length-n DHT code c_n is a pair of functions (f_n, g_n), where

(i) $f_n : \mathcal{U}^n \to \mathcal{P}(\mathcal{X}^n)$ denotes the encoding function;

(ii) $g_n : \mathcal{V}^n \times \mathcal{Y}^n \to \hat{\mathcal{H}}$ denotes a deterministic decision function specified by an acceptance region (for null hypothesis H_0) $\mathcal{A}_n \subseteq \mathcal{V}^n \times \mathcal{Y}^n$ as $g_n(\mathbf{v}, \mathbf{y}) = 1 - \mathbb{1}_{\{(\mathbf{v},\mathbf{y}) \in \mathcal{A}_n\}}$, $\forall (\mathbf{v}, \mathbf{y}) \in \mathcal{V}^n \times \mathcal{Y}^n$.

We emphasize at this point that there is no loss in generality in restricting our attention to a deterministic decision function for the objective of characterizing the error-exponents trade-off in HT (for e.g., see Lemma 3 in [24]).

A code $c_n = (f_n, g_n)$ induces the joint PMFs $P_{\mathbf{UVXY}\hat{H}}^{(c_n)}$ and $Q_{\mathbf{UVXY}\hat{H}}^{(c_n)}$ under the null and alternative hypotheses, respectively, where

$$P_{\mathbf{UVXY}\hat{H}}^{(c_n)}(\mathbf{u}, \mathbf{v}, \mathbf{x}, \mathbf{y}, \hat{h}) := P_{UV}^{\otimes n}(\mathbf{u}, \mathbf{v}) \, f_n(\mathbf{x}|\mathbf{u}) \, P_{Y|X}^{\otimes n}(\mathbf{y}|\mathbf{x}) \, \mathbb{1}_{\left\{g_n(\mathbf{v}, \mathbf{y}) = \hat{h}\right\}}, \tag{3}$$

and

$$Q_{\mathbf{UVXY}\hat{H}}^{(c_n)}(\mathbf{u}, \mathbf{v}, \mathbf{x}, \mathbf{y}, \hat{h}) := Q_{UV}^{\otimes n}(\mathbf{u}, \mathbf{v}) \, f_n(\mathbf{x}|\mathbf{u}) \, P_{Y|X}^{\otimes n}(\mathbf{y}|\mathbf{x}) \, \mathbb{1}_{\left\{g_n(\mathbf{v}, \mathbf{y}) = \hat{h}\right\}}, \tag{4}$$

respectively. For a given code c_n, the type I and type II error probabilities are $\alpha_n(c_n) := \mathbb{P}_{P^{(c_n)}}(\hat{H} = 1)$ and $\beta_n(c_n) := \mathbb{P}_{Q^{(c_n)}}(\hat{H} = 0)$ respectively. The following definition formally states the error-exponents trade-off we aim to characterize.

Definition 1 (**Error-exponent region**)**.** *An error-exponent pair $(\kappa_\alpha, \kappa_\beta) \in \mathbb{R}_{\geq 0}^2$ is said to be achievable if there exists a sequence of codes $\{c_n\}_{n \in \mathbb{N}}$ such that*

$$\liminf_{n \to \infty} -\frac{1}{n} \log \alpha_n(c_n) \geq \kappa_\alpha, \tag{5a}$$

$$\liminf_{n \to \infty} -\frac{1}{n} \log \beta_n(c_n) \geq \kappa_\beta. \tag{5b}$$

The error-exponent region $\bar{\mathcal{R}}$ is the closure of the set of all achievable error-exponent pairs $(\kappa_\alpha, \kappa_\beta)$. Set $\mathcal{R} := \{(\kappa_\alpha, \kappa(\kappa_\alpha)) : \kappa_\alpha \in (0, \kappa_\alpha^\star)\}$, where $\kappa_\alpha^\star = \inf\{\kappa_\alpha : \kappa(\kappa_\alpha) = 0\}$ and $\kappa(\kappa_\alpha) := \sup\{\kappa_\beta : (\kappa_\alpha, \kappa_\beta) \in \bar{\mathcal{R}}\}$.

We are interested in a computable characterization of $\mathcal{R}$, which pertains to the region of positive error-exponents (i.e., excluding the boundary points corresponding to Stein's exponent). To this end, we present two inner bounds on $\mathcal{R}$ in the next section.

3. Main Results

In this section, we obtain two inner bounds on $\mathcal{R}$, first using a separation-based scheme which performs independent HT and channel coding, termed the SHTCC scheme, and the second via a joint HT and channel coding scheme that uses hybrid coding for communication between the observer and the decision maker.

3.1. Inner Bound on $\mathcal{R}$ via SHTCC Scheme

Let $\mathcal{S} = \mathcal{X}$ and $P_{SXY} = P_{SX}P_{Y|X}$ be a PMF under which $S - X - Y$ forms a Markov chain. For $x \in \mathcal{X}$, let $\Lambda_{x, P_{SXY}}(y) := \log\left(P_{Y|X=x}(y) / P_{Y|S=x}(y)\right)$ and define

$$E_{\mathrm{sp}}(P_{SX}, \theta) := \sum_{s \in \mathcal{S}} P_S(s) \psi^*_{P_{Y|S=s}, \Lambda_{s, P_{SXY}}}(\theta),$$

where the rate function ψ^* is defined in (1). For a fixed P_{SX} and $R \geq 0$, let

$$E_{\mathrm{ex}}(R, P_{SX}) := \max_{\rho \geq 1} -\rho R - \rho \log \left(\sum_{s, x, \tilde{x}} P_S(s) P_{X|S}(x|s) P_{X|S}(\tilde{x}|s) \right.$$

$$\left. \left(\sum_y \left(P_{Y|X}(y|x) P_{Y|X}(y|\tilde{x}) \right)^{\frac{1}{2}} \right)^{\frac{1}{\rho}} \right),$$

denote the expurgated exponent [38,44]. Let $\mathcal{W}$ be a finite set and $\mathfrak{F}$ denote the set of all continuous mappings from $\mathcal{P}(\mathcal{U})$ to $\mathcal{P}(\mathcal{W}|\mathcal{U})$, where $\mathcal{P}(\mathcal{W}|\mathcal{U})$ is the set of all conditional distributions $P_{W|U}$. Set $\theta_{\mathrm{l}}(P_{SX}) := \sum_{s \in \mathcal{S}} P_S(s) D\left(P_{Y|S=s} || P_{Y|X=s}\right)$, $\theta_{\mathrm{u}}(P_{SX}) :=$

$\sum_{s \in \mathcal{S}} P_S(s) D(P_{Y|X=s} || P_{Y|S=s})$, $\Theta(P_{SX}) := (-\theta_l(P_{SX}), \theta_u(P_{SX}))$. Denote an arbitrary element of $\mathfrak{F} \times \mathbb{R}_{\geq 0} \times \mathcal{P}(\mathcal{S} \times \mathcal{X}) \times \Theta(P_{SX})$ by $(\omega, R, P_{SX}, \theta)$, and set

$$\mathcal{L}(\kappa_\alpha) := \left\{ \begin{matrix} (\omega, R, P_{SX}, \theta) : \ \zeta(\kappa_\alpha, \omega) - \rho(\kappa_\alpha, \omega) \leq R < I_P(X; Y|S), \ P_{SXY} = P_{SX} P_{Y|X} \\ \min\{ E_{\mathrm{sp}}(P_{SX}, \theta), E_{\mathrm{ex}}(R, P_{SX}), E_{\mathrm{b}}(\kappa_\alpha, \omega, R) \} \geq \kappa_\alpha \end{matrix} \right\},$$

$$\hat{\mathcal{L}}(\kappa_\alpha, \omega) := \left\{ P_{\hat{U}\hat{V}\hat{W}} : \ D(P_{\hat{U}\hat{V}\hat{W}} || P_{UV\hat{W}}) \leq \kappa_\alpha, P_{\hat{W}|\hat{U}} = \omega(P_{\hat{U}}), P_{UV\hat{W}} = P_{UV} P_{\hat{W}|\hat{U}} \right\}, \quad (6\mathrm{a})$$

$$E_{\mathrm{b}}(\kappa_\alpha, \omega, R) := \begin{cases} R - \zeta(\kappa_\alpha, \omega) + \rho(\kappa_\alpha, \omega), & \text{if } 0 \leq R < \zeta(\kappa_\alpha, \omega), \\ \infty, & \text{otherwise}, \end{cases}$$

$$\zeta(\kappa_\alpha, \omega) := \max_{P_{\hat{U}\hat{W}} : \ \exists \ P_{\hat{V}}, P_{\hat{U}\hat{V}\hat{W}} \in \hat{\mathcal{L}}(\kappa_\alpha, \omega)} I_P(\hat{U}; \hat{W}), \quad (6\mathrm{b})$$

$$\rho(\kappa_\alpha, \omega) := \min_{P_{\hat{V}\hat{W}} : \ \exists \ P_{\hat{U}}, P_{\hat{U}\hat{V}\hat{W}} \in \hat{\mathcal{L}}(\kappa_\alpha, \omega)} I_P(\hat{V}; \hat{W}), \quad (6\mathrm{c})$$

$$E_1(\kappa_\alpha, \omega) := \min_{(P_{\tilde{U}\tilde{V}\tilde{W}}, Q_{\tilde{U}\tilde{V}\tilde{W}}) \in \mathcal{T}_1(\kappa_\alpha, \omega)} D(P_{\tilde{U}\tilde{V}\tilde{W}} || Q_{\tilde{U}\tilde{V}\tilde{W}}),$$

$$E_2(\kappa_\alpha, \omega, R) := \begin{cases} \min\limits_{(P_{\tilde{U}\tilde{V}\tilde{W}}, Q_{\tilde{U}\tilde{V}\tilde{W}}) \in \mathcal{T}_2(\kappa_\alpha, \omega)} D(P_{\tilde{U}\tilde{V}\tilde{W}} || Q_{\tilde{U}\tilde{V}\tilde{W}}) + E_{\mathrm{b}}(\kappa_\alpha, \omega, R), & \text{if } R < \zeta(\kappa_\alpha, \omega), \\ \infty, & \text{otherwise}, \end{cases}$$

$$E_3(\kappa_\alpha, \omega, R, P_{SX}) := \begin{cases} \min\limits_{(P_{\tilde{U}\tilde{V}\tilde{W}}, Q_{\tilde{U}\tilde{V}\tilde{W}}) \in \mathcal{T}_3(\kappa_\alpha, \omega)} D(P_{\tilde{U}\tilde{V}\tilde{W}} || Q_{\tilde{U}\tilde{V}\tilde{W}}) + E_{\mathrm{b}}(\kappa_\alpha, \omega, R) + E_{\mathrm{ex}}(R, P_{SX}), \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if } R < \zeta(\kappa_\alpha, \omega), \\ \min\limits_{(P_{\tilde{U}\tilde{V}\tilde{W}}, Q_{\tilde{U}\tilde{V}\tilde{W}}) \in \mathcal{T}_3(\kappa_\alpha, \omega)} D(P_{\tilde{U}\tilde{V}\tilde{W}} || Q_{\tilde{U}\tilde{V}\tilde{W}}) + \rho(\kappa_\alpha, \omega) + E_{\mathrm{ex}}(R, P_{SX}), \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{otherwise}, \end{cases}$$

$$E_4(\kappa_\alpha, \omega, R, P_{SX}, \theta) := \begin{cases} \min\limits_{P_{\hat{V}} : P_{\hat{U}\hat{V}\hat{W}} \in \hat{\mathcal{L}}(\kappa_\alpha, \omega)} D(P_{\hat{V}} || Q_V) + E_{\mathrm{b}}(\kappa_\alpha, \omega, R) + E_{\mathrm{m}}(P_{SX}, \theta) - \theta, \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if } R < \zeta(\kappa_\alpha, \omega), \\ \min\limits_{P_{\hat{V}} : P_{\hat{U}\hat{V}\hat{W}} \in \hat{\mathcal{L}}(\kappa_\alpha, \omega)} D(P_{\hat{V}} || Q_V) + \rho(\kappa_\alpha, \omega) + E_{\mathrm{m}}(P_{SX}, \theta) - \theta, \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{otherwise}, \end{cases}$$

where,

$$\mathcal{T}_1(\kappa_\alpha, \omega) := \left\{ \begin{matrix} (P_{\tilde{U}\tilde{V}\tilde{W}}, Q_{\tilde{U}\tilde{V}\tilde{W}}) : \ P_{\tilde{U}\tilde{W}} = P_{\hat{U}\hat{W}}, \ P_{\tilde{V}\tilde{W}} = P_{\hat{V}\hat{W}}, \ Q_{\tilde{U}\tilde{V}\tilde{W}} \\ := Q_{UV} P_{\tilde{W}|\tilde{U}} \text{ for some } P_{\hat{U}\hat{V}\hat{W}} \in \hat{\mathcal{L}}(\kappa_\alpha, \omega) \end{matrix} \right\}, \quad (6\mathrm{d})$$

$$\mathcal{T}_2(\kappa_\alpha, \omega) := \left\{ \begin{matrix} (P_{\tilde{U}\tilde{V}\tilde{W}}, Q_{\tilde{U}\tilde{V}\tilde{W}}) : \ P_{\tilde{U}\tilde{W}} = P_{\hat{U}\hat{W}}, \ P_{\tilde{V}} = P_{\hat{V}}, \ H_P(\tilde{W}|\tilde{V}) \geq H_P(\hat{W}|\hat{V}), \\ Q_{\tilde{U}\tilde{V}\tilde{W}} := Q_{UV} P_{\tilde{W}|\tilde{U}} \text{ for some } P_{\hat{U}\hat{V}\hat{W}} \in \hat{\mathcal{L}}(\kappa_\alpha, \omega) \end{matrix} \right\},$$

$$\mathcal{T}_3(\kappa_\alpha, \omega) := \left\{ \begin{matrix} (P_{\tilde{U}\tilde{V}\tilde{W}}, Q_{\tilde{U}\tilde{V}\tilde{W}}) : \ P_{\tilde{U}\tilde{W}} = P_{\hat{U}\hat{W}}, \ P_{\tilde{V}} = P_{\hat{V}}, \ Q_{\tilde{U}\tilde{V}\tilde{W}} := Q_{UV} P_{\tilde{W}|\tilde{U}} \\ \text{for some } P_{\hat{U}\hat{V}\hat{W}} \in \hat{\mathcal{L}}(\kappa_\alpha, \omega) \end{matrix} \right\}.$$

We have the following lower bound for $\kappa(\kappa_\alpha)$, which translates to an inner bound for $\mathcal{R}$.

Theorem 1 (Inner bound via SHTCC scheme). $\kappa(\kappa_\alpha) \geq \kappa_{\mathrm{s}}^\star(\kappa_\alpha)$, *where*

$$\kappa_{\mathrm{s}}^\star(\kappa_\alpha) := \max_{(\omega, R, P_{SX}, \theta) \in \mathcal{L}(\kappa_\alpha)} \min \{ E_1(\kappa_\alpha, \omega), E_2(\kappa_\alpha, \omega, R), E_3(\kappa_\alpha, \omega, R, P_{SX}),$$

$$E_4(\kappa_\alpha, \omega, R, P_{SX}, \theta) \}. \quad (7)$$

The proof of Theorem 1 is presented in Section 4.1. The SHTCC scheme, which achieves the error-exponent pair $(\kappa_\alpha, \kappa_{\mathrm{s}}^\star(\kappa_\alpha))$, is a coding scheme analogous to separate source and channel coding for the lossy transmission of a source over a communication channel with correlated side-information at the receiver [45], however, with the objective of reliable HT. In this scheme, the source samples are first compressed to an index, which acts as the message

to be transmitted over the channel. But, in contrast to standard communication problems, there is a need to protect certain messages more reliably than others; hence, an unequal error-protection scheme [39,42] is used. To describe briefly, the SHTCC scheme involves (*i*) the quantization and binning of **u** sequences, whose type $P_{\mathbf{u}}$ is within a κ_α-neighborhood (in terms of KL divergence) of P_U, using **V** as side information at the decision maker for decoding, and (*ii*) unequal error-protection channel coding scheme in [39] for protecting a special message which informs the decision maker that $P_{\mathbf{u}}$ lies outside the κ_α-neighborhood of P_U. The output of the channel decoder is processed by an empirical conditional entropy decoder which recovers the quantization codeword with the least conditional entropy with **V**. Since this decoder depends only on the empirical distributions of the observations, it is universal and useful in the hypothesis testing context, where multiple distributions are involved (as was first noted in [8]). The various factors E_1 to E_4 in (7) have natural interpretations in terms of events that could possibly result in a hypothesis testing error. Specifically, E_1 and E_2 correspond to the error events arising due to quantization and binning, respectively, while E_3 and E_4 correspond to the error events of wrongly decoding an ordinary channel codeword and special message codeword, respectively.

Remark 1 (Generalization of Han–Kobayashi inner bound). *In Theorem 1 in [14], Han and Kobayashi obtained an inner bound on $\mathcal{R}$ for DHT over a noiseless channel. At a high level, their coding scheme involves type-based quantization of $\mathbf{u} \in \mathcal{U}^n$ sequences, whose type $P_{\mathbf{u}}$ lies within a κ_α-neighborhood of P_U, where κ_α is the desired type I error-exponent. As a corollary, Theorem 1 recovers the lower bound for $\kappa(\kappa_\alpha)$ obtained in [14] by (i) setting $E_{\mathrm{ex}}(R, P_{SX})$, $E_{\mathrm{m}}(P_{SX}, \theta)$ and $E_{\mathrm{m}}(P_{SX}, \theta) - \theta$ to ∞, which hold when the channel is noiseless, and (ii) maximizing over the set $\{(\omega, R, P_{SX}, \theta) \in \mathfrak{F} \times \mathbb{R}_{\geq 0} \times \mathcal{P}(\mathcal{S} \times \mathcal{X}) \times \Theta(P_{SX}) : \zeta(\kappa_\alpha, \omega) \leq R < I_P(X; Y|S), P_{SXY} := P_{SX}P_{Y|X}\} \subseteq \mathcal{L}(\kappa_\alpha)$ in (7). Then, note that the terms $E_2(\kappa_\alpha, \omega, R)$, $E_3(\kappa_\alpha, \omega, R, P_{SX})$ and $E_4(\kappa_\alpha, \omega, R, P_{SX}, \theta)$ all equal ∞, and thus the inner bound in Theorem 1 reduces to that given in Theorem 1 in [14].*

Remark 2 (Improvement via time-sharing). *Since the lower bound on $\kappa(\kappa_\alpha)$ in Theorem 1 is not necessarily concave, a tighter bound can be obtained using the technique of time-sharing similar to Theorem 3 in [14]. We omit its description, as it is cumbersome, although straightforward.*

Theorem 1 also recovers the lower bound for the optimal type II error-exponent for a fixed type I error probability constraint established in Theorem 2 in [10] by letting $\kappa_\alpha \to 0$. The details are provided in Appendix A. Further, specializing the lower bound in Theorem 1 to the case of TAI, i.e., when $Q_{UV} = P_U P_V$, we obtain the following corollary which characterizes the optimal type II error-exponent for TAI established in Proposition 7 in [10] as a special case.

Corollary 1 (Inner bound for TAI). *Let $P_{UV} \in \mathcal{P}(\mathcal{U} \times \mathcal{V})$ be an arbitrary distribution and $Q_{UV} = P_U P_V$. Then,*

$$\kappa(\kappa_\alpha) \geq \kappa_{\mathrm{s}}^\star(\kappa_\alpha) \geq \kappa_{\mathrm{i}}^\star(\kappa_\alpha), \tag{8}$$

where

$$\kappa_{\mathrm{i}}^\star(\kappa_\alpha) := \max_{(\omega, P_{SX}, \theta) \in \mathcal{L}^\star(\kappa_\alpha)} \min\left\{E_1^{\mathrm{i}}(\kappa_\alpha, \omega), E_2^{\mathrm{i}}(\kappa_\alpha, \omega, P_{SX}), E_3^{\mathrm{i}}(\kappa_\alpha, \omega, P_{SX}, \theta)\right\},$$

$$\mathcal{L}^\star(\kappa_\alpha) := \left\{\begin{array}{l} (\omega, P_{SX}, \theta) \in \mathfrak{F} \times \mathcal{P}(\mathcal{S} \times \mathcal{X}) \times \Theta(P_{SX}) : \zeta(\kappa_\alpha, \omega) < I_P(X; Y|S), \\ P_{SXY} := P_{SX}P_{Y|X}, \ \min\left\{E_{\mathrm{sp}}(P_{SX}, \theta), E_{\mathrm{ex}}(\zeta(\kappa_\alpha, \omega), P_{SX})\right\} \geq \kappa_\alpha \end{array}\right\}, \tag{9}$$

$$E_1^{\mathrm{i}}(\kappa_\alpha, \omega) := \min_{P_{\hat{V}\hat{W}}: \exists P_{\hat{U}\hat{V}\hat{W}} \in \hat{\mathcal{L}}(\kappa_\alpha, \omega)} I_P(\hat{V}; \hat{W}) + D(P_{\hat{V}}||P_V),$$

$$E_2^{\mathrm{i}}(\kappa_\alpha, \omega, P_{SX}) := \rho(\kappa_\alpha, \omega) + E_{\mathrm{ex}}(\zeta(\kappa_\alpha, \omega), P_{SX}),$$

$$E_3^{\mathrm{i}}(\kappa_\alpha, \omega, P_{SX}, \theta) := \rho(\kappa_\alpha, \omega) + E_{\mathrm{sp}}(P_{SX}, \theta) - \theta,$$

and $\hat{\mathcal{L}}(\kappa_\alpha, \omega)$, $\zeta(\kappa_\alpha, \omega)$ and $\rho(\kappa_\alpha, \omega)$ are defined in (6a), (6b) and (6c), respectively. In particular,

$$\lim_{\kappa_\alpha \to 0} \kappa(\kappa_\alpha) = \kappa_s^\star(0) = \kappa_i^\star(0) = \max_{\substack{P_{W|U}: I_P(U;W) \leq C(P_{Y|X}), \\ P_{UVW} = P_{UV} P_{W|U}}} I_P(V;W), \tag{10}$$

where $|\mathcal{W}| \leq |\mathcal{U}| + 1$ and $C(P_{Y|X})$ denotes the capacity of the channel $P_{Y|X}$.

The proof of Corollary 1 is given in Section 4.2. Its achievability follows from a special case of the SHTCC scheme without binning at the encoder.

Next, we consider *testing against dependence* (TAD) for which Q_{UV} is an arbitrary joint distribution and $P_{UV} = Q_U Q_V$. Theorem 1 specialized to TAD gives the following corollary.

Corollary 2 (Inner bound for TAD). *Let $Q_{UV} \in \mathcal{P}(\mathcal{U} \times \mathcal{V})$ be an arbitrary distribution and $P_{UV} = Q_U Q_V$. Then,*

$$\kappa(\kappa_\alpha) \geq \kappa_s^\star(\kappa_\alpha) = \kappa_d^\star(\kappa_\alpha) := \max_{\substack{(\omega, P_{SX}, \theta) \\ \in \mathcal{L}^\star(\kappa_\alpha)}} \min\left\{ E_1^d(\kappa_\alpha, \omega), E_2^d(\kappa_\alpha, \omega, P_{SX}), E_3^d(P_{SX}, \theta) \right\}, \tag{11}$$

where

$$E_1^d(\kappa_\alpha, \omega) := \min_{\substack{(P_{\tilde{U}\tilde{V}\tilde{W}}, Q_{\tilde{U}\tilde{V}\tilde{W}}) \\ \in \mathcal{T}_1(\kappa_\alpha, \omega)}} D(P_{\tilde{U}\tilde{V}\tilde{W}} || Q_{\tilde{U}\tilde{V}\tilde{W}}) \geq \min_{\substack{(P_{\hat{V}\hat{W}}, Q_{V\hat{W}}): P_{\hat{U}\hat{V}\hat{W}} \in \hat{\mathcal{L}}(\kappa_\alpha, \omega), \\ Q_{UV\hat{W}} = Q_{UV} P_{\hat{W}|\hat{U}}}} D(P_{\hat{V}\hat{W}} || Q_{V\hat{W}}),$$

$$E_2^d(\kappa_\alpha, \omega, P_{SX}) := E_{\text{ex}}(\zeta(\kappa_\alpha, \omega), P_{SX}),$$
$$E_3^d(P_{SX}, \theta) := E_{\text{sp}}(P_{SX}, \theta) - \theta,$$

and $\hat{\mathcal{L}}(\kappa_\alpha, \omega)$, $\mathcal{T}_1(\kappa_\alpha, \omega)$ and $\mathcal{L}^\star(\kappa_\alpha)$ are given in (6a), (6d) and (9), respectively. In particular,

$$\lim_{\kappa_\alpha \to 0} \kappa(\kappa_\alpha) \geq \kappa_s^\star(0) = \kappa_d^\star(0) \geq \kappa_{\text{TAD}}^\star, \tag{12}$$

where

$$\kappa_{\text{TAD}}^\star = \max_{\substack{(P_{W|U}, P_{SX}): \\ I_Q(W;U) \leq I_P(X;Y|S), \\ Q_{UVW} = Q_{UV} P_{W|U}, \\ P_{SXY} = P_{SX} P_{Y|X}}} \min\left\{ D(Q_V Q_W || Q_{VW}), E_{\text{ex}}(I_Q(U;W), P_{SX}), \theta_1(P_{SX}) \right\},$$

and $|\mathcal{W}| \leq |\mathcal{U}| + 1$.

The proof of Corollary 2 is given in Section 4.3. Note that the expression for $\kappa_s^\star(\kappa_\alpha)$ given in (11) is relatively simpler to compute compared to that in Theorem 1. This will be handy in showing that the JHTCC scheme strictly outperforms the SHTCC scheme, which we highlight via an example in Section 3.3 below.

3.2. Inner Bound via JHTCC Scheme

It is well known that joint source-channel coding schemes offer advantages over separation-based coding schemes in several information theoretic problems, such as the transmission of correlated sources over a multiple-access channel [40,46] and the error-exponent in the lossless or lossy transmission of a source over a noisy channel [42,47]. Recently, it was shown via an example in [10] that joint schemes also achieve a strictly larger type II error-exponent in DHT problems compared to a separation-based scheme in some scenarios. Motivated by this, we present an inner bound on $\mathcal{R}$ using a generalization of the JHTCC scheme in [10].

Let $\mathcal{W}$ and $\mathcal{S}$ be arbitrary finite sets, and $\mathfrak{F}'$ denote the set of all continuous mappings from $\mathcal{P}(\mathcal{U} \times \mathcal{S})$ to $\mathcal{P}(\mathcal{W} | \mathcal{U} \times \mathcal{S})$, where $\mathcal{P}(\mathcal{W} | \mathcal{U} \times \mathcal{S})$ is the set of all conditional

distributions $P_{W|US}$. Let $\left(P_S, \omega'(\cdot, P_S), P_{X|USW}, P_{X'|US}\right)$ denote an arbitrary element of $\mathcal{P}(\mathcal{S}) \times \mathfrak{F}' \times \mathcal{P}(\mathcal{X}|\mathcal{U} \times \mathcal{S} \times \mathcal{W}) \times \mathcal{P}(\mathcal{X}|\mathcal{U} \times \mathcal{S})$, and define

$$\mathcal{L}_h(\kappa_\alpha) := \left\{ \left(P_S, \omega'(\cdot, P_S), P_{X|USW}, P_{X'|US}\right) : E_{\mathsf{b}}'(\kappa_\alpha, \omega', P_S, P_{X|USW}) \geq \kappa_\alpha \right\},$$

$$\hat{\mathcal{L}}_h(\kappa_\alpha, \omega', P_S, P_{X|USW})$$

$$:= \left\{ \begin{array}{c} (P_{\hat{U}\hat{V}\hat{W}\hat{Y}S} : D(P_{\hat{U}\hat{V}\hat{W}\hat{Y}|S} || P_{UV\hat{W}Y|S} | P_S) \leq \kappa_\alpha, \ P_{SUV\hat{W}XY} := P_S P_{UV} P_{\hat{W}|\hat{U}S} P_{X|USW} P_{Y|X}, \\ P_{\hat{W}|\hat{U}S} = \omega'(P_{\hat{U}}, P_S) \end{array} \right\},$$

$$E_{\mathsf{b}}'(\kappa_\alpha, \omega', P_S, P_{X|USW}) := \rho'(\kappa_\alpha, \omega', P_S, P_{X|USW}) - \zeta_q'(\kappa_\alpha, \omega', P_S),$$

$$\zeta'(\kappa_\alpha, \omega', P_S) := \max_{\substack{P_{\hat{U}\hat{W}S} : \exists\, P_{\hat{V}\hat{Y}} \text{ s.t.} \\ P_{\hat{U}\hat{V}\hat{W}\hat{Y}S} \in \hat{\mathcal{L}}_h(\kappa_\alpha, \omega', P_S, P_{X|USW})}} I_P(\hat{U}; \hat{W}|S),$$

$$\rho'(\kappa_\alpha, \omega', P_S, P_{X|USW}) := \min_{\substack{P_{\hat{V}\hat{W}\hat{Y}S} : \exists\, P_{\hat{U}} \text{ s.t.} \\ P_{\hat{U}\hat{V}\hat{W}\hat{Y}S} \in \hat{\mathcal{L}}_h(\kappa_\alpha, \omega', P_S, P_{X|USW})}} I_P(\hat{Y}, \hat{V}; \hat{W}|S),$$

$$E_1'(\kappa_\alpha, \omega') := \min_{(P_{\tilde{U}\tilde{V}\tilde{W}\tilde{Y}S}, Q_{\tilde{U}\tilde{V}\tilde{W}\tilde{Y}S}) \in \mathcal{T}_1'(\kappa_\alpha, \omega')} D(P_{\tilde{U}\tilde{V}\tilde{W}\tilde{Y}|S} || Q_{\tilde{U}\tilde{V}\tilde{W}\tilde{Y}|S} | P_S),$$

$$E_2'(\kappa_\alpha, \omega', P_S, P_{X|USW}) := \min_{(P_{\tilde{U}\tilde{V}\tilde{W}\tilde{Y}S}, Q_{\tilde{U}\tilde{V}\tilde{W}\tilde{Y}S}) \in \mathcal{T}_2'(\kappa_\alpha, \omega', P_S, P_{X|USW})} D(P_{\tilde{U}\tilde{V}\tilde{W}\tilde{Y}|S} || Q_{\tilde{U}\tilde{V}\tilde{W}\tilde{Y}|S} | P_S)$$

$$+ E_{\mathsf{b}}'(\kappa_\alpha, \omega', P_S, P_{X|USW}),$$

$$E_3'(\kappa_\alpha, \omega', P_S, P_{X|USW}, P_{X'|US}) := \min_{P_{\hat{V}\hat{Y}S} : P_{\hat{U}\hat{V}\hat{W}\hat{Y}S} \in \hat{\mathcal{L}}_h(\kappa_\alpha, \omega', P_S, P_{X|USW})} D(P_{\hat{V}\hat{Y}|S} || Q_{VY'|S} | P_S)$$

$$+ E_{\mathsf{b}}'(\kappa_\alpha, \omega', P_S, P_{X|USW}),$$

$$Q_{SUVX'Y'} := P_S Q_{UV} P_{X'|US} P_{Y'|X'}, \ P_{Y'|X'} := P_{Y|X},$$

$$\mathcal{T}_1'(\kappa_\alpha, \omega', P_S, P_{X|USW}) := \left\{ \begin{array}{cl} (P_{\tilde{U}\tilde{V}\tilde{W}\tilde{Y}S}, & P_{\tilde{U}\tilde{W}S} = P_{\hat{U}\hat{W}S}, \ P_{\tilde{V}\tilde{W}\tilde{Y}S} = P_{\hat{V}\hat{W}\hat{Y}S}, \\ Q_{\tilde{U}\tilde{V}\tilde{W}\tilde{Y}S}) : & Q_{S\tilde{U}\tilde{V}\tilde{W}\tilde{X}\tilde{Y}} := P_S Q_{UV} P_{\tilde{W}|\tilde{U}S} P_{X|USW} P_{Y|X} \\ & \text{for some } P_{\hat{U}\hat{V}\hat{W}\hat{Y}S} \in \hat{\mathcal{L}}_h(\kappa_\alpha, \omega', P_S, P_{X|USW}) \end{array} \right\},$$

$$\mathcal{T}_2'(\kappa_\alpha, \omega', P_S, P_{X|USW}) := \left\{ \begin{array}{cl} (P_{\tilde{U}\tilde{V}\tilde{W}\tilde{Y}S}, & P_{\tilde{U}\tilde{W}S} = P_{\hat{U}\hat{W}S}, \ P_{\tilde{V}\tilde{Y}S} = P_{\hat{V}\hat{Y}S}, \\ Q_{\tilde{U}\tilde{V}\tilde{W}\tilde{Y}S}) : & H_P(\tilde{W}|\tilde{V}, \tilde{Y}, S) \geq H_P(\hat{W}|\hat{V}, \hat{Y}, S), \\ & Q_{S\tilde{U}\tilde{V}\tilde{W}\tilde{X}\tilde{Y}} := P_S Q_{UV} P_{\tilde{W}|\tilde{U}S} P_{X|USW} P_{Y|X} \\ & \text{for some } P_{\hat{U}\hat{V}\hat{W}\hat{Y}S} \in \hat{\mathcal{L}}_h(\kappa_\alpha, \omega', P_S, P_{X|USW}) \end{array} \right\}.$$

Then, we have the following result.

Theorem 2 (Inner bound via JHTCC scheme).

$$\kappa(\kappa_\alpha) \geq \max\{\kappa_{\mathsf{h}}^\star(\kappa_\alpha), \kappa_{\mathsf{u}}^\star(\kappa_\alpha)\}, \tag{13}$$

where

$$\kappa_{\mathsf{h}}^\star(\kappa_\alpha) := \max_{(P_S, \omega', P_{X|USW}, P_{X'|US}) \in \mathcal{L}_h(\kappa_\alpha)} \min \left\{ E_1'(\kappa_\alpha, \omega'), \ E_2'(\kappa_\alpha, \omega', P_S, P_{X|USW}), \right.$$

$$\left. E_3'(\kappa_\alpha, \omega', P_S, P_{X|USW}, P_{X'|US}) \right\},$$

$$\kappa_{\mathsf{u}}^\star(\kappa_\alpha) := \max_{(P_S, P_{X|US}) \in \mathcal{P}(\mathcal{S}) \times \mathcal{P}(\mathcal{X}|\mathcal{S} \times \mathcal{U})} \kappa_{\mathsf{u}}(\kappa_\alpha, P_S, P_{X|US}),$$

$$\kappa_{\mathsf{u}}(\kappa_\alpha, P_S, P_{X|US}) := \min_{P_S P_{\hat{V}\hat{Y}} : D\left(P_{\hat{V}\hat{Y}|S} || P_{VY|S} | P_S\right) \leq \kappa_\alpha} D\left(P_{\hat{V}\hat{Y}|S} || Q_{VY|S} | P_S\right),$$

$$P_{SUVXY} = P_S P_{UV} P_{X|US} P_{Y|X} \quad and \quad Q_{SUVXY} = P_S Q_{UV} P_{X|US} P_{Y|X}.$$

The proof of Theorem 2 is given in Section 4.4, and utilizes a generalization of hybrid coding scheme [40] to achieve the stated inner bound. Specifically, the error-exponent pair $(\kappa_\alpha, \kappa_h^\star(\kappa_\alpha))$ is achieved using type-based hybrid coding, while $(\kappa_\alpha, \kappa_u^\star(\kappa_\alpha))$ is realized by uncoded transmission, in which the channel input **X** is generated as the output of a DMC $P_{X|U}$ with input **U** (along with time sharing). In standard hybrid coding, the source sequence is first quantized via joint typicality and the channel input is then chosen as a function of both the original source sequence and its quantization. At the decoder, the quantized codeword is first recovered using the channel output and side information via joint typicality decoding, and an estimate of the source sequence is output as a function of the channel output and recovered codeword. The quantization part forms the *digital* part of the scheme, while the use of the source sequence for encoding and channel output for decoding comprises the *analog* part. The scheme derives its name from these joint hybrid digital-analog operations. In the HT context considered here, the aforementioned source quantization is replaced by a type-based quantization at the encoder, and the joint typicality decoder is replaced by a universal empirical conditional entropy decoder. We note that Theorem 2 recovers the lower bound on the optimal type II error-exponent proved in Theorem 5 in [10]. The details are provided in Appendix B.

Next, we provide a comparison between the SHTCC and JHTCC bounds via an example as mentioned earlier.

3.3. Comparison of Inner Bounds

We compare the inner bounds established in Theorem 1 and Theorem 2 for a simple setting of TAD over a BSC. For this purpose, we will use the inner bound $\kappa_d^\star(\kappa_\alpha)$ stated in Corollary 2 and $\kappa_u^\star(\kappa_\alpha)$ that is achieved by uncoded transmission. Our objective is to illustrate that the JHTCC scheme achieves a strictly tighter bound on $\mathcal{R}$ compared to the SHTCC scheme, at least for some points of the trade-off.

Example 1. *Let* $p, q \in [0, 0.5]$, $\mathcal{U} = \mathcal{V} = \mathcal{X} = \mathcal{Y} = \mathcal{S} = \{0, 1\}$,

$$Q_{UV} = \begin{bmatrix} q & 0.5 - q \\ 0.5 - q & q \end{bmatrix}, \quad P_{Y|X} = \begin{bmatrix} 1 - p & p \\ p & 1 - p \end{bmatrix}, \text{ and } P_{UV} = Q_U Q_V.$$

A comparison of the inner bounds achieved by the SHTCC and JHTCC schemes for the above example are shown in Figures 2 and 3, where we plot the error-exponents trade-off achieved by uncoded transmission (a lower bound for the JHTCC scheme), and the expurgated exponent at a zero rate:

$$E_{\text{ex}}(0) := \max_{P_{SX} \in \mathcal{P}(\mathcal{S} \times \mathcal{X})} E_{\text{ex}}(P_{SX}, 0) = -0.25 \log(4p(1 - p)),$$

which is an upper bound on $\kappa_d^\star(\kappa_\alpha)$ for any $\kappa_\alpha \geq 0$. To compute $E_{\text{ex}}(0)$, we used the closed-form expression for $E_{\text{ex}}(\cdot)$ given in Problem 10.26(c) in [38]. Clearly, it can be seen that the JHTCC scheme outperforms SHTCC scheme for κ_α below a threshold, which depends on the source and channel distributions. In particular, the threshold below which improvement is seen is reduced when the channel or the source becomes more uniform. The former behavior can be seen directly by comparing the subplots in Figures 2 and 3, while the latter can be noted by comparing Figure 2a with Figure 3a, or Figure 2b with Figure 3b.
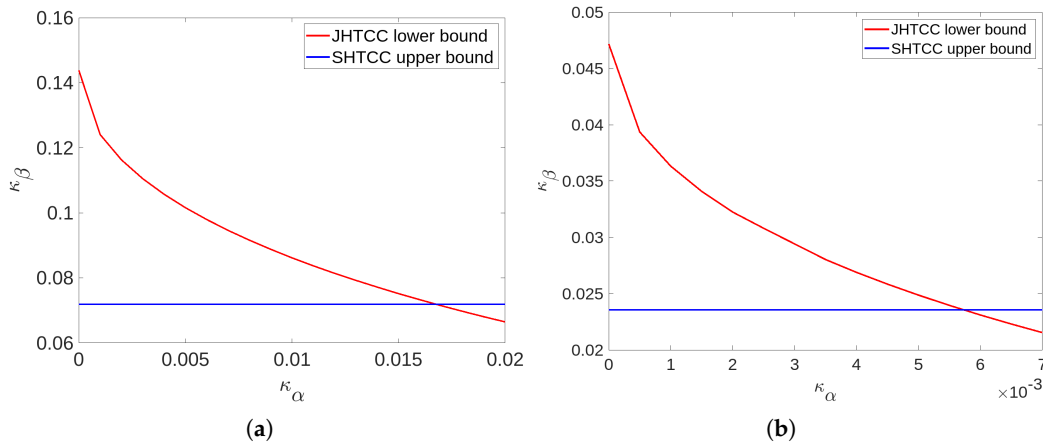
Figure 2. Comparison of the error-exponents trade-off achieved by the SHTCC and JHTCC schemes for TAD over a BSC in Example 1 with parameters $p = 0.25, q = 0$ for (**a**) and $p = 0.35, q = 0$ for (**b**). The red curve shows $(\kappa_\alpha, \kappa_u^\star(\kappa_\alpha))$ pairs achieved by uncoded transmission while the blue line plots $(\kappa_\alpha, E_{ex}(0))$. The joint scheme clearly achieves a better error-exponent trade-off for values of κ_α below a threshold which depends on the transition kernel of the channel. In particular, a more uniform channel results in a lesser threshold.
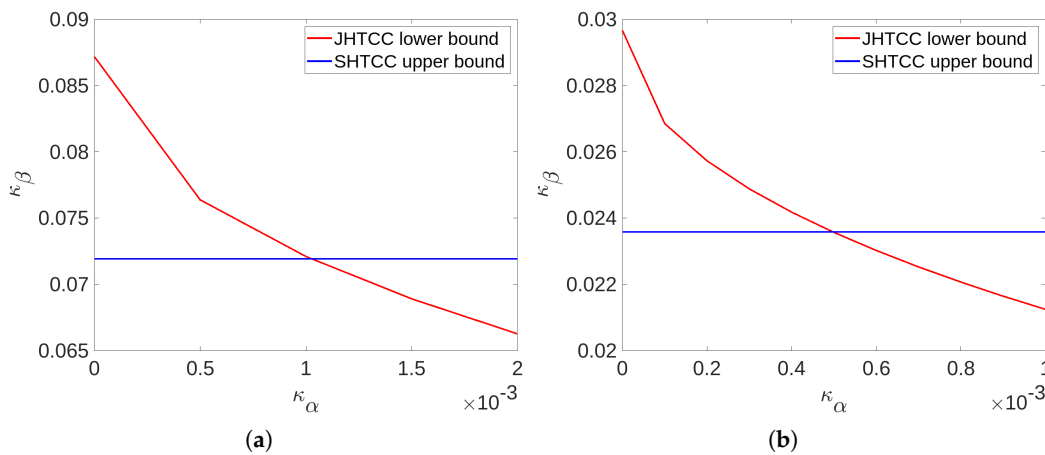

Figure 3. Comparison of the error-exponents trade-off achieved by the SHTCC and JHTCC schemes for Example 1 with parameters $p = 0.25, q = 0.05$ for (**a**) and $p = 0.35, q = 0.05$ for (**b**). The JHTCC scheme improves over the separation based scheme for small values of κ_α; however, the region of improvement is reduced compared to Figure 2 as the source is more uniformly distributed.

4. Proofs

4.1. Proof of Theorem 1

We will show the achievability of the error-exponent pair $(\kappa_\alpha, \kappa_s^\star(\kappa_\alpha))$ by constructing a suitable ensemble of HT codes, and showing that the expected type I and type II error probabilities (over this ensemble) satisfy (5) for the pair $(\kappa_\alpha, \kappa_s^\star(\kappa_\alpha))$. Then, an expurgation argument [44] will be used to show the existence of a HT code that satisfies (5) for the same error-exponent pair, thus showing that $(\kappa_\alpha, \kappa_s^\star(\kappa_\alpha)) \in \mathcal{R}$ as desired.

Let $n \in \mathbb{N}$, $|\mathcal{W}| < \infty$, $\kappa_\alpha > 0$, $(\omega, R, P_{SX}, \theta) \in \mathcal{L}(\kappa_\alpha)$, $R' := \zeta(\kappa_\alpha, \omega)$, and $\eta > 0$ be a small number. Additionally, suppose that $R \geq 0$ satisfies

$$\zeta(\kappa_\alpha, \omega) - \rho(\kappa_\alpha, \omega) \leq R < I_P(X; Y|S), \tag{14}$$

where $\zeta(\kappa_\alpha, \omega)$ and $\rho(\kappa_\alpha, \omega)$ are defined in (6b) and (6c), respectively. The SHTCC scheme is as follows:

Encoding: The observer's encoder is composed of two stages, a *source encoder* followed by a *channel encoder*.

Source encoder: The source encoding comprises a quantization scheme followed by binning to reduce the rate if necessary.

Quantization codebook: Let

$$\mathcal{D}_n(P_U, \eta) := \left\{ P_{\hat{U}} \in \mathcal{T}(\mathcal{U}^n) : D(P_{\hat{U}} || P_U) \leq \kappa_\alpha + \eta \right\}. \tag{15}$$

Consider some ordering on the types in $\mathcal{D}_n(P_U, \eta)$ and denote the elements as $P_{\hat{U}_i}$ for $i \in \left[|\mathcal{D}_n(P_U, \eta)| \right]$. For each type $P_{\hat{U}_i} \in \mathcal{D}_n(P_U, \eta)$, $i \in \left[|\mathcal{D}_n(P_U, \eta)| \right]$, choose a joint type variable $P_{\hat{U}_i \hat{W}_i} \in \mathcal{T}(\mathcal{U}^n \times \mathcal{W}^n)$ such that

$$D\left(P_{\hat{W}_i | \hat{U}_i} || P_{W_i | U} | P_{\hat{U}_i} \right) \leq \frac{\eta}{3}, \tag{16a}$$

$$I_P\left(\hat{U}_i; \hat{W}_i \right) \leq R' + \frac{\eta}{3}, \tag{16b}$$

where $P_{W_i | U} = \omega(P_{\hat{U}_i})$. Note that this is always possible for n sufficiently large by the definition of R'.

Let

$$\mathcal{D}_n(P_{UW}, \eta) := \left\{ P_{\hat{U}_i \hat{W}_i} : i \in \left[|\mathcal{D}_n(P_U, \eta)| \right] \right\}, \tag{17a}$$

$$R'_i := I_P\left(\hat{U}_i; \hat{W}_i \right) + (\eta/3), i \in \left[|\mathcal{D}_n(P_U, \eta)| \right], \tag{17b}$$

$$\mathcal{M}'_i := \left[1 + \sum_{k=1}^{i-1} e^{nR'_k} : \sum_{k=1}^{i} e^{nR'_k} \right], \tag{17c}$$

and $\mathbb{B}_{W,n} = \left\{ \mathbf{W}(j), 1 \leq j \leq \sum_{i=1}^{|\mathcal{D}_n(P_U, \eta)|} |\mathcal{M}'_i| \right\}$ denote a random quantization codebook such that the codeword $\mathbf{W}(j) \sim \mathsf{Unif}\left[\mathcal{T}_n(P_{\hat{W}_i}) \right]$, if $j \in \mathcal{M}'_i$ for some $i \in \left[|\mathcal{D}_n(P_U, \eta)| \right]$. Denote a realization of $\mathbb{B}_{W,n}$ by $\mathcal{B}_{W,n} = \left\{ \mathbf{w}(j) \in \mathcal{W}^n, 1 \leq j \leq \sum_{i=1}^{|\mathcal{D}_n(P_U, \eta)|} |\mathcal{M}'_i| \right\}$.

Quantization scheme: For a given codebook $\mathcal{B}_{W,n}$ and $\mathbf{u} \in \mathcal{T}_n(P_{\hat{U}_i})$ such that $P_{\hat{U}_i} \in \mathcal{D}_n(P_U, \eta)$ for some $i \in \left[|\mathcal{D}_n(P_U, \eta)| \right]$, let

$$\tilde{M}(\mathbf{u}, \mathcal{B}_{W,n}) := \left\{ j \in \mathcal{M}'_i : \mathbf{w}(j) \in \mathcal{B}_{W,n}, (\mathbf{u}, \mathbf{w}(j)) \in \mathcal{T}_n(P_{\hat{U}_i \hat{W}_i}), P_{\hat{U}_i \hat{W}_i} \in \mathcal{D}_n(P_{UW}, \eta) \right\}.$$

If $|\tilde{M}(\mathbf{u}, \mathcal{B}_{W,n})| \geq 1$, let $M'(\mathbf{u}, \mathcal{B}_{W,n})$ denote an index selected uniformly at random from the set $\tilde{M}(\mathbf{u}, \mathcal{B}_{W,n})$, otherwise, set $M'(\mathbf{u}, \mathcal{B}_{W,n}) = 0$. Denoting the support of $M'(\mathbf{u}, \mathcal{B}_{W,n})$ by $\mathcal{M}'$, we have for sufficiently large n that

$$|\mathcal{M}'| \leq 1 + \sum_{i=1}^{|\mathcal{D}_n(P_U, \eta)|} e^{nR'_i} \leq 1 + |\mathcal{D}_n(P_U, \eta)| e^{\max_{P_{\hat{U}\hat{W}} \in \mathcal{D}_n(P_{UW}, \eta)} nI(\hat{U}; \hat{W}) + (n\eta/3)} \leq e^{n(R' + \eta)}, \tag{18}$$

where the last inequality uses (16b) and $|\mathcal{D}_n(P_U, \eta)| \leq (n+1)^{|\mathcal{U}|}$.

Binning: If $|\mathcal{M}'| > |\mathcal{M}|$, then the source encoder performs binning as described below. Let $R_n := \log\left(e^{nR} / |\mathcal{D}_n(P_U, \eta)| \right)$, $\mathcal{M}_i := [1 + (i-1)R_n : iR_n]$, $i \in \left[|\mathcal{D}_n(P_U, \eta)| \right]$, and $\mathcal{M} := \{0\} \cup \left\{ \cup_{i \in [|\mathcal{D}_n(P_U, \eta)|]} \mathcal{M}_i \right\}$. Note that

$$e^{nR_n} \geq e^{nR - |\mathcal{U}| \log(n+1)}. \tag{19}$$

Let $f_{\mathbb{B}}$ denote the random binning function such that for each $j \in \mathcal{M}'_i$, $f_{\mathbb{B}}(j) \sim \mathsf{Unif}\left[|\mathcal{M}_i| \right]$ for $i \in \left[|\mathcal{D}_n(P_U, \eta)| \right]$, and $f_{\mathbb{B}}(0) = 0$ with probability one. Denote a realization of $f_{\mathbb{B}}(j)$ by f_b, where $f_b : \mathcal{M}' \to \mathcal{M}$. Given a codebook $\mathcal{B}_{W,n}$ and binning function f_b, the source encoder outputs $M = f_b(M'(\mathbf{u}, \mathcal{B}_{W,n}))$ for $\mathbf{u} \in \mathcal{U}^n$. If $|\mathcal{M}'| \leq |\mathcal{M}|$, then f_b is taken to be the identity map (no binning), and in this case, $M = M'(\mathbf{u}, \mathcal{B}_{W,n})$.

Channel codebook: Let $\mathbb{B}_{X,n} := \{\mathbf{X}(m) \in \mathcal{X}^n, m \in \mathcal{M}\}$ denote a random channel codebook generated as follows. Without loss of generality, denote the elements of the set $\mathcal{S} = \mathcal{X}$ as $1, \ldots, |\mathcal{X}|$. The codeword length n is divided into $|\mathcal{S}| = |\mathcal{X}|$ blocks, where the length of the first block is $\lceil P_S(1)n \rceil$, the second block is $\lceil P_S(2)n \rceil$, so on and so forth, and the length of the last block is chosen such that the total length is n. For $i \in [|\mathcal{X}|]$, let $k_i := \sum_{l=1}^{i-1} \lceil P_S(l)n \rceil + 1$ and $\bar{k}_i := \sum_{l=1}^{i} \lceil P_S(l)n \rceil$, where the empty sum is defined to be zero. Let $\mathbf{s} \in \mathcal{X}^n$ be such that $s_{k_i}^{\bar{k}_i} = i$, i.e., the elements of $\mathbf{s}$ equal i in the ith block for $i \in [|\mathcal{X}|]$. Let $\mathbf{X}(0) = \mathbf{s}$ with probability one, and the remaining codewords $\mathbf{X}(m)$, $m \in \mathcal{M} \backslash \{0\}$ be constant composition codewords [38] selected such that $X_{k_i}^{\bar{k}_i}(m) \sim \mathsf{Unif}\big[\mathcal{T}_{\lceil P_S(i)n \rceil}\big(\hat{P}_{X|S}(\cdot|i)\big)\big]$, where $\hat{P}_{X|S}$ is such that $\mathcal{T}_{\lceil P_S(i)n \rceil}\big(\hat{P}_{X|S}(\cdot|i)\big)$ is non-empty and $D(\hat{P}_{X|S}||P_{X|S}|P_S) \leq \frac{\eta}{3}$. Denote a realization of $\mathbb{B}_{X,n}$ by $\mathcal{B}_{X,n} := \{\mathbf{x}(m) \in \mathcal{X}^n, m \in \mathcal{M}\}$. Note that for $m \in \mathcal{M} \backslash \{0\}$ and large n, the codeword pair $(\mathbf{x}(0), \mathbf{x}(m))$ has joint type (approx) $P_{\mathbf{x}(0)\mathbf{x}(j)} = \hat{P}_{SX} := P_S \hat{P}_{X|S}$.

Channel encoder: For a given $\mathcal{B}_{X,n}$, the channel encoder outputs $\mathbf{x} = \mathbf{x}(m)$ for output m from the source encoder. Denote this map by $f_{\mathcal{B}_{X,n}} : \mathcal{M} \to \mathcal{X}^n$.

Encoder: Denote by $f_n : \mathcal{U}^n \to \mathcal{P}(\mathcal{X}^n)$ the encoder induced by all the above operations, i.e., $f_n(\cdot|\mathbf{u}) = f_{\mathcal{B}_{X,n}} \circ f_b\big(M'(\mathbf{u}, \mathcal{B}_{W,n})\big)$.

Decision function: The decision function consists of three parts, a channel decoder, a source decoder and a tester.

Channel decoder: The channel decoder first performs a Neyman–Pearson test on the channel output $\mathbf{y}$ according to $\tilde{\Pi}_\theta : \mathcal{Y}^n \to \{0, 1\}$, where

$$\tilde{\Pi}_\theta(\mathbf{y}) := \mathbb{1}\left(\sum_{i=1}^{n} \log\left(\frac{P_{Y|X}(y_i|s_i)}{P_{Y|S}(y_i|s_i)}\right) \geq n\theta\right). \tag{20}$$

If $\tilde{\Pi}_\theta(\mathbf{y}) = 1$, then $\hat{M} = 0$. Else, for a given $\mathcal{B}_{X,n}$, maximum likelihood (ML) decoding is done on the remaining set of codewords $\{\mathbf{x}(m), m \in \mathcal{M} \backslash \{0\}\}$, and $\hat{M}$ is set equal to the ML estimate. Denote the channel decoder induced by the above operations by $g_{\mathcal{B}_{X,n}}$, where $g_{\mathcal{B}_{X,n}} : \mathcal{Y}^n \to \mathcal{M}$.

For a given codebook $\mathcal{B}_{X,n}$, the channel encoder–decoder pair described above induces a distribution

$$P_{\mathbf{XY}\hat{M}|M}^{(\mathcal{B}_{X,n})}(m, \mathbf{x}, \mathbf{y}, \hat{m}|m) := \mathbb{1}_{\left\{f_{\mathcal{B}_{X,n}}(m)=\mathbf{x}\right\}} P_{Y|X}^{\otimes n}(\mathbf{y}|\mathbf{x}) \mathbb{1}_{\left\{\hat{m}=g_{\mathcal{B}_{X,n}}\right\}}.$$

Note that $P_{\mathbf{x}(0)\mathbf{x}(m)} = \hat{P}_{SX}$, $\mathbf{Y} \sim \prod_{i=1}^{|\mathcal{X}|} P_{Y|X}^{\otimes \lceil P_S(i)n \rceil}(\cdot|i)$ for $M = 0$ and $\mathbf{Y} \sim \prod_{i=1}^{|\mathcal{X}|} P_{Y|S}^{\otimes \lceil P_S(i)n \rceil}(\cdot|i)$ for $M = m \neq 0$. Then, it follows by an application of Proposition A1 proved in Appendix C that for any $\mathcal{B}_{X,n}$ and n sufficiently large, the Neyman–Pearson test in (20) yields

$$\mathbb{P}_{P^{(\mathcal{B}_{X,n})}}\big(\hat{M} = 0|M = m\big) \leq e^{-n\big(E_{\mathsf{sp}}(P_{SX},\theta)-\eta\big)}, \ m \in \mathcal{M} \backslash \{0\}, \tag{21a}$$

$$\mathbb{P}_{P^{(\mathcal{B}_{X,n})}}\big(\hat{M} \neq 0|M = 0\big) \leq e^{-n\big(E_{\mathsf{sp}}(P_{SX},\theta)-\theta-\eta\big)}. \tag{21b}$$

Moreover, given $\hat{M} \neq 0$, a random coding argument over the ensemble of $\mathbb{B}_X^n$ (see Exercise 10.18, 10.24 in [38,44]) shows that there exists a deterministic codebook $\mathcal{B}_{X,n}$ such that (21a) and (21b) holds, and the ML decoding described above asymptotically achieves

$$\mathbb{P}_{P^{(\mathcal{B}_{X,n})}}\big(\hat{M} \neq m|M = m \neq 0, \hat{M} \neq 0\big) \leq e^{-n\big(E_{\mathsf{ex}}(R,P_{SX})-\eta\big)}. \tag{22}$$

This deterministic codebook $\mathcal{B}_{X,n}$ is used for channel coding.

Source decoder: For a given codebook $\mathcal{B}_{W,n}$ and inputs $\hat{M} = \hat{m}$ and $\mathbf{V} = \mathbf{v}$, the source decoder first decodes for the quantization codeword $\mathbf{w}(\hat{m}')$ (if required) using the empirical conditional entropy decoder, and then declares the output $\hat{H}$ of the hypothesis test based on $\mathbf{w}(\hat{m}')$ and $\mathbf{v}$. More specifically, if binning is not performed, i.e., if $|\mathcal{M}| \geq |\mathcal{M}'|$, $\hat{M}' = \hat{m}$.

Otherwise, $\hat{M}' = \hat{m}'$, where $\hat{m}' = 0$ if $\hat{m} = 0$ and $\hat{m}' = \arg\min_{j:f_b(j)=\hat{m}} H_e(\mathbf{w}(j)|\mathbf{v})$ otherwise. Denote the source decoder induced by the above operations by $g_{\mathcal{B}_{W,n}} : \mathcal{M} \times \mathcal{V}^n \to \mathcal{M}'$.

Testing and Acceptance region: If $\hat{m}' = 0$, $\hat{H} = 1$ is declared. Otherwise, $\hat{H} = 0$ or $\hat{H} = 1$ is declared depending on whether $(\hat{m}', \mathbf{v}) \in \mathcal{A}_n$ or $(\hat{m}', \mathbf{v}) \notin \mathcal{A}_n$, respectively, where $\mathcal{A}_n$ denotes the acceptance region for H_0 as specified next. For a given codebook $\mathcal{B}_{W,n}$, let $\mathcal{O}_{m'}$ denote the set of $\mathbf{u}$ such that the source encoder outputs m', $m' \in \mathcal{M}'\backslash\{0\}$. For each $m' \in \mathcal{M}'\backslash\{0\}$ and $\mathbf{u} \in \mathcal{O}_{m'}$, let

$$\mathcal{Z}_{m'}(\mathbf{u}) = \{\mathbf{v} \in \mathcal{V}^n : (\mathbf{w}(m'), \mathbf{u}, \mathbf{v}) \in \mathcal{J}_n(\kappa_\alpha + \eta, P_{W_{m'}UV})\},$$

where $\mathcal{J}_n(r, P_X) := \{\mathbf{x} \in \mathcal{X}^n : D(P_{\mathbf{x}}||P_X) \le r\}$,

$$P_{UVW_{m'}} := P_{UV}P_{W_{m'}|U} \text{ and } P_{W_{m'}|U} = \omega(P_{\mathbf{u}}). \tag{23}$$

For $m' \in \mathcal{M}'\backslash\{0\}$, set $\mathcal{Z}_{m'} := \{\mathbf{v} : \mathbf{v} \in \mathcal{Z}_{m'}(\mathbf{u}) \text{ for some } \mathbf{u} \in \mathcal{O}_{m'}\}$, and define the acceptance region for H_0 at the decision maker as $\mathcal{A}_n := \cup_{m' \in \mathcal{M}'\backslash 0} m' \times \mathcal{Z}_{m'}$ or equivalently as $\mathcal{A}_n^e := \cup_{m' \in \mathcal{M}'\backslash 0} \mathcal{O}_{m'} \times \mathcal{Z}_{m'}$. Note that $\mathcal{A}_n$ is the same as the acceptance region for H_0 in Theorem 1 in [14]. Denote the decision function induced by $g_{\mathcal{B}_{X,n}}$, $g_{\mathcal{B}_{W,n}}$ and $\mathcal{A}_n$ by $g_n : \mathcal{Y}^n \times \mathcal{V}^n \to \hat{\mathcal{H}}$.

Induced probability distribution: The PMFs induced by a code $c_n = (f_n, g_n)$ with respect to codebook $\mathcal{B}_n := (\mathcal{B}_{W,n}, f_b, \mathcal{B}_{X,n})$ under H_0 and H_1 are

$$P^{(\mathcal{B}_n, c_n)}_{\mathbf{UV}M'M\mathbf{XY}\hat{M}\hat{M}'\hat{H}}(\mathbf{u}, \mathbf{v}, m', m, \mathbf{x}, \mathbf{y}, \hat{m}, \hat{m}', \hat{h})$$
$$:= P^{\otimes n}_{UV}(\mathbf{u}, \mathbf{v}) \, \mathbb{1}_{\left\{M'\left(\mathbf{u}, \mathcal{B}_{W,n}\right)=m', \, f_b(m')=m\right\}} P^{(\mathcal{B}_{X,n})}_{\mathbf{XY}\hat{M}|M}(\mathbf{x}, \mathbf{y}, \hat{m}|m) \, \mathbb{1}_{\left\{g_{\mathcal{B}_{W,n}}(m, \mathbf{v})=\hat{m}', \, \hat{h}=\mathbb{1}_{\left\{(\hat{m}', \mathbf{v})\in\mathcal{A}_n^c\right\}}\right\}},$$

$$Q^{(\mathcal{B}_n, c_n)}_{\mathbf{UV}M'M\mathbf{XY}\hat{M}\hat{M}'\hat{H}}(\mathbf{u}, \mathbf{v}, m', m, \mathbf{x}, \mathbf{y}, \hat{m}, \hat{m}', \hat{h})$$
$$:= Q^{\otimes n}_{UV}(\mathbf{u}, \mathbf{v}) \, \mathbb{1}_{\left\{M'\left(\mathbf{u}, \mathcal{B}_{W,n}\right)=m', \, f_b(m')=m\right\}} P^{(\mathcal{B}_{X,n})}_{\mathbf{XY}\hat{M}|M}(\mathbf{x}, \mathbf{y}, \hat{m}|m) \, \mathbb{1}_{\left\{g_{\mathcal{B}_{W,n}}(m, \mathbf{v})=\hat{m}', \, \hat{h}=\mathbb{1}_{\left\{(\hat{m}', \mathbf{v})\in\mathcal{A}_n^c\right\}}\right\}},$$

respectively. For simplicity, we will denote the above distributions by $P^{(\mathcal{B}_n)}$ and $Q^{(\mathcal{B}_n)}$. Let $\mathbb{B}_n := (\mathbb{B}_{W,n}, f_{\mathbb{B}}, \mathcal{B}_{X,n})$, $\mathfrak{B}_n$, and μ_n denote the random codebook, its support, and the probability measure induced by its random construction, respectively. Additionally, define $\bar{\mathbb{P}}_{P^{(\mathbb{B}_n)}} := \mathbb{E}_{\mu_n}\left[\mathbb{P}_{P^{(\mathbb{B}_n)}}\right]$ and $\bar{\mathbb{P}}_{Q^{(\mathbb{B}_n)}} := \mathbb{E}_{\mu_n}\left[\mathbb{P}_{Q^{(\mathbb{B}_n)}}\right]$.

Analysis of the type I and type II error probabilities: We analyze the type I and type II error probabilities averaged over the random ensemble of quantization and binning codebooks $(\mathbb{B}_W, f_{\mathbb{B}})$. Then, an expurgation technique [44] guarantees the existence of a sequence of deterministic codebooks $\{\mathcal{B}_n\}_{n\in\mathbb{N}}$ and a code $\{c_n = (f_n, g_n)\}_{n\in\mathbb{N}}$ that achieves the lower bound given in Theorem 1.

Type I error probability: In the following, random sets where the randomness is induced due to $\mathbb{B}_n$ will be written using blackboard bold letters, e.g., $\mathbb{A}_n$ for the random acceptance region for H_0. Note that a type I error can occur only under the following events:

(i) $\mathcal{E}_{\mathrm{EE}} := \bigcup_{P_{\hat{U}} \in \mathcal{D}_n(P_U, \eta)} \bigcup_{\mathbf{u} \in \mathcal{T}_n(P_{\hat{U}})} \mathcal{E}_{\mathrm{EE}}(\mathbf{u})$, where

$$\mathcal{E}_{\mathrm{EE}}(\mathbf{u}) := \{\nexists\, j \in \mathcal{M}'\backslash\{0\} \text{ s.t. } (\mathbf{u}, \mathbf{W}(j)) \in \mathcal{T}_n(P_{\hat{U}_i\hat{W}_i}), P_{\hat{U}_i} = P_{\mathbf{u}},$$
$$P_{\hat{U}_i\hat{W}_i} \in \mathcal{D}_n(P_{UW}, \eta)\},$$

(ii) $\mathcal{E}_{\mathrm{NE}} := \{\hat{M}' = M' \text{ and } (\hat{M}', \mathbf{V}) \notin \mathbb{A}_n\}$,
(iii) $\mathcal{E}_{\mathrm{OCE}} := \{M' \neq 0, \hat{M} \neq M \text{ and } (\hat{M}', \mathbf{V}) \notin \mathbb{A}_n\}$,
(iv) $\mathcal{E}_{\mathrm{SCE}} := \{M' = M = 0, \hat{M} \neq M \text{ and } (\hat{M}', \mathbf{V}) \notin \mathbb{A}_n\}$,
(v) $\mathcal{E}_{\mathrm{BE}} := \{M' \neq 0, \hat{M} = M, \hat{M}' \neq M' \text{ and } (\hat{M}', \mathbf{V}) \notin \mathbb{A}_n\}$.

Here, $\mathcal{E}_{\mathrm{EE}}$ corresponds to the event that there does not exist a quantization codeword corresponding to atleast one sequence $\mathbf{u}$ of type $P_{\mathbf{u}} \in \mathcal{D}_n(P_U, \eta)$; $\mathcal{E}_{\mathrm{NE}}$ corresponds to the event, in which, there is neither an error at the channel decoder nor at the empirical conditional entropy decoder; $\mathcal{E}_{\mathrm{OCE}}$ and $\mathcal{E}_{\mathrm{SCE}}$ corresponds to the case, in which there is an error at the channel decoder (hence also at the empirical conditional entropy decoder); and $\mathcal{E}_{\mathrm{BE}}$ corresponds to the case that there is an error (due to binning) only at the empirical conditional entropy decoder. For the event $\mathcal{E}_{\mathrm{EE}}$, it follows from a slight generalization of the type-covering lemma (Lemma 9.1 in [38]) that

$$\bar{\mathbb{P}}_{P(\mathbb{B}_n)}(\mathcal{E}_{\mathrm{EE}}) \leq e^{-e^{n\Omega(\eta)}}. \tag{24}$$

Since $e^{n\Omega(\eta)}/n \xrightarrow{(n)} \infty$ for $\eta > 0$, the event $\mathcal{E}_{\mathrm{EE}}$ may be safely ignored from the analysis of the error-exponents. Given that $\mathcal{E}_{\mathrm{EE}}^c$ holds for some $\mathcal{B}_{W,n}$, it follows from Equation 4.22 in [14] that

$$\bar{\mathbb{P}}_{P(\mathbb{B}_n)}(\mathcal{E}_{\mathrm{NE}}|\mathcal{E}_{\mathrm{EE}}^c) \leq e^{-n\kappa_\alpha}, \tag{25}$$

for sufficiently large n since the acceptance region is the same as that in Theorem 1 in [14].

Next, consider the event $\mathcal{E}_{\mathrm{OCE}}$. We have for sufficiently large n that

$$
\begin{aligned}
\bar{\mathbb{P}}_{P(\mathbb{B}_n)}(\mathcal{E}_{\mathrm{OCE}}) &\leq \bar{\mathbb{P}}_{P(\mathbb{B}_n)}(M' \neq 0)\bar{\mathbb{P}}_{P(\mathbb{B}_n)}(\hat{M} \neq M|M' \neq 0) \\
&\overset{(a)}{\leq} \bar{\mathbb{P}}_{P(\mathbb{B}_n)}(\hat{M} \neq M|M \neq 0) \\
&\leq \bar{\mathbb{P}}_{P(\mathbb{B}_n)}(\hat{M} = 0|M \neq 0) + \bar{\mathbb{P}}_{P(\mathbb{B}_n)}(\hat{M} \neq M|M \neq 0, \hat{M} \neq 0) \\
&\overset{(b)}{\leq} e^{-n(E_{\mathrm{m}}(P_{SX}, \theta) - \eta)} + e^{-n(E_{\mathrm{ex}}(R, P_{SX}) - \eta)} \\
&= e^{-n(\min\{E_{\mathrm{m}}(P_{SX}, \theta), E_{\mathrm{ex}}(R, P_{SX})\} - \eta)},
\end{aligned}
\tag{26}
$$

where

(a)   holds since the event $\{M' \neq 0\}$ is equivalent to $\{M \neq 0\}$;
(b)   holds due to (21a) and (22), which holds for $\mathcal{B}_{X,n}$.

Additionally, the probability of $\mathcal{E}_{\mathrm{SCE}}$ can be upper bounded as

$$
\begin{aligned}
\bar{\mathbb{P}}_{P(\mathbb{B}_n)}(\mathcal{E}_{\mathrm{SCE}}) &\leq \bar{\mathbb{P}}_{P(\mathbb{B}_n)}(M' = 0) \\
&\leq \bar{\mathbb{P}}_{P(\mathbb{B}_n)}(M' = 0|\mathbf{U} \in \mathcal{D}_n(P_U, \eta)) + \bar{\mathbb{P}}_{P(\mathbb{B}_n)}(\mathbf{U} \notin \mathcal{D}_n(P_U, \eta)) \\
&= \bar{\mathbb{P}}_{P(\mathbb{B}_n)}(\mathcal{E}_{\mathrm{EE}}) + \bar{\mathbb{P}}_{P(\mathbb{B}_n)}(\mathbf{U} \notin \mathcal{D}_n(P_U, \eta)) \\
&\leq e^{-n\kappa_\alpha},
\end{aligned}
\tag{27}
$$

where (27) is due to (24), the definition of $\mathcal{D}_n(P_U, \eta)$ in (15) and Lemma 2.2 and Lemma 2.6 in [38].

Finally, consider the event $\mathcal{E}_{\mathrm{BE}}$. Note that this event occurs only when $|\mathcal{M}| \leq |\mathcal{M}'|$. Additionally, $M = 0$ iff $M' = 0$, and hence $M' \neq 0$ and $\hat{M} = M$ implies that $\hat{M} \neq 0$. Let

$$\mathcal{D}_n(P_{VW}, \eta) := \left\{ P_{\hat{V}\hat{W}} : \begin{array}{l} \exists\, (\mathbf{w}, \mathbf{u}, \mathbf{v}) \in \underset{m' \in \mathcal{M}'\backslash\{0\}}{\cup} \mathcal{J}_n(\kappa_\alpha + \eta, P_{W_{m'}UV}), P_{W_{m'}UV} \text{ satisfies} \\ (23) \text{ and } P_{\mathbf{wuv}} = P_{\hat{W}\hat{U}\hat{V}} \end{array} \right\}.$$

We have

$$\bar{\mathbb{P}}_{P(\mathbb{B}_n)}(\mathcal{E}_{\mathrm{BE}}) = \bar{\mathbb{P}}_{P(\mathbb{B}_n)}(\mathcal{E}_{\mathrm{BE}}, (M', \mathbf{V}) \in \mathbb{A}_n) + \bar{\mathbb{P}}_{P(\mathbb{B}_n)}(\mathcal{E}_{\mathrm{BE}}, (M', \mathbf{V}) \notin \mathbb{A}_n). \tag{28}$$

The second term in (28) can be upper bounded as

$$\bar{\mathbb{P}}_{P(\mathbb{B}_n)}(\mathcal{E}_{\mathrm{BE}}, (M', \mathbf{V}) \notin \mathbb{A}_n) \leq \bar{\mathbb{P}}_{P(\mathbb{B}_n)}((M', \mathbf{V}) \notin \mathbb{A}_n, \mathcal{E}_{\mathrm{EE}}) + \bar{\mathbb{P}}_{P(\mathbb{B}_n)}((M', \mathbf{V}) \notin \mathbb{A}_n, \mathcal{E}_{\mathrm{EE}}^c)$$

$$\leq e^{-e^{n\Omega(\eta)}} + \bar{\mathbb{P}}_{P(\mathbb{B}_n)}\big((M', \mathbf{V}) \notin \mathbb{A}_n | \mathcal{E}_{\text{EE}}^c\big)$$

$$\leq e^{-e^{n\Omega(\eta)}} + \bar{\mathbb{P}}_{P(\mathbb{B}_n)}\big((\mathbf{U}, \mathbf{V}) \notin \mathbb{A}_n^e\big)$$

$$\leq e^{-e^{n\Omega(\eta)}} + e^{-n\kappa_\alpha}, \tag{29}$$

where the inequality in (29) follows from Equation (4.22) in [14] for sufficiently large n since the acceptance region $\mathbb{A}_n^e$ is the same as that in [14]. To bound the first term in (28), define $\mathcal{D}_n(P_V, \eta) := \{P_{\hat{V}} : \exists\, P_{\hat{V}\hat{W}} \in \mathcal{D}_n(P_{VW}, \eta)\}$, and observe that since $(M', \mathbf{V}) \in \mathbb{A}_n$ implies $M' \neq 0$, we have

$$\bar{\mathbb{P}}_{P(\mathbb{B}_n)}\big(\mathcal{E}_{\text{BE}}, (M', \mathbf{V}) \in \mathbb{A}_n\big)$$

$$= \sum_{(m',m) \in \mathcal{M}' \times \mathcal{M}} \bar{\mathbb{P}}_{P(\mathbb{B}_n)}\big(\mathcal{E}_{\text{BE}}, (M', \mathbf{V}) \in \mathbb{A}_n, M = m, M' = m'\big)$$

$$= \sum_{(m',m) \in \mathcal{M}' \times \mathcal{M}} \bar{\mathbb{P}}_{P(\mathbb{B}_n)}\big(M = m, M' = m', \hat{M} = M\big)$$

$$\qquad\qquad \bar{\mathbb{P}}_{P(\mathbb{B}_n)}\big(\hat{M}' \neq M', (\hat{M}', \mathbf{V}) \notin \mathbb{A}_n, (M', \mathbf{V}) \in \mathbb{A}_n | M' = m', M = m, \hat{M} = M\big)$$

$$\leq \sum_{(m',m) \in \mathcal{M}' \times \mathcal{M}} \bar{\mathbb{P}}_{P(\mathbb{B}_n)}\big(M = m, M' = m', \hat{M} = M\big)$$

$$\qquad\qquad \bar{\mathbb{P}}_{P(\mathbb{B}_n)}\big(\hat{M}' \neq M', (M', \mathbf{V}) \in \mathbb{A}_n | M' = m', M = m, \hat{M} = M\big) \tag{30}$$

$$\overset{(a)}{=} \bar{\mathbb{P}}_{P(\mathbb{B}_n)}\big(\hat{M}' \neq M', (M', \mathbf{V}) \in \mathbb{A}_n | M' = 1, M = 1, \hat{M} = M\big)$$

$$\overset{(b)}{\leq} \sum_{P_{\mathbf{v}} \in \mathcal{D}_n(P_V, \eta)} \sum_{\mathbf{v} \in P_{\mathbf{v}}} \bar{\mathbb{P}}_{P(\mathbb{B}_n)}\big(\mathbf{V} = \mathbf{v} | M' = 1\big)$$

$$\qquad\qquad \bar{\mathbb{P}}_{P(\mathbb{B}_n)}\big(\exists\, j \in f_{\mathbb{B}}^{-1}(1),\ j \neq 1,\ H_e(\mathbf{W}(j)|\mathbf{v}) \leq H_e(\mathbf{W}(1)|\mathbf{v}) | M' = 1, \mathbf{V} = \mathbf{v}\big), \tag{31}$$

where (a) follows since by the symmetry of the source encoder, binning function and random codebook construction, the term in (30) is independent of (m, m'); (b) holds since $(M', \mathbf{V}) \in \mathbb{A}_n$ implies that $P_{\mathbf{v}} \in \mathcal{D}_n(P_V, \eta)$ and $(\mathbf{V}, \mathbb{B}_W) - M' - (M, \hat{M})$ form a Markov chain. Defining $P_{\hat{V}} = P_{\mathbf{v}}$, and the event $\mathcal{E}_1' := \{M' = 1, \mathbf{V} = \mathbf{v}\}$, we obtain

$$\bar{\mathbb{P}}_{P(\mathbb{B}_n)}\Big(\exists\, j \in f_{\mathbb{B}}^{-1}(1),\ j \neq 1,\ H_e(\mathbf{W}(j)|\mathbf{v}) \leq H_e(\mathbf{W}(1)|\mathbf{v}) \,\big|\, \mathcal{E}_1'\Big)$$

$$= \sum_{j \in \mathcal{M}' \setminus \{0,1\}} \bar{\mathbb{P}}_{P(\mathbb{B}_n)}\big(f_{\mathbb{B}}(j) = 1, H_e(\mathbf{W}(j)|\mathbf{v}) \leq H_e(\mathbf{W}(1)|\mathbf{v}) \,\big|\, \mathcal{E}_1'\big)$$

$$\overset{(a)}{\leq} \frac{1}{e^{nR_n}} \sum_{j \in \mathcal{M}' \setminus \{0,1\}} \bar{\mathbb{P}}_{P(\mathbb{B}_n)}\big(H_e(\mathbf{W}(j)|\mathbf{v}) \leq H_e(\mathbf{W}(1)|\mathbf{v}) \,\big|\, \mathcal{E}_1'\big)$$

$$\overset{(b)}{\leq} \frac{1}{e^{nR_n}} \sum_{j \in \mathcal{M}' \setminus \{0,1\}} \sum_{P_{\hat{W}}:P_{\hat{V}\hat{W}} \in \mathcal{D}_n(P_{VW}, \eta)} \sum_{\mathbf{w}:(\mathbf{v},\mathbf{w}) \in \mathcal{T}_n(P_{\hat{V}\hat{W}})} \bar{\mathbb{P}}_{P(\mathbb{B}_n)}\big(\mathbf{W}(1) = \mathbf{w} \,\big|\, \mathcal{E}_1'\big)$$

$$\qquad\qquad \sum_{\tilde{\mathbf{w}} \in \mathcal{T}_n(P_{\hat{W}}):H_e(\tilde{\mathbf{w}}|\mathbf{v}) \leq H(\hat{W}|\hat{V})} \bar{\mathbb{P}}_{P(\mathbb{B}_n)}\big(\mathbf{W}(j) = \tilde{\mathbf{w}} \,\big|\, \mathcal{E}_1' \cup \{\mathbf{W}(1) = \mathbf{w}\}\big)$$

$$\overset{(c)}{\leq} \frac{1}{e^{nR_n}} \sum_{j \in \mathcal{M}' \setminus \{0,1\}} \sum_{P_{\hat{W}}:P_{\hat{V}\hat{W}} \in \mathcal{D}_n(P_{VW}, \eta)} \sum_{\mathbf{w}:(\mathbf{v},\mathbf{w}) \in \mathcal{T}_n(P_{\hat{V}\hat{W}})} \bar{\mathbb{P}}_{P(\mathbb{B}_n)}\big(\mathbf{W}(1) = \mathbf{w} \,\big|\, \mathcal{E}_1'\big)$$

$$\qquad\qquad \sum_{\tilde{\mathbf{w}} \in \mathcal{T}_n(P_{\hat{W}}):H_e(\tilde{\mathbf{w}}|\mathbf{v}) \leq H(\hat{W}|\hat{V})} 2\, \bar{\mathbb{P}}_{P(\mathbb{B}_n)}\big(\mathbf{W}(j) = \tilde{\mathbf{w}}\big), \tag{32}$$

where

(a)   follows since $f_{\mathbb{B}}(\cdot)$ is the uniform binning function independent of $\mathbb{B}_{W,n}$;

(b)   holds due to the fact that if $P_{\mathbf{v}} \in \mathcal{D}_n(P_V, \eta)$, then $M' = 1$ implies that $(\mathbf{W}(1), \mathbf{v}) \in \mathcal{T}_n(P_{\hat{V}\hat{W}})$ with probability one for some $P_{\hat{V}\hat{W}} \in \mathcal{D}_n(P_{VW}, \eta)$;

(c)   holds since $\bar{\mathbb{P}}_{P(\mathbb{B}_n)}\big(\mathbf{W}(j) = \tilde{\mathbf{w}} \,\big|\, \mathcal{E}_1' \cup \{\mathbf{W}(1) = \mathbf{w}\}\big) \leq 2\, \bar{\mathbb{P}}_{P(\mathbb{B}_n)}\big(\mathbf{W}(j) = \tilde{\mathbf{w}}\big)$, which follows similarly to Equation (101) in [10].

Continuing, we can write for sufficiently large n,

$$\bar{\mathbb{P}}_{P(\mathbb{B}_n)}\left(\exists\, j \in f_{\mathbb{B}}^{-1}(1),\, j \neq 1, H_e(\mathbf{W}(j)|\mathbf{v}) \leq H_e(\mathbf{W}(1)|\mathbf{v}) \mid \mathcal{E}_1'\right)$$

$$\overset{(a)}{\leq} \frac{1}{e^{nR_n}} \sum_{j \in \mathcal{M}' \setminus \{0,1\}} \sum_{P_{\hat{W}}:P_{\hat{V}\hat{W}} \in \mathcal{D}_n(P_{VW},\eta)} \sum_{\mathbf{w}:(\mathbf{v},\mathbf{w}) \in \mathcal{T}_n(P_{\hat{V}\hat{W}})} \bar{\mathbb{P}}_{P(\mathbb{B}_n)}\left(\mathbf{W}(1) = \mathbf{w} \mid \mathcal{E}_1'\right)$$
$$\sum_{\tilde{\mathbf{w}} \in \mathcal{T}_n(P_{\hat{W}}):H_e(\tilde{\mathbf{w}}|\mathbf{v}) \leq H(\hat{W}|\hat{V})} 2\, e^{-n(H(\hat{W})-\eta)}$$

$$\overset{(b)}{\leq} \frac{1}{e^{nR_n}} \sum_{j \in \mathcal{M}' \setminus \{0,1\}} \sum_{P_{\hat{W}}:P_{\hat{V}\hat{W}} \in \mathcal{D}_n(P_{VW},\eta)} \sum_{\mathbf{w}:(\mathbf{v},\mathbf{w}) \in \mathcal{T}_n(P_{\hat{V}\hat{W}})} \bar{\mathbb{P}}_{P(\mathbb{B}_n)}\left(\mathbf{W}(1) = \mathbf{w} \mid \mathcal{E}_1'\right)$$
$$(n+1)^{|\mathcal{V}||\mathcal{W}|} e^{nH(\hat{W}|\hat{V})} 2\, e^{-n(H(\hat{W})-\eta)}$$

$$\leq \frac{1}{e^{nR_n}} \sum_{j \in \mathcal{M}' \setminus \{0,1\}} \sum_{P_{\hat{W}}:P_{\hat{V}\hat{W}} \in \mathcal{D}_n(P_{VW},\eta)} 2\,(n+1)^{|\mathcal{V}||\mathcal{W}|}\, e^{-n\left(I(\hat{W};\hat{V})-\eta\right)}$$

$$\overset{(c)}{\leq} \frac{1}{e^{nR_n}} \sum_{j \in \mathcal{M}' \setminus \{0,1\}} 2\,(n+1)^{|\mathcal{W}|}\,(n+1)^{|\mathcal{V}||\mathcal{W}|}\, e^{-n\left(\min_{P_{\hat{V}\hat{W}} \in \mathcal{D}_n(P_{VW},\eta)} I(\hat{W};\hat{V})-\eta\right)}$$

$$\overset{(d)}{\leq} e^{-n(R-R'+\rho_n-\eta_n')}, \tag{33}$$

where $\rho_n := \min_{P_{\hat{V}\hat{W}} \in \mathcal{D}_n(P_{VW},\eta)} I(\hat{V};\hat{W})$ and $\eta_n' := 3\eta + o(1)$. In the above,

(a) used Lemma 2.3 in [38] and the fact that the codewords are chosen uniformly at random from $\mathcal{T}_n(P_{\hat{W}})$;

(b) follows since the total number of sequences $\tilde{\mathbf{w}} \in \mathcal{T}_n(P_{\hat{W}})$ such that $P_{\tilde{\mathbf{w}}\mathbf{v}} = P_{\tilde{W}\tilde{V}}$ and $H(\tilde{W}|\tilde{V}) \leq H(\hat{W}|\hat{V})$ is upper bounded by $e^{nH(\hat{W}|\hat{V})}$, and $|\mathcal{T}(\mathcal{W}^n \times \mathcal{V}^n)| \leq (n+1)^{|\mathcal{V}||\mathcal{W}|}$;

(c) holds due to Lemma 2.2 in [38];

(d) follows from $R' := \zeta(\kappa_\alpha$, (14), (18) and (19).

Thus, since $\rho_n \to \rho(\kappa_\alpha,\omega) + O(\eta)$, we have from (28), (29), (31), (33) for large enough n that

$$\bar{\mathbb{P}}_{P(\mathbb{B}_n)}(\mathcal{E}_{\text{BE}}) \leq e^{-n(\min\{\kappa_\alpha, R-\zeta(\kappa_\alpha,\omega)+\rho(\kappa_\alpha,\omega)-O(\eta)\})}. \tag{34}$$

By choice of $(\omega, P_{SX}, \theta) \in \mathcal{L}(\kappa_\alpha)$, it follows from (24), (25), (26), (27) and (34) that the type I error probability is upper bounded by $e^{-n(\kappa_\alpha-O(\eta))}$ for large n.

Type II error probability: We analyze the type II error probability averaged over $\mathbb{B}_n$. A type II error can occur only under the following events:

(i) $\mathcal{E}_a := \left\{ \begin{array}{l} \hat{M} = M, \hat{M}' = M' \neq 0, (\mathbf{U}, \mathbf{V}, \mathbf{W}(M')) \in \mathcal{T}_n\left(P_{\hat{U}\hat{V}\hat{W}}\right) \\ \text{s.t. } P_{\hat{U}\hat{W}} \in \mathcal{D}_n(P_{UW},\eta) \text{ and } P_{\hat{V}\hat{W}} \in \mathcal{D}_n(P_{VW},\eta) \end{array} \right\}$,

(ii) $\mathcal{E}_b := \left\{ \begin{array}{l} M' \neq 0, \hat{M} = M, \hat{M}' \neq M', f_{\mathbb{B}}(\hat{M}') = f_{\mathbb{B}}(M'), (\mathbf{U}, \mathbf{V}, \mathbf{W}(M'), \\ \mathbf{W}(\hat{M}')) \in \mathcal{T}_n\left(P_{\hat{U}\hat{V}\hat{W}\hat{W}_d}\right) \text{ s.t. } P_{\hat{U}\hat{W}} \in \mathcal{D}_n(P_{UW},\eta), \\ P_{\hat{V}\hat{W}_d} \in \mathcal{D}_n(P_{VW},\eta) \text{ and } H_e(\mathbf{W}(\hat{M}')|\mathbf{V}) \leq H_e(\mathbf{W}(M')|\mathbf{V}) \end{array} \right\}$,

(iii) $\mathcal{E}_c := \left\{ \begin{array}{l} M' \neq 0, \hat{M} \neq M \text{ or } 0, (\mathbf{U}, \mathbf{V}, \mathbf{W}(M'), \mathbf{W}(\hat{M}')) \in \mathcal{T}_n\left(P_{\hat{U}\hat{V}\hat{W}\hat{W}_d}\right) \text{ s.t.} \\ P_{\hat{U}\hat{W}} \in \mathcal{D}_n(P_{UW},\eta) \text{ and } P_{\hat{V}\hat{W}_d} \in \mathcal{D}_n(P_{VW},\eta) \end{array} \right\}$,

(iv) $\mathcal{E}_d := \left\{ M = M' = 0, \hat{M} \neq M, (\mathbf{V}, \mathbf{W}(\hat{M}')) \in \mathcal{T}_n\left(P_{\hat{V}\hat{W}_d}\right) \text{ s.t. } P_{\hat{V}\hat{W}_d} \in \mathcal{D}_n(P_{VW},\eta) \right\}$.

Similar to (24), it follows that $\bar{\mathbb{P}}_{Q^{(\mathbb{B}_n)}}(\mathcal{E}_{EE}) \leq e^{-e^{n\Omega(\eta)}}$. Hence, we may assume that $\mathcal{E}_{EE}^c$ holds for the type II error-exponent analysis. It then follows from the analysis in Equations (4.23)–(4.27) in [14] that for sufficiently large n,

$$\bar{\mathbb{P}}_{Q^{(\mathbb{B}_n)}}(\mathcal{E}_a | \mathcal{E}_{EE}^c) \leq e^{-n(E_1(\kappa_\alpha,\omega)-O(\eta))}.$$

The analysis of the error events $\mathcal{E}_b$, $\mathcal{E}_c$ and $\mathcal{E}_d$ follows similarly to that in the proof of Theorem 2 in [10], and results in

$$-\frac{1}{n}\log\left(\bar{\mathbb{P}}_{Q^{(\mathbb{B}_n)}}(\mathcal{E}_b)\right)$$

$$\gtrsim \begin{cases} \min\limits_{(P_{\tilde{U}\tilde{V}\tilde{W}},Q_{\tilde{U}\tilde{V}\tilde{W}})\in\mathcal{T}_2(\kappa_\alpha,\omega)} D(P_{\tilde{U}\tilde{V}\tilde{W}}||Q_{\tilde{U}\tilde{V}\tilde{W}}) + E_b(\kappa_\alpha,\omega,R) - O(\eta), & \text{if } R < \zeta(\kappa_\alpha,\omega) + \eta, \\ \infty, & \text{otherwise}, \end{cases}$$

$$= E_2(\kappa_\alpha,\omega,R) - O(\eta).$$

$$\frac{-1}{n}\log\left(\bar{\mathbb{P}}_{Q^{(\mathbb{B}_n)}}(\mathcal{E}_c)\right)$$

$$\gtrsim \begin{cases} \min\limits_{(P_{\tilde{U}\tilde{V}\tilde{W}},Q_{\tilde{U}\tilde{V}\tilde{W}})\in\mathcal{T}_3(\kappa_\alpha,\omega)} D(P_{\tilde{U}\tilde{V}\tilde{W}}||Q_{\tilde{U}\tilde{V}\tilde{W}}) + E_b(\kappa_\alpha,\omega,R) & \text{if } R < \zeta(\kappa_\alpha,\omega) + \eta \\ \qquad\qquad +E_{ex}(R,P_{SX}) - O(\eta), & \\ \min\limits_{(P_{\tilde{U}\tilde{V}\tilde{W}},Q_{\tilde{U}\tilde{V}\tilde{W}})\in\mathcal{T}_3(\kappa_\alpha,\omega)} D(P_{\tilde{U}\tilde{V}\tilde{W}}||Q_{\tilde{U}\tilde{V}\tilde{W}}) + \rho(\kappa_\alpha,\omega) & \text{otherwise}, \\ \qquad\qquad +E_{ex}(R,P_{SX}) - O(\eta), & \end{cases}$$

$$= E_3(\kappa_\alpha,\omega,R,P_{SX}) - O(\eta).$$

$$\frac{-1}{n}\log\left(\bar{\mathbb{P}}_{Q^{(\mathbb{B}_n)}}(\mathcal{E}_d)\right)$$

$$\gtrsim \begin{cases} \min\limits_{P_{\tilde{V}}:P_{\tilde{V}\tilde{W}}\in\mathcal{D}_n(P_{VW},\eta)} D(P_{\tilde{V}}||Q_V) + E_b(\kappa_\alpha,\omega,R) & \text{if } R < \zeta(\kappa_\alpha,\omega) + \eta, \\ \qquad\qquad +E_{sp}(P_{SX},\theta) - \theta - O(\eta), & \\ \min_{P_{\tilde{V}}:P_{\tilde{V}\tilde{W}}\in\mathcal{D}_n(P_{VW},\eta)} D(P_{\tilde{V}}||Q_V) + \rho(\kappa_\alpha,\omega) & \text{otherwise}, \\ \qquad\qquad +E_{sp}(P_{SX},\theta) - \theta - O(\eta), & \end{cases}$$

$$= E_4(\kappa_\alpha,\omega,R,P_{SX},\theta) - O(\eta).$$

Since the exponent of the type II error probability is lower bounded by the minimum of the exponent of the type II error-causing events, we have shown from the above that for a fixed $(\omega,R,P_{SX},\theta) \in \mathcal{L}(\kappa_\alpha)$ and sufficiently large n,

$$\bar{\mathbb{P}}_{P^{(\mathbb{B}_n)}}(\hat{H}=1) \leq e^{-n(\kappa_\alpha-O(\eta))}, \tag{35a}$$

$$\bar{\mathbb{P}}_{Q^{(\mathbb{B}_n)}}(\hat{H}=0) \leq e^{-n(\bar{\kappa}_s(\kappa_\alpha,\omega,R,P_{SX},\theta)-O(\eta))}, \tag{35b}$$

where

$$\bar{\kappa}_s(\kappa_\alpha,\omega,R,P_{SX},\theta) := \min\left\{E_1(\kappa_\alpha,\omega), E_2(\kappa_\alpha,\omega,R), E_3(\kappa_\alpha,\omega,R,P_{SX}), E_4(\kappa_\alpha,\omega,R,P_{SX},\theta)\right\}.$$

Expurgation: To complete the proof, we extract a deterministic codebook $\mathcal{B}_n^\star$ that satisfies

$$\mathbb{P}_{P^{(\mathcal{B}_n^\star)}}(\hat{H}=1) \leq e^{-n(\kappa_\alpha-O(\eta))},$$

$$\mathbb{P}_{Q^{(\mathcal{B}_n^\star)}}(\hat{H}=0) \leq e^{-n(\bar{\kappa}_s(\kappa_\alpha,\omega,R,P_{SX},\theta)-O(\eta))}.$$

For this purpose, remove a set $\mathfrak{B}_n' \subset \mathfrak{B}_n$ of highest type I error probability codebooks such that the remaining set $\mathfrak{B}_n \backslash \mathfrak{B}_n'$ has a probability of $\tau \in (0.25, 0.5)$, i.e., $\mu_n(\mathfrak{B}_n \backslash \mathfrak{B}_n') = \tau$. Then, it follows from (35a) and (35b) that for all $\mathcal{B}_n \in \mathfrak{B}_n \backslash \mathfrak{B}_n'$,

$$\mathbb{P}_{P^{(\mathcal{B}_n)}}(\hat{H}=1) \leq 2e^{-n(\kappa_\alpha-O(\eta))},$$

$$\tilde{\mathbb{P}}_{Q^{(\mathbb{B}_n)}}\left(\hat{H}=0\right) \leq 4e^{-n\left(\bar{\kappa}_s(\kappa_\alpha,\omega,R,P_{SX},\theta)-O(\eta)\right)},$$

where $\tilde{\mathbb{P}}_{Q^{(\mathbb{B}_n)}} = \frac{1}{\tau}\mathbb{E}_{\mu_n}\left[\mathbb{P}_{Q^{\mathbb{B}_n}}\mathbb{1}_{\{\mathbb{B}_n \in \mathfrak{B}_n \setminus \mathfrak{B}_n'\}}\right]$ is a PMF. Perform one more similar expurgation step to obtain $\mathcal{B}_n^\star = (\mathcal{B}_{W,n}^\star, f_b^\star, \mathcal{B}_{X,n}^\star) \in \mathfrak{B}_n \setminus \mathfrak{B}_n'$ such that for all sufficiently large n

$$\mathbb{P}_{P(\mathcal{B}_n^\star)}\left(\hat{H}=1\right) \leq 2e^{-n(\kappa_\alpha - O(\eta))} \leq e^{-n\left(\kappa_\alpha - O(\eta) - (\log 2/n)\right)},$$

$$\mathbb{P}_{Q(\mathcal{B}_n^\star)}\left(\hat{H}=0\right) \leq 4e^{-n\left(\bar{\kappa}_s(\kappa_\alpha,\omega,R,P_{SX},\theta)-O(\eta)\right)} \leq e^{-n\left(\bar{\kappa}_s(\kappa_\alpha,\omega,R,P_{SX},\theta)-O(\eta)-(\log 4/n)\right)}.$$

Maximizing over $(\omega, R, P_{SX}, \theta) \in \mathcal{L}(\kappa_\alpha)$ and noting that $\eta > 0$ is arbitrary completes the proof.

4.2. Proof of Corollary 1

Consider $(\omega, P_{SX}, \theta) \in \mathcal{L}^\star(\kappa_\alpha)$ and $R = \zeta(\kappa_\alpha, \omega)$. Then, $(\omega, R, P_{SX}, \theta) \in \mathcal{L}(\kappa_\alpha)$. Additionally, for any $(P_{\tilde{U}\tilde{V}\tilde{W}}, Q_{\tilde{U}\tilde{V}\tilde{W}}) \in \mathcal{T}_1(\kappa_\alpha, \omega)$, we have

$$\begin{aligned}
D(P_{\tilde{U}\tilde{V}\tilde{W}}||Q_{\tilde{U}\tilde{V}\tilde{W}}) &= D(P_{\tilde{U}\tilde{W}}||Q_{\tilde{U}\tilde{W}}) + D\left(P_{\tilde{V}|\tilde{U}\tilde{W}}||Q_{\tilde{V}|\tilde{U}\tilde{W}}|P_{\tilde{U}\tilde{W}}\right) \\
&\overset{(a)}{\geq} D\left(P_{\tilde{V}|\tilde{U}\tilde{W}}||P_V|P_{\tilde{U}\tilde{W}}\right) \\
&= D\left(P_{\tilde{V}\tilde{U}\tilde{W}}||P_V P_{\tilde{U}\tilde{W}}\right) \\
&\overset{(b)}{\geq} D\left(P_{\tilde{V}\tilde{W}}||P_V P_{\tilde{W}}\right) \\
&\overset{(c)}{=} D\left(P_{\hat{V}\hat{W}}||P_V P_{\hat{W}}\right) \\
&= I_P(\hat{V};\hat{W}) + D(P_{\hat{V}}||P_V),
\end{aligned} \tag{36}$$

where (a) is due to the non-negativity of KL divergence and since $Q_{\tilde{V}|\tilde{U}\tilde{W}} = P_V$; (b) is because of the monotonicity of KL divergence Theorem 2.14 in [43]; (c) follows since for $(P_{\tilde{U}\tilde{V}\tilde{W}}, Q_{\tilde{U}\tilde{V}\tilde{W}}) \in \mathcal{T}_1(\kappa_\alpha, \omega)$, $P_{\tilde{V}\tilde{W}} = P_{\hat{V}\hat{W}}$ for some $P_{\hat{U}\hat{V}\hat{W}} \in \hat{\mathcal{L}}(\kappa_\alpha, \omega)$. Minimizing over all $P_{\hat{U}\hat{V}\hat{W}} \in \hat{\mathcal{L}}(\kappa_\alpha, \omega)$ yields that

$$\begin{aligned}
E_1(\kappa_\alpha, \omega) &= \min_{(P_{\tilde{U}\tilde{V}\tilde{W}}, Q_{\tilde{U}\tilde{V}\tilde{W}}) \in \mathcal{T}_1(\kappa_\alpha, \omega)} D(P_{\tilde{U}\tilde{V}\tilde{W}}||Q_{\tilde{U}\tilde{V}\tilde{W}}) \\
&\geq \min_{P_{\hat{U}\hat{V}\hat{W}} \in \hat{\mathcal{L}}(\kappa_\alpha, \omega)} \left[I_P(\hat{V};\hat{W}) + D(P_{\hat{V}}||P_V)\right] \\
&= \min_{P_{\hat{V}\hat{W}}: P_{\hat{U}\hat{V}\hat{W}} \in \hat{\mathcal{L}}(\kappa_\alpha, \omega)} \left[I_P(\hat{V};\hat{W}) + D(P_{\hat{V}}||P_V)\right] := E_1^{\mathrm{i}}(\kappa_\alpha, \omega),
\end{aligned}$$

where the inequality above follows from (36). Next, since $\zeta(\kappa_\alpha, \omega) = R$, we have that $E_2(\kappa_\alpha, \omega, R) = \infty$. Additionally, by the non-negativity of KL divergence

$$\begin{aligned}
E_3(\kappa_\alpha, \omega, R, P_{SX}) &= \min_{\substack{(P_{\tilde{U}\tilde{V}\tilde{W}}, Q_{\tilde{U}\tilde{V}\tilde{W}}) \\ \in \mathcal{T}_3(\kappa_\alpha, \omega)}} D(P_{\tilde{U}\tilde{V}\tilde{W}}||Q_{\tilde{U}\tilde{V}\tilde{W}}) + \rho(\kappa_\alpha, \omega) + E_{\mathrm{ex}}(R, P_{SX}) \\
&\geq \rho(\kappa_\alpha, \omega) + E_{\mathrm{ex}}(\zeta(\kappa_\alpha, \omega), P_{SX}) := E_2^{\mathrm{i}}(\kappa_\alpha, \omega, P_{SX}), \\
E_4(\kappa_\alpha, \omega, P_{SX}, \theta) &= \min_{P_{\hat{V}}: P_{\hat{U}\hat{V}\hat{W}} \in \hat{\mathcal{L}}(\kappa_\alpha, \omega)} D(P_{\hat{V}}||P_V) + \rho(\kappa_\alpha, \omega) + E_{\mathrm{m}}(P_{SX}, \theta) - \theta \\
&= \rho(\kappa_\alpha, \omega) + E_{\mathrm{m}}(P_{SX}, \theta) - \theta := E_3^{\mathrm{i}}(\kappa_\alpha, \omega, P_{SX}, \theta),
\end{aligned}$$

where the final equality is since $P_{UV}P_{W|U} \in \hat{\mathcal{L}}(\kappa_\alpha, \omega)$ for $P_{W|U} := \omega(P_U)$. The claim in (8) now follows from Theorem 1.

Next, we prove (10). Note that $\hat{\mathcal{L}}(0, \omega) = \{P_{UVW} = P_{UV}P_{W|U} : P_{W|U} = \omega(P_U)\}$ and $\mathcal{L}^\star(0) = \{(\omega, P_{SX}, \theta) \in \mathfrak{F} \times \mathcal{P}(\mathcal{S} \times \mathcal{X}) \times \Theta(P_{SX}) : I_P(U;W) < I_P(X;Y|S), P_{W|U} = $

$\omega(P_U)$, $P_{SXY} := P_{SX}P_{Y|X}\}$ since $E_{sp}(P_{SX}, \theta) \geq 0$ and $E_{ex}(I_P(U; W), P_{SX}) \geq 0$. Hence, we have

$$E_1^i(0, \omega) \geq \min_{P_{\hat{U}\hat{V}\hat{W}} \in \hat{\mathcal{L}}(0, \omega)} I_P(\hat{V}; \hat{W}) = I_P(V; W).$$

Additionally, $\rho(0, \omega) = I_P(V; W)$, $E_2^i(0, \omega, P_{SX}) \geq \rho(0, \omega)$ and $E_3^i(0, \omega, P_{SX}, \theta) \geq \rho(0, \omega)$. By choosing $P_{XS} = P_X^\star P_S$ where $P_X^\star$ is the capacity achieving input distribution, we have $I_P(X; Y|S) = C$. Then, it follows from (8) and the continuity of $E_1^i(\kappa_\alpha, \omega)$, $E_2^i(\kappa_\alpha, \omega, P_{SX})$ and $E_3^i(\kappa_\alpha, \omega, P_{SX}, \theta)$ in κ_α that $\lim_{\kappa_\alpha \to 0} \kappa(\kappa_\alpha) \geq \kappa_i^\star(0)$. On the other hand, $\lim_{\kappa_\alpha \to 0} \kappa(\kappa_\alpha) \leq \kappa_i^\star(0)$ follows from the converse proof in Proposition 7 in [10]. The proof of the cardinality bound $|\mathcal{W}| \leq |\mathcal{U}| + 1$ follows from a standard application of the Eggleston–Fenchel–Carathéodory theorem (Theorem 18 in [48]), thus completing the proof.

4.3. Proof of Corollary 2

Specializing Theorem 1 to TAD, note that $\rho(\kappa_\alpha, \omega) = 0$ since $P_{\hat{U}\hat{V}\hat{W}} = Q_U Q_V P_{\hat{W}|\hat{U}} \in \hat{\mathcal{L}}(\kappa_\alpha, \omega)$ and $I_P(\hat{V}; \hat{W}) = 0$. Additionally, for $R \geq \zeta(\kappa_\alpha, \omega)$, $E_b(\kappa_\alpha, \omega, R) = \infty$. Hence,

$$\mathcal{L}(\kappa_\alpha) = \left\{ (\omega, R, P_{SX}, \theta) : \begin{array}{c} \zeta(\kappa_\alpha, \omega) \leq R < I_P(X; Y|S), P_{SXY} = P_{SX}P_{Y|X}, \\ \min\{E_{sp}(P_{SX}, \theta), E_{ex}(R, P_{SX})\} \geq \kappa_\alpha \end{array} \right\},$$

$$\hat{\mathcal{L}}(\kappa_\alpha, \omega) := \left\{ P_{\hat{U}\hat{V}\hat{W}} : D(P_{\hat{U}\hat{V}\hat{W}} || P_{UV\hat{W}}) \leq \kappa_\alpha, P_{\hat{W}|\hat{U}} = \omega(P_{\hat{U}}), P_{UV\hat{W}} = Q_U Q_V P_{\hat{W}|\hat{U}} \right\}.$$

Then, we have

$$E_1(\kappa_\alpha, \omega) := E_1^d(\kappa_\alpha, \omega) := \min_{(P_{\tilde{U}\tilde{V}\tilde{W}}, Q_{\tilde{U}\tilde{V}\tilde{W}}) \in \mathcal{T}_1(\kappa_\alpha, \omega)} D(P_{\tilde{U}\tilde{V}\tilde{W}} || Q_{\tilde{U}\tilde{V}\tilde{W}})$$

$$\overset{(a)}{\geq} \min_{(P_{\tilde{U}\tilde{V}\tilde{W}}, Q_{\tilde{U}\tilde{V}\tilde{W}}) \in \mathcal{T}_1(\kappa_\alpha, \omega)} D(P_{\tilde{V}\tilde{W}} || Q_{\tilde{V}\tilde{W}})$$

$$\overset{(b)}{=} \min_{\substack{(P_{\hat{V}\hat{W}}, Q_{V\hat{W}}) : P_{\hat{U}\hat{V}\hat{W}} \in \hat{\mathcal{L}}(\kappa_\alpha, \omega), \\ Q_{UV\hat{W}} = Q_{UV}P_{\hat{W}|\hat{U}}}} D(P_{\hat{V}\hat{W}} || Q_{V\hat{W}}), \tag{37}$$

where (a) follows due to the data-processing inequality for KL divergence Theorem 2.15 in [43]; (b) is since $(P_{\tilde{U}\tilde{V}\tilde{W}}, Q_{\tilde{U}\tilde{V}\tilde{W}}) \in \mathcal{T}_1(\kappa_\alpha, \omega)$ implies that $P_{\tilde{V}\tilde{W}} = P_{\hat{V}\hat{W}}$ and $Q_{\tilde{U}\tilde{V}\tilde{W}} = Q_{UV}P_{\hat{W}|\hat{U}}$ for some $P_{\hat{U}\hat{V}\hat{W}} \in \hat{\mathcal{L}}(\kappa_\alpha, \omega)$. Next, note that since $R \geq \zeta(\kappa_\alpha, \omega)$, $E_2(\kappa_\alpha, \omega, R) = \infty$. Additionally,

$$E_3(\kappa_\alpha, \omega, R, P_{SX}) = \min_{(P_{\tilde{U}\tilde{V}\tilde{W}}, Q_{\tilde{U}\tilde{V}\tilde{W}}) \in \mathcal{T}_3(\kappa_\alpha, \omega)} D(P_{\tilde{U}\tilde{V}\tilde{W}} || Q_{\tilde{U}\tilde{V}\tilde{W}}) + E_{ex}(R, P_{SX}) \tag{38a}$$

$$\overset{(a)}{=} E_{ex}(R, P_{SX}), \tag{38b}$$

$$E_4(\kappa_\alpha, \omega, P_{SX}, \theta) = \min_{P_{\hat{V}} : P_{\hat{U}\hat{V}\hat{W}} \in \hat{\mathcal{L}}(\kappa_\alpha, \omega)} D(P_{\hat{V}} || Q_V) + E_m(P_{SX}, \theta) - \theta$$

$$\overset{(b)}{=} E_m(P_{SX}, \theta) - \theta =: E_3^d(P_{SX}, \theta), \tag{38c}$$

where

(a)   is obtained by taking $P_{\hat{U}\hat{V}\hat{W}} = Q_U Q_V P_{W|U} \in \hat{\mathcal{L}}(\kappa_\alpha, \omega)$ and $P_{W|U} = \omega(Q_U)$ in the definition of $\mathcal{T}_3(\kappa_\alpha, \omega)$. This implies that $(P_{\tilde{U}\tilde{V}\tilde{W}}, Q_{\tilde{U}\tilde{V}\tilde{W}}) = (Q_{UV}P_{W|U}, Q_{UV}P_{W|U}) \in \mathcal{T}_3(\kappa_\alpha, \omega)$, and hence that the first term in the right hand side (RHS) of (38a) is zero;

(b)   is due to $Q_U Q_V P_{W|U} \in \hat{\mathcal{L}}(\kappa_\alpha, \omega)$ for $P_{W|U} = \omega(Q_U)$.

Since $E_{ex}(R, P_{SX})$ is a non-increasing function of R and $R \geq \zeta(\kappa_\alpha, \omega)$, selecting $R = \zeta(\kappa_\alpha, \omega)$ maximizes $E_3(\kappa_\alpha, \omega, R, P_{SX})$. Then, (11) follows from (37), (38b) and (38c).

Next, we prove (12). Note that $\zeta(0,\omega) = I_Q(U;W)$, where $Q_{UW} = Q_U P_{W|U}$, $P_{W|U} = \omega(Q_U)$, and since $E_{\mathrm{sp}}(P_{SX},\theta) \geq 0$ and $E_{\mathrm{ex}}(I_Q(U;W), P_{SX}) \geq 0$,

$$\mathcal{L}^\star(0) = \begin{cases} (\omega, P_{SX}, \theta) \in \mathfrak{F} \times \mathcal{P}(\mathcal{S} \times \mathcal{X}) \times \Theta(P_{SX}) : \ I_Q(U;W) < I_P(X;Y|S), \\ Q_{UVW} = Q_{UV} P_{W|U}, \ P_{W|U} = \omega(Q_U), P_{SXY} := P_{SX} P_{Y|X} \end{cases}.$$

Additionally, $\hat{\mathcal{L}}(0,\omega) = \{Q_U Q_V P_{W|U} : P_{W|U} = \omega(Q_U)\}$. By choosing $\theta = -\theta_1(P_{SX})$ (defined above (6a)) that maximizes $E_3^{\mathrm{d}}(P_{SX},\theta)$, we have

$$E_1^{\mathrm{d}}(0,\omega) \geq \min_{\substack{(P_{\hat{V}\hat{W}}, Q_{V\hat{W}}): \ P_{\hat{U}\hat{V}\hat{W}} \in \hat{\mathcal{L}}(0,\omega), \\ Q_{UV\hat{W}} = Q_{UV} P_{\hat{W}|\hat{U}}}} D(P_{\hat{V}\hat{W}} || Q_{V\hat{W}})$$

$$= \min_{\substack{(P_{W|U}, P_{SX}): \ I_Q(U;W) \leq I_P(X;Y|S), \\ Q_{UVW} = Q_{UV} P_{W|U}, P_{SXY} = P_{SX} P_{Y|X}}} D(Q_V Q_W || Q_{VW}), \tag{39a}$$

$$E_2^{\mathrm{d}}(0,\omega,P_{SX}) = E_{\mathrm{ex}}(I_Q(U;W), P_{SX}), \tag{39b}$$

$$E_3^{\mathrm{d}}(P_{SX}, -\theta_1(P_{SX})) = E_{\mathrm{m}}(P_{SX}, -\theta_1(P_{SX})) + \theta_1(P_{SX}) = \theta_1(P_{SX}), \tag{39c}$$

where (39c) is due to $E_{\mathrm{m}}(P_{SX}, -\theta_1(P_{SX})) = 0$. The latter in turn follows similar to (A10) and (A11) from the definition of $E_{\mathrm{m}}(\cdot,\cdot)$. From (11), (39a, 39b, 39c), and the continuity of $E_1^{\mathrm{d}}(\kappa_\alpha, \omega)$, $E_2^{\mathrm{d}}(\kappa_\alpha, \omega, P_{SX})$ in κ_α, (12) follows. The proof of the cardinality bound $|\mathcal{W}| \leq |\mathcal{U}| + 1$ in the RHS of (39a) follows via a standard application of the Eggleston–Fenchel–Carathéodory Theorem (Theorem 18 in [48]). To see this, note that it is sufficient to preserve $\{Q_U(u), u \in \mathcal{U}\}$, $D(Q_V Q_W || Q_{VW})$ and $H_Q(U|W)$, all of which can be written as a linear combination of functionals of $Q_{U|W}(\cdot|w)$ with weights $Q_W(w)$. Thus, it requires $|\mathcal{U}| - 1$ points to preserve $\{Q_U(u), u \in \mathcal{U}\}$ and one each for $D(Q_V Q_W || Q_{VW})$ and $H_Q(U|W)$. This completes the proof.

4.4. Proof of Theorem 2

We will show that the error-exponent pairs $(\kappa_\alpha, \kappa_{\mathrm{h}}^\star(\kappa_\alpha))$ and $(\kappa_\alpha, \kappa_{\mathrm{u}}^\star(\kappa_\alpha))$ are achieved by a hybrid coding scheme and uncoded transmission scheme, respectively. First, we describe the hybrid coding scheme.

Let $n \in \mathbb{N}$, $|\mathcal{W}| < \infty$, $\kappa_\alpha > 0$, and $(P_S, \omega'(\cdot, P_S), P_{X|USW}, P_{X'|US}) \in \mathcal{L}_h(\kappa_\alpha)$. Further, let $\eta > 0$ be a small number, and choose a sequence $\mathbf{s} \in \mathcal{T}_n(P_{\hat{S}})$, where $P_{\hat{S}}$ satisfies $D(P_{\hat{S}} || P_S) \leq \eta$. Set $R' := \zeta'(\kappa_\alpha, \omega', P_{\hat{S}})$.

Encoding: The encoder performs type-based quantization followed by hybrid coding [40]. The details are as follows:

Quantization codebook: Let $\mathcal{D}_n(P_U, \eta)$ be as defined in (15). Consider some ordering on the types in $\mathcal{D}_n(P_U, \eta)$ and denote the elements as $P_{\hat{U}_i}$, $i \in [|\mathcal{D}_n(P_U, \eta)|]$. For each joint type $P_{\hat{S}\hat{U}_i}$ such that $P_{\hat{U}_i} \in \mathcal{D}_n(P_U, \eta)$ and $\hat{S} \perp\!\!\!\perp \hat{U}_i$, choose a joint type variable $P_{\hat{S}\hat{U}_i\hat{W}_i}$, $P_{\hat{W}_i} \in \mathcal{T}(\mathcal{W}^n)$, such that $D(P_{\hat{W}_i|\hat{U}_i\hat{S}} || P_{W_i|U\hat{S}} | P_{\hat{U}_i\hat{S}}) \leq \eta/3$ and $I(\hat{S}, \hat{U}_i; \hat{W}_i) \leq R' + (\eta/3)$, where $P_{W_i|U,S} = \omega'(P_{\hat{U}_i}, P_{\hat{S}})$. Define $\mathcal{D}_n(P_{SUW}, \eta) := \{P_{\hat{S}\hat{U}_i\hat{W}_i} : i \in [|\mathcal{D}_n(P_U, \eta)|]\}$, $R_i' := I_P(\hat{S}, \hat{U}_i; \hat{W}_i) + (\eta/3)$ for $i \in [|\mathcal{D}_n(P_U, \eta)|]$ and $\mathcal{M}_i' := [1 + \sum_{m=1}^{i-1} e^{nR_m'} : \sum_{m=1}^{i} e^{nR_m'}]$, $i \in [|\mathcal{D}_n(P_U, \eta)|]$. Let $\mathbb{B}_{W,n} = \{\mathbf{W}(j) \in \mathcal{W}^n, 1 \leq j \leq \sum_{i=1}^{|\mathcal{D}_n(P_U,\eta)|} e^{nR_i'}\}$ denote a random quantization codebook such that for $i \in [|\mathcal{D}_n(P_U, \eta)|]$, each codeword $\mathbf{W}(j)$, $j \in \mathcal{M}_i'$, is independently selected from $\mathcal{T}_n(P_{\hat{W}_i})$ according to uniform distribution, i.e., $\mathbf{W}(j) \sim \mathsf{Unif}[\mathcal{T}_n(P_{\hat{W}_i})]$. Let $\mathcal{B}_{W,n}$ denote a realization of $\mathbb{B}_{W,n}$.

Type-based hybrid coding: For $\mathbf{u} \in \mathcal{T}_n(P_{\hat{U}_i})$ such that $P_{\hat{U}_i} \in \mathcal{D}_n(P_U, \eta)$ for some $i \in [|\mathcal{D}_n(P_U, \eta)|]$, let

$$\bar{M}(\mathbf{u}, \mathcal{B}_{W,n}) := \left\{ j \in \mathcal{M}_i' : \ \mathbf{w}(j) \in \mathcal{B}_{W,n}, \ (\mathbf{s}, \mathbf{u}, \mathbf{w}(j)) \in \mathcal{T}_n(P_{\hat{S}\hat{U}_i\hat{W}_i}), \ P_{\hat{S}\hat{U}_i\hat{W}_i} \in \mathcal{D}_n(P_{SUW}, \eta) \right\}.$$

If $|\bar{M}(\mathbf{u}, \mathcal{B}_{W,n})| \geq 1$, let $M'(\mathbf{u}, \mathcal{B}_{W,n})$ denote an index selected uniformly at random from the set $\bar{M}(\mathbf{u}, \mathcal{B}_{W,n})$; otherwise, set $M'(\mathbf{u}, \mathcal{B}_{W,n}) = 0$. Given $\mathcal{B}_{W,n}$ and $\mathbf{u} \in \mathcal{U}^n$, the quantizer outputs $M' = M'(\mathbf{u}, \mathcal{B}_{W,n})$, where the support of M' is $\mathcal{M}' := \{0\} \bigcup_{i=1}^{|\mathcal{D}_n(P_U, \eta)|} \mathcal{M}'_i$. Note that for sufficiently large n, it follows similarly to (18) that $|\mathcal{M}'| \leq e^{n(R'+\eta)}$. For a given $\mathcal{B}_{W,n}$ and $\mathbf{u} \in \mathcal{U}^n$, the encoder transmits $\mathbf{X} \sim P_{X|USW}^{\otimes n}(\cdot|\mathbf{u}, \mathbf{s}, \mathbf{w}(m'))$ if $M' = m' \neq 0$, and $\mathbf{X}' \sim P_{X'|US}^{\otimes n}(\cdot|\mathbf{u}, \mathbf{s})$ if $M' = 0$.

Acceptance region: For a given codebook $\mathcal{B}_{W,n}$ and $m' \in \mathcal{M}'\backslash\{0\}$, let $\mathcal{O}_{m'}$ denote the set of $\mathbf{u}$ such that $M'(\mathbf{u}, \mathcal{B}_{W,n}) = m'$. For each $m' \in \mathcal{M}'\backslash\{0\}$ and $\mathbf{u} \in \mathcal{O}_{m'}$, set

$$\mathcal{Z}'_{m'}(\mathbf{u}) = \left\{ (\mathbf{v}, \mathbf{y}) \in \mathcal{V}^n \times \mathcal{Y}^n : (\mathbf{s}, \mathbf{u}, \bar{\mathbf{w}}_{m'}, \mathbf{v}, \mathbf{y}) \in \mathcal{J}_n(\kappa_\alpha + \eta, P_{\hat{S}UW_{m'}VY}) \right\},$$

where recall that $\mathcal{J}_n(r, P_X) := \{\mathbf{x} \in \mathcal{X}^n : D(P_{\mathbf{x}} || P_X) \leq r\}$, and

$$P_{\hat{S}UW_{m'}VXY} = P_{\hat{S}} P_{UV} P_{W_{m'}|U\hat{S}} P_{X|U\hat{S}W_{m'}} P_{Y|X}, \tag{40a}$$

$$P_{W_{m'}|U\hat{S}} = \omega'(P_{\mathbf{u}}, P_{\hat{S}}) \text{ and } P_{X|U\hat{S}W_{m'}} = P_{X|USW}. \tag{40b}$$

For $m' \in \mathcal{M}'\backslash\{0\}$, define $\mathcal{Z}'_{m'} := \{(\mathbf{v}, \mathbf{y}) : (\mathbf{v}, \mathbf{y}) \in \mathcal{Z}'_{m'}(\mathbf{u}) \text{ for some } \mathbf{u} \in \mathcal{O}_{m'}\}$. The acceptance region for H_0 is given by $\mathcal{A}_n := \cup_{m' \in \mathcal{M}'\backslash 0} \mathbf{s} \times m' \times \mathcal{Z}'_{m'}$ or equivalently as $\mathcal{A}_n^e := \cup_{m' \in \mathcal{M}'\backslash 0} \mathbf{s} \times \mathcal{O}_{m'} \times \mathcal{Z}'_{m'}$.

Decoding: Given codebook $\mathcal{B}_{W,n}$, $\mathbf{Y} = \mathbf{y}$, and $\mathbf{V} = \mathbf{v}$, if $(\mathbf{v}, \mathbf{y}) \in \bigcup_{m' \in \mathcal{M}'\backslash\{0\}} \mathcal{Z}'_{m'}$, then $\hat{M}' = \hat{m}'$, where $\hat{m}' := \arg\min_{j \in \mathcal{M}'\backslash 0} H_e(\mathbf{w}(j)|\mathbf{v}, \mathbf{y}, \mathbf{s})$. Otherwise, $\hat{M}' = 0$. Denote the decoder induced by the above operations by $g_{\mathcal{B}_{W,n}} : \mathcal{S}^n \times \mathcal{V}^n \times \mathcal{Y}^n \to \mathcal{M}'$.

Testing: If $\hat{M}' = 0$, $\hat{H} = 1$ is declared. Otherwise, $\hat{H} = 0$ or $\hat{H} = 1$ is declared depending on whether $(\mathbf{s}, \hat{m}', \mathbf{v}, \mathbf{y}) \in \mathcal{A}_n$ or $(\mathbf{s}, \hat{m}', \mathbf{v}, \mathbf{y}) \notin \mathcal{A}_n$, respectively. Denote the decision function induced by $g_{\mathcal{B}_{W,n}}$ and $\mathcal{A}_n$ by $g_n : \mathcal{S}^n \times \mathcal{V}^n \times \mathcal{Y}^n \to \hat{\mathcal{H}}$.

Induced probability distribution: The PMFs induced by a code $c_n = (f_n, g_n)$ with respect to codebook $\mathcal{B}_{W,n}$ under H_0 and H_1 are

$P_{\mathbf{UV}M'\mathbf{XY}\hat{M}'\hat{H}}^{(\mathcal{B}_{W,n}, c_n)}(\mathbf{u}, \mathbf{v}, m', \mathbf{x}, \mathbf{y}, \hat{m}', \hat{h})$

$$:= \begin{cases} P_{UV}^{\otimes n}(\mathbf{u}, \mathbf{v}) \, \mathbb{1}_{\left\{M'(\mathbf{u}, \mathcal{B}_{W,n}) = m'\right\}} P_{X|USW}^{\otimes n}(\mathbf{x}|\mathbf{s}, \mathbf{u}, \mathbf{w}(m')) P_{Y|X}^{\otimes n}(\mathbf{y}|\mathbf{x}) \\ \qquad \mathbb{1}_{\left\{g_{\mathcal{B}_{W,n}}(\mathbf{v}, \mathbf{y}, \mathbf{s}) = \hat{m}'\right\}} \mathbb{1}_{\left\{\hat{h} = \mathbb{1}_{\left\{(\mathbf{s}, \hat{m}', \mathbf{v}, \mathbf{y}) \in \mathcal{A}_n^c\right\}}\right\}}, & \text{if } m' \neq 0, \\ P_{UV}^{\otimes n}(\mathbf{u}, \mathbf{v}) \, \mathbb{1}_{\left\{M'(\mathbf{u}, \mathcal{B}_{W,n}) = m'\right\}} P_{X'|US}^{\otimes n}(\mathbf{x}|\mathbf{s}, \mathbf{u}) P_{Y|X}^{\otimes n}(\mathbf{y}|\mathbf{x}) \\ \qquad \mathbb{1}_{\left\{g_{\mathcal{B}_{W,n}}(\mathbf{v}, \mathbf{y}, \mathbf{s}) = \hat{m}'\right\}} \mathbb{1}_{\left\{\hat{h} = \mathbb{1}_{\left\{(\mathbf{s}, \hat{m}', \mathbf{v}, \mathbf{y}) \in \mathcal{A}_n^c\right\}}\right\}}, & \text{otherwise,} \end{cases}$$

and

$Q_{\mathbf{UV}M'\mathbf{XY}\hat{M}'\hat{H}}^{(\mathcal{B}_{W,n}, c_n)}(\mathbf{u}, \mathbf{v}, m', \mathbf{x}, \mathbf{y}, \hat{m}', \hat{h})$

$$:= \begin{cases} Q_{UV}^{\otimes n}(\mathbf{u}, \mathbf{v}) \, \mathbb{1}_{\left\{M'(\mathbf{u}, \mathcal{B}_{W,n}) = m'\right\}} P_{X|USW}^{\otimes n}(\mathbf{x}|\mathbf{s}, \mathbf{u}, \mathbf{w}(m')) P_{Y|X}^{\otimes n}(\mathbf{y}|\mathbf{x}) \\ \qquad \mathbb{1}_{\left\{g_{\mathcal{B}_{W,n}}(\mathbf{v}, \mathbf{y}, \mathbf{s}) = \hat{m}'\right\}} \mathbb{1}_{\left\{\hat{h} = \mathbb{1}_{\left\{(\mathbf{s}, \hat{m}', \mathbf{v}, \mathbf{y}) \in \mathcal{A}_n^c\right\}}\right\}}, & \text{if } m' \neq 0, \\ Q_{UV}^{\otimes n}(\mathbf{u}, \mathbf{v}) \, \mathbb{1}_{\left\{M'(\mathbf{u}, \mathcal{B}_{W,n}) = m'\right\}} P_{X'|US}^{\otimes n}(\mathbf{x}|\mathbf{s}, \mathbf{u}) P_{Y|X}^{\otimes n}(\mathbf{y}|\mathbf{x}) \\ \qquad \mathbb{1}_{\left\{g_{\mathcal{B}_{W,n}}(\mathbf{v}, \mathbf{y}, \mathbf{s}) = \hat{m}'\right\}} \mathbb{1}_{\left\{\hat{h} = \mathbb{1}_{\left\{(\mathbf{s}, \hat{m}', \mathbf{v}, \mathbf{y}) \in \mathcal{A}_n^c\right\}}\right\}}, & \text{otherwise,} \end{cases}$$

respectively. For brevity, we will denote $\mathcal{B}_{W,n}$ by $\mathcal{B}_n$, $\mathbb{B}_{W,n}$ by $\mathbb{B}_n$, and the above probability distributions by $P^{(\mathcal{B}_n)}$ and $Q^{(\mathcal{B}_n)}$. Let $\mathfrak{B}_n$ and μ_n stand for the support and probability measure of $\mathbb{B}_n$, respectively, and set $\bar{\mathbb{P}}_{P^{(\mathbb{B}_n)}} := \mathbb{E}_{\mu_n}[\mathbb{P}_{P^{(\mathbb{B}_n)}}]$, $\bar{\mathbb{P}}_{Q^{(\mathbb{B}_n)}} := \mathbb{E}_{\mu_n}[\mathbb{P}_{Q^{(\mathbb{B}_n)}}]$

Analysis of the type I and type II error probabilities: We analyze the expected type I and type II error probabilities, where the expectation is with respect to the randomness of $\mathbb{B}_n$, followed by the expurgation technique to extract a sequence of deterministic codebooks $\{\mathcal{B}_n\}_{n\in\mathbb{N}}$ and a code $\{c_n = (f_n, g_n)\}_{n\in\mathbb{N}}$ that achieves the lower bound in Theorem 2.

Type I error probability: Denoting by $\mathbb{A}_n$ the random acceptance region for H_0, note that a type I error can occur only under the following events:

(i) $\quad \mathcal{E}'_{\text{EE}} := \bigcup\limits_{P_{\hat{U}} \in \mathcal{D}_n(P_U, \eta)} \bigcup\limits_{\mathbf{u} \in \mathcal{T}_n(P_{\hat{U}})} \mathcal{E}'_{\text{EE}}(\mathbf{u})$, where

$$\mathcal{E}'_{\text{EE}}(\mathbf{u}) := \left\{ \begin{array}{c} \nexists\, j \in \mathcal{M}' \backslash \{0\} \text{ s.t. } (\mathbf{s}, \mathbf{u}, \mathbf{W}(j)) \in \mathcal{T}_n(P_{\hat{S}\hat{U}_i\hat{W}_i}), \ P_{\hat{S}\hat{U}_i} = P_{\mathbf{su}}, \\ P_{\hat{S}\hat{U}_i\hat{W}_i} \in \mathcal{D}_n(P_{SUW}, \eta) \end{array} \right\},$$

(ii) $\quad \mathcal{E}'_{\text{NE}} := \{\hat{M}' = M' \text{ and } (\mathbf{s}, \hat{M}', \mathbf{V}, \mathbf{Y}) \notin \mathbb{A}_n\}$,

(iii) $\quad \mathcal{E}'_{\text{ODE}} := \{M' \neq 0, \hat{M}' \neq M' \text{ and } (\mathbf{s}, \hat{M}', \mathbf{V}, \mathbf{Y}) \notin \mathbb{A}_n\}$,

(iv) $\quad \mathcal{E}'_{\text{SDE}} := \{M' = 0, \hat{M}' \neq M' \text{ and } (\mathbf{s}, \hat{M}', \mathbf{V}, \mathbf{Y}) \notin \mathbb{A}_n\}$.

By definition of R'_i, we have, similar to (24), the following:

$$\bar{\mathbb{P}}_{P\mathbb{B}_n}\left(\mathcal{E}'_{\text{EE}}\right) \leq e^{-e^{n\Omega(\eta)}}. \tag{41}$$

Next, the event $\mathcal{E}'_{\text{NE}}$ can be upper bounded as

$$\begin{aligned} \bar{\mathbb{P}}_{P\mathbb{B}_n}\left(\mathcal{E}'_{\text{NE}} | \mathcal{E}'^c_{\text{EE}}\right) &\leq \bar{\mathbb{P}}_{P\mathbb{B}_n}\left((\mathbf{s}, \hat{M}', \mathbf{V}, \mathbf{Y}) \notin \mathbb{A}_n | \hat{M}' = M', \mathcal{E}'^c_{EE}\right) \\ &= 1 - \bar{\mathbb{P}}_{P\mathbb{B}_n}\left((\mathbf{s}, \mathbf{U}, \mathbf{V}, \mathbf{Y}) \in \mathbb{A}^e_n | \mathcal{E}'^c_{EE}\right). \end{aligned} \tag{42}$$

For $\mathbf{u} \in \mathbb{O}_{m'}$, note that, similar to Equation 4.17 in [14], we have

$$\bar{\mathbb{P}}_{P\mathbb{B}_n}\left((\mathbf{V}, \mathbf{Y}) \in \mathbb{Z}'_{m'}(\mathbf{u}) | \mathbf{U} = \mathbf{u}, \mathbf{W}(m') = \bar{\mathbf{w}}_{m'}, \mathcal{E}'^c_{EE}\right) \geq 1 - e^{-n\left(\kappa_\alpha + \frac{\eta}{3} - D(P_{\mathbf{u}} \| P_U)\right)}.$$

From this and (15), we obtain, similar to Equation (4.22) in [14] that

$$\bar{\mathbb{P}}_{P\mathbb{B}_n}\left((\mathbf{s}, \mathbf{U}, \mathbf{V}, \mathbf{Y}) \in \mathbb{A}^n_e | \mathcal{E}'^c_{EE}\right) \geq 1 - e^{-n\kappa_\alpha}. \tag{43}$$

Substituting (43) in (42) yields

$$\bar{\mathbb{P}}_{P\mathbb{B}_n}\left(\mathcal{E}'_{\text{NE}} | \mathcal{E}'^c_{\text{EE}}\right) \leq e^{-n\kappa_\alpha}. \tag{44}$$

Next, we bound the probability of the event $\mathcal{E}'_{\text{ODE}}$ as follows:

$$\begin{aligned} &\bar{\mathbb{P}}_{P\mathbb{B}_n}\left(\mathcal{E}'_{\text{ODE}}\right) \\ &= \bar{\mathbb{P}}_{P\mathbb{B}_n}\left(M' \neq 0, \hat{M}' \neq M', (\mathbf{s}, M', \mathbf{V}, \mathbf{Y}) \in \mathbb{A}_n, (\mathbf{s}, \hat{M}', \mathbf{V}, \mathbf{Y}) \notin \mathbb{A}_n\right) \\ &\qquad\qquad + \bar{\mathbb{P}}_{P\mathbb{B}_n}\left(M' \neq 0, \hat{M}' \neq M', (\mathbf{s}, M', \mathbf{V}, \mathbf{Y}) \notin \mathbb{A}_n, (\mathbf{s}, \hat{M}', \mathbf{V}, \mathbf{Y}) \notin \mathbb{A}_n\right) \\ &\leq \bar{\mathbb{P}}_{P\mathbb{B}_n}\left(M' \neq 0, \hat{M}' \neq M', (\mathbf{s}, M', \mathbf{V}, \mathbf{Y}) \in \mathbb{A}_n, (\mathbf{s}, \hat{M}', \mathbf{V}, \mathbf{Y}) \notin \mathbb{A}_n\right) \\ &\qquad\qquad + \bar{\mathbb{P}}_{P\mathbb{B}_n}\left(M' \neq 0, \hat{M}' \neq M', (\mathbf{s}, M', \mathbf{V}, \mathbf{Y}) \notin \mathbb{A}_n\right) \\ &\overset{(a)}{\leq} \bar{\mathbb{P}}_{P\mathbb{B}_n}\left(M' \neq 0, \hat{M}' \neq M', (\mathbf{s}, M', \mathbf{V}, \mathbf{Y}) \in \mathbb{A}_n, (\mathbf{s}, \hat{M}', \mathbf{V}, \mathbf{Y}) \notin \mathbb{A}_n\right) + e^{-e^{n\Omega(\eta)}} + e^{-n\kappa_\alpha} \quad (45) \\ &\leq \bar{\mathbb{P}}_{P\mathbb{B}_n}\left(\hat{M}' \neq M' | M' \neq 0, (\mathbf{s}, M', \mathbf{V}, \mathbf{Y}) \in \mathbb{A}_n\right) + e^{-e^{n\Omega(\eta)}} + e^{-n\kappa_\alpha}, \\ &\overset{(b)}{=} \bar{\mathbb{P}}_{P\mathbb{B}_n}\left(\hat{M}' \neq M' | M' \neq 0, \hat{M}' \neq 0, (\mathbf{s}, M', \mathbf{V}, \mathbf{Y}) \in \mathbb{A}_n\right) \tag{46} \\ &\overset{(c)}{\leq} e^{-n\left(\rho'(\kappa_\alpha, \omega', P_S, P_{X|USW}) - \zeta'(\kappa_\alpha, \omega', P_{\hat{S}}) - O(\eta)\right)}, \tag{47} \end{aligned}$$

where (a) follows similar to (29) using (41) and (43); (b) is since $(\mathbf{s}, M', \mathbf{V}, \mathbf{Y}) \in \mathbb{A}_n$ implies that $\hat{M}' \neq 0$; and (c) follows similar to (33). Further,

$$
\begin{aligned}
\bar{\mathbb{P}}_{P\mathbb{B}_n}\left(\mathcal{E}'_{\text{SDE}}\right) &\leq \bar{\mathbb{P}}_{P\mathbb{B}_n}\left(M' = 0\right) \\
&\leq \bar{\mathbb{P}}_{P\mathbb{B}_n}\left(M' = 0 | \mathcal{E}'^c_{EE}\right) + \bar{\mathbb{P}}_{P\mathbb{B}_n}\left(\mathcal{E}'_{EE}\right) \\
&= \sum_{\mathbf{u}: P_{\mathbf{u}} \notin \mathcal{D}_n(P_U, \eta)} P_U^{\otimes n}(\mathbf{u}) + \bar{\mathbb{P}}_{P\mathbb{B}_n}\left(\mathcal{E}'_{EE}\right) \\
&\leq e^{-n\kappa_\alpha} + e^{-e^{n\Omega(\eta)}},
\end{aligned}
\tag{48}
$$

where the penultimate equality is since given $\mathcal{E}'^c_{EE}$, $M' = 0$ occurs only for $\mathbf{U} = \mathbf{u}$ such that $P_{\mathbf{u}} \notin \mathcal{D}_n(P_U, \eta)$, and the final inequality follows from (41), the definition of $\mathcal{D}_n(P_U, \eta)$ and Lemma 1.6 in [38]. From (41), (44), (47) and (48), the expected type I error probability satisfies $e^{-n(\kappa_\alpha - O(\eta))}$ for sufficiently large n via the union bound.

Type II error probability: Next, we analyze the expected type II error probability over $\mathbb{B}_n$. Let

$$
\mathcal{D}_n(P_{SVWY}, \eta) := \left\{
\begin{array}{l}
P_{\hat{S}\hat{V}\hat{W}\hat{Y}} : \quad \exists\, (\mathbf{s}, \mathbf{u}, \mathbf{v}, \bar{\mathbf{w}}, \mathbf{y}) \in \underset{m' \in \mathcal{M}'\backslash\{0\}}{\cup} \mathcal{J}_n\left(\kappa_\alpha + \eta, P_{\hat{S}UVW_{m'}Y}\right), \\
\qquad P_{\hat{S}UVW_{m'}Y} \text{ satisfies (40) and } P_{\mathbf{suv}\bar{\mathbf{w}}\mathbf{y}} = P_{\hat{S}\hat{U}\hat{V}\hat{W}\hat{Y}}
\end{array}
\right\},
$$

$$
\mathcal{F}'_{1,n}(\eta) := \left\{
P_{\hat{S}\tilde{U}\tilde{V}\tilde{W}\tilde{Y}} \in \mathcal{T}(\mathcal{S}^n \times \mathcal{U}^n \times \mathcal{V}^n \times \mathcal{W}^n \times \mathcal{Y}^n) \; : \;
\begin{array}{l}
P_{\hat{S}\tilde{U}\tilde{W}} \in \mathcal{D}_n(P_{SUW}, \eta), \\
P_{\hat{S}\tilde{V}\tilde{W}\tilde{Y}} \in \mathcal{D}_n(P_{SVWY}, \eta)
\end{array}
\right\}.
$$

A type II error can occur only under the following events:

(a)　$\mathcal{E}'_a := \left\{
\begin{array}{l}
\hat{M}' = M' \neq 0, (\mathbf{s}, \mathbf{U}, \mathbf{V}, \mathbf{W}(M'), \mathbf{Y}) \in \mathcal{T}_n(P_{\hat{S}\hat{U}\hat{V}\hat{W}\hat{Y}}) \\
\text{s.t. } P_{\hat{U}\hat{W}} \in \mathcal{D}_n(P_{SUW}, \eta) \text{ and } P_{\hat{S}\hat{V}\hat{W}\hat{Y}} \in \mathcal{D}_n(P_{SVWY}, \eta)
\end{array}
\right\}$,

(b)　$\mathcal{E}'_b := \left\{
\begin{array}{l}
M' \neq 0, \hat{M}' \neq M', (\mathbf{s}, \mathbf{U}, \mathbf{V}, \mathbf{W}(M'), \mathbf{Y}, \mathbf{W}(\hat{M}')) \in \mathcal{T}_n\left(P_{\hat{S}\hat{U}\hat{V}\hat{W}\hat{Y}\hat{W}_d}\right) \\
\text{s.t. } P_{\hat{S}\hat{U}\hat{W}} \in \mathcal{D}_n(P_{SUW}, \eta), P_{\hat{S}\hat{V}\hat{W}_d\hat{Y}} \in \mathcal{D}_n(P_{SVWY}, \eta), \\
\text{and } H_e\left(\mathbf{W}(\hat{M}') | \mathbf{s}, \mathbf{V}, \mathbf{Y}\right) \leq H_e\left(\mathbf{W}(M') | \mathbf{s}, \mathbf{V}, \mathbf{Y}\right)
\end{array}
\right\}$,

(c)　$\mathcal{E}'_c := \left\{
\begin{array}{l}
M' = 0, \hat{M}' \neq M', (\mathbf{s}, \mathbf{V}, \mathbf{Y}, \mathbf{W}(\hat{M}')) \in \mathcal{T}_n\left(P_{\hat{S}\hat{V}\hat{Y}\hat{W}_d}\right) \\
\text{s.t. } P_{\hat{S}\hat{V}\hat{W}_d\hat{Y}} \in \mathcal{D}_n(P_{SVWY}, \eta)
\end{array}
\right\}$.

Considering the event $\mathcal{E}'_a$, we have

$$
\begin{aligned}
&\bar{\mathbb{P}}_{Q\mathbb{B}_n}\left(\mathcal{E}'_a\right) \\
&\leq \sum_{\substack{P_{\hat{S}\tilde{U}\tilde{V}\tilde{W}\tilde{Y}} \\ \in \mathcal{F}'_{1,n}(\eta)}} \sum_{\substack{(\mathbf{u}, \mathbf{v}, \bar{\mathbf{w}}, \mathbf{y}): \\ (\mathbf{s}, \mathbf{u}, \mathbf{v}, \bar{\mathbf{w}}, \mathbf{y}) \\ \in \mathcal{T}_n(P_{\hat{S}\tilde{U}\tilde{V}\tilde{W}\tilde{Y}})}} \sum_{m' \in \mathcal{M}'\backslash\{0\}} \bar{\mathbb{P}}_{Q\mathbb{B}_n}\left(\mathbf{U} = \mathbf{u}, \mathbf{V} = \mathbf{v}, M' = m', \mathbf{W}(m') = \bar{\mathbf{w}}, \mathbf{Y} = \mathbf{y} | \mathbf{S} = \mathbf{s}\right) \\
&\leq \sum_{P_{\hat{S}\tilde{U}\tilde{V}\tilde{W}\tilde{Y}} \in \mathcal{F}'_{1,n}(\eta)} \sum_{\substack{(\mathbf{u}, \mathbf{v}, \bar{\mathbf{w}}, \mathbf{y}): \\ (\mathbf{s}, \mathbf{u}, \mathbf{v}, \bar{\mathbf{w}}, \mathbf{y}) \in \mathcal{T}_n(P_{\hat{S}\tilde{U}\tilde{V}\tilde{W}\tilde{Y}})}} \sum_{m' \in \mathcal{M}'\backslash\{0\}} \bar{\mathbb{P}}_{Q\mathbb{B}_n}\left(\mathbf{U} = \mathbf{u}, \mathbf{V} = \mathbf{v}, M' = m' | \mathbf{S} = \mathbf{s}\right) \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \bar{\mathbb{P}}_{Q\mathbb{B}_n}\left(\mathbf{W}(m') = \bar{\mathbf{w}} | \mathbf{U} = \mathbf{u}, \mathbf{V} = \mathbf{v}, M' = m', \mathbf{S} = \mathbf{s}\right) \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \bar{\mathbb{P}}_{Q\mathbb{B}_n}\left(\mathbf{Y} = \mathbf{y} | \mathbf{U} = \mathbf{u}, \mathbf{V} = \mathbf{v}, M' = m', \mathbf{W}(m') = \bar{\mathbf{w}}, \mathbf{S} = \mathbf{s}\right) \\
&\overset{(a)}{\leq} \sum_{P_{\hat{S}\tilde{U}\tilde{V}\tilde{W}\tilde{Y}} \in \mathcal{F}'_{1,n}(\eta)} \sum_{\substack{(\mathbf{u}, \mathbf{v}, \bar{\mathbf{w}}, \mathbf{y}): \\ (\mathbf{s}, \mathbf{u}, \mathbf{v}, \bar{\mathbf{w}}, \mathbf{y}) \in \mathcal{T}_n(P_{\hat{S}\tilde{U}\tilde{V}\tilde{W}\tilde{Y}})}} e^{-n\left(H(\tilde{U},\tilde{V}) + D\left(P_{\tilde{U}\tilde{V}} || Q_{UV}\right)\right)} e^{-n\left(H(\tilde{W}|\hat{S},\tilde{U}) - \eta\right)} \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad e^{-n\left(H(\tilde{Y}|\tilde{U},\hat{S},\tilde{W}) + D\left(P_{\tilde{Y}|\tilde{U}\hat{S}\tilde{W}} || P_{Y|USW} | P_{\tilde{U}\hat{S}\tilde{W}}\right)\right)} \\
&\leq \sum_{P_{\hat{S}\tilde{U}\tilde{V}\bar{W}\tilde{Y}} \in \mathcal{F}'_{1,n}(\eta)} e^{nH(\tilde{U},\tilde{V},\tilde{W},\tilde{Y}|\hat{S})} e^{-n\left(H(\tilde{U},\tilde{V}) + D\left(P_{\tilde{U}\tilde{V}} || Q_{UV}\right)\right)} e^{-n\left(H(\tilde{W}|\hat{S},\tilde{U}) - \eta\right)}
\end{aligned}
$$

$$e^{-n\left(H(\tilde{Y}|\tilde{U},\hat{S},\tilde{W}) + D\left(P_{\tilde{Y}|\tilde{U}\hat{S}\tilde{W}}||P_{Y|USW}|P_{\tilde{U}\hat{S}\tilde{W}}\right)\right)}$$

$$\leq e^{-n\mathsf{E}'_{1,n}}, \tag{49}$$

where

$$
\begin{aligned}
\mathsf{E}'_{1,n} &:= \min_{P_{\hat{S}\tilde{U}\tilde{V}\tilde{W}\tilde{Y}} \in \mathcal{F}'_{1,n}(\eta)} H(\tilde{U},\tilde{V}) + D(P_{\tilde{U}\tilde{V}}||Q_{UV}) + H(\tilde{W}|\hat{S},\tilde{U}) - \eta + H(\tilde{Y}|\tilde{U},\hat{S},\tilde{W}) \\
&\quad + D\left(P_{\tilde{Y}|\tilde{U}\hat{S}\tilde{W}}||P_{Y|USW}|P_{\tilde{U}\hat{S}\tilde{W}}\right) - H(\tilde{U},\tilde{V},\tilde{W},\tilde{Y}|\hat{S}) - \frac{1}{n}||\mathcal{U}||\mathcal{V}||\mathcal{W}||\mathcal{Y}|\log(n+1) \\
&\gtrsim \min_{\substack{(P_{\tilde{U}\tilde{V}\tilde{W}\tilde{Y}S},Q_{\tilde{U}\tilde{V}\tilde{W}\tilde{Y}S}) \\ \in \mathcal{T}'_1(\kappa_\alpha,\omega',P_S,P_{X|USW})}} D(P_{\tilde{U}\tilde{V}\tilde{W}\tilde{Y}|S}||Q_{UVWY|S}|P_S) - O(\eta) \\
&= E'_1(\kappa_\alpha,\omega') - O(\eta).
\end{aligned}
$$

For the inequality in (a) above, we used $\sum \bar{\mathbb{P}}_{Q^{\mathbb{B}_n}}\left(M'=m'|\mathbf{U}=\mathbf{u},\mathbf{V}=\mathbf{v},\mathbf{S}=\mathbf{s}\right) \leq 1$ and

$$\bar{\mathbb{P}}_{Q^{\mathbb{B}_n}}\left(\mathbf{W}(m')=\bar{\mathbf{w}}|\mathbf{U}=\mathbf{u},\mathbf{V}=\mathbf{v},\mathbf{S}=\mathbf{s},M'=m'\right) \leq \begin{cases} e^{-n\left(H(\tilde{W}|\hat{S},\tilde{U})-\eta\right)}, & \text{if } \bar{\mathbf{w}} \in \mathcal{T}_n(\tilde{W}), \\ 0, & \text{otherwise,} \end{cases}$$

which in turn follows from the fact that given $M'=m'$ and $\mathbf{U}=\mathbf{u}$, $\mathbf{W}(m')$ is uniformly distributed in the set $\mathcal{T}_n\left(P_{\tilde{W}|\hat{S}\tilde{U}},\mathbf{s},\mathbf{u}\right)$ and that for sufficiently large n $\left|\mathcal{T}_n\left(P_{\tilde{W}|\hat{S}\tilde{U}},\mathbf{s},\mathbf{u}\right)\right| \geq e^{n\left(H(\tilde{W}|\hat{S},\tilde{U})-\eta\right)}$.

Next, we analyze the probability of the event $\mathcal{E}'_b$. Let

$$\mathcal{F}'_{2,n}(\eta) := \left\{ P_{\hat{S}\tilde{U}\tilde{V}\tilde{W}\tilde{Y}\bar{W}_d} : \begin{array}{l} P_{\hat{S}\tilde{U}\tilde{W}} \in \mathcal{D}_n(P_{SUW},\eta), P_{\hat{S}\tilde{V}\bar{W}_d\tilde{Y}} \in \mathcal{D}_n(P_{SVWY},\eta) \\ H(\bar{W}_d|\hat{S},\tilde{V},\tilde{Y}) \leq H(\tilde{W}|\hat{S},\tilde{V},\tilde{Y}) \end{array} \right\}.$$

Then,

$$
\begin{aligned}
&\bar{\mathbb{P}}_{Q^{\mathbb{B}_n}}\left(\mathcal{E}'_b\right) \\
&\leq \sum_{\substack{P_{\hat{S}\tilde{U}\tilde{V}\tilde{W}\tilde{Y}\bar{W}_d} \\ \in \mathcal{F}'_{2,n}(\eta)}} \sum_{\substack{(\mathbf{u},\mathbf{v},\bar{\mathbf{w}},\mathbf{y},\mathbf{w}'): \\ (\mathbf{s},\mathbf{u},\mathbf{v},\bar{\mathbf{w}},\mathbf{y},\mathbf{w}') \\ \in \mathcal{T}_n(P_{\hat{S}\tilde{U}\tilde{V}\tilde{W}\tilde{Y}\bar{W}_d})}} \sum_{\substack{m' \in \\ \mathcal{M}'\backslash\{0\}}} \bar{\mathbb{P}}_{Q^{\mathbb{B}_n}}\left(\mathbf{U}=\mathbf{u},\mathbf{V}=\mathbf{v},M'=m',\mathbf{W}(m')=\bar{\mathbf{w}},\mathbf{Y}=\mathbf{y}|\mathbf{S}=\mathbf{s}\right) \\
&\qquad\qquad\qquad\qquad \sum_{\hat{m}' \in \mathcal{M}'\backslash\{0,m'\}} \bar{\mathbb{P}}_{Q^{\mathbb{B}_n}}\left(\bar{\mathbf{W}}(\hat{m}')=\mathbf{w}'|\mathbf{U}=\mathbf{u},M'=m',\mathbf{W}(m')=\bar{\mathbf{w}},\mathbf{S}=\mathbf{s}\right) \\
&\leq \sum_{\substack{P_{\hat{S}\tilde{U}\tilde{V}\tilde{W}\tilde{Y}\bar{W}_d} \\ \in \mathcal{F}'_{2,n}(\eta)}} \sum_{\substack{(\mathbf{u},\mathbf{v},\bar{\mathbf{w}},\mathbf{y}): \\ (\mathbf{s},\mathbf{u},\mathbf{v},\bar{\mathbf{w}},\mathbf{y}) \in \mathcal{T}_n(P_{\hat{S}\tilde{U}\tilde{V}\tilde{W}\tilde{Y}})}} e^{-n\left(H(\tilde{U},\tilde{V})+D\left(P_{\tilde{U}\tilde{V}}||Q_{UV}\right)\right)} e^{-n\left(H(\tilde{W}|\hat{S},\tilde{U})-\eta\right)} \\
&\qquad\qquad e^{-n\left(H(\tilde{Y}|\tilde{U},\hat{S},\tilde{W})+D\left(P_{\tilde{Y}|\tilde{U}\hat{S}\tilde{W}}||P_{Y|USW}|P_{\tilde{U}\hat{S}\tilde{W}}\right)\right)} e^{n\left(\zeta'(\kappa_\alpha,\omega',P_{\hat{S}})+\eta\right)} \frac{2e^{nH(\bar{W}_d|\hat{S},\tilde{V},\tilde{Y})}}{e^{n\left(H(\bar{W}_d)-\eta\right)}} \\
&\leq e^{-n\mathsf{E}'_{2,n}}, \tag{50}
\end{aligned}
$$

where

$$
\begin{aligned}
\mathsf{E}'_{2,n} &\gtrsim \min_{(P_{\tilde{U}\tilde{V}\tilde{W}\tilde{Y}S},Q_{\tilde{U}\tilde{V}\tilde{W}\tilde{Y}S}) \in \mathcal{T}'_2(\kappa_\alpha,\omega',P_S,P_{X|USW})} D(P_{\tilde{U}\tilde{V}\tilde{W}\tilde{Y}|S}||Q_{UVWY|S}|P_S) + \rho'(\kappa_\alpha,\omega',P_S,P_{X|USW}) \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad - \zeta'(\kappa_\alpha,\omega',P_S) - O(\eta) \\
&= E'_2(\kappa_\alpha,\omega',P_S,P_{X|USW}) - O(\eta).
\end{aligned}
$$

Finally, considering the event $\mathcal{E}'_c$, we have

$$\bar{\mathbb{P}}_{Q^{\mathbb{B}_n}}\left(\mathcal{E}'_c\right) = \sum_{\substack{\mathbf{u}\in\mathcal{T}_n(P_{\tilde{U}}):\\P_{\tilde{U}}\in\mathcal{D}_n(P_U,\eta)}} \bar{\mathbb{P}}_{Q^{\mathbb{B}_n}}\left(\mathbf{U}=\mathbf{u},\mathcal{E}'_{\text{EE}},\mathcal{E}'_c|\mathbf{S}=\mathbf{s}\right) + \sum_{\substack{\mathbf{u}\in\mathcal{T}_n(P_{\tilde{U}}):\\P_{\tilde{U}}\notin\mathcal{D}_n(P_U,\eta)}} \bar{\mathbb{P}}_{Q^{\mathbb{B}_n}}\left(\mathbf{U}=\mathbf{u},\mathcal{E}'_c|\mathbf{S}=\mathbf{s}\right).$$

The first term in the RHS decays double exponentially as $e^{-e^{n\Omega(\eta)}}$, while the second term can be handled as follows:

$$\sum_{\substack{\mathbf{u}\in\mathcal{T}_n(P_{\tilde{U}}):P_{\tilde{U}}\notin\mathcal{D}_n(P_U,\eta)}} \bar{\mathbb{P}}_{Q^{\mathbb{B}_n}}\left(\mathbf{U}=\mathbf{u},\mathcal{E}'_c|\mathbf{S}=\mathbf{s}\right)$$

$$\leq \sum_{\substack{\mathbf{u}\in\mathcal{T}_n(P_{\tilde{U}}):\\P_{\tilde{U}}\notin\mathcal{D}_n(P_U,\eta)}} \sum_{\substack{(\mathbf{v},\mathbf{y},\mathbf{w}'):\\(\mathbf{s},\mathbf{v},\mathbf{y},\mathbf{w}')\in\mathcal{T}_n\left(P_{\hat{S}\tilde{V}\tilde{Y}\tilde{W}_d}\right),\\P_{\hat{S}\tilde{V}\tilde{W}_d\tilde{Y}}\in\mathcal{D}_n(P_{SVWY},\eta)}} \sum_{\hat{m}\in\mathcal{M}\setminus\{0\}} \bar{\mathbb{P}}_{Q^{\mathbb{B}_n}}\left(\mathbf{U}=\mathbf{u},\mathbf{V}=\mathbf{v},M'=0,\mathbf{Y}=\mathbf{y}|\mathbf{S}=\mathbf{s}\right)$$

$$\sum_{\hat{m}'\in\mathcal{M}\setminus\{0\}} \bar{\mathbb{P}}_{Q^{\mathbb{B}_n}}\left(\mathbf{W}(\hat{m}')=\bar{\mathbf{w}}\right)$$

$$\leq \sum_{\substack{P_{\tilde{U}\hat{S}\tilde{V}\tilde{W}_d\tilde{Y}}\in\mathcal{D}_n(P_U,\eta)^c\\\times\mathcal{D}_n(P_{SVWY},\eta)}} e^{nH(\tilde{U},\tilde{V},\tilde{Y}|\hat{S})}e^{-n\left(H(\tilde{U},\tilde{V},\tilde{Y}|\hat{S})+D\left(P_{\tilde{U}\tilde{V}\tilde{Y}|\hat{s}}||Q_{UVY'|\hat{s}}|P_{\hat{s}}\right)\right)}\frac{e^{nH(\tilde{W}_d|\hat{S},\tilde{V},\tilde{Y})}e^{n(R'+\eta)}}{e^{n\left(H(\tilde{W}_d)-\eta\right)}}$$

$$\leq e^{-n\mathsf{E}'_{3,n}}, \tag{51}$$

where

$$\mathsf{E}'_{3,n} \gtrsim \min_{\substack{P_{\tilde{V}\tilde{Y}S}:P_{\tilde{U}\tilde{V}\tilde{W}\tilde{Y}S}\in\\\hat{\mathcal{L}}_h(\kappa_\alpha,\omega',P_S,P_{X|USW})}} D\left(P_{\tilde{V}\tilde{Y}|S}||Q_{VY'|S}|P_S\right) + \rho'(\kappa_\alpha,\omega',P_S,P_{X|USW}) - \zeta'(\kappa_\alpha,\omega',P_S) - O(\eta)$$

$$= E'_3(\kappa_\alpha,\omega',P_S,P_{X|USW},P_{X'|US}) - O(\eta).$$

Since the exponent of the type II error probability is lower bounded by the minimum of the exponent of the type II error-causing events, it follows from (49), (50) and (51) that for a fixed $\left(P_S,\omega'(\cdot,P_S),P_{X|USW},P_{X'|US}\right)\in\mathcal{L}_h(\kappa_\alpha)$

$$\bar{\mathbb{P}}_{P^{(\mathbb{B}_n)}}\left(\hat{H}=1\right)\leq e^{-n(\kappa_\alpha-O(\eta))}, \tag{52a}$$

$$\bar{\mathbb{P}}_{Q^{(\mathbb{B}_n)}}\left(\hat{H}=0\right)\leq e^{-n\left(\bar{\kappa}_h(\kappa_\alpha,\omega',P_S,P_{X|USW},P_{X'|US})-O(\eta)\right)}, \tag{52b}$$

where $\bar{\kappa}_h = \min\left\{E'_1(\kappa_\alpha,\omega'),E'_2(\kappa_\alpha,\omega',P_S,P_{X|USW}),E'_3(\kappa_\alpha,\omega',P_S,P_{X|USW},P_{X'|US})\right\}$. Performing expurgation as in the proof of Theorem 1 to obtain a deterministic codebook $\mathcal{B}_n$ satisfying (52a, 52b), maximizing over $\left(P_S,\omega'(\cdot,P_S),P_{X|USW},P_{X'|US}\right)\in\mathcal{L}_h(\kappa_\alpha)$ and noting that $\eta>0$ is arbitrary yields $\kappa(\kappa_\alpha)\geq\kappa^\star_{\mathsf{h}}(\kappa_\alpha)$.

Finally, we show that $\kappa(\kappa_\alpha)\geq\kappa^\star_{\mathsf{u}}(\kappa_\alpha)$, which will complete the proof. Fix $P_{X|US}$ and let $P_{UVXY}:=P_{UV}P_{X|US}P_{Y|X}$ and $Q_{UVXY}:=Q_{UV}P_{X|US}P_{Y|X}$. Consider an uncoded transmission scheme in which the channel input $\mathbf{X}\sim f_n(\cdot|\mathbf{u})=P^{\otimes n}_{X|US}(\cdot|\mathbf{u},\mathbf{s})$. Let the decision rule g_n be specified by the acceptance region $\mathcal{A}_n=\left\{(\mathbf{s},\mathbf{v},\mathbf{y}):D\left(P_{\mathbf{vy}|\mathbf{s}}||P_{VY|S}|P_\mathbf{s}\right)\leq\kappa_\alpha+\eta\right\}$ for some small $\eta>0$. Then, it follows from Lemma 2.6 in [42] that for sufficiently large n,

$$\alpha_n(f_n,g_n)=P^{\otimes n}_{VY|S}(\mathcal{A}^c_n|\mathbf{s})\leq e^{-n\kappa_\alpha},$$

$$\beta_n(f_n,g_n)=Q^{\otimes n}_{VY|S}(\mathcal{A}_n|\mathbf{s})\leq e^{-n(\kappa^\star_{\mathsf{u}}(\kappa_\alpha)-O(\eta))}.$$

The proof is complete by noting that $\eta>0$ is arbitrary.

5. Conclusions

This work explored the trade-off between the type I and type II error-exponents for distributed hypothesis testing over a noisy channel. We proposed a separate hypothesis

testing and channel coding scheme as well as a joint scheme utilizing hybrid coding, and analyzed their performance resulting in two inner bounds on the error-exponents trade-off. The separate scheme recovers some of the existing bounds in the literature as special cases. We also showed via an example of testing against dependence that the joint scheme strictly outperforms the separate scheme at some points of the error-exponents trade-off. An interesting avenue for future research is the exploration of novel outer bounds that could shed light on the scenarios where the separate or joint schemes are tight.

Author Contributions: Conceptualization, S.S. and D.G.; writing—original draft preparation, S.S. and D.G.; supervision, D.G. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: Not applicable.

Abbreviations

The following abbreviations are used in this manuscript:

HT	Hypothesis testing
DHT	Distributed hypothesis testing
TAD	Testing against dependence
TAI	Testing against independence
SHTCC	Separate hypothesis testing and channel coding
JHTCC	Joint hypothesis testing and channel coding

Appendix A. Proof that Theorem 1 Recovers Theorem 2 in [10]

We prove that $\lim_{\kappa_\alpha \to 0} \kappa_s^\star(\kappa_\alpha) = \kappa_s$, where κ_s is the lower bound on the type II error-exponent for a fixed type I error probability constraint and unit bandwidth ratio established in Theorem 2 in [10]. Note that $\hat{\mathcal{L}}(0, \omega) = \{P_{UVW} = P_{UV}P_{W|U}, P_{W|U} = \omega(P_U)\}$, $\zeta(0, \omega) = I_P(U; W)$, and $\rho(0, \omega) = I_P(V; W)$. The result then follows from Theorem 1 by noting that $\hat{\mathcal{L}}(\kappa_\alpha, \omega)$, $\zeta(\kappa_\alpha, \omega)$ and $\rho(\kappa_\alpha, \omega)$ are continuous in κ_α and the fact that $E_{sp}(P_{SX}, \theta)$, $E_{ex}(R, P_{SX})$ and $E_b(\kappa_\alpha, \omega, R)$ are all greater than or equal to zero.

Appendix B. Proof that Theorem 2 Recovers Theorem 5 in [10]

We show that $\lim_{\kappa_\alpha \to 0} \kappa_h^\star(\kappa_\alpha) = \kappa_h$, where κ_h is as defined in Theorem 5 in [10]. Note that $\hat{\mathcal{L}}_h(0, \omega', P_S, P_{X|USW}) := \{P_{UV\hat{W}YS} : P_{SUVWXY} := P_S P_{UV} P_{W|US} P_{X|USW} P_{Y|X}, P_{W|US} = \omega'(P_U, P_S)\}$, $\zeta'(0, \omega', P_S) := I_P(U; W|S)$, $\rho(0, \omega', P_S, P_{X|USW}) = I_P(Y, V; W|S)$,

$$\mathcal{L}_h(0) := \{(P_S, \omega'(\cdot, P_S), P_{X|USW}, P_{X'|US}) : I_P(U; W|S) < I_P(Y, V; W|S)\},$$

and $E'_b(0, \omega', P_S, P_{X|USW}) = I_P(Y, V; W|S) - I_P(U; W|S)$. The result then follows from Theorem 2 via the continuity of $\hat{\mathcal{L}}_h(\kappa_\alpha, \cdot, \cdot, \cdot)$, $\zeta'(\kappa_\alpha, \cdot, \cdot)$, $\rho(\kappa_\alpha, \cdot, \cdot, \cdot)$, $\mathcal{L}_h(\kappa_\alpha)$ and $E'_b(\kappa_\alpha, \cdot, \cdot, \cdot)$ in κ_α.

Appendix C. An Auxiliary Result

Here, we prove a result which was used in the proof of Theorem 1, namely Proposition A1 given below. For this purpose, we require a few properties of log-moment generating function, which we briefly review next.

Lemma A1. *(Theorem 15.3, Theorem 15.6 in [43])*

(i) $\psi_{P_Z,f}(0) = 0$ and $\psi'_{P_Z,f}(0) = \mathbb{E}_{P_Z}[f(Z)]$, where $\psi'_{P_Z,f}(\lambda)$ denotes the derivative of $\psi_{P_Z,f}(\lambda)$ with respect to λ.

(ii) $\psi_{P_Z,f}(\lambda)$ is a strictly convex function in λ.

(iii) $\psi^*_{P_Z,f}(\theta)$ is strictly convex and strictly positive in θ except $\psi^*_{P_Z,f}(\mathbb{E}_{P_Z}[Z]) = 0$.

Proposition A1 is basically a characterization of the error-exponent region of a hypothesis testing problem, which we introduce next. Let $P_{X_0 X_1} \in \mathcal{P}(\mathcal{X}^2)$ be an arbitrary joint PMF, and consider a sequence of pairs of n-length sequences $(\tilde{\mathbf{x}}, \mathbf{x}')$ such that

$$P_{\tilde{\mathbf{x}}\mathbf{x}'}(\tilde{x}, x') \xrightarrow{(n)} P_{X_0 X_1}(\tilde{x}, x'), \ \forall \ (\tilde{x}, x') \in \mathcal{X}^2. \tag{A1}$$

Consider the following HT:

$$H_0 : \mathbf{Y} \sim P^{\otimes n}_{Y|X}(\cdot|\tilde{\mathbf{x}}), \tag{A2a}$$

$$H_1 : \mathbf{Y} \sim P^{\otimes n}_{Y|X}(\cdot|\mathbf{x}'). \tag{A2b}$$

With the achievability of an error-exponent pair $(\kappa_\alpha, \kappa_\beta)$ defined similar to Definition 1, consider the error-exponent region of interest

$$\mathcal{R}'(P_{X_0 X_1}) := \{(\kappa_\alpha, \kappa'(\kappa_\alpha, P_{X_0 X_1})) : \kappa_\alpha \in (0, \kappa'^\star_\alpha)\},$$

where $\kappa'(\kappa_\alpha, P_{X_0 X_1}) := \sup\{\kappa_\beta : (\kappa_\alpha, \kappa_\beta) \text{ is achievable for HT in (A2)}\}$ and $\kappa'^\star_\alpha := \inf\{\kappa_\alpha : \kappa'(\kappa_\alpha, P_{X_0 X_1}) = 0\}$. We mention at this point that the notation $\mathcal{R}'(P_{X_0 X_1})$ is justified as the error-exponent region for the above hypothesis test depends on $(\tilde{\mathbf{x}}, \mathbf{x}')$ only through its limiting joint type $P_{X_0 X_1}$, as will be evident later. Given this, the following proposition provides a single-letter characterization of $\mathcal{R}'(P_{X_0 X_1})$.

Proposition A1.

$$\mathcal{R}'(P_{X_0 X_1}) = \bigcup_{\theta \in \mathcal{I}(P_{X_0 X_1})} \left(\mathbb{E}_{P_{X_0 X_1}}\left[\psi^*_{P_{Y|X}(\cdot|X_0), \Pi_{X_0, X_1}}(\theta)\right], \ \mathbb{E}_{P_{X_0 X_1}}\left[\psi^*_{P_{Y|X}(\cdot|X_0), \Pi_{X_0, X_1}}(\theta)\right] - \theta \right)$$

where $\Pi_{\tilde{x}, x'}(y) := \log\left(P_{Y|X}(y|x')/P_{Y|X}(y|\tilde{x})\right)$ for $(\tilde{x}, x') \in \mathcal{X}^2$,

$$\mathcal{I}(P_{X_0 X_1}) := \left(-d_{\min}(P_{X_0 X_1}), d_{\max}(P_{X_0 X_1})\right),$$
$$d_{\min}(P_{X_0 X_1}) := \mathbb{E}_{P_{X_0 X_1}}\left[D\left(P_{Y|X}(\cdot|X_0)||P_{Y|X}(\cdot|X_1)\right)\right]$$
$$d_{\max}(P_{X_0 X_1}) := \mathbb{E}_{P_{X_0 X_1}}\left[D\left(P_{Y|X}(\cdot|X_1)||P_{Y|X}(\cdot|X_0)\right)\right].$$

Proof. Let $(\tilde{\mathbf{x}}, \mathbf{x}') \in \mathcal{X}^n \times \mathcal{X}^n$ be sequences that satisfy (A1). For simplicity, we will denote $d_{\max}(P_{X_0 X_1})$ and $d_{\min}(P_{X_0 X_1})$ by $d_{\max}$ and $d_{\min}$, respectively.
Achievability: We will show that for $-d_{\min} < \theta < d_{\max}$,

$$\kappa'\left(\mathbb{E}_{P_{X_0 X_1}}\left[\psi^*_{P_{Y|X}(\cdot|X_0), \Pi_{X_0, X_1}}(\theta)\right], P_{X_0 X_1}\right) \geq \mathbb{E}_{P_{X_0 X_1}}\left[\psi^*_{P_{Y|X}(\cdot|X_0), \Pi_{X_0, X_1}}(\theta)\right] - \theta.$$

Consider the Neyman–Pearson test given by $g_n(\mathbf{y}) = \mathbb{1}_{\left\{\Pi^{(n)}_{\tilde{\mathbf{x}}, \mathbf{x}'}(\mathbf{y}) \geq n\theta\right\}}$, where $\Pi^{(n)}_{\tilde{\mathbf{x}}, \mathbf{x}'}(\mathbf{y}) := \sum_{i=1}^n \Pi_{\tilde{x}_i, x'_i, P_{Y|X}}(y_i)$. Observe that the type I error probability can be upper bounded for $\theta > -d_{\min}$ and sufficiently large n as follows:

$$\alpha_n(g_n) = \mathbb{P}_{P^{\otimes n}_{Y|X}(\cdot|\tilde{\mathbf{x}})}\left(\Pi^{(n)}_{\tilde{\mathbf{x}}, \mathbf{x}'}(\mathbf{Y}) \geq n\theta\right)$$

$$\overset{(a)}{\leq} e^{-\sup_{\lambda \geq 0} n\theta\lambda - \psi_{P^{\otimes n}_{Y|X}(\cdot|\tilde{\mathbf{x}}), \Pi^{(n)}_{\tilde{\mathbf{x}}, \mathbf{x}'}}(\lambda)}$$

$$\overset{(b)}{=} e^{-\sup_{\lambda \in \mathbb{R}} n\left(\theta\lambda - \frac{1}{n}\psi_{P_{Y|X}^{\otimes n}(\cdot|\tilde{\mathbf{x}}),\Pi_{\tilde{\mathbf{x}},\mathbf{x}'}^{(n)}}(\lambda)\right)}, \tag{A3}$$

where (a) follows from the Chernoff bound, and (b) holds because for $\theta > -d_{\min}$ and sufficiently large n, the supremum in (A3) always occurs at $\lambda \geq 0$. To see this, note that the term $l_n(\lambda) := \theta\lambda - n^{-1}\psi_{P_{Y|X}^{\otimes n}(\cdot|\tilde{\mathbf{x}}),\Pi_{\tilde{\mathbf{x}},\mathbf{x}'}^{(n)}}(\lambda)$ is a concave function of λ by Lemma A1 (i). Additionally, denoting its derivative with respect to λ by $l'_n(\lambda)$, we have

$$l'_n(0) = \theta - \frac{1}{n}\mathbb{E}_{P_{Y|X}^{\otimes n}(\cdot|\tilde{\mathbf{x}})}\left[\Pi_{\tilde{\mathbf{x}},\mathbf{x}'}^{(n)}\right] \tag{A4}$$

$$= \theta - \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{P_{Y|X}(\cdot|\tilde{x}_i)}\left[\log\left(P_{Y|X}(Y_i|x_i')/P_{Y|X}(Y_i|\tilde{x}_i)\right)\right]$$

$$\xrightarrow{(n)} \theta + d_{\min} > 0, \tag{A5}$$

where (A4) follows from Lemma A1 (iii), and (A5) is due to the absolute continuity assumption, $P_{Y|X}(\cdot|x) \ll P_{Y|X}(\cdot|x')$, $\forall\,(x,x') \in \mathcal{X}^2$ on the channel, and (A1). Thus, by the concavity of $l_n(\lambda)$, its supremum has to occur at $\lambda \geq 0$. Simplifying the term within the exponent in (A3), we obtain

$$\frac{1}{n}\psi_{P_{Y|X}^{\otimes n}(\cdot|\mathbf{x}),\Pi_{\tilde{\mathbf{x}},\mathbf{x}'}^{(n)}}(\lambda) = \sum_{\tilde{x},x'}P_{\tilde{\mathbf{x}}\mathbf{x}'}(\tilde{x},x')\log\left(\mathbb{E}_{P_{Y|X}(\cdot|\tilde{x})}\left[\left(P_{Y|X}(Y|x')/P_{Y|X}(Y|\tilde{x})\right)^{\lambda}\right]\right) \tag{A6}$$

$$\xrightarrow{(n)} \mathbb{E}_{P_{X_0X_1}}\left[\log\left(\mathbb{E}_{P_{Y|X}(\cdot|X_0)}\left[e^{\lambda\Pi_{X_0,X_1}(Y)}\right]\right)\right], \tag{A7}$$

where (A7) follows from (A1) and the absolute continuity assumption on $P_{Y|X}$. Substituting (A7) in (A3) and from (1), we obtain for arbitrarily small (but fixed) $\delta > 0$ and sufficiently large n, that

$$\alpha_n(g_n) \leq e^{-\sup_{\lambda \in \mathbb{R}}\left(n\left(\theta\lambda - \mathbb{E}_{P_{X_0X_1}}\left[\log\left(\mathbb{E}_{P_{Y|X}(\cdot|X_0)}\left[e^{\lambda\Pi_{X_0,X_1}(Y)}\right]\right)\right]\right) - \delta\right)}$$

$$= e^{-n\left(\mathbb{E}_{P_{X_0X_1}}\left[\sup_{\lambda \in \mathbb{R}}\left(\theta\lambda - \mathbb{E}_{P_{Y|X}(\cdot|X_0)}\left(e^{\lambda\Pi_{X_0,X_1}(Y)}\right)\right)\right] - \delta\right)}$$

$$= e^{-n\left(\mathbb{E}_{P_{X_0X_1}}\left[\psi^*_{P_{Y|X}(\cdot|X_0),\Pi_{X_0,X_1}}(\theta)\right] - \delta\right)}. \tag{A8}$$

Similarly, it can be shown that for $\theta < d_{\max}$,

$$\beta_n(g_n) \leq e^{-n\left(\mathbb{E}_{P_{X_0X_1}}\left[\psi^*_{P_{Y|X}(\cdot|X_1),\Pi_{X_0,X_1}}(\theta)\right] - \delta\right)}. \tag{A9}$$

Moreover, for $(\tilde{x},x') \in \mathcal{X}^2$, we have

$$e^{\psi_{P_{Y|X}(\cdot|x'),\Pi_{\tilde{x},x'}}(\lambda)} = \sum_{y\in\mathcal{Y}}P_{Y|X}^{\lambda+1}(\cdot|x')/P_{Y|X}^{\lambda}(\cdot|\tilde{x}) = e^{\psi_{P_{Y|X}(\cdot|\tilde{x}),\Pi_{\tilde{x},x'}}(\lambda+1)}.$$

It follows that

$$\psi^*_{P_{Y|X}(\cdot|x'),\Pi_{\tilde{x},x'}}(\theta) := \sup_{\lambda\in\mathbb{R}}\lambda\theta - \psi_{P_{Y|X}(\cdot|x'),\Pi_{\tilde{x},x'}}(\lambda)$$

$$= \sup_{\lambda\in\mathbb{R}}\lambda\theta - \psi_{P_{Y|X}(\cdot|\tilde{x}),\Pi_{\tilde{x},x'}}(\lambda+1)$$

$$= \psi^*_{P_{Y|X}(\cdot|\tilde{x}),\Pi_{\tilde{x},x'}}(\theta) - \theta.$$

Hence,

$$\mathbb{E}_{P_{X_0 X_1}}\left[\psi^*_{P_{Y|X}(\cdot|X_1),\Pi_{X_0,X_1}}(\theta)\right] = \mathbb{E}_{P_{X_0 X_1}}\left[\psi^*_{P_{Y|X}(\cdot|X_0),\Pi_{X_0,X_1}}(\theta)\right] - \theta.$$

From this, (A8) and (A9), we obtain for $-d_{\min} < \theta < d_{\max}$ that

$$\kappa'\left(\mathbb{E}_{P_{X_0 X_1}}\left[\psi^*_{P_{Y|X}(\cdot|X_0),\Pi_{X_0,X_1}}(\theta)\right] - \delta, P_{X_0 X_1}\right) \ge \mathbb{E}_{P_{X_0 X_1}}\left[\psi^*_{P_{Y|X}(\cdot|X_0),\Pi_{X_0,X_1}}(\theta)\right] - \theta - \delta.$$

Then, the proof of achievability is completed by noting that $\delta > 0$ is arbitrary and $\kappa'(\kappa_\alpha, P_{X_0 X_1})$ is a continuous function of κ_α for a fixed $P_{X_0 X_1}$.

Converse: Let $\mathcal{I}_n(\tilde{x}, x') := \{i \in [n] : \tilde{x}_i = \tilde{x} \text{ and } x'_i = x'\}$. For any $\theta \in \mathbb{R}$ and decision function g_n, we have from Theorem 14.9 in [43] that

$$\alpha_n(g_n) + e^{-n\theta}\beta_n(g_n) \ge \mathbb{P}_{P^{\otimes n}_{Y|X}(\cdot|\tilde{\mathbf{x}})}\left(\log\left(P^{\otimes n}_{Y|X}(\mathbf{Y}|\mathbf{x}')/P^{\otimes n}_{Y|X}(\mathbf{Y}|\tilde{\mathbf{x}})\right) \ge n\theta\right).$$

Simplifying the RHS above, we obtain

$$\mathbb{P}_{P^{\otimes n}_{Y|X}(\cdot|\tilde{\mathbf{x}})}\left(\log\left(P^{\otimes n}_{Y|X}(\mathbf{Y}|\mathbf{x}')/P^{\otimes n}_{Y|X}(\mathbf{Y}|\tilde{\mathbf{x}})\right) \ge n\theta\right)$$

$$= \mathbb{P}_{P^{\otimes n}_{Y|X}(\cdot|\tilde{\mathbf{x}})}\left(\sum_{\tilde{x},x'}\sum_{i\in\mathcal{I}_n(\tilde{x},x')}\log\left(P_{Y|X}(Y_i|x'_i)/P_{Y|X}(Y_i|\tilde{x}_i)\right) \ge n\theta\right)$$

$$= \mathbb{P}_{P^{\otimes n}_{Y|X}(\cdot|\tilde{\mathbf{x}})}\left(\sum_{\tilde{x},x'}\sum_{i\in\mathcal{I}_n(\tilde{x},x')}\log\left(P_{Y|X}(Y_i|x'_i)/P_{Y|X}(Y_i|\tilde{x}_i)\right) \ge \sum_{(\tilde{x},x')\in\mathcal{X}^2} n P_{\tilde{\mathbf{x}}\mathbf{x}'}(\tilde{x},x')\theta\right)$$

$$\stackrel{(a)}{\ge} \mathbb{P}_{P^{\otimes n}_{Y|X}(\cdot|\tilde{\mathbf{x}})}\left(\bigcap_{\tilde{x},x'}\left(\sum_{i\in\mathcal{I}_n(\tilde{x},x')}\log\left(P_{Y|X}(Y_i|x'_i)/P_{Y|X}(Y_i|\tilde{x}_i)\right) \ge n P_{\tilde{\mathbf{x}}\mathbf{x}'}(\tilde{x},x')\theta\right)\right)$$

$$\stackrel{(b)}{=} \prod_{(\tilde{x},x')\in\mathcal{X}^2} \mathbb{P}_{P^{\otimes n}_{Y|X}(\cdot|\tilde{\mathbf{x}})}\left(\sum_{i\in\mathcal{I}_n(\tilde{x},x')}\log\left(P_{Y|X}(Y_i|x'_i)/P_{Y|X}(Y_i|\tilde{x}_i)\right) \ge n P_{\tilde{\mathbf{x}}\mathbf{x}'}(\tilde{x},x')\theta\right),$$

where

(a)   follows since

$$\bigcap_{\tilde{x},x'}\left\{\sum_{i\in\mathcal{I}_n(\tilde{x},x')}\log\left(P_{Y|X}(Y_i|x'_i)/P_{Y|X}(Y_i|\tilde{x}_i)\right) \ge n P_{\tilde{\mathbf{x}}\mathbf{x}'}(\tilde{x},x')\theta\right\}$$

$$\subseteq \left\{\sum_{\tilde{x},x'}\sum_{i\in\mathcal{I}_n(\tilde{x},x')}\log\left(P_{Y|X}(Y_i|x'_i)/P_{Y|X}(Y_i|\tilde{x}_i)\right) \ge \sum_{(\tilde{x},x')\in\mathcal{X}^2} n P_{\tilde{\mathbf{x}}\mathbf{x}'}(\tilde{x},x')\theta\right\};$$

(b)   is due to the independence of the events $\big\{\sum_{i\in\mathcal{I}_n(\tilde{x},x')}\log\left(P_{Y|X}(Y_i|x'_i)/P_{Y|X}(Y_i|\tilde{x}_i)\right) \ge n P_{\tilde{\mathbf{x}}\mathbf{x}'}(\tilde{x},x')\theta\big\}$ for different $(\tilde{x},x')\in\mathcal{X}^2$.

Define $b_{\tilde{x},x'}(\theta) := \min_{\tilde{Q}_{\tilde{x}}\in\mathcal{P}(\mathcal{Y}):\mathbb{E}_{\tilde{Q}_{\tilde{x}}}[\log(P_{Y|X}(Y|x')/P_{Y|X}(Y|\tilde{x}))]\ge\theta} D\left(\tilde{Q}_x||P_{Y|X}(\cdot|\tilde{x})\right)$. Then, for arbitrary $\delta > 0$, $\delta' > \delta$ and sufficiently large n, we can write

$$\alpha_n + e^{-n\theta}\beta_n \stackrel{(a)}{\ge} \prod_{(\tilde{x},x')\in\mathcal{X}^2} e^{-n P_{\tilde{\mathbf{x}}\mathbf{x}'}(\tilde{x},x')\left(b_{\tilde{x},x'}(\theta)+\delta\right)}$$

$$\stackrel{(b)}{\ge} \prod_{(\tilde{x},x')\in\mathcal{X}^2} e^{-n P_{\tilde{\mathbf{x}}\mathbf{x}'}(\tilde{x},x')\left(\psi^*_{P_{Y|X}(\cdot|\tilde{x}),\Pi_{\tilde{x},x'}}(\theta)+\delta\right)}$$

$$\stackrel{(c)}{=} e^{-n\left(\mathbb{E}_{P_{X_0 X_1}}\left[\psi^*_{P_{Y|X}(\cdot|X_0),\Pi_{X_0,X_1}}(\theta)\right]+\delta'\right)},$$

where (a) follows from Theorem 15.9 in [43]; (b) follows since $b_{\tilde{x},x'}(\theta) = \psi^*_{P_{Y|X}(\cdot|\tilde{x}),\Pi_{\tilde{x},x'}}(\theta)$ from Theorem 15.6 in [43] and Theorem 15.11 in [43]; and (c) is due to (A1). The equation above implies that

$$\limsup_{n\to\infty} \min\left\{-\frac{1}{n}\log\alpha_n, -\frac{1}{n}\log\beta_n + \theta\right\} \le \mathbb{E}_{P_{X_0 X_1}}\left[\psi^*_{P_{Y|X}(\cdot|X_0),\Pi_{X_0,X_1}}(\theta)\right] + \delta'.$$

Hence, if $-\log(\alpha_n)/n > \mathbb{E}_{P_{X_0 X_1}}\left[\psi^*_{P_{Y|X}(\cdot|X_0),\Pi_{X_0,X_1}}(\theta)\right] + \delta'$ for all sufficiently large n, then

$$\limsup_{n\to\infty} -\frac{1}{n}\log\beta_n \le \mathbb{E}_{P_{X_0 X_1}}\left[\psi^*_{P_{Y|X}(\cdot|X_0),\Pi_{X_0,X_1}}(\theta)\right] - \theta + \delta'.$$

Since δ (and δ') is arbitrary, this implies via the continuity of $\kappa'(\kappa_\alpha, P_{X_0 X_1})$ in κ_α that

$$\kappa'\left(\mathbb{E}_{P_{X_0 X_1}}\left[\psi^*_{P_{Y|X}(\cdot|X_0),\Pi_{X_0,X_1}}(\theta)\right], P_{X_0 X_1}\right) \le \mathbb{E}_{P_{X_0 X_1}}\left[\psi^*_{P_{Y|X}(\cdot|X_0),\Pi_{X_0,X_1}}(\theta)\right] - \theta.$$

To complete the proof, we need to show that θ can be restricted to lie in $\mathcal{I}(P_{X_0 X_1})$. Toward this, it suffices to show the following:

(i) $\mathbb{E}_{P_{X_0 X_1}}\left[\psi^*_{P_{Y|X}(\cdot|X_0),\Pi_{X_0,X_1}}(-d_{\min})\right] = 0,$

(ii) $\mathbb{E}_{P_{X_0 X_1}}\left[\psi^*_{P_{Y|X}(\cdot|X_0),\Pi_{X_0,X_1}}(d_{\max})\right] = d_{\max}$, and

(iii) $\mathbb{E}_{P_{X_0 X_1}}\left[\psi^*_{P_{Y|X}(\cdot|X_0),\Pi_{X_0,X_1}}(\theta)\right]$ and $\mathbb{E}_{P_{X_0 X_1}}\left[\psi^*_{P_{Y|X}(\cdot|X_0),\Pi_{X_0,X_1}}(\theta)\right] - \theta$ are convex functions of θ.

We have

$$\mathbb{E}_{P_{X_0 X_1}}\left[\psi^*_{P_{Y|X}(\cdot|X_0),\Pi_{X_0,X_1}}(-d_{\min})\right]$$

$$:= \sup_{\lambda\in\mathbb{R}} -\lambda\, \mathbb{E}_{P_{X_0 X_1}}\left[D\left(P_{Y|X}(\cdot|X_0)||P_{Y|X}(\cdot|X_1)\right)\right] - \mathbb{E}_{P_{X_0 X_1}}\left[\psi_{P_{Y|X}(\cdot|X_0),\Pi_{X_0,X_1}}(\lambda)\right]$$

$$\le \sum_{\tilde{x},x'} P_{X_0 X_1}(\tilde{x},x')\left[\sup_{\lambda_{\tilde{x},x'}\in\mathbb{R}} -\lambda_{\tilde{x},x'} D\left(P_{Y|X}(\cdot|\tilde{x})||P_{Y|X}(\cdot|x')\right) - \psi_{P_{Y|X}(\cdot|\tilde{x}),\Pi_{\tilde{x},x'}}(\lambda_{\tilde{x},x'})\right]$$

$$= 0, \tag{A10}$$

where (A10) follows since each term inside the square braces in the penultimate equation is zero, which in turn follows from Lemma A1 (iii). Additionally,

$$\mathbb{E}_{P_{X_0 X_1}}\left[\psi^*_{P_{Y|X}(\cdot|X_0),\Pi_{X_0,X_1}}(-d_{\min})\right] = \sum_{\tilde{x},x'} P_{X_0 X_1}(\tilde{x},x')\,\psi^*_{P_{Y|X}(\cdot|\tilde{x}),\Pi_{\tilde{x},x'}}(-d_{\min}) \ge 0, \tag{A11}$$

where (A11) follows from the non-negativity of $\psi^*_{P_{Y|X}(\cdot|\tilde{x}),\Pi_{\tilde{x},x'}}$ for every $(\tilde{x},x') \in \mathcal{X}^2$ stated in Lemma A1 (iii). Combining (A10) and (A11) proves (i). We also have

$$\mathbb{E}_{P_{X_0 X_1}}\left[\psi^*_{P_{Y|X}(\cdot|X_0),\Pi_{X_0,X_1}}(d_{\max})\right] - d_{\max} = \mathbb{E}_{P_{X_0 X_1}}\left[\psi^*_{P_{Y|X}(\cdot|X_1),\Pi_{X_0,X_1}}(d_{\max})\right] = 0, \tag{A12}$$

where the final equality follows similarly to the proof of (i). This proves (ii). Finally, (iii) follows from Lemma A1 (iii) and the fact that a weighted sum of convex functions is convex provided the weights are non-negative, thus completing the proof. $\quad\square$

References

1. Neyman, J.; Pearson, E. On the Problem of the Most Efficient Tests of Statistical Hypotheses. *Philos. Trans. R. Soc. Lond.* **1933**, *231*, 289–337.
2. Chernoff, H. A measure of asymptotic efficiency for tests of a hypothesis based on a sum of observations. *Ann. Math. Statist.* **1952**, *23*, 493–507. [CrossRef]
3. Hoeffding, W. Asymptotically optimal tests for multinominal distributions. *Ann. Math. Statist.* **1965**, *36*, 369–400. [CrossRef]
4. Blahut, R.E. Hypothesis Testing and Information Theory. *IEEE Trans. Inf. Theory* **1974**, *20*, 405–417. [CrossRef]
5. Tuncel, E. On error-exponents in hypothesis testing. *IEEE Trans. Inf. Theory* **2005**, *51*, 2945–2950. [CrossRef]
6. Ahlswede, R.; Csiszár, I. Hypothesis Testing with Communication Constraints. *IEEE Trans. Inf. Theory* **1986**, *32*, 533–542. [CrossRef]
7. Han, T.S. Hypothesis Testing with Multiterminal Data Compression. *IEEE Trans. Inf. Theory* **1987**, *33*, 759–772. [CrossRef]
8. Shimokawa, H.; Han, T.S.; Amari, S. Error Bound of Hypothesis Testing with Data Compression. In Proceedings of the IEEE International Symposium on Information Theory, Trondheim, Norway, 27 June–1 July 1994; p. 114.
9. Rahman, M.S.; Wagner, A.B. On the Optimality of Binning for Distributed Hypothesis Testing. *IEEE Trans. Inf. Theory* **2012**, *58*, 6282–6303. [CrossRef]
10. Sreekumar, S.; Gündüz, D. Distributed Hypothesis Testing Over Discrete Memoryless Channels. *IEEE Trans. Inf. Theory* **2020**, *66*, 2044–2066. [CrossRef]
11. Salehkalaibar, S.; Wigger, M. Distributed Hypothesis Testing Based on Unequal-Error Protection Codes. *IEEE Trans. Inf. Theory* **2020**, *66*, 4150–4182. [CrossRef]
12. Berger, T. Decentralized estimation and decision theory. In Proceedings of the IEEE 7th Spring Workshop on IInformation Theory, Mt. Kisco, NY, USA, September 1979.
13. Shalaby, H.M.H.; Papamarcou, A. Multiterminal Detection with Zero-Rate Data Compression. *IEEE Trans. Inf. Theory* **1992**, *38*, 254–267. [CrossRef]
14. Han, T.S.; Kobayashi, K. Exponential-Type Error Probabilities for Multiterminal Hypothesis Testing. *IEEE Trans. Inf. Theory* **1989**, *35*, 2–14. [CrossRef]
15. Gündüz, D.; Kurka, D.B.; Jankowski, M.; Amiri, M.M.; Ozfatura, E.; Sreekumar, S. Communicate to Learn at the Edge. *IEEE Commun. Mag.* **2020**, *58*, 14–19. [CrossRef]
16. Gündüz, D.; Qin, Z.; Aguerri, I.E.; Dhillon, H.S.; Yang, Z.; Yener, A.; Wong, K.K.; Chae, C.B. Beyond Transmitting Bits: Context, Semantics, and Task-Oriented Communications. *IEEE J. Sel. Areas Commun.* **2023**, *41*, 5–41. [CrossRef]
17. Zhao, W.; Lai, L. Distributed Testing Against Independence with Multiple Terminals. In Proceedings of the 52nd Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, USA, 30 September–3 October 2014; pp. 1246–1251.
18. Wigger, M.; Timo, R. Testing Against Independence with Multiple Decision Centers. In Proceedings of the International Conference on Signal Processing and Communications, Bangalore, India, 12–15 June 2016; pp. 1–5.
19. Salehkalaibar, S.; Wigger, M.; Wang, L. Hypothesis Testing In Multi-Hop Networks. *IEEE Trans. Inf. Theory* **2019**, *65*, 4411–4433. [CrossRef]
20. Zaidi, A.; Aguerri, I.E. Optimal Rate-Exponent Region for a Class of Hypothesis Testing Against Conditional Independence Problems. In Proceedings of the 2019 IEEE Information Theory Workshop (ITW), Visby, Sweden, 25–28 August 2019; pp. 1–5.
21. Zaidi, A. Rate-Exponent Region for a Class of Distributed Hypothesis Testing Against Conditional Independence Problems. *IEEE Trans. Inf. Theory* **2023**, *69*, 703–718. [CrossRef]
22. Mhanna, M.; Piantanida, P. On Secure Distributed Hypothesis Testing. In Proceedings of the IEEE International Symposium on Information Theory, Hong Kong, China, 14–19 June 2015; pp. 1605–1609.
23. Sreekumar, S.; Gündüz, D. Testing Against Conditional Independence Under Security Constraints. In Proceedings of the IEEE Int. Symp. Inf. Theory (ISIT), Vail, CO, USA, 17–22 June 2018; pp. 181–185.
24. Sreekumar, S.; Cohen, A.; Gündüz, D. Privacy-Aware Distributed Hypothesis Testing. *Entropy* **2020**, *22*, 665. [CrossRef] [PubMed]
25. Gilani, A.; Amor, S.B.; Salehkalaibar, S.; Tan, V. Distributed Hypothesis Testing with Privacy Constraints. *Entropy* **2019**, *21*, 478. [CrossRef] [PubMed]
26. Katz, G.; Piantanida, P.; Debbah, M. Distributed Binary Detection with Lossy Data Compression. *IEEE Trans. Inf. Theory* **2017**, *63*, 5207–5227. [CrossRef]
27. Xiang, Y.; Kim, Y.H. Interactive hypothesis testing with communication constraints. In Proceedings of the 50th Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, USA, 1–5 October 2012; pp. 1065–1072.
28. Xiang, Y.; Kim, Y.H. Interactive hypothesis testing against Independence. In Proceedings of the IEEE International Symposium on Information Theory, Istanbul, Turkey, 7–12 July 2013; pp. 2840–2844.
29. Tian, C.; Chen, J. Successive Refinement for Hypothesis Testing and Lossless One-Helper Problem. *IEEE Trans. Inf. Theory* **2008**, *54*, 4666–4681. [CrossRef]
30. Haim, E.; Kochman, Y. On Binary Distributed Hypothesis Testing. *arXiv* **2017**, arXiv:1801.00310.
31. Weinberger, N.; Kochman, Y. On the Reliability Function of Distributed Hypothesis Testing Under Optimal Detection. *IEEE Trans. Inf. Theory* **2019**, *65*, 4940–4965. [CrossRef]
32. Hadar, U.; Liu, J.; Polyanskiy, Y.; Shayevitz, O. Error Exponents in Distributed Hypothesis Testing of Correlations. In Proceedings of the IEEE International Symposium on Information Theory (ISIT), Paris, France, 7–12 July 2019; pp. 2674–2678.

33. Watanabe, S. Neyman-Pearson Test for Zero-Rate Multiterminal Hypothesis Testing. *IEEE Trans. Inf. Theory* **2018**, *64*, 4923–4939. [CrossRef]

34. Xu, X.; Huang, S.L. On Distributed Learning With Constant Communication Bits. *IEEE J. Sel. Areas Inf. Theory* **2022**, *3*, 125–134. [CrossRef]

35. Salehkalaibar, S.; Tan, V.Y.F. Distributed Sequential Hypothesis Testing With Zero-Rate Compression. In Proceedings of the 2021 IEEE Information Theory Workshop (ITW), Kanazawa, Japan, 17–21 October 2021; pp. 1–5.

36. Sreekumar, S.; Gündüz, D. Strong Converse for Testing Against Independence over a Noisy channel. In Proceedings of the IEEE International Symposium on Information Theory (ISIT), Los Angeles, CA, USA, 21–26 June 2020; pp. 1283–1288.

37. Salehkalaibar, S.; Wigger, M. Distributed Hypothesis Testing with Variable-Length Coding. *IEEE J. Sel. Areas Inf. Theory* **2020**, *1*, 681–694. [CrossRef]

38. Csiszár, I.; Körner, J. *Information Theory: Coding Theorems for Discrete Memoryless Systems*; Cambridge University Press: Cambridge, UK, 2011.

39. Borade, S.; Nakiboğlu, B.; Zheng, L. Unequal Error Protection: An Information-Theoretic Perspective. *IEEE Trans. Inf. Theory* **2009**, *55*, 5511–5539. [CrossRef]

40. Minero, P.; Lim, S.H.; Kim, Y.H. A Unified Approach to Hybrid Coding. *IEEE Trans. Inf. Theory* **2015**, *61*, 1509–1523. [CrossRef]

41. Weinberger, N.; Kochman, Y.; Wigger, M. Exponent Trade-off for Hypothesis Testing Over Noisy Channels. In Proceedings of the IEEE International Symposium on Information Theory (ISIT), Paris, France, 7–12 July 2019; pp. 1852–1856.

42. Csiszár, I. On the error-exponent of source-channel transmission with a distortion threshold. *IEEE Trans. Inf. Theory* **1982**, *28*, 823–828. [CrossRef]

43. Polyanskiy, Y.; Wu, Y. *Information Theory: From Coding to Learning*; Cambridge University Press: Cambridge, UK, 2012. Available online: https://people.lids.mit.edu/yp/homepage/data/itbook-export.pdf (accessed on 10 December 2022).

44. Gallager, R. A simple derivation of the coding theorem and some applications. *IEEE Trans. Inf. Theory* **1965**, *11*, 3–18. [CrossRef]

45. Merhav, N.; Shamai, S. On joint source-channel coding for the Wyner-Ziv source and the Gelfand-Pinsker channel. *IEEE Trans. Inf. Theory* **2003**, *49*, 2844–2855. [CrossRef]

46. Cover, T.; Gamal, A.E.; Salehi, M. Multiple Access Channels with Arbitrarily Correlated Sources. *IEEE Trans. Inf. Theory* **1980**, *26*, 648–657. [CrossRef]

47. Csiszár, I. Joint Source-Channel Error Exponent. *Prob. Control Inf. Theory* **1980**, *9*, 315–328.

48. Eggleston, H.G. *Convexity*, 6th ed.; Cambridge University Press: Cambridge, UK, 1958.