Imperial College London

Department of Electrical and Electronic Engineering

Distributed Hypothesis Testing Under Privacy Constraints

Sreejith Sreekumar

August 2019

Supervised by Dr. Deniz Gündüz

A thesis submitted in fulfilment of the requirements for the degree of
Doctor of Philosophy in Electrical and Electronic Engineering of Imperial College London
and the Diploma of Imperial College London

Abstract

Hypothesis Testing (HT) is one of the central topics of study in statistics. Traditionally, it is assumed that the data on which the hypothesis test is to be performed is available unaltered to the decision maker or *detector* that performs the hypothesis test. However, this is seldom observed in practice, and often the data is observed remotely, and needs to be communicated to the detector over a noisy communication channel, such as a wired or a wireless communication network. The performance of a hypothesis test obviously depends on how accurately the observed data is communicated to the detector, i.e., less distortion of the data implies better performance. However, in many situations less distortion also implies reduced privacy (security) for the observer as there is the threat of leaking sensitive information to the detector (external eavesdropper). The privacy (security) threat is increasingly becoming an important concern due to the availability of affordable large scale computing resources.

In this dissertation, we study HT in a distributed setting, in which the data is observed at a remote node, referred to as *observer*, and communicated over a noisy channel to the detector, which has access to its own correlated side-information. Considering a hypothesis test on the joint distribution of the observer's data and detector's side information, we first study the optimal trade-off between the *type I* and *type II error-exponents*, i.e., the trade-off between the asymptotic exponential rate of decay of the type I and type II error probabilities with respect to the number of observed data samples, and establish single-letter inner bounds on this trade-off. Of special interest is the asymmetric case of characterizing the optimal type II error-exponent for a fixed non-zero constraint on the type I error probability, for which we obtain exact single-letter characterization in some special cases. We also investigate the aspects of data privacy in the above setting with a rate-limited noiseless channel by exploring the trade-off between rate, type II error-exponent and privacy. Finally, considering an eavesdropper with access to correlated side-information, we study the trade-off between rate, type II error-exponent and security when the detector and eavesdropper are connected to the observer via a noisy broadcast channel.

Declaration of Originality

I hereby certify that this thesis is the result of my own work under the guidance of my Ph.D. advisor, Dr. Deniz Gündüz. Any ideas or quotations from the work of other people are appropriately referenced.

Imperial College London
London, United Kingdom
1 August 2019

Sreejith Sreekumar

Copyright Declaration

Acknowledgements

First of all, I would like to express my deepest gratitude to my supervisor, Dr. Deniz Gündüz, for giving me constant guidance and precious advices during my doctorate. Thanks to his vast knowledge and kind help, he has been an incredible source of motivation and inspiration for me, both from a professional and a human point of view. He has been a friend more than a supervisor, and has helped me navigate the tough times. Moreover, I would like to acknowledge the financial support provided by the department of Electrical and Electronics Engineering of Imperial College London for funding my Ph.D. studies, and for enabling me to study in such an excellent research environment. Special thanks are to Prof. Asaf Cohen, with whom I had the opportunity of having several stimulating discussions, which contributed immensely to this thesis. I would also like to thank my examiners, Prof. Ramji Venkataramanan and Prof. Wei Dai, for thoroughly reading this thesis, and for the interesting and useful discussion we had during the viva examination, which greatly improved the quality of this work. I am also very thankful to the Intelligent Systems and Networks group's administrators, Patrick and Joan, who have always helped me without hesitation, always willing to go one step further to help me.

I have to thank all the people of the Information Processing and Communications Lab, whom I had the good fortune of sharing the lab space with. I cherish the numerous interesting conversations and the countless coffee breaks and dinners with them: David, Morteza, Zayd, Kiarash, Giulio, Samuel, Junlin, Rui, Joan, Qianqian, Mohammad, Buraq, Elif, Can, Ecenaz, Cristina, Yasin, Borzoo, Mihajlo, Emre, Mahdi, Nan, Eirina; as well as the numerous visiting students. Thanks also to all of my new and old friends, with whom I have shared the best experiences of these years.

London, United Kingdom
1 August 2019

Sreejith

To my family,

my greatest source of inspiration.

Contents

List of Figures

List of Acronyms

HT	Hypothesis Testing
TACI	Testing against Conditional Independence
TAI	Testing Against Independence
DMC	Discrete Memoryless Channel
JSCC	Joint Source Channel Coding
MT-JSCC	Multi Terminal Joint Source Channel Coding
BT	Berger-Tung
SHTCC	Separate Hypothesis Testing and Channel Coding
JHTCC	Joint Hypothesis Testing and Channel Coding
SCS	Shannon Cipher System
MAC	Multiple Access Channel
IoT	Internet of Things
i.i.d.	independent and identically distributed
r.v.	random variable
W.l.o.g.	Without loss of generality
iff	if and only if
KL	Kullback-Leibler
NP	Neyman-Pearson

Chapter 1

Introduction

Given data samples, hypothesis testing (HT) deals with the problem of ascertaining the true assumption or hypothesis about the data from among a set of hypotheses. Such tests have traditionally assumed the availability of the data at a single place, for example, in the case of a statistician performing the test on empirical data available from an experiment. Among the various statistical problems encountered in science, the problem of ascertaining the true probability distribution of the data arises quite frequently. Given n samples of data X^n, drawn independent and identically distributed (i.i.d.) from alphabet $\mathcal{X}$ according to distribution P_X or Q_X, the problem of identifying the true underlying distribution is a binary HT problem with the null and alternate hypothesis given by

$$H_0 : X^n \sim P_{X^n} := \prod_{i=1}^{n} P_X, \tag{1.1}$$

$$\text{and } H_1 : X^n \sim Q_{X^n} := \prod_{i=1}^{n} Q_X, \tag{1.2}$$

respectively. Let H and $\hat{H}$ denote the r.v.'s corresponding to the true hypothesis and the decision of the HT, respectively, with support $\hat{\mathcal{H}} = \mathcal{H} := \{0, 1\}$. The detector that performs the HT is specified by a decision rule (possibly stochastic) $g^{(n)} : \mathcal{X}^n \mapsto \hat{\mathcal{H}}$ with output $\hat{H}$, where 0 and 1 denotes the decision H_0 and H_1, respectively. There are two kinds of errors possible in this test, namely, the error that the detector decides in favour of H_1 given that the true distribution is P_X and the error that the detector decides in favour of H_0 given that the true distribution is Q_X. The first kind of error is known in the literature as the type I error, while the second one is known as the type II error. Denoting an arbitrary stochastic decision rule that specifies $g^{(n)}$ by $P_{\hat{H}|X^n}$,

the type I and type II error probabilities can be written[1] as

$$\bar{\alpha}^{(n)}\left(g^{(n)}\right) := P_{\hat{H}}(1) := \sum_{x^n \in \mathcal{X}^n} P_{X^n}(x^n)\ P_{\hat{H}|X^n}(1|x^n),$$

and

$$\bar{\beta}^{(n)}\left(g^{(n)}\right) := Q_{\hat{H}}(0) := \sum_{x^n \in \mathcal{X}^n} Q_{X^n}(x^n)\ P_{\hat{H}|X^n}(0|x^n),$$

respectively. Obviously, there is a trade-off between the type I and type II error probabilities. For $\epsilon \in [0,1]$, let

$$\beta^{(n)}(\epsilon) = \left\{ \min_{g^{(n)}} \bar{\beta}^{(n)}\left(g^{(n)}\right)\ \text{s.t.}\ \bar{\alpha}^{(n)}\left(g^{(n)}\right) \leq \epsilon \right\}, \tag{1.3}$$

denote the minimum type II error probability that can be achieved for a given constraint ϵ on the type I error probability. The optimal performance in HT is first characterized in the seminal paper of Neyman and Pearson [1], where the well-known *Neyman-Pearson* (NP) test that achieves the optimal trade-off between the type I and type II error probabilities was proposed. For a discrete alphabet $\mathcal{X}^n$, the NP test is given by

$$P_{\hat{H}|X^n}(0|x^n) = \begin{cases} 1 & \text{if } \log\left(\frac{P_{X^n}(x^n)}{Q_{X^n}(x^n)}\right) > \theta, \\ p_0 & \text{if } \log\left(\frac{P_{X^n}(x^n)}{Q_{X^n}(x^n)}\right) = \theta, \\ 0 & \text{if } \log\left(\frac{P_{X^n}(x^n)}{Q_{X^n}(x^n)}\right) < \theta, \end{cases} \tag{1.4}$$

for some $\theta \in \mathbb{R}$. For any $\epsilon \in [0,1]$, there exists a NP test that achieves a type I-type II error probability pair $\left(\epsilon, \beta^{(n)}(\epsilon)\right)$, for some unique $\theta \in \mathbb{R}$ and p_0 chosen such that

$$\epsilon = P_{X^n}\left(\log\left(\frac{P_{X^n}(x^n)}{Q_{X^n}(x^n)}\right) > \theta\right) + p_0\ P_{X^n}\left(\log\left(\frac{P_{X^n}(x^n)}{Q_{X^n}(x^n)}\right) = \theta\right). \tag{1.5}$$

It is also well known that both the type I and type II error probabilities can be simultaneously made to decay to zero exponentially as n tends to infinity [2], provided

[1] We will consider discrete alphabets with finite support in this thesis unless specified otherwise.

that $P_X \neq Q_X$. Given this, the performance of HT can be analyzed in the following regimes:

(i) **Stein's regime**

The goal here is to characterize the optimal type II error-exponent for a given constraint ϵ on the type I error probability[2], i.e. to characterize $\liminf_{n\to\infty} \frac{-1}{n} \log(\beta^{(n)}(\epsilon))$ for all $\epsilon \in [0, 1]$.

(ii) **Chernoff's regime**

The objective here is to characterize the maximum value of type II error-exponent for a given constraint κ on the type I error-exponent, i.e., maximize $\liminf_{n\to\infty} \frac{-1}{n} \log\left(\bar{\beta}^{(n)}(g^{(n)})\right)$ over $g^{(n)}$ such that $\bar{\alpha}^{(n)}(g^{(n)}) \leq e^{-n\kappa}$ for sufficiently large n.

In the Stein's regime, the maximum value of the type II error-exponent is characterized by the now ubiquitous Chernoff-Stein's lemma [3] [4] as

$$\lim_{n\to\infty} \frac{1}{n} \log(\beta_n(\epsilon)) = D(P_X||Q_X), \tag{1.6}$$

where $D(P_X||Q_X)$ is the Kullback-Leibler (KL) divergence between probability distributions P_X and Q_X defined as

$$D(P_X||Q_X) := \sum_{x\in\mathcal{X}} P_X(x) \log\left(\frac{P_X(x)}{Q_X(x)}\right). \tag{1.7}$$

As is evident from (1.6), the optimal achievable type II error-exponent is independent of ϵ. This property is referred to as the *strong converse* for HT in the literature. On the other hand, the optimal trade-off between the type I and type II error-exponents is established in [5] and [2]. This trade-off can also be stated in terms of the Legendre-Fenchel transform of the log-moment generating function (Log-MGF) of the random variable (r.v.) $\log\left(\frac{P_X(X)}{Q_X(X)}\right)$ [6, Theorem 15.1].

The situation however becomes much more complicated when the samples are not directly accessible to the detector, e.g., post data compression or if the data is to be

[2]This is the more commonly used convention, but one may also consider the dual problem of maximizing the type I error-exponent for a given constraint on the type II error probability.

communicated from a remote node over a wireless network to the detector. With the explosion of data generation and consumption in today's world, data is often generated at multiple nodes that are geographically separated. Even if the data is generated at a single node, the sheer size of the data may necessitate distributed storage at different locations. For example, consider the data flow in social networks today. Data is generated by different users in a distributed manner, which is then stored for later use in servers that may be geographically separated. Sensor networks provide another example that involves distributed data generation. Data is gathered at multiple remote sensors or nodes and transmitted over noisy links to another node for further processing. Cloud computing and the Internet of things (IoT) are future emerging technological applications that demand distributed storage and processing. Machine learning is yet another example where distributed statistical inference plays a key role in a variety of contexts like anomaly or fraud detection. In such distributed scenarios, the problem of identifying the joint statistics of data leads naturally to the problem of distributed HT over noisy channels, which is one of the central topics of study in this thesis.

While exchange of data among participants is necessary to perform distributed statistical inference, there are other inherent challenges associated with communicating data over a public network. With adversarial cyber-attacks becoming frequent, the protection of sensitive data is no longer a luxury but a necessity for corporate and governmental institutions as well as individuals. There are basically two kinds of adversarial attacks, passive and active. In the former, the adversary just eavesdrops or taps onto public data exchanged between two parties and draws inferences about the data. In the active mode, the adversary, in addition to eavesdropping, may also potentially distort the signals in order to disrupt the communication or mislead the legitimate receiver to draw incorrect inferences. Traditionally, secure communication in the presence of a passive adversary is ensured by using cryptographic techniques whose success rely on the limited computational power available to the adversary. In such cryptographic techniques, the two legitimate parties involved in the communication are assumed to have access to a private key or a random source from which both parties can generate a common key, using which encryption and decryption is then done at the sender and receiver, respectively. The security of the communication depends on

the fact that it is computationally (almost) impossible for the adversary to decrypt the messages without the key. Information theoretic security provides another notion of measuring the robustness of communication that does not assume any limitation on the computational power of the adversary. Here, as before, in order to have provable security, the legitimate receiver must have access to some additional resource that is not available to the adversary like a shared secret key, a better channel or better side-information (correlated with the message) relative to the adversary.

While attacks from external third parties is a major challenge to be dealt with, attacks threatening the privacy of the data are not necessarily carried out by third party intruders or eavesdroppers. There are many cases in which the legitimate receiver of the information is also a potential adversary. This is the case for personal information collected by various companies or governmental agencies, that are being used for purposes outside their initial intentions. For example, various financial transactions of a consumer (through a bank, an online seller, etc.) can be used for advertising particular products, or health-related consumption information, from medical purchases to unhealthy eating patterns, can be used by insurance companies for setting differential premiums. Similarly, posts on social media, or location information provided to a mobile app can be exploited by employers to track activities of their employees. Data privacy is increasingly becoming another dominant theme in today's inter-connected world.

In general, any disclosure of personal data to legitimate entities to receive some utility in return, e.g., in the form of better services or increased social connectivity, come at the expense of a possible loss of privacy, which may have unintended and potentially adverse effects to users in the future. By revealing no information at all, perfect privacy can be achieved; however, at the cost of zero utility. With the increasing computational power of data mining and machine learning algorithms, companies are able to store and process an increasing amount of data collected from individual consumers, enabling them to provide increasingly personalized services, which are often attractive to consumers, who may not be completely aware of the privacy implications. We often do not read the fine print before allowing apps or online services to collect, often times, seemingly irrelevant information for the service they offer. Therefore, privacy

and utility are increasingly becoming contradictory objectives.

It is essential to note at this point that the problem of privacy against legitimate receivers of data is fundamentally different from that of information security. While conventional cryptographic protocols are used to protect the information flow (between a sender and a receiver) from a malicious adversary, the problem of privacy is due to the very nature of this information flow. In other words, the legitimate receiver is a potential adversary in the context of privacy, and in most cases the privacy leakage occurs through unintended parallel channels through statistical inference attacks. Therefore, encryption of data is irrelevant in this context, since the concern of privacy still exists once data is decrypted by the legitimate receiver.

The privacy problem arises in many applications. The basic idea behind most existing privacy-preserving mechanisms is to distort the original data prior to sharing it. This distortion may be obtained by perturbing data, lowering data precision, aggregating data, etc. The underlying idea in all these approaches is that the data is distorted in such a way that while receiving certain amount of utility, the leakage of private information is kept as low as possible. It is clear that the trade-off between utility and privacy depends on the system at hand. While the metric of utility usually accompanies the system model, there is no standard way of measuring data privacy in the literature. As a result, there exists several metrics such as differential privacy, mutual information leakage, total variation distance, average distortion, etc., to quantify the leakage of privacy in the literature [7]. The choice of the privacy measure employed depends on the application at hand and the kind of privacy guarantees required.

1.1 Objectives

In this dissertation, we first study a distributed HT problem over a noisy channel from the point of view of type I and type II error-exponents trade-off. This system model will be introduced in detail in Chapter 2. Herein, two separate nodes, one referred to as the *observer* and the other as detector, each observe correlated data that is i.i.d. across samples. The observer communicates its observations to the detector over a noisy channel, in order that the detector perform a hypothesis test on the

joint distribution of its own observations with that of the observer. This model is an extension of the *one sided compression model* studied in the classic paper of Ahlswede and Csiszar [4], where the rate-limited error-free channel is replaced by a noisy one. Our aim is to obtain computable single-letter bounds on the trade-off between the two error-exponents, and identify special cases where an optimal characterization exists. Towards this, we first study a corner point of this trade-off that corresponds to the asymmetric case of maximizing the type II error-exponent in the Stein's regime (or zero type I error-exponent), and later consider the general symmetric trade-off in the Chernoff's regime. We subsequently introduce an additional privacy constraint and analyze the trade-off between the communication rate, type II error-exponent and privacy in the Stein's regime for the case of a rate-limited noiseless channel. We also consider a distributed HT problem in the presence of an eavesdropper with the requirement that the observer's data is kept as secure as possible from the eavesdropper, and analyze the trade-off between the rate, error-exponent and security achieved in the Stein's regime.

1.2 Outline and Contributions

This dissertation is divided into six chapters. In the following, we outline the content and results of each chapter, as well as the corresponding publications.

Chapter 2

In Chapter 2, we introduce the system model for distributed HT over a noisy channel problem mentioned above, and obtain two lower bounds on the optimal type II error-exponent in the Stein's regime, one using a separation based scheme that performs separate hypothesis testing and channel coding, and the other using a joint scheme that uses hybrid coding for communication between the observer and the detector. Optimal single-letter characterization of the type II error-exponent is established for the special case of *testing against conditional independence* (TACI) of the observer's observations from those of the detector, given some additional side-information at

the detector. The content of this chapter has been published or is under review for publication in:

- Sreejith Sreekumar and Deniz Gündüz, "Distributed hypothesis testing over discrete memoryless channels", *arXiv:1802.07665 [cs.IT]*, Revised for IEEE Transactions on Information Theory.
- Sreejith Sreekumar and Deniz Gündüz, "Distributed hypothesis testing over noisy channels", *IEEE International Symposium on Information Theory (ISIT), Aachen, Germany*, June 2017.
- Sreejith Sreekumar and Deniz Gündüz, "Hypothesis testing over a noisy channel", *IEEE International Symposium on Information Theory (ISIT), Paris, France*, July 2019.

Chapter 3

In Chapter 3, we generalize the setting studied in Chapter 2 to that of the trade-off between both the type I and type II error-exponents. The content of this chapter is under review for possible publication in:

- Sreejith Sreekumar and Deniz Gündüz, "Distributed hypothesis testing over a noisy channel: Error-exponents trade-off", *arXiv:1908.07521 [stat.OT]*, Submitted to IEEE Transactions on Information Theory.

Chapter 4

In Chapter 4, we consider the distributed HT problem studied in Chapter 2 with an additional privacy constraint. We focus on the case of a rate-limited noiseless channel and obtain an inner bound on the trade-off between communication rate, type II error-exponent and privacy in the Stein's regime, and identify scenarios where optimal single-letter characterizations can be established. The results in this chapter have been published or is under review for publication in:

- Sreejith Sreekumar, Asaf Cohen and Deniz Gündüz, "Distributed hypothesis testing under privacy constraints", *in IEEE Information Theory Workshop (ITW)*, Guangzhou, China, November 2018.

- Sreejith Sreekumar, Asaf Cohen and Deniz Gündüz, "Privacy-aware distributed hypothesis testing", *arXiv:1807.02764 [cs.IT]*, Submitted to IEEE Transactions on Information Theory.

Chapter 5

In Chapter 5, we consider the same setting as in Chapter 2, but with an additional eavesdropper, against whom the observer's observations need to be protected. Considering a broadcast channel connecting the observer to the detector and eavesdropper, the trade-off between rate, type II error-exponent and security achieved is studied. The results in this chapter have been partially published in:

- Sreejith Sreekumar and Deniz Gündüz, "Testing against conditional independence under security constraints", in *IEEE International Symposium on Information Theory (ISIT)*, Vail, USA, June 2018.

Chapter 6

Finally, in Chapter 6 we provide the conclusions of the research presented in this dissertation, and discuss potential research directions that can be considered in the future, as well as some open questions.

Chapter 2

Distributed HT over a Noisy Channel: Stein's regime

2.1 Overview

In this chapter, we study a distributed binary HT problem involving two parties, an observer and a detector. The observer observes a discrete memoryless source and communicates its observations to the detector over a discrete memoryless channel (DMC). The detector observes another discrete memoryless source correlated with that at the observer, and performs a binary HT on the joint distribution of the two sources using its own observations and the information received from the observer. The trade-off between the type I error probability and the type II error-exponent of the HT is explored. Single-letter lower bounds on the optimal type II error-exponent are obtained by using three different coding schemes, a *separate HT and channel coding* (SHTCC) scheme, a local decision scheme, and a *joint HT and channel coding* (JHTCC) scheme based on hybrid coding. Exact single-letter characterization of the same is established for the special case of *testing against conditional independence* (TACI), and it is shown to be achieved by the SHTCC scheme. To the best of our knowledge, the trade-off between type I and type II error-exponents in distributed HT over a noisy channel, of which the trade-off studied in this chapter is a corner point, has not been explored before.

2.2 Introduction

In modern communication networks like in sensor networks, cloud computing and Internet of things (IoT), data is gathered at multiple remote nodes, referred to as

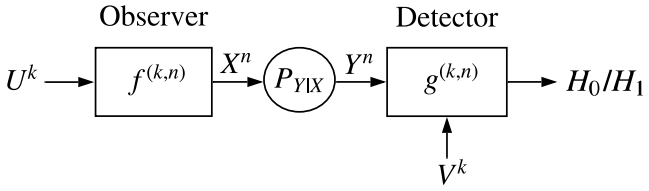


FIGURE 2.1: Distributed HT over a DMC.

observers, and transmitted over noisy links to another node for further processing. Often, there is some prior statistical knowledge available about the data, for example, that the joint probability distribution of the data belongs to a certain prescribed set. In such scenarios, it is of interest to identify the true underlying probability distribution, and this naturally leads to the problem of distributed HT over noisy channels. The simplest case of such a scenario is depicted in Fig. 2.1, where there is a single observer and two possibilities for the joint distribution of the data. The observer observes k i.i.d. data samples U^k, and communicates its observation to the detector by n uses of the DMC, characterized by the conditional distribution $P_{Y|X}$. The detector performs a binary hypothesis test on the joint distribution of the data (U^k, V^k) to decide between them, based on the channel outputs Y^n as well as its own observations V^k. The null and the alternate hypothesis of the hypothesis test are given[1] by

$$H_0 : (U^k, V^k) \sim \prod_{i=1}^{k} P_{UV}, \tag{2.1a}$$

and

$$H_1 : (U^k, V^k) \sim \prod_{i=1}^{k} Q_{UV}, \tag{2.1b}$$

respectively. Our goal is to characterize the optimal exponential rate of decay of the type II error probability asymptotically, known as the *type II error-exponent (henceforth, also referred to as error-exponent)* for a prescribed constraint on the type I error probability for the above hypothesis test.

[1] Although a r.v. is specified together with its probability distribution, here, we abuse the notation for ease of exposition, and denote the observations at the observer and detector under both the null and alternate hypothesis by (U^n, V^n), with probability distribution $\prod_{i=1}^{n} P_{UV}$ and $\prod_{i=1}^{n} Q_{UV}$, respectively. This terminology is used throughout the thesis, except in Chapter 3.

2.3 Previous Work and Our Contributions

The study of distributed statistical inference under communication constraints was conceived by Berger in [8]. In [8], and in the follow up literature summarized below, communication from the observers to the detector are assumed to be over rate-limited error-free channel. Some of the fundamental results in this setting for the case of a single observer was established by Ahlswede and Csiszár in [4]. They obtained a tight single-letter characterization of the optimal error-exponent for a special case of HT known as *testing against independence* (TAI), in which, $Q_{UV} = P_U \times P_V$. Furthermore, the authors established a lower bound on the optimal error-exponent for the general HT case, and proved a *strong converse* result, which states that the optimal achievable error-exponent is independent of the constraint on the type I error probability. A tighter lower bound for the general HT problem is established by Han [9], which recovers the corresponding lower bound in [4]. Han also considered complete data compression in a related setting where either U, or V, or both (also referred to as two-sided compression setting) are compressed and communicated to the detector using a message set of size two. It is shown that, asymptotically, the optimal error-exponent achieved in these three settings are equal. In contrast, a single-letter characterization of the optimal error-exponent for even the TAI with two-sided compression and general rate constraints remains open till date. Shalaby et al. [10] extended the complete data compression result of Han to show that the optimal error-exponent is not improved even if the rate constraint is relaxed to that of zero-rate compression (sub-exponential message set with respect to blocklength k). Shimokawa et al. [11] obtained a tighter lower bound on the optimal error-exponent for general HT by considering quantization and binning at the encoder along with a minimum empirical-entropy decoder. Rahman and Wagner [12] studied the setting with multiple observers, in which, they showed that for the case of a single-observer, the *quantize-bin-test* scheme achieves the optimal error-exponent for *testing against conditional independence* (TACI), in which, $V = (E, Z)$ and $Q_{UEZ} = P_{UZ}P_{E|Z}$. Extensions of the distributed HT problem has also been considered in several other interesting scenarios involving multiple detectors [13], multiple observers [14], interactive HT [15, 16], collaborative HT [17], HT with lossy source reconstruction [18], HT over a multi-hop relay network [19], etc., in which, the

authors obtain a single-letter characterization of the optimal error-exponent in some special cases.

In contrast, HT in distributed settings that involve communication over noisy channels has not been considered until now. In noiseless rate-limited settings, the encoder can reliably communicate its observation subject to a rate constraint. However, this is no longer the case in noisy settings, which complicates the study of error-exponents in HT. Since the capacity of the channel $P_{Y|X}$, denoted by $C(P_{Y|X})$, quantifies the maximum rate of reliable communication over the channel, it is reasonable to expect that it plays a role in the characterization of the optimal error-exponent similar to the rate-constraint R in the noiseless setting. Another measure of the noisiness of the channel is the so-called *reliability function* $E(R, P_{Y|X})$ [20], which is defined as the maximum achievable exponential decay rate of the probability of error (asymptotically) with respect to the blocklength for message rate of R. It appears natural that the reliability function plays a role in the characterization of the achievable error-exponent for distributed HT over a noisy channel. Indeed, in Theorem 2.2 given below, we provide a lower bound on the optimal error-exponent that depends on the *expurgated exponent* at rate R, $E_x(R, P_{Y|X})$, which is a lower bound on $E(R, P_{Y|X})$ [21]. However, surprisingly, it will turn out that the reliability function does not play a role in the characterization of the error-exponent for TACI in the regime of vanishing type I error probability constraint.

The goal of this chapter is to study the best attainable error-exponent for distributed HT over a DMC with a single observer and obtain a computable characterization of the same. Although a complete solution is not to be expected for this problem (since even the corresponding noiseless case is still open), the aim is to provide an achievable scheme for the general problem, and to identify special cases in which a tight characterization can be obtained. In the sequel, we first introduce a separation based scheme that performs independent hypothesis testing and channel coding, which we refer to as the *separate hypothesis testing and channel coding* (SHTCC) scheme. This scheme combines the Shimokawa-Han-Amari scheme [11], which is the best known coding scheme till date for distributed HT over a rate-limited noiseless channel, with the channel coding scheme that achieves the expurgated exponent [21] [20] of the

channel along with the best channel coding error-exponent for a single special message. The channel coding scheme is based on the Borade-Nakiboğlu-Zheng unequal error-protection scheme [22]. As we show later, the SHTCC scheme achieves the optimal error-exponent for TACI.

Our second scheme is a zero-rate compression scheme referred to as the *local decision* scheme, in which, the observer makes a tentative guess on the true hypothesis based on its own observation, and communicates its one bit decision to the detector. It is shown in [23] that the local decision scheme achieves the optimal error-exponent for HT over a noisy channel, when the detector does not have access to side-information V^k. Although the above mentioned separation based schemes are attractive due to their modular design, *joint source channel coding* (JSCC) schemes are known to outperform separation based schemes in several different contexts, for example, the error exponent for reliable transmission of a source over a DMC [24], reliable transmission of correlated sources over a multiple-access channel [25], etc., to name a few. While in separation based schemes coding is usually performed by first quantizing the observed source sequence to an index, and transmitting the channel codeword corresponding to that index (independent of the source sequence), JSCC schemes allow the channel codeword to be dependent on the source sequence, in addition to the quantization index. Motivated by this, we propose a third scheme, referred to as the *joint HT and channel coding* (JHTCC) scheme, based on *hybrid coding* [26] for the communication between the observer and the detector.

The main contributions in this chapter can be summarized as follows.

(i) We propose three different coding schemes (namely, SHTCC, local decision and JHTCC) for distributed HT over a DMC, and analyze the error-exponents achieved by these schemes.

(ii) We obtain an exact single-letter characterization of the optimal error-exponent for the special case of TACI with a vanishing type I error probability constraint, and show that it is achievable by the SHTCC scheme.

The rest of the chapter is organized as follows. In Section 2.4, we introduce the notations, detailed system model and definitions. Following this, we introduce the main results in Section 2.5 and 2.6. The achievable schemes are presented in Section 2.5 and the optimality results for special cases are discussed in Section 2.6. The proofs of the results are presented immediately after the statement or in Appendix A. Finally, Section 2.7 concludes the chapter.

2.4 Preliminaries

2.4.1 Notations

Random variables (r.v.'s) are denoted by capital letters (e.g., X), their realizations by the corresponding lower case letters (e.g., x), and their support by calligraphic letters (e.g., $\mathcal{X}$). The cardinality of a finite set $\mathcal{X}$ is denoted by $|\mathcal{X}|$. $X - Y - Z$ denotes that X, Y and Z form a Markov chain. For $m \in \mathbb{Z}^+$, X^m denotes the sequence $X_1, \ldots, X_m$. Following the notation in [20], for a probability distribution P_X on r.v. X, $T^m_{P_X}$ and $T^m_{[P_X]_\delta}$ (or $T^m_{[X]_\delta}$) denote the set of sequences $x^m \in \mathcal{X}^m$ of type (or empirical distribution) P_X and the set of strongly [2] P_X-typical sequences, respectively. The set of all possible types of sequences of length m with alphabet $\mathcal{X}$ is denoted by $\mathcal{T}^m_{\mathcal{X}}$, and $\cup_{m\in\mathbb{Z}^+}\mathcal{T}^m_{\mathcal{X}}$ is denoted by $\mathcal{T}_{\mathcal{X}}$. Similar notations apply for pair's and other larger combinations of r.v.'s, e.g., $T^m_{P_{XY}}$ $T^m_{[P_{XY}]_\delta}$, $\mathcal{T}^m_{\mathcal{XY}}$, $\mathcal{T}_{\mathcal{XY}}$, etc.. The standard information theoretic quantities like Kullback-Leibler (KL) divergence between distributions P_X and Q_X, the entropy of X with distribution P_X, the conditional entropy of X given Y and the mutual information between X and Y with joint distribution P_{XY}, are denoted by $D(P_X||Q_X)$, $H_{P_X}(X)$, $H_{P_{XY}}(X|Y)$ and $I_{P_{XY}}(X;Y)$, respectively. When the distribution of the r.v.'s involved are clear from the context, the last three quantities are denoted simply by $H(X)$, $H(X|Y)$ and $I(X;Y)$, respectively. Given realizations $X^m = x^m$ and $Y^m = y^m$, $H_e(x^m|y^m)$ denotes the conditional empirical entropy defined

[2] The set of strongly P_X-typical sequences, $T^m_{[P_X]_\delta}$, is the set of sequences $x^m \in \mathcal{X}^m$ such that $|\frac{1}{m}\sum_{i=1}^m \mathbb{1}(x_i = x) - P_X(x)| \leq \delta,\ \forall x \in \mathcal{X}$, and in addition no $x \in \mathcal{X}$ with $P_X(x) = 0$ occurs in x^m.

as

$$H_e(x^m|y^m) := H_{P_{\tilde{X}\tilde{Y}}}(\tilde{X}|\tilde{Y}), \tag{2.2}$$

where $P_{\tilde{X}\tilde{Y}}$ denote the joint type of (x^m, y^m), and $:=$ represents equality by definition (throughout this chapter). For $a \in \mathbb{R}^+$, $[a]$ denotes the set of integers $\{1, 2, \ldots, \lceil a \rceil\}$. All logarithms considered in this chapter are with respect to the base e. For any set $\mathcal{G}$, $\mathcal{G}^c$ denotes the set complement. $a_k \xrightarrow{(k)} b$ represents $\lim_{k\to\infty} a_k = b$. Similar notations are used for inequalities that hold asymptotically, e.g., , $a_k \overset{(k)}{\geq} b$ denotes $\lim_{k\to\infty} a_k \geq b$. $\mathbb{P}(\mathcal{E})$ denotes the probability of event $\mathcal{E}$. For functions $f_1 : \mathcal{A} \mapsto \mathcal{B}$ and $f_2 : \mathcal{B} \mapsto \mathcal{C}$, $f_2 \circ f_1$ denotes function composition. Finally, $\mathbb{1}(\cdot)$ denotes the indicator function, and $O(\cdot)$ and $o(\cdot)$ denote the standard asymptotic notation.

2.4.2 Problem formulation

All the r.v.'s considered in this chapter are discrete with finite support. Unless specified otherwise, we will denote the probability distribution of a r.v. Z under the null and alternate hypothesis by P_Z and Q_Z, respectively. Let $k, n \in \mathbb{Z}^+$ be arbitrary. As shown in Fig. 2.1, the encoder (at the observer) observes U^k, and transmits codeword $X^n = f^{(k,n)}(U^k)$, where $f^{(k,n)} : \mathcal{U}^k \mapsto \mathcal{X}^n$ represents the encoding function (possibly stochastic). Let $\tau := \frac{n}{k}$ denote the *bandwidth ratio*. The channel output Y^n is given by the probability law

$$P_{Y^n|X^n}(y^n|x^n) = \prod_{j=1}^{n} P_{Y|X}(y_j|x_j), \tag{2.3}$$

i.e., the channels between the observers and the detector are independent of each other and memoryless. Depending on the received symbols Y^n and its own observations V^k, the detector makes a decision between the two hypotheses H_0 and H_1 given in (2.1). Let $H \in \{0, 1\}$ denote the actual hypothesis and $\hat{H} \in \{0, 1\}$ denote the output of the hypothesis test, where 0 and 1 denote H_0 and H_1, respectively, and $\mathcal{A}_{(k,n)} \subseteq \mathcal{Y}^n \times \mathcal{V}^k$ denote the acceptance region for H_0. Then, the decision rule $g^{(k,n)} : \mathcal{Y}^n \times \mathcal{V}^k \mapsto \{0, 1\}$

is given by

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

Let

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

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

Definition 2.1. An error-exponent κ is (τ, ϵ) achievable if there exists a sequence of integers k, corresponding sequences of encoding function $f^{(k,n_k)}$ and decision rules $g^{(k,n_k)}$ such that $n_k \leq \tau k$, $\forall\ k$,

$$\liminf_{k\to\infty} \frac{-1}{k} \log\left(\beta\left(k, n_k, f^{(k,n_k)}, g^{(k,n_k)}\right)\right) \geq \kappa, \tag{2.4a}$$

$$\text{and } \limsup_{k\to\infty} \alpha\left(k, n_k, f^{(k,n_k)}, g^{(k,n_k)}\right) \leq \epsilon. \tag{2.4b}$$

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

$$\kappa(\tau, \epsilon) := \sup\{\kappa' : \kappa' \text{ is } (\tau, \epsilon) \text{ achievable}\}. \tag{2.5}$$

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

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

expression given by

$$\lim_{\epsilon \to 0} \kappa(\tau, \epsilon) = \sup_{\substack{P_{X^n|U^k} \in \mathcal{P}_{\mathcal{X}^n|\mathcal{U}^k}, \\ k, n \leq \tau k}} \frac{1}{k} D\left(P_{Y^n V^k} || Q_{Y^n V^k}\right). \tag{2.6}$$

However, the above expression is intractable as it does not single-letterize in general. Moreover, the encoder and the detector for such a scheme would be computationally complex to implement from a practical viewpoint. Also, analysis of such a scheme is prohibitively complex as it involves optimization over large dimensional probability simplexes, when k and n are large. In the next section, we establish three different single-letter lower bounds on $\kappa(\tau, \epsilon)$ by using the SHTCC, local decision and JHTCC schemes, respectively. These schemes will be explained in detail in the sections below, but we summarize them briefly here. The SHTCC scheme is a separation based scheme that performs separate HT and channel coding. The local decision scheme is basically a one bit scheme that performs a local HT at the observer based on U^k and communicates its decision to the detector, which makes the final decision based on the information received from the observer and its own side-information. Finally, the JHTCC scheme utilizes hybrid coding [26] for communication between the observer and the detector, thus performing joint HT and channel coding.

2.5 Achievable schemes

In [11], Shimokawa et al. obtained a lower bound on the optimal error-exponent for distributed HT over a rate-limited noiseless channel by using a coding scheme that involves quantization and binning at the encoder. In this scheme, the type[3] of the observed sequence $U^k = u^k$ is transmitted by the encoder to the detector, which is useful to improve the performance of the hypothesis test. In fact, in order to achieve the error-exponent proposed in [11], it is sufficient to send a message indicating whether U^k is typical or not, rather than sending the exact type of U^k. Although it is not possible to get perfect reliability for messages transmitted over a noisy channel, intuitively, it is desirable to protect the typicality information about the observed sequence as reliably

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

as possible. Based on this intuition, we next propose the SHTCC scheme that performs independent HT and channel coding and protects the message indicating whether U^k is typical or not, as reliably as possible.

2.5.1 SHTCC Scheme:

In the SHTCC scheme, the encoding and decoding functions are restricted to be of the form $f^{(k,n)} = f_c^{(k,n)} \circ f_s^{(k)}$ and $g^{(k,n)} = g_s^{(k)} \circ g_c^{(k,n)}$, respectively. The source encoder $f_s^{(k)} : \mathcal{U}^k \mapsto \mathcal{M} = \{0, 1, \cdots, \lceil e^{kR} \rceil\}$ generates an index $M = f_s^{(k)}(U^k)$ and the channel encoder $f_c^{(k,n)} : \mathcal{M} \mapsto \tilde{\mathcal{C}} = \{X^n(j),\ j \in [0 : \lceil e^{kR} \rceil]\}$ generates the channel-input codeword $X^n = f_c^{(k,n)}(M)$. Note that the rate of this coding scheme is $\frac{kR}{n} = \frac{R}{\tau}$ bits per channel use. The channel decoder $g_c^{(k,n)} : \mathcal{Y}^n \mapsto \mathcal{M}$ maps the channel-output Y^n into an index $\hat{M} = g_c^{(k,n)}(Y^n)$, and $g_s^{(k)} : \mathcal{M} \times \mathcal{V}^k \mapsto \{0, 1\}$ outputs the result of the HT as $\hat{H} = g_s^{(k)}(\hat{M}, V^k)$. Note that $f_c^{(k,n)}$ depends on U^k only through the output of $f_s^{(k)}(U^k)$ and $g_c^{(k,n)}$ depends on V^k only through Y^n. Hence, the scheme is modular in the sense that $(f_c^{(k,n)}, g_c^{(k,n)})$ can be designed independent of $(f_s^{(k)}, g_s^{(k)})$. In other words, any good channel coding scheme may be used in conjunction with a good compression scheme. If U^k is not typical according to P_U, $f_s^{(k)}$ outputs a *special* message, referred to as the *error* message, denoted by $M = 0$, to inform the detector to declare $\hat{H} = 1$. There is obviously a trade-off between the reliability of the error message and the other messages in channel coding. The best known reliability for protecting a single *special* message when the other messages $M \in [e^{nR}]$ of rate R, referred to as *ordinary* messages, are required to be communicated reliably is given by the *red-alert exponent* in [22]. The red-alert exponent is defined as

$$E_m(R, P_{Y|X}) := \max_{\substack{P_{SX}:\ \mathcal{S}=\mathcal{X},\\ I(X;Y|S)=R,\\ S-X-Y}} \sum_{s\in\mathcal{S}} P_S(s)\ D\left(P_{Y|S=s} || P_{Y|X=s}\right). \tag{2.7}$$

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

error-exponent equal to $E_x(R, P_{Y|X})$ is achievable, where

$$E_x(R, P_{Y|X}) := \max_{P_X} \max_{\rho \geq 1} \left\{ -\rho\ R - \rho\ \log \left(\sum_{x,\tilde{x}} P_X(x) P_X(\tilde{x}) \left(\sum_y \sqrt{P_{Y|X}(y|x) P_{Y|X}(y|\tilde{x})} \right)^{\frac{1}{\rho}} \right) \right\}, \tag{2.8}$$

is the *expurgated* exponent at rate R [21] [20]. Let

$$E_m(P_{SX}, P_{Y|X}) := \sum_{s \in \mathcal{S}} P_S(s)\ D\left(P_{Y|S=s} || P_{Y|X=s}\right), \tag{2.9}$$

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

$$E_x(R, P_{SX}, P_{Y|X}) := \max_{\rho \geq 1} \left\{ -\rho\ R - \rho\ \log \left(\sum_{s,x,\tilde{x}} P_S(s) P_{X|S}(x|s) P_{X|S}(\tilde{x}|s) \left(\sum_y \sqrt{P_{Y|X}(y|x) P_{Y|X}(y|\tilde{x})} \right)^{\frac{1}{\rho}} \right) \right\}. \tag{2.10}$$

Although Borade et al.'s scheme is concerned only with the reliability of the special message, it is not hard to see using the technique of *random-coding* that for a fixed distribution P_{SX}, there exists a codebook $\tilde{C}$, and encoder and decoder as in Borade et al.'s scheme, such that the rate is $0 \leq R \leq I(X;Y|S)$ and the special message achieves a reliability equal to $E_m(P_{SX}, P_{Y|X})$, while the ordinary messages achieve a reliability equal to $E_x(R, P_{SX}, P_{Y|X})$. Note that $E_m(P_{SX}, P_{Y|X})$ and $E_x(R, P_{SX}, P_{Y|X})$ denote Borade et al.'s red-alert exponent and the expurgated exponent with fixed distribution P_{SX}, respectively, and that both are inter-dependent through P_{SX}. Thus, varying P_{SX} provides a trade-off between the reliability for the ordinary messages and the special message. We will use Borade et al.'s scheme for channel coding in the SHTCC scheme, such that the error message and the other messages correspond to the special and ordinary messages, respectively. The SHTCC scheme will be described in detail in Appendix A.1. We next state a lower bound on the optimal error-exponent $\kappa(\tau, \epsilon)$ that is achieved by the SHTCC scheme. For brevity, we will use the shorter notations C,

$E_m(P_{SX})$ and $E_x(R, P_{SX})$ instead of $C(P_{Y|X})$, $E_m(P_{SX}, P_{Y|X})$ and $E_x(R, P_{SX}, P_{Y|X})$, respectively.

Theorem 2.2. *For $\tau \geq 0$, $\kappa(\tau, \epsilon) \geq \kappa_s(\tau)$, $\forall\ \epsilon \in (0, 1]$, where*

$$\begin{aligned}&\kappa_s(\tau)\\ &:= \sup_{\substack{(P_{W|U}, P_{SX}, R)\\ \in\ \mathcal{B}(\tau, P_{Y|X})}} \min\Big\{E_1(P_{W|U}),\ E_2(P_{W|U}, P_{SX}, \tau),\ E_3(P_{W|U}, P_{SX}, \tau),\\ &\qquad\qquad E_4(P_{W|U}, P_{SX}, \tau)\Big\},\end{aligned} \tag{2.11}$$

where

$$\begin{aligned}&\mathcal{B}\left(\tau, P_{Y|X}\right)\\ &:= \left\{\begin{array}{ll}(P_{W|U}, P_{SX}, R): & P_{UVWSXY}(P_{W|U}, P_{SX}) := P_{UV} P_{W|U} P_{SX} P_{Y|X},\\ & \mathcal{S} = \mathcal{X},\ I_P(U; W|V) \leq R < \tau I_P(X; Y|S)\end{array}\right\},\end{aligned} \tag{2.12}$$

$$E_1(P_{W|U}) := \min_{P_{\tilde{U}\tilde{V}\tilde{W}} \in \mathcal{T}_1(P_{UW}, P_{VW})} D(P_{\tilde{U}\tilde{V}\tilde{W}} || Q_{UVW}), \tag{2.13}$$

$$\begin{aligned}&E_2(P_{W|U}, P_{SX}, R)\\ &:= \begin{cases}\min\limits_{P_{\tilde{U}\tilde{V}\tilde{W}} \in \mathcal{T}_2(P_{UW}, P_V)} D(P_{\tilde{U}\tilde{V}\tilde{W}} || Q_{UVW}) + R & \\ \qquad\qquad -I_P(U; W|V), & \text{if } I_P(U; W) > R,\\ \qquad\qquad \infty, & \text{otherwise,}\end{cases}\end{aligned} \tag{2.14}$$

$$\begin{aligned}&E_3(P_{W|U}, P_{SX}, R, \tau)\\ &:= \begin{cases}\min\limits_{P_{\tilde{U}\tilde{V}\tilde{W}} \in \mathcal{T}_3(P_{UW}, P_V)} D(P_{\tilde{U}\tilde{V}\tilde{W}} || Q_{UVW}) + R & \\ \qquad\qquad -I_P(U; W|V) + \tau E_x\left(\frac{R}{\tau}, P_{SX}\right), & \text{if } I_P(U; W) > R,\\ \min\limits_{P_{\tilde{U}\tilde{V}\tilde{W}} \in \mathcal{T}_3(P_{UW}, P_V)} D(P_{\tilde{U}\tilde{V}\tilde{W}} || Q_{UVW}) + I_P(V; W) & \\ \qquad\qquad +\tau E_x\left(\frac{R}{\tau}, P_{SX}\right), & \text{otherwise,}\end{cases}\end{aligned} \tag{2.15}$$

$$\begin{aligned}&E_4(P_{W|U}, P_{SX}, R, \tau)\\ &:= \begin{cases}D(P_V || Q_V) + R - I_P(U; W|V) + \tau E_m\left(P_{SX}\right), & \text{if } I_P(U; W) > R,\\ D(P_V || Q_V) + I_P(V; W) + \tau E_m\left(P_{SX}\right), & \text{otherwise,}\end{cases}\end{aligned} \tag{2.16}$$

$$Q_{UVW} := Q_{UV} P_{W|U},$$
$$\mathcal{T}_1(P_{UW}, P_{VW}) := \{P_{\tilde{U}\tilde{V}\tilde{W}} \in \mathcal{T}_{\mathcal{U}\mathcal{V}\mathcal{W}} : P_{\tilde{U}\tilde{W}} = P_{UW},\ P_{\tilde{V}\tilde{W}} = P_{VW}\},$$
$$\mathcal{T}_2(P_{UW}, P_V) := \{P_{\tilde{U}\tilde{V}\tilde{W}} \in \mathcal{T}_{\mathcal{U}\mathcal{V}\mathcal{W}} : P_{\tilde{U}\tilde{W}} = P_{UW},\ P_{\tilde{V}} = P_V,\ H(\tilde{W}|\tilde{V}) \geq H_P(W|V)\},$$
$$\mathcal{T}_3(P_{UW}, P_V) := \{P_{\tilde{U}\tilde{V}\tilde{W}} \in \mathcal{T}_{\mathcal{U}\mathcal{V}\mathcal{W}} : P_{\tilde{U}\tilde{W}} = P_{UW},\ P_{\tilde{V}} = P_V\}.$$

The proof of Theorem 2.2 is given in Appendix A.1. Although the expression $\kappa_s(\tau)$ in Theorem 2.2 appears complicated, the terms $E_1(P_{W|U})$ to $E_4(P_{W|U}, P_{SX}, R, \tau)$ can be understood to correspond to distinct events that can possibly lead to a type II error. Note that $E_1(P_{W|U})$ and $E_2(P_{W|U}, P_{SX}, R)$ are the same terms appearing in the error-exponent achieved by the Shimokawa et al.'s scheme [11] for the noiseless channel setting, while $E_3(P_{W|U}, P_{SX}, R, \tau)$ and $E_4(P_{W|U}, P_{SX}, R, \tau)$ are additional terms introduced due to the noisiness of the channel. $E_3(P_{W|U}, P_{SX}, R, \tau)$ corresponds to the event that U^k is P_U-typical, an error occurs at the channel decoder and the detector decides $\hat{H} = 0$, whereas $E_4(P_{W|U}, P_{SX}, R, \tau)$ is due to the event that U^k is not P_U-typical, an error occurs at the channel decoder and the detector decides $\hat{H} = 0$. Note that, in general, $E_m(P_{SX})$ can take the value of ∞ and when this happens, the term $\tau E_m(P_{SX})$ becomes undefined for $\tau = 0$. In this case, we define $\tau E_m(P_{SX}) := 0$.

Remark 2.3. In the SHTCC scheme, although we use Borade et al.'s scheme for channel coding, that is concerned specifically with the protection of a special message when the ordinary message rate is R, any other channel coding scheme with the same rate can be employed. For instance, the ordinary message can be transmitted with an error-exponent equal to the reliability function $E(R, P_{Y|X})$ [20] of the channel $P_{Y|X}$ at rate R, while the special message achieves the maximum reliability possible subject to this constraint. However, it should be noted that a computable characterization of neither $E(R, P_{Y|X})$ (for all values of R) nor the associated best reliability achievable for a single message is known in general.

2.5.2 Local Decision Scheme (Zero-Rate Compression Scheme)

The SHTCC scheme described above is a two stage scheme in which the observer communicates a compressed version W^k of U^k using a channel code of rate $\frac{R}{\tau}$ bits per

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

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

$$\beta_0 := \beta_0(P_U, P_V, Q_{UV}) := \min_{\substack{P_{\tilde{U}\tilde{V}}: \\ P_{\tilde{U}}=P_U,\ P_{\tilde{V}}=P_V}} D(P_{\tilde{U}\tilde{V}}||Q_{UV}), \tag{2.17}$$

$$\text{and } E_c := E_c(P_{Y|X}) := D(P_{Y|X=a}||P_{Y|X=b}), \tag{2.18}$$

where a and b denote channel input symbols that satisfy

$$(a, b) = \arg\max_{(x,x') \in \mathcal{X} \times \mathcal{X}} D(P_{Y|X=x}||P_{Y|X=x'}). \tag{2.19}$$

Note that β_0 denotes the optimal error-exponent for distributed HT over a noiseless channel, when the communication rate-constraint is zero [9] [10]. We define

$$\kappa_0(\tau) := \begin{cases} D(P_V||Q_V) & , \text{ if } \tau = 0, \\ \min\left(\beta_0, \tau E_c + D(P_V||Q_V)\right), & \text{otherwise,} \end{cases} \tag{2.20}$$

We have the following result.

Proposition 2.4. For $\tau \geq 0$, $\kappa(\tau, \epsilon) \geq \kappa_0(\tau),\ \forall\ \epsilon \in (0, 1]$.

Proof. Let $k \in \mathbb{Z}^+$ and $n_k = \lfloor \tau k \rfloor$. For $\tau = 0$, Proposition 2.4 follows from Stein's lemma [4] applied to i.i.d. sequence V^k available at the detector. Assume $\tau > 0$. Fix $\delta > 0$ (a small number). We define the functions $f_s^{(k)}$ and $f_c^{(k,n_k)}$ for the encoder $f^{(k,n_k)}$

as follows:

$$f_s^{(k)}(u^k) = \begin{cases} 0, \text{ if } P_{u^k} \in T^k_{[P_U]_\delta}, \\ 1, \text{ otherwise,} \end{cases} \tag{2.21}$$

and

$$f_c^{(k,n_k)}\left(f_s^{(k)}(u^k)\right) = \begin{cases} a^n, \text{ if } f_s^{(k)}(u^k) = 0, \\ b^n, \text{ otherwise.} \end{cases} \tag{2.22}$$

Here, a^{n_k} and b^{n_k} denote the codewords formed by repeating the symbols a and b from the channel input alphabet $\mathcal{X}$, which are defined in (2.19). Let the functions $g_s^{(k)}$ and $g_c^{(k,n_k)}$ of the decision rule $g^{(k,n_k)}$ be defined by

$$g_c^{(k,n_k)}(y^{n_k}) = \begin{cases} 0, \text{ if } y^{n_k} \in T^{n_k}_{[P_{Y|X=a}]_\delta}, \\ 1, \text{ otherwise,} \end{cases}$$

and

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

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

$$\begin{aligned} \lim_{k\to\infty} \mathbb{P}(U^k \in T^k_{[P_U]_\delta} | H = 0) = 1, \\ \lim_{k\to\infty} \mathbb{P}(V^k \in T^k_{[P_V]_\delta} | H = 0) = 1, \\ \text{and } \lim_{k\to\infty} \mathbb{P}(Y^{n_k} \in T^{n_k}_{[P_{Y|X=a}]_\delta} | H = 0) = 1. \end{aligned}$$

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

$$\begin{aligned} \hat{\mathcal{E}}_1 &:= \{U^k \in T^k_{[P_U]_\delta},\ V^k \in T^k_{[P_V]_\delta} \text{ and } Y^{n_k} \in T^{n_k}_{[P_{Y|X=a}]_\delta}\}, \\ \hat{\mathcal{E}}_2 &:= \{U^k \notin T^k_{[P_U]_\delta},\ V^k \in T^k_{[P_V]_\delta} \text{ and } Y^{n_k} \in T^{n_k}_{[P_{Y|X=a}]_\delta}\}. \end{aligned}$$

It follows from the zero-rate compression result in [9] that for sufficiently large k,

$$\mathbb{P}(\hat{\mathcal{E}}_1 | H = 1) \leq e^{-k(\beta_0 - O(\delta))} \tag{2.23}$$

Also, for sufficiently large k, we can write

$$\begin{aligned}\mathbb{P}(\hat{\mathcal{E}}_2|H=1) &\leq \mathbb{P}(V^k \in T^k_{[P_V]_\delta}|H=1)\ \mathbb{P}\left(Y^{n_k} \in T^{n_k}_{[P_{Y|X=a}]_\delta}|U^k \notin T^k_{[P_U]_\delta}\right) \\ &= \mathbb{P}(V^k \in T^k_{[P_V]_\delta}|H=1)\ \mathbb{P}\left(Y^{n_k} \in T^{n_k}_{[P_{Y|X=a}]_\delta}|X^{n_k}=b^{n_k}\right) \\ &\leq e^{-k(D(P_V||Q_V)-O(\delta))} \cdot e^{-n_k(E_c-O(\delta))}. \qquad (2.24)\end{aligned}$$

Here, (2.24) follows from Lemma 2.2 and Lemma 2.6 in [20]. By the union bound, it follows that

$$\beta(k, n_k, f^{(k,n_k)}, g^{(k,n_k)}) \leq \mathbb{P}(\hat{\mathcal{E}}_1|H=1) + \mathbb{P}(\hat{\mathcal{E}}_2|H=1),$$

from which it follows that,

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

The result follows since $\delta > 0$ is arbitrary. □

The local decision scheme would be particularly useful when the communication channel is very noisy, so that reliable communication is not possible at any positive rate. In [23], it is shown that a local decision-like scheme achieves the optimal error-exponent for HT over a DMC, i.e., when the detector has no side-information. Moreover, it is also proved that optimal error-exponent is not improved if the type I error probability constraint is relaxed; and hence, strong converse holds. In the limiting case of zero channel capacity, i.e., $C(P_{Y|X}) = 0$, it is intuitive to expect that communication from the observer to the detector does not improve the achievable error-exponent for distributed HT. In Appendix A.3, we show that this is indeed the case in a strong converse sense, i.e., the optimal error-exponent depends only on the side-information V^k, and is given by $D(P_V||Q_V)$, for any constraint $\epsilon \in (0,1)$ on the type I error probability. This is in contrast to the zero-rate compression case considered in [9], where one bit of communication between the observer and detector can achieve a strictly positive error-exponent, in general.

Remark 2.5. As $D(P_V||Q_V)$ characterizes the optimal error-exponent when the channel has zero capacity, the error-exponent is zero when $P_V = Q_V$. In fact, zero capacity for a DMC implies a stronger condition that it is not possible to transmit even a single bit reliably (asymptotically). Hence, $P_V = Q_V$ along with $C(P_{Y|X}) = 0$ implies that the minimum sum of type I and type II error probabilities achievable is equal to 1, which is the same as that achievable by randomly making a decision on the hypotheses. However, for a more general channel, e.g., a non-stationary memoryless channel $\{\prod_{i=1}^{n} P_{Y_i|X_i}\}_{n=1}^{\infty}$ such that the maximum mutual information rate $\frac{1}{n}\sum_{i=1}^{n} C(P_{Y_i|X_i})$, is of the order $\Theta\left(\frac{1}{\sqrt{n}}\right)$ bits, $\Theta(\sqrt{n})$ bits can be transmitted at arbitrarily small probability of error (asymptotically). Hence, for such channels (and $P_V = Q_V$), it is possible to drive the sum of type I and type II error probabilities arbitrarily close to zero even though the capacity is zero.

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

2.5.3 JHTCC Scheme

Hybrid coding is a form of JSCC introduced in [26] for the lossy transmission of sources over noisy networks. As the name suggests, hybrid coding is a combination of the digital and analog (uncoded) transmission schemes. For simplicity [4], we assume that $k = n$ ($\tau = 1$). In hybrid coding, the source U^n is first mapped to one of the codewords $\bar{W}^n$ within a compression codebook. Then, a symbol-by-symbol function (deterministic) of the $\bar{W}^n$ and U^n is transmitted as the channel codeword X^n. This procedure is reversed at the decoder, in which, the decoder first attempts to obtain an

[4]For the case $\tau \neq 1$, as mentioned in [26], we can consider hybrid coding over super symbols U^{k^*} and X^{n^*}, where k^* and n^* are some integers satisfying the constraint $n^* \leq \tau k^*$. This amounts to enlarging the source and side-information r.v.'s alphabets, and thus results in a harder optimization problem over the conditional probability distributions $P_{\bar{W}|U^{k^*}S}$ and $P_{X^{n^*}|U^{k^*}S\bar{W}}$ given in Theorem 2.6. However, we omit its description since the technique is standard and only adds notational clutter.

estimate $\hat{\bar{W}}^n$ of $\bar{W}^n$ using the channel output Y^n and its own correlated side information V^n. Then, the reconstruction $\hat{U}^n$ of the source is obtained as a symbol-by-symbol function of the reconstructed codeword, Y^n and V^n. In this subsection, we propose a lower bound on the optimal error-exponent that is achieved by a scheme that utilizes hybrid coding for the communication between the observer and the detector, which we refer to as the JHTCC scheme. Post estimation of $\hat{\bar{W}}^n$, the detector performs the hypothesis test using $\hat{\bar{W}}^n$, Y^n and V^n, instead of estimating $\hat{U}^n$ as is done in JSCC problems. We will in fact consider a slightly generalized form of hybrid coding in that the encoder and detector is allowed to perform "time-sharing" according to a sequence S^n that is known a priori to both parties. Also, the input X^n is allowed to be generated according to an arbitrary memoryless stochastic function instead of a deterministic function. The JHTCC scheme is described in detail in Appendix A.2. Next, we state a lower bound on $\kappa(\tau,\epsilon)$ that is achieved by the JHTCC scheme.

Theorem 2.6. *$\kappa(1,\epsilon) \geq \kappa_h, \ \forall\ \epsilon \in (0,1]$, where*

$$\kappa_h := \sup_{\mathbf{b}\ \in\ \mathcal{B}_h} \min \Big\{ E_1'(P_S, P_{\bar{W}|US}, P_{X|US\bar{W}}),\ E_2'(P_S, P_{\bar{W}|US}, P_{X|US\bar{W}}), \\ E_3'(P_S, P_{\bar{W}|US}, P_{X'|US}, P_{X|US\bar{W}}) \Big\}, \tag{2.25}$$

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

$$E_1'\left(P_S, P_{\bar{W}|US}, P_{X|US\bar{W}}\right) := \min_{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} \in \mathcal{T}_1'\left(\hat{P}_{US\bar{W}}, \hat{P}_{VS\bar{W}Y}\right)} D\left(P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} || \hat{Q}_{UVS\bar{W}Y}\right), \tag{2.26}$$

$$E_2'\left(P_S, P_{\bar{W}|US}, P_{X|US\bar{W}}\right) := \min_{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} \in \mathcal{T}_2'\left(\hat{P}_{US\bar{W}}, \hat{P}_{VS\bar{W}Y}\right)} D\left(P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} || \hat{Q}_{UVS\bar{W}Y}\right) \\ + I_{\hat{P}}(\bar{W};V,Y|S) - I_{\hat{P}}(U;\bar{W}|S), \tag{2.27}$$

$$E_3'\left(P_S, P_{\bar{W}|US}, P_{X'|US}, P_{X|US\bar{W}}\right) := D(\hat{P}_{VSY} || \check{Q}_{VSY}) + I_{\hat{P}}(\bar{W};V,Y|S) \\ - I_{\hat{P}}(U;\bar{W}|S), \tag{2.28}$$

$$\hat{Q}_{UVS\bar{W}X'XY}(P_S, P_{\bar{W}|US}, P_{X'|US}, P_{X|US\bar{W}}) := Q_{UV}P_S P_{\bar{W}|US}P_{X'|US}P_{X|US\bar{W}}P_{Y|X}, \tag{2.29}$$

$$\check{Q}_{UVSX'XY}(P_S, P_{X'|US}) := Q_{UV} P_S P_{X'|US} \mathbb{1}(X = X') P_{Y|X},$$
$$\mathcal{T}_1'(\hat{P}_{US\bar{W}}, \hat{P}_{VS\bar{W}Y}) := \{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} \in \mathcal{T}_{\mathcal{U}\mathcal{V}\mathcal{S}\mathcal{W}\mathcal{Y}} : P_{\tilde{U}\tilde{S}\tilde{W}} = \hat{P}_{US\bar{W}},\ P_{\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} = \hat{P}_{VS\bar{W}Y}\},$$
$$\mathcal{T}_2'(\hat{P}_{US\bar{W}}, \hat{P}_{VS\bar{W}Y}) := \{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} \in \mathcal{T}_{\mathcal{U}\mathcal{V}\mathcal{S}\mathcal{W}\mathcal{Y}} : P_{\tilde{U}\tilde{S}\tilde{W}} = \hat{P}_{US\bar{W}},\ P_{\tilde{V}\tilde{S}\tilde{Y}} = \hat{P}_{VSY},$$
$$H(\tilde{W}|\tilde{V}, \tilde{S}, \tilde{Y}) \geq H_{\hat{P}}(\bar{W}|V, S, Y)\}.$$

The proof of Theorem 2.6 is given in Appendix A.2. The different factors inside the minimum in (2.25) can be intuitively understood to be related to the various events that could possibly lead to a type 2 error. More specifically, let the event that the encoder is unsuccessful in finding a codeword $\bar{W}^n$ in the quantization codebook that is typical with U^n be referred to as the *encoding error*, and the event that a wrong codeword $\hat{\bar{W}}^n$ (unintended by the encoder) is reconstructed at the detector be referred to as the *decoding error*. Then, $E_1'(P_S, P_{\bar{W}|US}, P_{X|US\bar{W}})$ is related to the event that neither the encoding nor the decoding error occurs, while $E_2'(P_S, P_{\bar{W}|US}, P_{X|US\bar{W}})$ and $E_3'(P_S, P_{\bar{W}|US}, P_{X'|US}, P_{X|US\bar{W}})$ are related to the events that only the decoding error and both the encoding and decoding errors occur, respectively. From Theorem 2.2, Theorem 2.6 and Proposition 2.4, we have the following corollary.

Corollary 2.7.

$$\kappa(1, \epsilon) \geq \max\left(\kappa_h, \kappa_0(1), \kappa_s(1)\right),\ \forall \epsilon \in (0, 1], \tag{2.30}$$

where $\kappa_h, \kappa_0(1)$ and $\kappa_s(1)$ refer to the lower bound on the error-exponent achieved by the JHTCC (Theorem 2.6), local decision (Proposition 2.4) and SHTCC (Theorem 2.2) schemes, respectively.

Thus far, we obtained lower bounds on the optimal error-exponent for distributed HT over a DMC. However, obtaining tight computable outer bounds is a challenging open problem, and consequently, an exact computable characterization of the optimal error-exponent is unknown (even when the communication channel is noiseless). However, as we show in the next section, the problem does admit single-letter characterization for TACI.

2.6 Optimality result for TACI

Recall that for TACI, $V = (E, Z)$ and the joint distribution under the null and the alternate hypotheses are given by P_{UEZ} and $Q_{UEZ} = P_{UZ}P_{E|Z}$, respectively. Let

$$\kappa(\tau) = \lim_{\epsilon \to 0} \kappa(\tau, \epsilon). \tag{2.31}$$

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

Proposition 2.8. For TACI over a DMC $P_{Y|X}$,

$$\kappa(\tau) = \sup \left\{ \begin{array}{c} I(E;W|Z) : \exists\ W \text{ s.t. } I(U;W|Z) \le \tau C(P_{Y|X}), \\ (Z,E) - U - W,\ |\mathcal{W}| \le |\mathcal{U}| + 1. \end{array} \right\},\ \tau \ge 0. \tag{2.32}$$

Proof. For the proof of achievability, we will show that $\kappa_s(\tau)$ when specialized to TACI recovers (2.32). Let $\mu > 0$ be a arbitrarily small positive number, and

$$\begin{aligned} &\mathcal{B}'\left(\tau, P_{Y|X}\right) \\ &:= \left\{ \begin{array}{c} (P_{W|U}, P_{SX}, R_m) :\ \mathcal{S} = \mathcal{X},\ P_{UEZWSXY}(P_{W|U}, P_{SX}) := P_{UEZ}P_{W|U}P_{SX}P_{Y|X}, \\ I(U;W|Z) \le R_m := \tau I(X;Y|S) - \mu < \tau I(X;Y|S) \end{array} \right\}. \end{aligned} \tag{2.33}$$

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

$$\begin{aligned} E_1(P_{W|U}) &= \min_{P_{\tilde{U}\tilde{E}\tilde{Z}\tilde{W}} \in \mathcal{T}_1(P_{UW}, P_{EZW})} D(P_{\tilde{U}\tilde{E}\tilde{Z}\tilde{W}} || P_Z P_{U|Z} P_{E|Z} P_{W|U}) \\ &\ge \min_{P_{\tilde{U}\tilde{E}\tilde{Z}\tilde{W}} \in \mathcal{T}_1(P_{UW}, P_{EZW})} D(P_{\tilde{E}\tilde{Z}\tilde{W}} || P_Z P_{E|Z} P_{W|Z}) \\ &= I(E;W|Z), \end{aligned} \tag{2.34}$$

where (2.34) follows from the log-sum inequality [20]. Also,

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

$$\min_{P_{\tilde{U}\tilde{E}\tilde{Z}\tilde{W}} \in \mathcal{T}_3(P_{UW}, P_{EZ})} D(P_{\tilde{U}\tilde{E}\tilde{Z}\tilde{W}} || P_Z P_{U|Z} P_{E|Z} P_{W|U}) + R_m - I(U;W|E,Z) + \tau E_x\left(\frac{R_m}{\tau}, P_{SX}\right)$$
$$\geq I(U;W|Z) - I(U;W|E,Z) = I(E;W|Z), \tag{2.35}$$

$$\min_{P_{\tilde{U}\tilde{E}\tilde{Z}\tilde{W}} \in \mathcal{T}_3(P_{UW}, P_{EZ})} D(P_{\tilde{U}\tilde{E}\tilde{Z}\tilde{W}} || P_Z P_{U|Z} P_{E|Z} P_{W|U}) + I(E,Z;W) + \tau E_x\left(\frac{R_m}{\tau}, P_{SX}\right)$$
$$\geq I(E;W|Z), \tag{2.36}$$

$$D(P_{EZ}||P_{EZ}) + R_m - I(U;W|E,Z) + \tau E_m\left(P_{SX}\right)$$
$$\geq I(U;W|Z) - I(U;W|E,Z) = I(E;W|Z), \tag{2.37}$$

$$D(P_{EZ}||P_{EZ}) + I(E,Z;W) + \tau E_m\left(P_{SX}\right) \geq I(E;W|Z), \tag{2.38}$$

where in (2.35)-(2.38), we used the non-negativity of KL-divergence, $E_x(\cdot,\cdot)$ and $E_m(\cdot)$. Thus, from (2.35)-(2.38), it follows that

$$E_3(P_{W|U}, P_{SX}, R_m, \tau) \geq I(E;W|Z), \tag{2.39}$$
$$\text{and } E_4(P_{W|U}, P_{SX}, R_m, \tau) \geq I(E;W|Z). \tag{2.40}$$

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

$$\kappa(\tau,\epsilon)$$
$$\geq \sup_{(P_{W|U}, P_{SX}, R_m) \in \mathcal{B}} \min\Big\{E_1(P_{W|U}), E_2(P_{W|U}, P_{SX}, R_m), E_3(P_{W|U}, P_{SX}, R_m, \tau), E_4(P_{W|U}, P_{SX}, R_m, \tau)\Big\}$$
$$\geq \sup_{(P_{W|U}, P_{SX}, R_m) \in \mathcal{B}} I(E;W|Z)$$
$$\geq \sup_{(P_{W|U}, P_{SX}, R_m) \in \mathcal{B}'} I(E;W|Z) \tag{2.41}$$
$$= \sup_{P_{W|U}: I(W;U|Z) \leq \tau C(P_{Y|X}) - \mu} I(E;W|Z), \tag{2.42}$$

where (2.41) follows from the fact that $\mathcal{B}' \subseteq \mathcal{B}$; and (2.42) follows by maximizing over all P_{SX} and noting that $\sup_{P_{XS}} I(X;Y|S) = C(P_{Y|X})$. The proof of achievability is complete by noting that $\mu > 0$ is arbitrary and $I(E;W|Z)$ and $I(U;W|Z)$ are continuous functions of $P_{W|U}$.

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

$$\limsup_{k\to\infty} \alpha\left(k, n_k, f^{(k,n_k)}, g^{(k,n_k)}\right) = 0, \tag{2.43}$$

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

$$\begin{aligned}
&\limsup_{k\to\infty} \frac{-1}{k} \log\left(\beta\left(k, n_k, f^{(k,n_k)}, g^{(k,n_k)}\right)\right) \\
&\leq \limsup_{k\to\infty} \frac{1}{k} D\left(P_{Y^{n_k}E^kZ^k} || Q_{Y^{n_k}E^kZ^k}\right) &(2.44)\\
&= \limsup_{n\to\infty} \frac{1}{k} I(Y^{n_k}; E^k|Z^k) &(2.45)\\
&= H(E|Z) - \liminf_{k\to\infty} \frac{1}{k} H(E^k|Y^{n_k}, Z^k), &(2.46)
\end{aligned}$$

where (2.45) follows since $Q_{Y^{n_k}E^kZ^k} = P_{Y^{n_k}Z^k} P_{E^k|Z^k}$. Now, let T be a r.v. uniformly distributed over $[k]$ and independent of all the other r.v.'s $(U^k, E^k, Z^k, X^{n_k}, Y^{n_k})$. Define an auxiliary r.v. $W := (W_T, T)$, where $W_i := (Y^{n_k}, E^{i-1}, Z^{i-1}, Z_{i+1}^k)$, $i \in [k]$. Then, the last term can be single-letterized as follows.

$$\begin{aligned}
H(E^k|Y^{n_k}, Z^k) &= \sum_{i=1}^k H(E_i|E^{i-1}, Y^{n_k}, Z^k) \\
&= \sum_{i=1}^k H(E_i|Z_i, W_i) \\
&= kH(E_T|Z_T, W_T, T) \\
&= kH(E|Z, W). &(2.47)
\end{aligned}$$

Substituting (2.47) in (2.46), we obtain

$$\limsup_{k\to\infty} \frac{-1}{k} \log\left(\beta\left(k, n_k, f_1^{(k,n_k)}, g^{(k,n_k)}\right)\right) \leq I(E;W|Z). \tag{2.48}$$

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

$$I(U^k; Y^{n_k}) - I(U^k; Z^k) \leq I(X^{n_k}; Y^{n_k}). \tag{2.49}$$

The R.H.S. of (2.49) can be upper bounded due to the memoryless nature of the channel as

$$I(X^{n_k}; Y^{n_k}) \leq n_k \max_{P_X} I(X; Y) = n_k C(P_{Y|X}), \tag{2.50}$$

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

$$\begin{aligned}
I(U^k; Y^{n_k}) - I(U^k; Z^k) &= I(U^k; Y^{n_k}|Z^k) && (2.51)\\
&= \sum_{i=1}^{k} I(Y^{n_k}; U_i|U^{i-1}, Z^k)\\
&= \sum_{i=1}^{k} I(Y^{n_k}, U^{i-1}, Z^{i-1}, Z_{i+1}^k; U_i|Z_i) && (2.52)\\
&= \sum_{i=1}^{k} I(Y^{n_k}, U^{i-1}, Z^{i-1}, Z_{i+1}^k, E^{i-1}; U_i|Z_i) && (2.53)\\
&\geq \sum_{i=1}^{k} I(Y^{n_k}, Z^{i-1}, Z_{i+1}^k, E^{i-1}; U_i|Z_i)\\
&= \sum_{i=1}^{k} I(W_i; U_i|Z_i) = kI(W_T; U_T|Z_T, T)\\
&= kI(W_T, T; U_T|Z_T) && (2.54)\\
&= kI(W; U|Z).
\end{aligned}$$

Here, (2.51) follows due to $Z^k - U^k - Y^{n_k}$; (2.52) follows since the sequences (U^k, Z^k) are memoryless; (2.53) follows since $E^{i-1} - (Y^{n_k}, U^{i-1}, Z^{i-1}, Z_{i+1}^k) - U_i$; (2.54) follows from the fact that T is independent of all the other r.v.'s. Finally, note that $(E, Z) - U - W$ holds and that the cardinality bound on W follows by standard arguments based on Caratheodory's theorem. This completes the proof of the converse, and hence of the proposition. □

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

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

2.7 Conclusions

In this chapter, we have studied the error-exponent achievable in a distributed HT problem over a DMC with side information available at the detector. We obtained single-letter lower bounds on the optimal error-exponent for general HT, and exact single-letter characterization for TACI. It is interesting to note from our results that the reliability function of the channel does not play a role in the characterization of the optimal error-exponent for TACI, and only the channel capacity matters. While a strong converse holds for distributed HT over a rate-limited noiseless channel [4], it remains an open question whether this property holds for noisy channels. As a first step, it is shown in [23] that this is indeed the case for HT over a DMC with no side-information. While we did not discuss the complexity of the schemes considered in this chapter, it is an important factor that needs to be taken into account in any practical implementation of these schemes. In this regard, it is evident that the local decision, SHTCC and JHTCC schemes are in increasing order of complexity.

Chapter 3

Distributed HT over a Noisy Channel: Chernoff's regime

3.1 Overview

In this chapter, we study the trade-off between both the type I and type II error exponents in the setting studied in Chapter 2. We will establish two inner bounds on this trade-off. The first inner bound is obtained using a combination of a type-based quantize-bin scheme and Borade et al.'s unequal error protection scheme, while the second inner bound is established using a novel type-based hybrid coding scheme. These bounds extend the achievability result of Han and Kobayashi obtained for the special case of a rate-limited noiseless channel to a noisy channel. For the special case of *testing for the marginal distribution* of the observer's observations with no side-information at the detector, we establish a single-letter characterization of the optimal trade-off between the two error-exponents. Our results imply that a "separation" holds in this case, in the sense that the optimal trade-off between the error-exponents is achieved by a scheme that performs independent HT and channel coding.

3.2 Introduction

Consider the distributed HT setting depicted in Fig. 3.1, which corresponds to the same system model in Chapter 2, but with a slightly different notation[1]. The k data samples observed by the observer, denoted by u^k, are communicated to the detector over a noisy DMC $P_{Y|X}$. Based on its own observations, denoted by v^k, and the channel

[1]This is done to avoid notational clutter and simplify the statement of the results in this chapter.

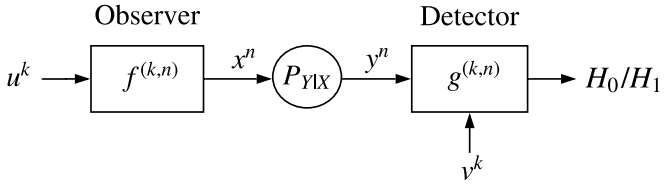


FIGURE 3.1: Distributed HT over a noisy channel.

output y^n, the detector performs the following hypothesis test on the joint probability distribution that generated (u^k, v^k):

$$H_0 : \prod_{i=1}^{k} P_{UV}(u,v), \ \forall \ (u,v) \in \mathcal{U} \times \mathcal{V}, \tag{3.1a}$$

$$H_1 : \prod_{i=1}^{k} P_{\bar{U}\bar{V}}(u,v), \ \forall \ (u,v) \in \mathcal{U} \times \mathcal{V}. \tag{3.1b}$$

Here, P_{UV} and $P_{\bar{U}\bar{V}}$ denote the joint probability distribution[2] from which the data is generated under the null and alternate hypothesis, respectively. Our goal is to characterize the performance of the above hypothesis test as measured by the type I and type II error exponents (see Definition 3.1 below).

3.3 Previous Work and Our Contributions

While the centralized setting in which all the data is available at a single location is well understood, thanks to [1–3,5], the optimal characterization of the error-exponents in distributed settings remain open except for some special cases. In the setting in Fig. 3.1 with a rate-limited noiseless channel, the trade-off between the communication rate, type I and type II error-exponents is explored in [27], where the authors establish an inner bound using a type-based quantization scheme. This problem is revisited recently in [28], where an inner bound is obtained using the technique of structured binning and analogy to the channel detection problem. The trade-off between the error-exponents has also been explored from an information-geometric perspective in the zero-rate compression scenario [29] [30], which provide further insights into the

[2]Note that the notation in this chapter is different from the other chapters in this thesis, as we study the trade-off between the error-exponents in the Chernoff regime as opposed to Stein's regime.

geometric properties of the optimal trade-off between the two error-exponents. While the above works focus on the asymptotic performance in distributed HT, a NP like test for zero-rate multiterminal HT is proposed in [31], which in addition to achieving the optimal trade-off between the two exponents, also achieves the optimal second order asymptotic performance among all symmetric (type-based) encoding schemes.

The main contributions in this chapter can be summarized as follows:

(i) In Theorem 3.7, we establish a single-letter characterization of the optimal trade-off between the type I and type II error-exponents[3] for the special case where the side-information v^k is absent and the hypothesis test is on the marginal distribution of u^k, referred to henceforth as the *non-distributed* setting.

(ii) We obtain an inner bound (Theorem 3.9) on the trade-off between the error-exponents in the distributed setting by using a SHTCC scheme that is a combination of a type-based quantize-bin strategy and unequal error-protection scheme in [22]. This result recovers the inner bound obtained in [27] for the case of a rate-limited noiseless channel, and the lower bound on the type II error-exponent in the Stein's regime, established in Theorem 2.2.

(iii) We obtain a second inner bound (Theorem 3.14) on the error-exponents trade-off by using a JHTCC scheme that is based on *hybrid coding*. This bound is at least as tight as that achieved by the SHTCC scheme in the Stein's regime.

The problem studied here has been investigated recently in [32], where an inner bound on the error-exponents is obtained using a combination of a type-based quantization scheme and unequal error protection scheme of [33] with two special messages. Our schemes differ from that in [32] in the following aspects: (i) In the SHTCC scheme, the encoder employs binning subsequent to quantization and Borade et al.'s unequal error protection with a single special message (in place of [33]); (ii) In the JHTCC scheme, the encoder uses a hybrid coding scheme that transmits the channel codeword generated as a function of the quantization codeword as well as the sequence u^k.

[3] A corner point of this trade-off, namely the optimal T2EE for a fixed non-zero constraint on the type I error probability, is established in [23].

The rest of the chapter is organized as follows. The problem formulation and definitions are introduced in Section 3.3.2. The main results are presented in Section 3.4 and 3.5. The proofs of the results are presented immediately after their statement or in Appendix B. Finally, Section 3.6 concludes the chapter.

3.3.1 Notations

The distribution of a r.v. X is denoted by P_X. Sets are denoted by calligraphic letters, e.g., the alphabet of a r.v. X is denoted by $\mathcal{X}$. The cartesian product of sets $\mathcal{X}$ and $\mathcal{Y}$ is denoted by $\mathcal{X} \times \mathcal{Y}$. The n-fold cartesian product of a set $\mathcal{X}$ is represented by $\mathcal{X}^n$. The set of probability distributions on $\mathcal{X}$ is denoted by $\mathcal{P}_{\mathcal{X}}$. We will extensively use the method of types [20]. Accordingly, the *type* (or empirical distribution) of a sequence $x^n \in \mathcal{X}^n$ is denoted by P_{x^n} or $P_{\tilde{X}}$, where $\tilde{X}$ denotes a r.v. with distribution equal to the empirical distribution of x^n. The type class of $P_{\tilde{X}}$, i.e., the set of sequences of length n with type $P_{\tilde{X}}$ is denoted by $\mathcal{T}_n(P_{\tilde{X}})$ or $\mathcal{T}_n(\tilde{X})$. The set of all possible types of sequences of length n with alphabet $\mathcal{X}$ is denoted by $\mathcal{T}_n(\mathcal{X})$. Similar notations will be used for pairs and larger combinations of sequences, e.g., the joint type of (x^n, y^n) is denoted by $P_{x^n y^n}$ or $P_{\tilde{X}\tilde{Y}}$, where $\tilde{X}\tilde{Y}$ is a r.v. with distribution $P_{x^n y^n}$. By abuse of notation, $P_{\tilde{X}} \in \mathcal{F}$, $\mathcal{F} \subseteq \mathcal{P}_{\mathcal{X}}$, will also be denoted by $\tilde{X} \in \mathcal{F}$, e.g., $P_{\tilde{X}} \in \mathcal{T}_n(\mathcal{X})$ by $\tilde{X} \in \mathcal{T}_n(\mathcal{X})$. For a given $x^n \in \mathcal{T}_n(P_{\tilde{X}})$, the conditional type class of x^n for a conditional type $P_{\tilde{Y}|\tilde{X}}$, i.e., the set of $y^n \in \mathcal{T}_n(P_{\tilde{Y}})$ such that $(x^n, y^n) \in \mathcal{T}_n(P_{\tilde{X}\tilde{Y}})$, is denoted by $\mathcal{T}_n(P_{\tilde{Y}|\tilde{X}}, x^n)$. The Shannon entropy of X, the mutual information between X and Y, and the KL divergence between X and $\hat{X}$ with same support $\mathcal{X}$ are denoted by $H(X)$, $I(X;Y)$ and $D(X||\hat{X})$ (or $D(P_X||P_{\hat{X}})$), respectively. The conditional divergence between two distributions $P_{X_1|X_2}$ and $P_{\bar{X}_1|\bar{X}_2}$ (defined on same alphabets) is denoted by $D\left(P_{X_1|X_2}||P_{\bar{X}_1|\bar{X}_2}\middle|P_{X_2}\right)$ or $D\left(X_1|X_2||\bar{X}_1|\bar{X}_2\middle|X_2\right)$ where,

$$\begin{aligned} D\left(P_{X_1|X_2}||P_{\bar{X}_1|\bar{X}_2}\middle|P_{X_2}\right) &:= D\left(X_1|X_2||\bar{X}_1|\bar{X}_2\middle|X_2\right) \\ &:= \sum_{x_2 \in \mathcal{X}_2} P_{X_2}(x_2) D\left(P_{X_1|X_2=x_2}||P_{\bar{X}_1|\bar{X}_2=x_2}\right). \end{aligned}$$

When $X_2 = \bar{X}_2$, the notation above is further simplified to $D\left(P_{X_1|X_2}||P_{\bar{X}_1|X_2}\right)$ or $D\left(X_1|X_2||\bar{X}_1|X_2\right)$. The set of r-divergent sequences from X is denoted by $\mathcal{J}_n^r(X)$, i.e.,

$$\mathcal{J}_n^r(X) = \{x^n \in \mathcal{X}^n : D(P_{x^n}||P_X) \leq r\}.$$

The limiting inequalities $\lim_{k\to\infty} a_k = b$, $\lim_{k\to\infty} a_k \geq b$, etc. are denoted by $a_k \xrightarrow{(k)} b$, $a_k \overset{(k)}{\geq} b_k$, etc., respectively. Probabilistic events are denoted by calligraphic letters, e.g., $\mathcal{E}$, and its probability by $\mathbb{P}(\mathcal{E})$. The complement of $\mathcal{E}$ is denoted by $\mathcal{E}^c$. Finally, the indicator function is denoted by $\mathbb{1}(\cdot)$ and the standard asymptotic notations of Big-o, Big-omega and Little-o are represented by $O(\cdot)$, $\Omega(\cdot)$ and $o(\cdot)$, respectively.

3.3.2 Problem formulation

All r.v.'s considered in this chapter are discrete with finite support unless specified otherwise, and all logarithms are with respect to the natural base e. Let $k, n \in \mathbb{Z}^+$. The encoder observes source sequence u^k, and transmits codeword $x^n = f^{(k,n)}(u^k)$, where $f^{(k,n)} : \mathcal{U}^k \mapsto \mathcal{X}^n$ represents the encoding function (possibly stochastic) of the observer. The channel output y^n of the DMC $P_{Y|X}$ given input x^n is generated according to the probability law given in (2.3). Depending on the received symbols y^n and side-information v^k observed at the detector, the detector makes a decision between the two hypotheses H_0 and H_1 given in (3.1). Let $P_{U^kV^kX^nY^n} := P_{U^kV^k}P_{X^n|U^k}P_{Y^n|X^n}$ and $P_{\bar{U}^k\bar{V}^k\bar{X}^n\bar{Y}^n} := P_{\bar{U}^k\bar{V}^k}P_{\bar{X}^n|\bar{U}^k}P_{\bar{Y}^n|\bar{X}^n}$ denote the probability distribution of the source sequence, side-information, channel input and channel output under the null and alternate hypothesis, respectively, where $P_{X^n|U^k}(x^n|u^k) = P_{\bar{X}^n|\bar{U}^k}(x^n|u^k) = \mathbb{P}(f^{(k,n)}(u^k) = x^n)$ for all $(u^k, x^n) \in \mathcal{U}^k \times \mathcal{X}^n$, and $P_{\bar{Y}^n|\bar{X}^n} := P_{Y^n|X^n}$. Let $H \in \{0, 1\}$ denote the actual hypothesis and $\hat{H} \in \{0, 1\}$ denote the output of the hypothesis test, where 0 and 1 denote H_0 and H_1, respectively. Let $\mathcal{A}_{k,n} \subseteq \mathcal{V}^k \times \mathcal{Y}^n$ denote the acceptance region for H_0. Then, the decision rule $g^{(k,n)} : \mathcal{V}^k \times \mathcal{Y}^n \mapsto \{0, 1\}$ is given by

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

Let

$$\alpha\left(k, n, f^{(k,n)}, g^{(k,n)}\right) := 1 - P_{V^k Y^n}\left(\mathcal{A}_{k,n}\right),$$
$$\text{and } \beta\left(k, n, f^{(k,n)}, g^{(k,n)}\right) := P_{\bar{V}^k \bar{Y}^n}\left(\mathcal{A}_{k,n}\right),$$

denote the type I and type II error probabilities for the encoding function $f^{(k,n)}$ and decision rule $g^{(k,n)}$, respectively. The following definition formally states the error-exponents trade-off we aim to characterize.

Definition 3.1. Let $\tau \in (0, \infty)$. An exponent pair $(\kappa_\alpha, \kappa_\beta)$ is τ-achievable if there exists sequences of integers k and n_k, corresponding sequence of encoding functions $f^{(k,n_k)}$ and decoding functions $g^{(n_k)}$, and $k_0 \in \mathbb{Z}^+$ such that

$$n_k \leq \tau k, \ \forall\ k \geq k_0, \tag{3.2a}$$

$$\alpha\left(k, n_k, f^{(k,n_k)}, g^{(k,n_k)}\right) \leq e^{-k\kappa_\alpha}, \ \forall\ k \geq k_0, \tag{3.2b}$$

$$\text{and } \liminf_{k\to\infty} -\frac{1}{k}\log\left(\beta\left(k, n_k, f^{(k,n_k)}, g^{(k,n_k)}\right)\right) \geq \kappa_\beta. \tag{3.2c}$$

Let

$$\kappa(\tau, \kappa_\alpha) := \sup\{\kappa_\beta : (\kappa_\alpha, \kappa_\beta) \text{ is } \tau\text{-achievable}\}.$$

For a fixed $\tau \in (0, \infty)$, we are interested in characterizing the boundary of the set of all τ-achievable $(\kappa_\alpha, \kappa_\beta)$ tuples defined as

$$\mathcal{R} := \{(\kappa_\alpha, \kappa(\tau, \kappa_\alpha)) : \kappa_\alpha \in (0, \infty]\}.$$

Towards this, we will first obtain a single-letter characterization of $\mathcal{R}$ in the non-distributed setting. This characterization will be subsequently used to obtain an inner bound on $\mathcal{R}$ in the distributed setting.

3.4 HT: Error exponents trade-off

In the non-distributed setting, the hypothesis test in (3.1) specializes to the following test:

$$H_0 : \prod_{i=1}^{k} P_U(u), \ \forall \ u \in \mathcal{U}, \tag{3.3a}$$

$$H_1 : \prod_{i=1}^{k} P_{\bar{U}}(u), \ \forall \ u \in \mathcal{U}. \tag{3.3b}$$

For brevity, we will denote the r.v. with distribution $P_{Y|X=x}$ by Y_x and the corresponding probability distribution by P_{Y_x} for all $x \in \mathcal{X}$. Let us define

$$\kappa_0 := \kappa_0(\tau, P_U, P_{\bar{U}}, P_{Y|X}) := \min\left(D(P_{\bar{U}}||P_U), \tau E_c\right),$$

where,

$$E_c := E_c(P_{Y|X}) := D(P_{Y_a}||P_{Y_b}), \tag{3.4}$$

$$\text{and } (a, b) := \underset{(x,x')\in\mathcal{X}\times\mathcal{X}}{\arg\max} \ D(P_{Y_x}||P_{Y_{x'}}). \tag{3.5}$$

It follows by interchanging P_U and $P_{\bar{U}}$ in Theorem 2 [23] that, we may restrict the range of κ_α within the interval $(0, \kappa_0]$ since $\kappa(\tau, \kappa_\alpha) = 0$ for $\kappa_\alpha \geq \kappa_0$. Hence, $\mathcal{R}$ can be redefined as $\mathcal{R} = \{(\kappa_\alpha, \ \kappa(\tau, \kappa_\alpha)) : \kappa_\alpha \in (0, \kappa_0]\}$.

In order to state our single-letter characterization of $\mathcal{R}$, we need some concepts regarding the log moment generating function (Log-MGF) of a r.v., which we briefly review below. For a given function $f : \mathcal{Z} \mapsto \mathbb{R}$ and a probability distribution P_Z on Z, the log-MGF of Z with respect to f, denoted by $\psi_{Z,f}(\lambda)$ is given by

$$\psi_{Z,f}(\lambda) := \psi_{P_Z,f}(\lambda) := \log\left(\mathbb{E}_{P_Z}\left(e^{\lambda f(Z)}\right)\right).$$

Let

$$\psi^*_{Z,f}(\theta) := \psi^*_{P_Z,f}(\theta) := \sup_{\lambda\in\mathbb{R}} \theta\lambda - \psi_{Z,f}(\lambda). \tag{3.6}$$

The following simple facts are straightforward to verify.

Lemma 3.1. *[6, Theorem 13.2, Theorem 13.3]*

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

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

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

We will assume[4] that

Assumption 3.2. $P_U \ll P_{\bar{U}}$, $P_{\bar{U}} \ll P_U$ and $P_{Y|X}$ is such that $P_{Y_x} \ll P_{Y_{x'}}$, $\forall\ (x, x') \in \mathcal{X} \times \mathcal{X}$.

When u^k is observed directly at the detector, a single-letter characterization of the optimal trade-off between the error-exponents is obtained in [2]. Below, we state an equivalent form of this characterization that is given in [6].

Theorem 3.3. *[6, Theorem 15.1] When* u^k *is observed directly at the detector, then for the HT given in* (3.3)*,*

$$\mathcal{R} = \left\{ \left(\psi^*_{U,f_{\mathcal{U}}}(\theta), \psi^*_{U,f_{\mathcal{U}}}(\theta) - \theta \right) :\ \theta \in \mathcal{I}(U, \bar{U}) \right\},$$

where, $f_{\mathcal{U}} : \mathcal{U} \mapsto \mathbb{R}^+$ *is defined as*

$$f_{\mathcal{U}}(u) := \log\left(\frac{P_{\bar{U}}(u)}{P_U(u)}\right), \tag{3.7}$$

and $\mathcal{I}(P_U, P_{\bar{U}}) := \big(-D(P_U||P_{\bar{U}}), D(P_{\bar{U}}||P_U) \big]$*. The decision rule that achieves the exponent pair* $\left(\psi^*_{U,f_{\mathcal{U}}}(\theta), \psi^*_{U,f_{\mathcal{U}}}(\theta) - \theta \right)$ *is the NP test [1] given by*

$$g^{(k)}_{\theta,\mathcal{U}}(u^k) = \mathbb{1}\left(\sum_{i=1}^{k} \log\left(\frac{P_{\bar{U}}(u_i)}{P_U(u_i)}\right) \geq k\theta \right). \tag{3.8}$$

[4]This technical condition ensures that for functions f and distributions P that we consider below, $\psi_{P,f}(\lambda) < \infty,\ \forall\ \lambda \in \mathbb{R}$.

To prove the main result, a strong converse result that follows from [6, Theorem 12.5] will turn out to be handy. We state it below for completeness. For the scenario where u^k is observed directly at the detector, let us denote the type I and type II error probabilities achieved by a decision rule (possibly stochastic) $\bar{g}^{(k)} : \mathcal{U}^k \mapsto \{0, 1\}$ by $\alpha'(\bar{g}^{(k)})$ and $\beta'(\bar{g}^{(k)})$, respectively. The following single-shot result provides a lower bound on a weighted sum of the type I and type II error probabilities[5].

Theorem 3.4. *[6, Theorem 12.5] For any $k \in \mathcal{Z}^+$ and any decision rule $\bar{g}^{(k)}$ as defined above,*

$$\alpha'\left(\bar{g}^{(k)}\right) + \gamma\beta'\left(\bar{g}^{(k)}\right) \geq P_{U^k}\left(\log\left(\frac{P_{U^k}(U^k)}{P_{\bar{U}^k}(U^k)}\right) \leq \log\gamma\right), \ \forall\ \gamma > 0,$$

where, $\alpha'(\bar{g}^{(k)})$ and $\beta'(\bar{g}^{(k)})$ denote the type I and type II error probabilities for decision rule $\bar{g}^{(k)}$.

We will also require a slight generalization of Theorem 3.3 for the case when the data samples are drawn from a product of finite non-identical distributions, i.e., the samples are independent, but not necessarily identically distributed. For an arbitrary given joint distribution $P_{X_0X_1} \in \mathcal{P}(\mathcal{X} \times \mathcal{X})$, let $\{(x_0^n, x_1^n)\}_{n\in\mathbb{Z}^+}$ denote a given pair of sequences such that

$$P_{x_0^n x_1^n}(x, x') \xrightarrow{(n)} P_{X_0X_1}(x, x'),\ \forall\ (x, x') \in \mathcal{X} \times \mathcal{X}. \tag{3.9}$$

Consider the following HT:

$$H_0 : Y^n \sim \prod_{i=1}^{n} P_{Y_{x_{0i}}}, \tag{3.10a}$$

$$H_1 : Y^n \sim \prod_{i=1}^{n} P_{Y_{x_{1i}}}. \tag{3.10b}$$

For a given decision rule $g^{(n)}(y^n) = 1 - \mathbb{1}\left(Y^n \in \mathcal{A}_n\right)$ with acceptance region $\mathcal{A}_n \subseteq \mathcal{Y}^n$ for H_0, let $\bar{\alpha}\left(n, g^{(n)}, x_0^n, x_1^n\right)$ and $\bar{\beta}\left(n, g^{(n)}, x_0^n, x_1^n\right)$ denote the type I and type II error

[5]Note that α denotes the complement of the type I error probability in [6, Theorem 12.5], whereas we use α' to denote the type I error probability.

probabilities, respectively, where

$$\bar{\alpha}\left(n, g^{(n)}, x_0^n, x_1^n\right) := 1 - \prod_{i=1}^{n} P_{Y_{x_{0i}}}\left(\mathcal{A}_n\right),$$
$$\text{and } \bar{\beta}\left(n, g^{(n)}, x_0^n, x_1^n\right) := \prod_{i=1}^{n} P_{Y_{x_{1i}}}\left(\mathcal{A}_n\right).$$

Definition 3.5. For a given joint distribution $P_{X_0X_1}$ and a pair of infinite sequences $\{(x_0^n, x_1^n)\}_{n\in\mathbb{Z}^+}$ such that (3.9) holds, an exponent pair $(\kappa_\alpha, \kappa_\beta)$ is achievable for the HT in (3.10) if there exists a sequence of decision rules $\{g^{(n)}\}_{n\in\mathbb{Z}^+}$ and $n_0 \in \mathbb{Z}^+$ such that

$$\bar{\alpha}\left(n, g^{(n)}, x_0^n, x_1^n\right) \leq e^{-n\kappa_\alpha},\ \forall\ n \geq n_0, \tag{3.11a}$$
$$\text{and } \liminf_{n\to\infty} -\frac{1}{n}\log\left(\bar{\beta}\left(n, g^{(n)}, x_0^n, x_1^n\right)\right) \geq \kappa_\beta. \tag{3.11b}$$

As will become evident later, the performance of the HT in (3.10) depends on $\{(x_0^n, x_1^n)\}_{n\in\mathbb{Z}^+}$ only through $P_{X_0X_1}$. Let

$$\bar{\kappa}_c(\kappa_\alpha, P_{X_0X_1}) := \sup\{\kappa_\beta : (\kappa_\alpha, \kappa_\beta) \text{ is achievable for HT in (3.10)}\}.$$
$$\text{and } \mathcal{R}_N(P_{X_0X_1}) := \{(\kappa_\alpha, \bar{\kappa}_c(\kappa_\alpha, P_{X_0X_1})) : \kappa_\alpha \in (0, \kappa_\alpha^*]\}.$$

where κ_α^* is the smallest number such that $\bar{\kappa}(\kappa_\alpha^*, P_{X_0X_1}) = 0$. The following proposition provides a single-letter characterization of $\mathcal{R}_N(P_{X_0X_1})$, and will be used later for obtaining a single-letter characterization of $\mathcal{R}$ in Theorem 3.7.

Proposition 3.6.

$$\mathcal{R}_N(P_{X_0X_1}) = \Bigg\{ \left(\mathbb{E}_{P_{X_0X_1}}\left(\psi^*_{Y_{X_0},\tilde{h}_{X_0,X_1}}(\theta)\right),\ \mathbb{E}_{P_{X_0X_1}}\left(\psi^*_{Y_{X_0},\tilde{h}_{X_0,X_1}}(\theta)\right) - \theta\right), \\ \theta \in \mathcal{I}(P_{X_0X_1}, P_{Y|X}) \Bigg\},$$

where, for each $(x, x') \in \mathcal{X}\times\mathcal{X}$, $\tilde{h}_{x,x'} : \mathcal{Y} \mapsto \mathbb{R}$ is given by

$$\tilde{h}_{x,x'}(y) := \log\left(\frac{P_{Y_{x'}}(y)}{P_{Y_x}(y)}\right), \tag{3.12}$$

and

$$\mathcal{I}(P_{X_0X_1}, P_{Y|X}) := \big(- D(P_{Y_{X_0}}||P_{Y_{X_1}}|P_{X_0X_1}), D(P_{Y_{X_1}}||P_{Y_{X_0}}|P_{X_0X_1})\big],$$

$$D(P_{Y_{X_0}}||P_{Y_{X_1}}|P_{X_0X_1}) := \sum_{(x_0,x_1)\in\mathcal{X}\times\mathcal{X}} P_{X_0X_1}(x_0,x_1) D(P_{Y_{x_0}}||P_{Y_{x_1}}),$$

$$D(P_{Y_{X_1}}||P_{Y_{X_0}}|P_{X_0X_1}) := \sum_{(x_0,x_1)\in\mathcal{X}\times\mathcal{X}} P_{X_0X_1}(x_0,x_1) D(P_{Y_{x_1}}||P_{Y_{x_0}}).$$

The decision rule that achieves the exponent pair $\left(\mathbb{E}_{P_{X_0X_1}}\left(\psi^*_{Y_{X_0},\tilde{h}_{X_0,X_1}}(\theta)\right),\ \mathbb{E}_{P_{X_0X_1}}\left(\psi^*_{Y_{X_0},\tilde{h}_{X_0,X_1}}(\theta)\right) - \theta\right)$ is the NP test given by

$$g^{(n)}_{\theta,\mathcal{Y}}(y^n) = \mathbb{1}\left(\sum_{i=1}^{n} \log\left(\frac{P_{Y_{x_{1i}}}(y_i)}{P_{Y_{x_{0i}}}(y_i)}\right) \geq n\theta\right). \tag{3.13}$$

Proof. The proof is given in Appendix B.1. □

The next theorem provides a single-letter characterization of $\kappa(\tau,\kappa_\alpha)$ and thereby of $\mathcal{R}$.

Theorem 3.7.

$$\kappa(\tau,\kappa_\alpha) = \sup\{\kappa_\beta : (\kappa_\alpha,\kappa_\beta) \in \mathcal{R}^*\}$$

where

$$\mathcal{R}^* := \bigcup_{\substack{P_{X_0X_1}\in \\ \mathcal{P}_{\mathcal{X}\times\mathcal{X}}}} \bigcup_{\substack{(\theta_0,\theta_1)\ \in \\ \mathcal{I}(P_U,P_{\bar{U}})\times\mathcal{I}(P_{X_0X_1},P_{Y|X})}} \left(\zeta_0(\theta_0,\theta_1,P_{X_0X_1}),\ \zeta_1(\theta_0,\theta_1,P_{X_0X_1})\right),$$

$$\zeta_0(\theta_0,\theta_1,P_{X_0X_1}) := \min\left\{\psi^*_{U,f_\mathcal{U}}(\theta_0),\ \tau\mathbb{E}_{P_{X_0X_1}}\left(\psi^*_{Y_{X_0},\tilde{h}_{X_0,X_1}}(\theta_1)\right)\right\},$$

$$\zeta_1(\theta_0,\theta_1,P_{X_0X_1}) := \min\left\{\psi^*_{U,f_\mathcal{U}}(\theta_0) - \theta_0,\ \tau\left(\mathbb{E}_{P_{X_0X_1}}\left(\psi^*_{Y_{X_0},\tilde{h}_{X_0,X_1}}(\theta_1)\right) - \theta_1\right)\right\},$$

where $f_\mathcal{U}$ and $\tilde{h}_{x,x'}$, $(x,x') \in \mathcal{X}\times\mathcal{X}$ are defined in (3.7) *and* (3.12)*, respectively.*

To provide some intuition about the terms appearing in $\zeta_0(\cdot)$ and $\zeta_1(\cdot)$ above, note that $\psi^*_{U,f_\mathcal{U}}(\theta_0)$ and $\psi^*_{U,f_\mathcal{U}}(\theta_0) - \theta_0$ are the same as the terms that appear in Theorem 3.3 which characterizes the error-exponent trade-off when u^k is directly available at the

detector. On the other hand, the second term within the minimum in $\zeta_0(\cdot)$ and $\zeta_1(\cdot)$ are additional factors introduced due to the noisiness of the communication channel.

Proof of Theorem 3.7. Achievability: Fix $P_{X_0X_1} \in \mathcal{P}_{\mathcal{X}\times\mathcal{X}}$. Let $n_k = \lfloor \tau k \rfloor$, and $x_0^{n_k}$ and $x_1^{n_k}$ be two arbitrary sequences from the set $\mathcal{X}^{n_k}$ such that (3.9) holds. Let $\theta_0 \in \mathcal{I}(P_U, P_{\bar{U}})$ and $\theta_1 \in \mathcal{I}(P_{X_0X_1}, P_{Y|X})$. The achievability scheme is as follows:
The encoder first locally performs the NP test $g^{(k)}_{\theta,\mathcal{U}}$ given in (3.8) on the observed samples u^k with $\theta = \theta_0$, and outputs the channel input codeword $f^{(k,n_k)}(u^k)$ according to the following rule:

$$f^{(k,n_k)}(u^k) = \begin{cases} x_0^{n_k}, & \text{if } g^{(k)}_{\theta_0,\mathcal{U}}(u^k) = 0, \\ x_1^{n_k}, & \text{otherwise.} \end{cases}$$

Based on the observed samples y^{n_k}, the detector outputs the decision of the HT according to the decision rule $g^{(n_k)} = g^{(n_k)}_{\theta_1,\mathcal{Y}}$ defined in (3.13). Let

$$\mathcal{A}^{(n_k)}_{\theta_1} = \left\{ y^{n_k} \in \mathcal{Y}^{n_k} : \sum_{i=1}^{n_k} \log\left(\frac{P_{Y|X=x_{1i}}(y_i)}{P_{Y|X=x_{0i}}(y_i)}\right) < n_k\theta_1 \right\}. \tag{3.14}$$

The type I error error probability can be upper bounded for sufficiently large k (and n_k) as follows:

$$\begin{aligned}
&\alpha\left(k, n_k, f^{(k,n_k)}, g^{(n_k)}\right) \\
&\le \mathbb{P}\left(g^{(k)}_{\theta_0,\mathcal{U}}(U^k) = 1\right) P_{Y^{n_k}|X^{n_k}=x_1^{n_k}}\left(\mathcal{Y}^{n_k} \backslash \mathcal{A}^{(n_k)}_{\theta_1}\right) \\
&\qquad + \mathbb{P}\left(g^{(k)}_{\theta_0,\mathcal{U}}(U^k) = 0\right) P_{Y^{n_k}|X^{n_k}=x_0^{n_k}}\left(\mathcal{Y}^{n_k} \backslash \mathcal{A}^{(n_k)}_{\theta_1}\right) \\
&\le \mathbb{P}\left(g^{(k)}_{\theta_0,\mathcal{U}}(U^k) = 1\right) + P_{Y^{n_k}|X^{n_k}=x_0^{n_k}}\left(\mathcal{Y}^{n_k} \backslash \mathcal{A}^{(n_k)}_{\theta_1}\right) \\
&\le e^{-k\left(\psi^*_{U,f_{\mathcal{U}}}(\theta_0)-\delta\right)} + e^{-n_k\left(\mathbb{E}_{P_{X_0X_1}}\left(\psi^*_{Y_{X_0},\tilde{h}_{X_0,X_1}}(\theta_1)\right)-\delta\right)},
\end{aligned} \tag{3.15}$$

where $\delta > 0$ is an arbitrary small, but fixed number. Similarly, the type II error probability can be upper bounded as follows:

$$\beta\left(k, n_k, f^{(k,n_k)}, g^{(n_k)}\right)$$

$$
\begin{aligned}
&\leq \mathbb{P}\left(g^{(k)}_{\theta_0,\mathcal{U}}(\bar{U}^k)=0\right) P_{Y^{n_k}|X^{n_k}=x_0^{n_k}}\left(\mathcal{A}^{(n_k)}_{\theta_1}\right) + \mathbb{P}\left(g^{(k)}_{\theta_0,\mathcal{U}}(\bar{U}^k)=1\right) \\
&\quad P_{Y^{n_k}|X^{n_k}=x_1^{n_k}}\left(\mathcal{A}^{(n_k)}_{\theta_1}\right) \\
&\leq \mathbb{P}\left(g^{(k)}_{\theta_0,\mathcal{U}}(\bar{U}^k)=0\right) + P_{Y^{n_k}|X^{n_k}=x_1^{n_k}}\left(\mathcal{A}^{(n_k)}_{\theta_1}\right) \\
&\leq e^{-k\left(\psi^*_{\bar{U},f_{\mathcal{U}}}(\theta_0)-\delta\right)} + e^{-n_k\left(\mathbb{E}_{P_{X_0X_1}}\left(\psi^*_{Y_{X_1},\tilde{h}_{X_0,X_1}}(\theta_1)\right)-\delta\right)} \\
&= e^{-k\left(\psi^*_{U,f_{\mathcal{U}}}(\theta_0)-\theta_0-\delta\right)} + e^{-n_k\left(\mathbb{E}_{P_{X_0X_1}}\left(\psi^*_{Y_{X_0},\tilde{h}_{X_0,X_1}}(\theta_1)\right)-\theta_1-\delta\right)}. \qquad (3.16)
\end{aligned}
$$

It follows from (3.15) and (3.16), respectively, that,

$$
\begin{aligned}
&\liminf_{k\to\infty} -\frac{1}{k}\log\left(\alpha\left(k,n_k,f^{(k,n_k)},g^{(n_k)}\right)\right) \\
&\geq \min\left(\psi^*_{U,f_{\mathcal{U}}}(\theta_0),\ \mathbb{E}_{P_{X_0X_1}}\left(\psi^*_{Y_{X_0},\tilde{h}_{X_0,X_1}}(\theta_1)\right)\right)-\delta \\
&= \zeta_0(\theta_0,\theta_1,P_{X_0X_1})-\delta,
\end{aligned}
$$

and

$$
\begin{aligned}
&\liminf_{k\to\infty} -\frac{1}{k}\log\left(\beta\left(k,n_k,f^{(k,n_k)},g^{(n_k)}\right)\right) \\
&\geq \min\left(\psi^*_{U,f_{\mathcal{U}}}(\theta_0)-\theta_0,\ \mathbb{E}_{P_{X_0X_1}}\left(\psi^*_{Y_{X_0},\tilde{h}_{X_0,X_1}}(\theta_1)\right)-\theta_1\right)-\delta \\
&= \zeta_1(\theta_0,\theta_1,P_{X_0X_1})-\delta.
\end{aligned}
$$

Since δ is arbitrary, it follows by varying $P_{X_0X_1}\in\mathcal{P}_{\mathcal{X}\times\mathcal{X}}$, $\theta_0\in\mathcal{I}(P_U,P_{\bar{U}})$ and $\theta_1\in\mathcal{I}(P_{X_0X_1},P_{Y|X})$ that

$$
\kappa(\tau,\kappa_\alpha)\geq\sup\{\kappa_\alpha:\ (\kappa_\alpha,\kappa_\beta)\in\mathcal{R}^*\}.
$$

This completes the proof of achievability.

Converse: From the proof of the converse part of Theorem 3.3, it follows that for $\theta_0\in\mathcal{I}(P_U,P_{\bar{U}})$,

$$
\mathcal{R}\subseteq\bigcup_{\theta_0\in\mathcal{I}(P_U,P_{\bar{U}})}\left(\psi^*_{U,f_{\mathcal{U}}}(\theta_0),\ \psi^*_{U,f_{\mathcal{U}}}(\theta_0)-\theta_0\right). \qquad (3.17)
$$

Also, note that for any encoding function $f^{(k,n_k)}$ and decoding function $g^{(n_k)}$ with decision region $\mathcal{A}^{(n_k)} \subseteq \mathcal{Y}^{n_k}$, we have

$$\begin{aligned}
&\alpha\left(k, n_k, f^{(k,n_k)}, g^{(n_k)}\right) \\
&= \sum_{u^k \in \mathcal{U}^k} P_U(u^k) \sum_{x^{n_k} \in \mathcal{X}^{n_k}} P_{X^{n_k}|U^k=u^k}(x^{n_k}) P_{Y^{n_k}|X^{n_k}=x^{n_k}}(\mathcal{Y}^{n_k} \backslash \mathcal{A}^{n_k}) \\
&\geq P_{Y^{n_k}|X^{n_k}=\bar{x}_0^{n_k}}(\mathcal{Y}^{n_k} \backslash \mathcal{A}^{n_k}), \qquad (3.18)
\end{aligned}$$

for some $\bar{x}_0^{n_k} \in \mathcal{X}^{n_k}$ that depends on $\mathcal{A}^{n_k}$. Similarly,

$$\begin{aligned}
\beta\left(k, n_k, f^{(k,n_k)}, g^{(n_k)}\right) &= \sum_{u^k \in \mathcal{U}^k} P_{\bar{U}}(u^k) \sum_{x^{n_k} \in \mathcal{X}^{n_k}} P_{X^{n_k}|U^k=u^k}(x^{n_k}) P_{Y^{n_k}|X^{n_k}=x^{n_k}}(\mathcal{A}^{n_k}) \\
&\geq P_{Y^{n_k}|X^{n_k}=\bar{x}_1^{n_k}}(\mathcal{A}^{n_k}), \qquad (3.19)
\end{aligned}$$

for some $\bar{x}_1^{n_k} \in \mathcal{X}^{n_k}$ (depends on $\mathcal{A}^{n_k}$). Let $\bar{P}_{X_0X_1}$ denote the joint type of the sequences $(\bar{x}_0^n, \bar{x}_1^n)$. Note that the R.H.S. of (3.18) and (3.19) correspond to the type I and type II error probabilities of the HT given in (3.10) with $n = n_k$, $x_0^{n_k} = \bar{x}_0^{n_k}$ and $x_1^{n_k} = \bar{x}_1^{n_k}$. Then, it follows from the converse part of the proof of Lemma 3.6 that if for some $\theta_1 \in \mathcal{I}(\bar{P}_{X_0X_1}, P_{Y|X})$ and all sufficiently large n_k, it holds that,

$$\alpha\left(k, n_k, f^{(k,n_k)}, g^{(n_k)}\right) < e^{-n_k\left(\mathbb{E}_{\bar{P}_{X_0X_1}}\left(\psi^*_{Y_{X_0},\tilde{h}_{X_0,X_1}}(\theta_1)\right)\right)}, \qquad (3.20)$$

then

$$\limsup_{k\to\infty} -\frac{1}{k}\log\left(\beta\left(k, n_k, f^{(k,n_k)}, g^{(n_k)}\right)\right) \leq \tau\left(\mathbb{E}_{\bar{P}_{X_0X_1}}\left(\psi^*_{Y_{X_0},\tilde{h}_{X_0,X_1}}(\theta_1) - \theta_1\right)\right). \qquad (3.21)$$

From (3.20) and (3.21), we have

$$\begin{aligned}
\mathcal{R} \subseteq \bigcup_{P_{X_0X_1} \in P_{\mathcal{X}\times\mathcal{X}}} \bigcup_{\theta_1 \in \mathcal{I}(P_{X_0X_1}, P_{Y|X})} \Bigg(&\tau\mathbb{E}_{P_{X_0X_1}}\left(\psi^*_{P_{Y_{X_0}},\tilde{h}_{X_0,X_1}}(\theta_1)\right), \\
&\tau\left(\mathbb{E}_{P_{X_0X_1}}\left(\psi^*_{P_{Y_{X_0}},\tilde{h}_{X_0,X_1}}(\theta_1)\right) - \theta_1\right)\Bigg). \qquad (3.22)
\end{aligned}$$

It follows from (3.17) and (3.22) that $(\kappa_\alpha, \kappa_\beta) \in \mathcal{R}$ only if there exists some $P_{X_0X_1}$ and $(\theta_0, \theta_1) \in \mathcal{I}(P_U, P_{\bar{U}}) \times \mathcal{I}(P_{X_0X_1}, P_{Y|X})$ such that

$$\begin{aligned}\kappa_\alpha &\leq \min\left\{\psi^*_{U,f_{\mathcal{U}}}(\theta_0),\ \tau\mathbb{E}_{P_{X_0X_1}}\left(\psi^*_{Y_{X_0},\tilde{h}_{X_0,X_1}}(\theta_1)\right)\right\},\\ \text{and } \kappa_\beta &\leq \min\left\{\psi^*_{U,f_{\mathcal{U}}}(\theta_0)-\theta_0,\ \tau\left(\mathbb{E}_{P_{X_0X_1}}\left(\psi^*_{Y_{X_0},\tilde{h}_{X_0,X_1}}(\theta_1)\right)-\theta_1\right)\right\},\end{aligned}$$

from which it follows that $\kappa(\tau, \kappa_\alpha) \leq \sup\{\kappa_\beta : (\kappa_\alpha, \kappa_\beta) \in \mathcal{R}^*\}$. This completes the proof. □

Remark 3.8. The optimal T2EE for a fixed constraint on the type I error probability can be recovered by taking limit $\theta_0 \to -D(P_U||P_{\bar{U}})$ and $\theta_1 \to -D(P_{Y_{X_0}}||P_{Y_{X_1}}|P_{X_0X_1})$. In this case,

$$\zeta_0\left(-D(P_U||P_{\bar{U}}), -D(P_{Y_{X_0}}||P_{Y_{X_1}}|P_{X_0X_1}), P_{X_0X_1}\right) = 0$$

and

$$\begin{aligned}&\zeta_1\left(-D(P_U||P_{\bar{U}}), -D(P_{Y_{X_0}}||P_{Y_{X_1}}|P_{X_0X_1}), P_{X_0X_1}\right)\\ &= \min\left\{D(P_U||P_{\bar{U}}), D(P_{Y_{X_0}}||P_{Y_{X_1}}|P_{X_0X_1})\right\}.\end{aligned}$$

Maximizing the second argument over all possible $P_{X_0X_1}$ yields,

$$\max_{P_{X_0X_1}\in\mathcal{P}_{\mathcal{X}\times\mathcal{X}}} D(P_{Y_{X_0}}||P_{Y_{X_1}}|P_{X_0X_1}) = E_c. \tag{3.23}$$

Hence, $\lim_{\kappa_\alpha\to 0}\kappa(\tau,\kappa_\alpha) = \kappa_0$, which is equal to the optimal T2EE established in Theorem 2 [23]. Note that $E_c < \infty$ under the assumption $P_{Y_x} << P_{Y_{x'}}$, $\forall\ (x, x') \in \mathcal{X}\times\mathcal{X}$.

3.5 Distributed HT: Error-exponents trade-off

In [27, Theorem 1], Han and Kobayashi obtained an inner bound on $\mathcal{R}$ in the distributed setting, where the communication channel is rate-limited and noiseless. At a high level, their coding scheme involved a type-based quantization of sequences u^k,

whose type P_{u^k} lies within a distance (in terms of KL divergence) equal to the desired type I error-exponent κ_α from P_U. The index of the codeword within the quantization codebook is revealed to the detector which takes the decision on the hypothesis based on the received index and side-information v^k. On the other hand, it is well-known from [11] that by performing binning subsequent to quantization and decoding using the side-information v^k can help reduce the communication rate to the detector. This enables better quantization of u^k and higher error-exponent in channel coding. However, these benefits come at a cost, as binning introduces additional errors that affect the type I and type II error probability. More specifically, a *binning error* occurs when the detector decodes the incorrect quantization codeword from the correctly decoded bin-index.

In this section, we will obtain inner bounds on $\mathcal{R}$ using a generalization of the SHTCC and JHTCC schemes in Chapter 2. In this chapter, we will refer to these generalized schemes also as SHTCC and JHTCC scheme, respectively.

3.5.1 SHTCC scheme:

The SHTCC scheme is a combination of a generalization of the Shimokawa-Han-Amari (SHA) scheme [11] and the Borade-Nakiboğlu-Zheng unequal error-protection scheme [22]. More specifically, the scheme involves

(i) quantization and binning of sequences u^k whose type P_{u^k} is within a distance of κ_α (in terms of KL-divergence) from P_U (instead of just the dominant type P_U as is done in the SHA scheme), and using the side-information v^k to decode for the quantization codeword from the (decoded) bin-index at the detector.

(ii) unequal error-protection channel coding scheme in [22] for protecting a special message which informs the detector that P_{u^k} is at a distance greater than κ_α from P_U.

Before, we state the inner bound on $\mathcal{R}$ achieved by the SHTCC scheme, some definitions are required. Let S denote a r.v. with support $\mathcal{S} = \mathcal{X}$, such that $S - X - Y$ and

$P_{SXY} = P_{SX}P_{Y|X}$. For $x \in \mathcal{X}$, we define

$$r_x(y) := \log\left(\frac{P_{Y|X=x}(y)}{P_{Y|S=x}(y)}\right), \tag{3.24}$$

and

$$E_m(P_{SX}, \theta) = \sum_{s\in\mathcal{S}} P_S(s)\psi^*_{P_{Y|S=s}, r_s}(\theta). \tag{3.25}$$

Let Ω denote the set of all continuous mappings from $\mathcal{P}_{\mathcal{U}}$ to $\mathcal{P}_{\mathcal{W}|\mathcal{U}}$, where $\mathcal{W}$ is an arbitrary finite set. Let

$$\theta_L(P_{SX}) := \sum_{s\in\mathcal{S}} P_S(s) D\left(P_{Y|S=s}||P_{Y|X=s}\right), \tag{3.26}$$

$$\theta_U(P_{SX}) := \sum_{s\in\mathcal{S}} P_S(s) D\left(P_{Y|X=s}||P_{Y|S=s}\right), \tag{3.27}$$

$$\Theta(P_{SX}) := (-\theta_L(P_{SX}), \theta_U(P_{SX})), \tag{3.28}$$

$$\begin{aligned}
&\mathcal{L}(\kappa_\alpha, \tau) \\
&:= \left\{ \begin{array}{l} (\omega, R, P_{SX}, \theta) \in \Omega \times \mathbb{R}^+ \times \mathcal{P}_{\mathcal{S}\mathcal{X}} \times \Theta(P_{SX}) : \zeta_q(\kappa_\alpha, \omega) - \rho(\kappa_\alpha, \omega) \le R < \\ \tau I(X;Y|S), \ \min\left(\tau E_m(P_{SX}, \theta), \tau E_x\left(\frac{R}{\tau}, P_{SX}\right), E_b(\kappa_\alpha, \omega, R)\right) \ge \kappa_\alpha \end{array} \right\}, \\
&\hat{\mathcal{L}}(\kappa_\alpha, \omega) \\
&:= \left\{ \hat{U}\hat{V}\hat{W} : D(\hat{U}\hat{V}\hat{W}||UVW) \le \kappa_\alpha, P_{W|U} = P_{\hat{W}|\hat{U}} = \omega(P_{\hat{U}}), V - U - W \quad \right\},
\end{aligned} \tag{3.29}$$

$$E_b(\kappa_\alpha, \omega, R) := \begin{cases} R - \zeta_q(\kappa_\alpha, \omega) + \rho(\kappa_\alpha, \omega) & \text{if } 0 \le R < \zeta_q(\kappa_\alpha, \omega), \\ \infty & \text{otherwise,} \end{cases}$$

$$\zeta_q(\kappa_\alpha, \omega) := \max_{\substack{\hat{U}\hat{W}:\ \exists \hat{V}, \\ \hat{U}\hat{V}\hat{W}\in\hat{\mathcal{L}}(\kappa_\alpha,\omega)}} I(\hat{U};\hat{W}),$$

$$\rho(\kappa_\alpha, \omega) := \min_{\substack{\hat{V}\hat{W}:\ \exists \hat{U}, \\ \hat{U}\hat{V}\hat{W}\in\hat{\mathcal{L}}(\kappa_\alpha,\omega)}} I(\hat{V};\hat{W}),$$

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

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

$$E_3(\kappa_\alpha,\omega,R,P_{SX},\tau)$$

$$:= \begin{cases} \min\limits_{\tilde{U}\tilde{V}\tilde{W}\in\mathcal{T}_3(\kappa_\alpha,\omega)} D(\tilde{U}\tilde{V}\tilde{W}||\bar{U}\bar{V}\bar{W}) + E_b(\kappa_\alpha,\omega,R) & \\ \qquad +\tau E_x\left(\frac{R}{\tau},P_{SX}\right), & \text{if } R<\zeta_q(\kappa_\alpha,\omega), \\ \min\limits_{\tilde{U}\tilde{V}\tilde{W}\in\mathcal{T}_3(\kappa_\alpha,\omega)} D(\tilde{U}\tilde{V}\tilde{W}||\bar{U}\bar{V}\bar{W}) + \rho(\kappa_\alpha,\omega) & \\ \qquad +\tau E_x\left(\frac{R}{\tau},P_{SX}\right), & \text{otherwise,} \end{cases}$$

$$E_4(\kappa_\alpha,\omega,R,P_{SX},\theta,\tau)$$

$$:= \begin{cases} \min\limits_{\hat{V}:\hat{U}\hat{V}\hat{W}\in\hat{\mathcal{L}}(\kappa_\alpha,\omega)} D(\hat{V}||\bar{V}) + E_b(\kappa_\alpha,\omega,R) & \\ \qquad +\tau\left(E_m(P_{SX},\theta)-\theta\right), & \text{if } R<\zeta_q(\kappa_\alpha,\omega), \\ \min\limits_{\hat{V}:\hat{U}\hat{V}\hat{W}\in\hat{\mathcal{L}}(\kappa_\alpha,\omega)} D(\hat{V}||\bar{V}) + \rho(\kappa_\alpha,\omega) & \\ \qquad +\tau\left(E_m(P_{SX},\theta)-\theta\right), & \text{otherwise,} \end{cases}$$

where,

$$P_{\bar{W}|\bar{U}} := P_{\tilde{W}|\tilde{U}},\ \bar{V}-\bar{U}-\bar{W},$$

$$\mathcal{T}_1(\kappa_\alpha,\omega) := \left\{ \tilde{U}\tilde{V}\tilde{W}: \begin{array}{c} P_{\tilde{U}\tilde{W}}=P_{\hat{U}\hat{W}},\ P_{\tilde{V}\tilde{W}}=P_{\hat{V}\hat{W}} \\ \text{for some } \hat{U}\hat{V}\hat{W}\in\hat{\mathcal{L}}(\kappa_\alpha,\omega) \end{array} \right\},$$

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

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

We have the following lower bound for $\kappa(\tau,\kappa_\alpha)$.

Theorem 3.9. *$\kappa(\tau,\kappa_\alpha)\geq\kappa_s^*(\tau,\kappa_\alpha)$, where*

$$\kappa_s^*(\tau,\kappa_\alpha) := \max_{\substack{(\omega,R,P_{SX},\theta)\\ \in\mathcal{L}(\kappa_\alpha,\tau)}} \min\Big\{ E_1(\kappa_\alpha,\omega), E_2(\kappa_\alpha,\omega,R), E_3(\kappa_\alpha,\omega,R,P_{SX},\tau),$$

$$E_4(\kappa_\alpha, \omega, R, P_{SX}, \theta, \tau)\Big\}. \tag{3.30}$$

The proof of Theorem 3.9 is presented in Appendix B.2. As a corollary, Theorem 3.9 recovers the lower bound for $\kappa(\tau, \kappa_\alpha)$ obtained in [27] for the case of a rate-limited noiseless channel by

1. setting $E_x\left(\frac{R}{\tau}, P_{SX}\right)$, $E_m(P_{SX}, \theta)$ and $E_m(P_{SX}, \theta) - \theta$ to ∞, which holds when the channel is noiseless.

2. maximizing over the set $\{(\omega, R, P_{SX}, \theta) \in \Omega \times \mathbb{R}^+ \times \mathcal{P}_{\mathcal{SX}} \times \Theta(P_{SX}) : \zeta_q(\kappa_\alpha, \omega) \leq R < \tau I(X;Y|S)\} \subseteq \mathcal{L}(\kappa_\alpha, \tau, P_{Y|X})$ in (3.30).

Then, note that the terms $E_2(\kappa_\alpha, \omega, R)$, $E_3(\kappa_\alpha, \omega, R, P_{SX}, \tau)$ and $E_4(\kappa_\alpha, \omega, R, P_{SX}, \theta, \tau)$ all equal ∞, and thus the inner bound in Theorem 3.9 reduces to that given in [27, Theorem 1].

Remark 3.10. Since the lower bound on $\kappa(\tau, \kappa_\alpha)$ in Theorem 3.9 is not necessarily concave, a tighter bound can be obtained using the technique of "time-sharing" similar to [27, Theorem 3]. We omit its description as it is cumbersome, although straightforward.

The terms $E_1(\cdot)$, $E_2(\cdot)$, $E_3(\cdot)$ and $E_4(\cdot)$ can be interpreted similarly to that done for Theorem 2.2. More specifically, $E_1(\cdot)$, which is identical to the lower bound on the error-exponent obtained in [27, Theorem 1] for the noiseless channel setting, corresponds to the event when there is no error in the encoding or decoding operation. On the other hand, $E_2(\cdot)$ correponds to the factor introduced due to a *binning* error event. Finally, $E_3(\cdot)$ and $E_4(\cdot)$ correspond to the factors introduced due to the erroneous decoding of the ordinary message and the special message in Borade et al.'s unequal error protection scheme (used for channel coding), respectively.

Specializing the lower bound in Theorem 3.9 to the case of TAI, we obtain the following.

Corollary 3.11. Let $P_{\bar{U}\bar{V}} = P_U P_V$. Then,

$$\kappa(\tau, \kappa_\alpha) \geq \max_{\substack{(\omega, R, P_{SX}, \theta) \in \\ \mathcal{L}^*(\kappa_\alpha, \tau)}} \min\left\{E_1^I(\kappa_\alpha, \omega), E_2^I(\kappa_\alpha, \omega, R, P_{SX}, \tau), E_3^I(\kappa_\alpha, \omega, R, P_{SX}, \theta, \tau)\right\},$$

where

$$\mathcal{L}^*(\kappa_\alpha, \tau) := \left\{ \begin{array}{l} (\omega, R, P_{SX}, \theta) \in \Omega \times \mathbb{R}^+ \times \mathcal{P}_{\mathcal{S}\mathcal{X}} \times \Theta(P_{SX}) : \zeta_q(\kappa_\alpha, \omega) \leq R < \\ \tau I(X;Y|S), \ \min\left(\tau E_m(P_{SX}, \theta), \tau E_x\left(\frac{R}{\tau}, P_{SX}\right)\right) \geq \kappa_\alpha \end{array} \right\},$$

$$E_1^I(\kappa_\alpha, \omega) := \min_{\hat{V}\hat{W}: \hat{U}\hat{V}\hat{W} \in \hat{\mathcal{L}}(\kappa_\alpha, \omega)} \left[I(\hat{V}; \hat{W}) + D(\hat{V}||V)\right],$$

$$E_2^I(\kappa_\alpha, \omega, R, P_{SX}, \tau) := \min_{\hat{V}: \hat{U}\hat{V}\hat{W} \in \hat{\mathcal{L}}(\kappa_\alpha, \omega)} D(\hat{V}||V) + \rho(\kappa_\alpha, \omega) + \tau E_x\left(\frac{R}{\tau}, P_{SX}\right),$$

$$E_3^I(\kappa_\alpha, \omega, P_{SX}, \theta, \tau) := \min_{\hat{V}: \hat{U}\hat{V}\hat{W} \in \hat{\mathcal{L}}(\kappa_\alpha, \omega)} D(\hat{V}||V) + \rho(\kappa_\alpha, \omega) + \tau\left(E_m(P_{SX}, \theta) - \theta\right),$$

and $\hat{\mathcal{L}}(\kappa_\alpha, \omega)$ is as defined in (3.29).

Proof. Note that $\mathcal{L}^*(\kappa_\alpha, \tau) \subseteq \mathcal{L}(\kappa_\alpha, \tau)$. Then, for any $\omega \in \mathcal{L}^*(\kappa_\alpha, \tau)$ and any $\tilde{U}\tilde{V}\tilde{W} \in \mathcal{T}_1(\kappa_\alpha, \omega)$,

$$\begin{aligned} D(\tilde{U}\tilde{V}\tilde{W}||\bar{U}\bar{V}\bar{W}) &= D(\tilde{U}\tilde{W}||\bar{U}\bar{W}) + D(\tilde{V}|\tilde{U}\tilde{W}||\bar{V}|\bar{U}\bar{W}) \\ &= D(\tilde{U}||\bar{U}) + D(\tilde{V}|\tilde{U}\tilde{W}||\bar{V}) \\ &\geq D(\tilde{U}||\bar{U}) + D(\tilde{V}|\tilde{W}||\bar{V}) \qquad (3.31) \\ &= D(\hat{U}||\bar{U}) + D(\hat{V}|\hat{W}||\bar{V}) \qquad (3.32) \\ &= D(\hat{U}||U) + I(\hat{V}; \hat{W}) + D(\hat{V}||V), \end{aligned}$$

where, in (3.31), we used the data processing (DPI) inequality for KL-divergence, and in (3.32), we used the fact that for $\tilde{U}\tilde{V}\tilde{W} \in \mathcal{T}_1(\kappa_\alpha, \omega)$, $P_{\tilde{U}\tilde{W}} = P_{\hat{U}\hat{W}}$ and $P_{\tilde{V}\tilde{W}} = P_{\hat{V}\hat{W}}$ for some $\hat{U}\hat{V}\hat{W} \in \hat{\mathcal{L}}(\kappa_\alpha, \omega)$. Minimizing over all $\hat{U}\hat{V}\hat{W} \in \hat{\mathcal{L}}(\kappa_\alpha, \omega)$ yields that

$$\begin{aligned} E_1(\kappa_\alpha, \omega) &= \min_{\tilde{U}\tilde{V}\tilde{W} \in \mathcal{T}_1(\kappa_\alpha, \omega)} D(\tilde{U}\tilde{V}\tilde{W}||\bar{U}\bar{V}\bar{W}) \\ &\geq \min_{\hat{U}\hat{V}\hat{W} \in \hat{\mathcal{L}}(\kappa_\alpha, \omega)} \left[I(\hat{V}; \hat{W}) + D(\hat{V}||V)\right] = E_1^I(\kappa_\alpha, \omega). \end{aligned}$$

Since $\zeta_q(\kappa_\alpha, \omega) \leq R$, we have that $E_2(\kappa_\alpha, \omega, R) = \infty$,

$$
\begin{aligned}
&E_3(\kappa_\alpha, \omega, R, P_{SX}, \tau) \\
&= \min_{\tilde{U}\tilde{V}\tilde{W} \in \mathcal{T}_2(\kappa_\alpha, \omega)} D(\tilde{U}\tilde{V}\tilde{W} || \bar{U}\bar{V}\bar{W}) + \rho(\kappa_\alpha, \omega) + \tau E_x\left(\frac{R}{\tau}, P_{SX}\right) \\
&\geq \min_{\hat{V}:\hat{U}\hat{V}\hat{W} \in \hat{\mathcal{L}}(\kappa_\alpha, \omega)} D(\hat{V}||V) + \rho(\kappa_\alpha, \omega) + \tau E_x\left(\frac{R}{\tau}, P_{SX}\right) \qquad (3.33) \\
&= E_2^I(\kappa_\alpha, \omega, R, P_{SX}, \tau),
\end{aligned}
$$

and

$$
\begin{aligned}
&E_4(\kappa_\alpha, \omega, R, P_{SX}, \theta, \tau) \\
&:= \min_{\hat{V}:\hat{U}\hat{V}\hat{W} \in \hat{\mathcal{L}}(\kappa_\alpha, \omega)} D(\hat{V}||\bar{V}) + \rho(\kappa_\alpha, \omega) + \tau\left(E_m(P_{SX}, \theta) - \theta\right) \\
&\geq \min_{\hat{V}:\hat{U}\hat{V}\hat{W} \in \hat{\mathcal{L}}(\kappa_\alpha, \omega)} D(\hat{V}||V) + \rho(\kappa_\alpha, \omega) + \tau\left(E_m(P_{SX}, \theta) - \theta\right) \qquad (3.34) \\
&= E_3^I(\kappa_\alpha, \omega, P_{SX}, \theta, \tau),
\end{aligned}
$$

where, to obtain (3.33) and (3.34), we used DPI for KL-divergence. This completes the proof. □

Corollary 3.12.

$$\lim_{\kappa_\alpha \to 0} \kappa_s^*(\tau, \kappa_\alpha) = \kappa_s(\tau),$$

where $\kappa_s(\tau)$ is the lower bound on the type II error-exponent for a fixed type I error probability constraint established in Theorem 2.2.

Proof. The result follows by noting that

$$
\begin{aligned}
\hat{\mathcal{L}}(0, \omega) &= \{UVW, P_{W|U} = \omega(P_U),\ V - U - W\}, \\
\zeta_q(0, \omega) &= I(U; W), \\
\rho(0, \omega) &= I(V; W),
\end{aligned}
$$

and the fact that $E_m(P_{SX},\theta), E_x\left(\frac{R}{\tau}, P_{SX}\right), E_b(\kappa_\alpha,\omega,R)$, and $E_m(P_{SX},\theta)-\theta$ for $\theta \in \Theta(P_{SX})$, are all greater than or equal to zero. □

The optimal T2EE for testing against independence in the Stein's regime (i.e. when $\kappa_\alpha \to 0$) can be recovered from Corollary 3.11 by taking the limit $\kappa_\alpha \to 0$.

Corollary 3.13. Let $P_{\bar{U}\bar{V}} = P_U P_V$. Then,

$$\lim_{\kappa_\alpha \to 0} \kappa(\tau,\kappa_\alpha) = \max_{\substack{W: W-U-V \\ I(U;W)\le \tau C}} I(V;W).$$

Proof. Note that

$$\hat{\mathcal{L}}(0,\omega) := \{UVW : P_{W|U} = \omega(P_U), V-U-W\},$$

and

$$\mathcal{L}^*(0,\tau) := \left\{ \begin{matrix} (\omega,R,P_{SX},\theta) \in \Omega \times \mathbb{R}^+ \times \mathcal{P}_{\mathcal{S}\mathcal{X}} \times \Theta(P_{SX}) : \\ I(U;W) \le R < \tau I(X;Y|S),\ P_{W|U} = \omega(P_U) \end{matrix} \right\} \tag{3.35}$$

Hence,

$$E_1^I(0,\omega) \ge \min_{\hat{U}\hat{V}\hat{W} \in \hat{\mathcal{L}}(0,\omega)} I(\hat{V};\hat{W}) = I(V;W), \tag{3.36}$$

for some $V-U-W$ such that $P_{W|U} = \omega(P_U)$. Also, we have

$$\rho(0,\omega) = I(V;W),$$

$$E_2^I(0,\omega,R,P_{SX},\tau) \ge \rho(0,\omega), \tag{3.37}$$

$$E_3^I(0,\omega,P_{SX},\theta,\tau) \ge \rho(0,\omega). \tag{3.38}$$

From (3.36)-(3.38), the result follows. □

3.5.2 JHTCC scheme

It is well known that joint source channel coding schemes outperforms separation based coding schemes in the context of reliable communication over a noisy channel [24–26]. Here, we obtain an inner bound on $\mathcal{R}$ using a generalization of the JHTCC scheme in Chapter 2.

For simplicity, we will assume that $k = n$, i.e., $\tau = 1$. Let Ω' denote the set of all continuous mappings from $\mathcal{P}_{\mathcal{U}} \times \mathcal{P}_{\mathcal{S}}$ to $\mathcal{P}_{\mathcal{W}'|\mathcal{US}}$, where $\mathcal{W}'$ is an arbitrary finite set. Let

$$\mathcal{L}_h(\kappa_\alpha) := \Big\{ \big(P_S, \omega'(\cdot, P_S), P_{X|USW'}, P_{X'|US}\big) \in \mathcal{P}_{\mathcal{S}} \times \Omega' \times \mathcal{P}_{\mathcal{X}|\mathcal{USW}'} \times \mathcal{P}_{\mathcal{X}|\mathcal{US}} : E'_b(\kappa_\alpha, \omega', P_S, P_{X|USW'}) > \kappa_\alpha \Big\},$$

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

$$:= \left\{ \begin{matrix} \hat{U}\hat{V}\hat{W}\hat{Y}S : \ D(\hat{U}\hat{V}\hat{W}\hat{Y}||UVW'Y|S) \le \kappa_\alpha, \ P_{SUVW'XY} := \\ P_S P_{UV} P_{W'|US} P_{X|USW'} P_{Y|X}, \ P_{W'|US} = P_{\hat{W}|\hat{U}S} = \omega'(P_{\hat{U}}, P_S) \end{matrix} \right\},$$

$$E'_b(\kappa_\alpha, \omega', P_S, P_{X|USW'}) := \rho'(\kappa_\alpha, \omega', P_S, P_{X|USW'}) - \zeta'_q(\kappa_\alpha, \omega', P_S),$$

$$\zeta'_q(\kappa_\alpha, \omega', P_S) := \max_{\substack{\hat{U}\hat{W}S: \ \exists \ \hat{V}\hat{Y}, \\ \hat{U}\hat{V}\hat{W}\hat{Y}S \in \\ \hat{\mathcal{L}}_h(\kappa_\alpha, \omega', P_S, P_{X|USW'})}} I(\hat{U}; \hat{W}|S),$$

$$\rho'(\kappa_\alpha, \omega', P_S, P_{X|USW'}) := \min_{\substack{\hat{V}\hat{W}\hat{Y}S: \ \exists \ \hat{U}, \\ \hat{U}\hat{V}\hat{W}\hat{Y}S \in \\ \hat{\mathcal{L}}_h(\kappa_\alpha, \omega', P_S, P_{X|USW'})}} I(\hat{Y}, \hat{V}; \hat{W}|S),$$

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

$$E'_2(\kappa_\alpha, \omega', P_S, P_{X|USW'})$$

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

$$E'_3(\kappa_\alpha, \omega', P_S, P_{X|USW'}, P_{X'|US})$$

$$:= \min_{\substack{\hat{V}\hat{Y}S: \hat{U}\hat{V}\hat{W}\hat{Y}S \ \in \\ \hat{\mathcal{L}}_h(\kappa_\alpha, \omega', P_S, P_{X|USW'})}} D(\hat{V}\hat{Y}||\bar{V}\check{Y}|S) + E'_b(\kappa_\alpha, \omega', P_S, P_{X|USW'}),$$

$$P_{S\bar{U}\bar{V}\bar{W}'\bar{X}\bar{Y}} := P_S P_{\bar{U}\bar{V}} P_{\bar{W}'|\bar{U}S} P_{\bar{X}|\bar{U}S\bar{W}'} P_{\bar{Y}|\bar{X}},$$

$P_{\bar{W}'|\bar{U}S} := P_{\tilde{W}|\tilde{U}S},\ P_{\bar{X}|\bar{U}S\bar{W}'} := P_{X|USW'},\ P_{\bar{Y}|\bar{X}} := P_{Y|X},$

$P_{S\bar{U}\bar{V}X'\check{Y}} := P_S P_{\bar{U}\bar{V}} P_{X'|US} P_{\check{Y}|X'},\ P_{\check{Y}|X'} := P_{Y|X},$

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

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

Then, we have the following result.

Theorem 3.14.

$\kappa(1, \kappa_\alpha) \geq \kappa_h^*(\kappa_\alpha),$ *where*

$$\kappa_h^*(\kappa_\alpha) := \max_{\substack{(P_S, \omega', P_{X|USW'}, P_{X'|US}) \\ \in\ \mathcal{L}_h(\kappa_\alpha)}} \min \Big\{ E_1'(\kappa_\alpha, \omega'),\ E_2'(\kappa_\alpha, \omega', P_S, P_{X|USW'}),$$
$$E_3'(\kappa_\alpha, \omega', P_S, P_{X|USW'}, P_{X'|US}) \Big\}.$$

The proof of Theorem 3.14 is given in Appendix B.3. It is easy to see that Theorem 3.14 recovers the lower bound on the type II error-exponent in the Stein's regime established in Theorem 2.6 as we show next.

Corollary 3.15.

$$\lim_{\kappa_\alpha \to 0} \kappa_h^*(\kappa_\alpha) = \kappa_h,$$

where κ_h is as given in Theorem 2.6.

Proof. The result follows by noting that

$\hat{\mathcal{L}}_h(0, \omega', P_S, P_{X|USW'})$

$$:= \left\{ \begin{array}{c} UVW'YS : P_{SUVW'XY} := P_S P_{UV} P_{W'|US} P_{X|USW'} P_{Y|X}, \\ P_{W'|US} = \omega'(P_U, P_S) \end{array} \right\}$$

$\zeta'_q(0, \omega', P_S) := I(U; W'|S)$,

$\rho(0, \omega', P_S, P_{X|USW'}) := I(Y, V; W'|S)$,

$\mathcal{L}_h(0) := \{\big(P_S, \omega'(\cdot, P_S), P_{X|USW'}, P_{X'|US}\big) \in \mathcal{P}_{\mathcal{S}} \times \Omega' \times \mathcal{P}_{\mathcal{X}|\mathcal{US}\mathcal{W}'} \times \mathcal{P}_{\mathcal{X}|\mathcal{US}} :$

$$I(U; W'|S) < I(Y, V; W'|S)\},$$

and $E'_b(0, \omega', P_S, P_{X|USW'}) := I(Y, V; W'|S) - I(U; W'|S)$.

□

3.6 Conclusions

In this chapter, we studied the trade-off between the exponents of the type I and type II error probabilities for distributed HT over a noisy channel with side-information at the detector. In the non-distributed setting, we obtained a single-letter characterization of the optimal trade-off between the error-exponents. The direct part of the proof shows that the optimal trade-off is achieved by a scheme, in which the observer performs an appropriate NP test locally and communicates the decision of the test to the detector using a suitable channel code, while the detector performs an appropriate NP test on the channel output. This implies that "separation" holds, in the sense that, there is no loss in optimality incurred by separating the tasks of HT and channel coding. For the distributed setting, we obtained inner bounds on the error-exponents trade-off using the SHTCC and JHTCC schemes. The latter bound is at least as good as the former when the type I error-exponent is zero. Exploring whether joint schemes offer strict advantage over separation based schemes is something worth investigating in the future. It would also be interesting to explore the trade-off between the error-exponents in a related setting, where the side-information also needs to be communicated to the detector over a noisy communication channel.

Chapter 4

Privacy-aware Distributed HT

4.1 Overview

In this chapter, we consider the distributed HT problem studied in Chapter 2, but with an additional privacy constraint. We focus on the case of a rate-limited noiseless communication channel between the observer and the detector. Thus, while the goal of the observer is to maximize the type II error-exponent of the test for a given type I error probability constraint, it also wants to keep a private part of its observations as oblivious to the detector as possible. Considering both equivocation and average distortion as possible measures of privacy, the trade-off between the communication rate from the observer to the detector, the type II error exponent, and privacy is studied. For the general HT problem, we establish single-letter inner bounds on both the rate-error exponent-equivocation and rate-error exponent-distortion trade-offs. Subsequently, single-letter characterizations for both trade-offs are obtained (i) for TACI; and (ii) when the communication rate constraint over the channel is zero. Finally, we show by providing a counterexample that, the strong converse which holds for distributed HT without a privacy constraint, does not hold when a privacy constraint is imposed. This implies that, in general, the rate-error exponent-equivocation and rate-error exponent-distortion trade-offs are not independent of the type I error probability constraint.

4.2 Introduction

Data inference and privacy are often contradicting objectives. In many multi-agent system, each agent/user reveals information about its data to a remote service, application or authority, which in turn, provides certain utility to the users based on their data. Many emerging networked systems can be thought of in this context, from social networks to smart grids and communication networks. While obtaining the promised utility is the main goal of the users, privacy of data that is shared is becoming increasingly important. Thus, it is critical that users reveal only the information relevant for obtaining the desired utility, while maximum possible privacy is retained for their sensitive information.

In distributed learning applications, typically the goal is to learn the joint probability distribution of data available at different locations. In such a scenario, the nodes communicate their observations to the detector, which then applies HT on the underlying joint distribution of the data based on its own observations and those received from other nodes. However, with the efficient data mining and machine learning algorithms available today, the detector can illegitimately infer some unintended private information from the data provided to it exclusively for HT purposes. Such threats are becoming increasingly imminent as large amounts of seemingly irrelevant yet sensitive data are collected from users, such as in medical research [34], social networks [35], online shopping [36] and smart grids [37]. Therefore, there is an inherent trade-off between the utility acquired by sharing data and the associated privacy leakage.

In this chapter, we consider the distributed HT problem studied in Chapter 2 for the case of a rate-limited noiseless communication channel, but with an additional privacy constraint. The detector performs a binary HT as given in (2.1). The performance of the HT is measured by the type II error exponent (or error-exponent henceforth) in the Stein's regime. While the goal is to maximize the performance of the HT, the observer also wants to maintain a certain level of privacy against the detector for some latent private data that is correlated with its observations. We are interested in characterizing the trade-off between the communication rate, error-exponent and the amount of information leakage of private data.

4.3 Previous Work and Our Contributions

Data privacy has been a hot topic of research in the past decade, spanning across multiple disciplines in computer and computational sciences. Several practical schemes have been proposed that deal with the protection or violation of data privacy in different contexts, e.g., see [38–43]. More relevant to the current setting, HT under mutual information and maximal leakage privacy constraints have been studied in [44] and [45], respectively, where the encoder uses a *memoryless privacy mechanism* to convey a noisy version of its observed data to the detector. The detector performs HT on the probability distribution of the observer's data, and the optimal privacy mechanism that maximizes the error-exponent while satisfying the privacy constraint is analyzed. Recently, a distributed version of this problem has been studied in [46], where the encoder applies a privacy mechanism to its observed data prior to further coding for compression, and the goal at the detector is to perform a HT on the joint distribution of its own observations with those of the observer. In contrast with [44], [45] and [46], we study *distributed HT with a privacy constraint*, but without considering a separate privacy mechanism at the encoder. In Section 4.4, we will further discuss the differences between the system model considered here and that of [46].

It is important to note here that the data privacy problem is fundamentally different from that of data security against an eavesdropper or an adversary. In data security, sensitive data is to be protected against an external malicious agent distinct from the legitimate parties in the system. The techniques for guaranteeing data security usually involve either cryptographic methods in which the legitimate parties are assumed to have additional resources unavailable to the adversary (e.g., a shared private key) or the availability of better communication channel conditions (e.g., using wiretap codes). However, in data privacy problems, the sensitive data is to be protected from the same legitimate party that receives the messages and provides the utility; and hence, the above mentioned techniques for guaranteeing data security are not applicable. Another model frequently used in the context of information-theoretic security assumes the availability of different side-information at the legitimate receiver and the eavesdropper [47, 48]. A distributed HT problem with security constraints formulated along these

lines is studied in [49], where the authors propose an inner bound on the rate-error exponent-equivocation trade-off. While our model is closely related to that in [49] when the side-information at the detector and eavesdropper coincide, there are some important differences which will be highlighted in Section 4.4.3.

Many different privacy measures have been considered in the literature to quantify the amount of private information leakage, such as k-anonymity [50], differential privacy [51], mutual information leakage [52–54], maximal leakage [55], and total variation distance [56] to count a few; see [7] for a detailed survey. Among these, mutual information between the private and revealed information (or, equivalently, the *equivocation* of private information given the revealed information) is perhaps the most commonly used measure in the information theoretic studies of privacy. It is well known that a necessary and sufficient condition to guarantee statistical independence between two random variables is to have zero mutual information between them. Furthermore, the average information leakage measured using an arbitrary privacy measure is upper bounded by a constant multiplicative factor of that measured by mutual information [53]. It is also shown in [52] that a differentially private scheme is not necessarily private when the information leakage is measured by mutual information. This is done by constructing an example that is differentially private, yet the mutual information leakage is arbitrarily high. Mutual information based measures have also been used in cryptographic security studies. For example, the notion of semantic security defined in [57] is shown to be equivalent to a measure based on mutual information in [58]. A rate-distortion approach to privacy is first explored by Yamamoto in [59] for a rate-constrained noiseless channel, where, in addition to a distortion constraint for legitimate data, a minimum distortion requirement is enforced for the private part. Recently, there have been several works that have used distortion as a security or privacy metric in several different contexts, such as side-information privacy in discriminatory lossy source coding [60] and rate distortion theory of secrecy systems [61], [62]. Distortion based measures have also been considered in steganography, for instance, in the context of watermarking systems in the presence of an attacker [63], [64]. In such systems, the goal of the encoder is to embed the watermark within the host data (covertext) such that the distortion between the covertext and the watermarked version (stegotext) is below a

certain threshold, while the aim of the attacker is to corrupt the stegotext to a permissible level of additional distortion such that correct decoding of the watermark is inhibited. That is, low distortion on the stegotext is treated as "uncorrupted" data, while high distortion is considered to be corrupted data.

In this chapter, we will consider both equivocation and average distortion as measures of privacy. In [65], error-exponent of a HT adversary is considered as a privacy measure. This can be considered as the opposite setting to ours, in the sense that, while the goal here is to increase the error-exponent under a privacy leakage constraint, the goal in [65] is to reduce the error-exponent under a constraint on possible transformations that can be applied on the data.

The amount of private information leakage that can be tolerated depends on the specific application at hand. While it may be possible to tolerate a moderate amount of information leakage in applications like online shopping or social networks, it may no longer be the case in matters related to information sharing among government agencies or corporations. While it is obvious that maximum privacy can be attained by revealing no information, this typically comes at the cost of zero utility. On the other hand, maximum utility can be achieved by revealing all the information, but at the cost of minimum privacy. Characterizing the optimal trade-off between the utility and the minimum privacy leakage between these two extremes is a fundamental and challenging research problem.

Main Contributions

The main contributions in this chapter are as follows.

(i) In Section 4.5, Theorem 4.4 (resp. Theorem 4.5), we establish a single-letter inner bound on the rate-error exponent-equivocation (resp. rate-error exponent-distortion) trade-off for distributed HT with a privacy constraint. The distortion and equivocation privacy constraints we consider, that is given in (4.5) and (4.6), respectively, are slightly stronger than what is usually considered in the literature (stated in (4.7) and (4.8), respectively).

(ii) Exact characterizations are obtained for some important special cases in Section 4.6. More specifically, a single-letter characterization of the optimal rate-error exponent-equivocation (resp. rate-error exponent-distortion) trade-off is established for:

   (a) TACI with a privacy constraint (for vanishing type I error probability constraint) in Section 4.6.1, Proposition 4.6 (resp. Proposition 4.7),

   (b) distributed HT with a privacy constraint for zero-rate compression ($R = 0$) in Section 4.6.2, Proposition 4.11 (resp. Proposition 4.10).

   Since the optimal trade-offs in Propositions 4.10 and 4.11 are independent of the constraint on the type I error probability, they are strong converse results in the context of HT.

(iii) Finally, in Section 4.7, we provide a counterexample showing that for positive rate $R > 0$, the strong converse result does not hold in general for TAI with a privacy constraint.

The organization of this chapter is as follows. Basic notations are introduced in Section 4.4.1. The problem formulation and associated definitions are given in Section 4.4.2. Main results are presented in Sections 4.5 to 4.7. The proofs of the results are presented either in Appendix C or immediately after the statement of the result. Finally, Section 4.8 concludes the chapter with some interesting avenues for future research.

4.4 Preliminaries

4.4.1 Notations

$X \perp Y$ denotes statistical independence of r.v.'s X and Y. $\xrightarrow{(n)}$ denotes asymptotic limit with respect to n, e.g., $a_n \xrightarrow{(n)} 0$ means that the sequence a_n tends to zero asymptotically with n. Similar notations apply for asymptotic inequalities, e.g. $a_n \overset{(n)}{\geq} b_n$, means that $a_n \geq b_n$ for sufficiently large n. $\mathbb{P}(\mathcal{E})$ denotes the probability of event $\mathcal{E}$.

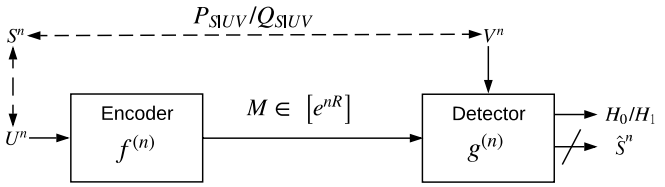


FIGURE 4.1: Distributed HT with a privacy constraint.

For positive real m, we define $[m] := \{1, \ldots, \lceil m \rceil\}$. For an arbitrary set $\mathcal{A}$, we denote its complement by $\mathcal{A}^c$, and for $\mathcal{A} \subseteq \mathbb{R}^n$, we denote its interior and closure by $int(\mathcal{A})$ and $cl(\mathcal{A})$ (with respect to the Euclidean metric), respectively. Whenever the range of the summation is not specified, this will mean summation over the entire support, e.g., $\sum_u$ denotes $\sum_{u \in \mathcal{U}}$, unless specified otherwise. Throughout this chapter, the base of the logarithms is taken to be e. For $a \in \mathbb{R}$, a^+ denotes $\max\{0, a\}$. For two probability distributions P and Q defined on a common support $\mathcal{X}$, $P << Q$ will mean that P is absolutely continuous with respect to Q. Finally, $O(\cdot)$, $o(\cdot)$ and $\Omega(\cdot)$ denotes the standard asymptotic notation of Big-O, Little-O and Big-Ω, respectively.

4.4.2 Problem formulation

Consider the HT setup illustrated in Fig. 4.1, where (U^n, V^n, S^n) denote n independent and identically distributed (i.i.d.) copies of triplet of r.v.'s (U, V, S). The encoder (observer) observes U^n and sends the message index $M := f^{(n)}(U^n)$, $M \in \mathcal{M}$, to the detector over an error-free channel using some encoding function (possibly stochastic) $f^{(n)} : \mathcal{U}^n \mapsto \mathcal{M}$. Given its own observation V^n, the detector performs the HT given in (2.1). Let H and $\hat{H}$ denote the r.v.'s corresponding to the true hypothesis and the output of the HT, respectively, with support $\mathcal{H} = \hat{\mathcal{H}} = \{0, 1\}$, where 0 denotes the null hypothesis and 1 the alternate hypothesis. Let $g^{(n)} : \mathcal{M} \times \mathcal{V}^n \mapsto \hat{\mathcal{H}} := \{0, 1\}$ denote the decision rule (possibly stochastic) at the detector with output $\hat{H}$.

The type I and type 2 error probability for an $(f^{(n)}, g^{(n)})$ pair are then given by

$$\bar{\alpha}\left(f^{(n)}, g^{(n)}\right) := \mathbb{P}(\hat{H} = 1 | H = 0) = P_{\hat{H}}(1),$$

and

$$\bar{\beta}\left(f^{(n)}, g^{(n)}\right) := \mathbb{P}(\hat{H} = 0 | H = 1) = Q_{\hat{H}}(0),$$

respectively, where

$$P_{\hat{H}}(1) = \sum_{u^n, m, v^n} \left[\prod_{i=1}^{n} P_{UV}(u_i, v_i)\right] \; P_{M|U^n}(m|u^n) \; P_{\hat{H}|MV^n}(1|m, v^n),$$

$$\text{and } Q_{\hat{H}}(0) = \sum_{u^n, m, v^n} \left[\prod_{i=1}^{n} Q_{UV}(u_i, v_i)\right] \; P_{M|U^n}(m|u^n) \; P_{\hat{H}|MV^n}(0|m, v^n).$$

The performance of HT is measured by the error-exponent achieved by the test in the Stein's regime, i.e., $\liminf_{n\to\infty} -\frac{1}{n}\log\left(\beta(f^{(n)}, \epsilon)\right)$, $\epsilon \in (0, 1)$, where

$$\begin{aligned} \beta\left(f^{(n)}, \epsilon\right) := \inf_{g^{(n)}} \bar{\beta}\left(f^{(n)},\ g^{(n)}\right), \\ \text{such that } \bar{\alpha}\left(f^{(n)},\ g^{(n)}\right) \leq \epsilon. \end{aligned} \tag{4.1}$$

Although the goal of the detector is to maximize the error-exponent achieved for the HT, it is also curious about the latent r.v. S^n that is correlated with the source U^n. S^n is referred to as the *private* part of U^n, which is distributed i.i.d. according to the joint distribution P_{SUV} and Q_{SUV} under the null and alternate hypothesis, respectively. It is desired to keep the private part as concealed as possible from the detector. We consider two measures of privacy for S^n at the detector. The first is the *equivocation* defined as $H(S^n|M, V^n)$. The second one is the *average distortion* between S^n and its reconstruction $\hat{S}^n$ at the detector, measured according to an arbitrary bounded additive distortion metric $d : \mathcal{S} \times \hat{\mathcal{S}} \mapsto [0, D_m]$ with multi-letter distortion defined as

$$d(s^n, \hat{s}^n) := \sum_{i=1}^{n} d(s_i, \hat{s}_i). \tag{4.2}$$

The goal is to ensure that the error-exponent for HT is maximized, while satisfying the constraints on the type I error probability ϵ and the privacy of S^n. In the sequel, we study the trade-off between the rate, error-exponent (henceforth also referred to simply

as the error exponent) and privacy achieved in the above setting. Before delving into that, a few definitions are in order.

Definition 4.1. For a given type I error probability constraint ϵ, a rate-error exponent-distortion tuple $(R, \kappa, \Delta_0, \Delta_1)$ is *achievable*, if there exists a sequence of encoding and decoding functions $f^{(n)} : \mathcal{U}^n \mapsto \mathcal{M}$, and $g^{(n)} : \mathcal{M} \times \mathcal{V}^n \mapsto \hat{\mathcal{H}}$ such that

$$\limsup_{n\to\infty} \frac{\log(|\mathcal{M}|)}{n} \leq R, \tag{4.3}$$

$$\liminf_{n\to\infty} \frac{-\log\left(\beta(f^{(n)}, \epsilon)\right)}{n} \geq \kappa, \tag{4.4}$$

and for any $\gamma > 0$, there exists an $n_0 \in \mathbb{Z}^+$ such that

$$\inf_{g_r^{(n)}} \mathbb{E}\left[d\left(S^n, \hat{S}^n\right)|H = i\right] \geq n\Delta_i - \gamma, \ \forall\ n \geq n_0, \ i = 0, 1, \tag{4.5}$$

where $\hat{S}^n = g_r^{(n)}(M, V^n)$, and $g_r^{(n)} : [e^{nR}] \times \mathcal{V}^n \mapsto \hat{\mathcal{S}}^n$ denotes an arbitrary reconstruction map (possibly stochastic) at the detector. The rate-error exponent-distortion region $\mathcal{R}_d(\epsilon)$ is the closure of the set of all such achievable $(R, \kappa, \Delta_0, \Delta_1)$ tuples for a given ϵ.

Definition 4.2. For a given type I error probability constraint ϵ, a rate-error exponent-equivocation $(R, \kappa, \Lambda_0, \Lambda_1)$ tuple is *achievable*, if there exists a sequence of encoding and decoding functions $f^{(n)} : \mathcal{U}^n \mapsto \mathcal{M}$ and $g^{(n)} : [e^{nR}] \times \mathcal{V}^n \mapsto \hat{\mathcal{H}}$ such that (4.3) and (4.4) are satisfied, and for any $\gamma > 0$, there exists a $n_0 \in \mathbb{Z}^+$ such that

$$H(S^n|M, V^n, H = i) \geq n\Lambda_i - \gamma, \ \forall\ n \geq n_0, \ i \in \{0, 1\}. \tag{4.6}$$

The rate-error exponent-equivocation region $\mathcal{R}_e(\epsilon)$ is the closure of the set of all such achievable $(R, \kappa, \Lambda_0, \Lambda_1)$ tuples for a given ϵ.

Note that the privacy measures considered in (4.5) and (4.6) are stronger than

$$\liminf_{n\to\infty} \inf_{g_r^{(n)}} \mathbb{E}\left[\frac{1}{n} d\left(S^n, \hat{S}^n\right)|H = i\right] \geq \Delta_i, \ i = 0, 1, \tag{4.7}$$

$$\text{and } \liminf_{n\to\infty} \frac{1}{n} H(S^n|M, V^n, H = i) \geq \Lambda_i, \ i = 0, 1, \tag{4.8}$$

respectively. To see this for the equivocation privacy measure, note that if $H(S^n|M, V^n, H = i) = n\Lambda_i^* - n^a$, $i = 0, 1$, for some $a \in (0, 1)$, then an equivocation pair $(\Lambda_0^*, \Lambda_1^*)$ is achievable under the constraint given in (4.8), while it is not achievable under the constraint given in (4.6).

4.4.3 Relation to Previous Work

Before stating our results, we briefly highlight the differences between our system model and the ones studied in [46] and [49]. In [46], the observer applies a privacy mechanism to the data before releasing it to the transmitter, which performs further encoding prior to transmission to the detector. More specifically, the observer checks if $U^n \in T^n_{[P_U]_\delta}$ and if successful, sends the output of a memoryless privacy mechanism applied to U^n, to the transmitter. Otherwise, it outputs a n-length zero-sequence. The privacy mechanism plays the role of randomizing the data (or adding noise) in order to achieve the desired privacy. This model is similar in spirit to the earlier works in [44] and [45], and the two stage encoding essentially results in a *separation between the problem of coding for privacy and coding for compression*. Similar privacy mechanisms that randomizes the data has also been used in other works that study utility-privacy trade-off like [66]. In our model, the tasks of coding for privacy and compression are done jointly (without a separate privacy mechanism) by utilizing all the available data samples U^n. Also, while we consider the equivocation (and average distortion) between the revealed information and the private part as the privacy measure, in [46], the mutual information between the observer's observations and the output of the memoryless mechanism is the privacy measure. Thus, for testing against independence in their model, a perfect privacy condition $\Lambda_0 = 0$ would imply that the error-exponent is also zero, since the output of the memoryless mechanism has to be independent of the observer's observations (under both hypotheses). However, as we show in Example 4.12 later, a positive error-exponent is achievable while guaranteeing perfect privacy in our model.

On the other hand, the difference between our model with equivocation as the privacy measure, and the security problem studied in [49] arises from the difference in

the privacy constraint imposed. More specifically, while in [49], the goal is to keep U^n private from an illegitimate eavesdropper, the objective here is to keep a r.v. S^n that is correlated with U^n private from the detector. Moreover, we consider the stronger privacy constraint given in (4.6) as opposed to (4.8) which is considered in [49]. To satisfy this stronger privacy constraint on S^n, we require that the aposteriori probability distribution of S^n given the observations (M, V^n) at the detector is close in some sense to a desired "target" memoryless distribution. To achieve this, we use a novel stochastic encoding scheme to induce the necessary randomness for S^n at the detector. Another difference is that the marginal distributions of U^n and the side-information at the eavesdropper are assumed to be the same under the null and alternate hypotheses in [49], which is not the case here.

Next, we state some supporting results that will be useful later for proving the main results.

4.4.4 Supporting Results

Let

$$g^{(n)}(m, v^n) = \mathbb{1}\left((m, v^n) \in \mathcal{A}_n^c\right) \tag{4.9}$$

denote the deterministic detector with acceptance region $\mathcal{A}_n \subseteq [e^{nR}] \times \mathcal{V}^n$ for H_0 and $\mathcal{A}_n^c$ for H_1. Then, the type I and type II error probabilities are given by

$$\bar{\alpha}\left(f^{(n)}, g^{(n)}\right) := P_{MV^n}(\mathcal{A}_n^c) = \mathbb{E}\left(\mathbb{1}(M, V^n) \in \mathcal{A}_n^c | H = 0\right), \tag{4.10}$$

$$\bar{\beta}\left(f^{(n)}, g^{(n)}\right) := Q_{MV^n}(\mathcal{A}_n) = \mathbb{E}\left(\mathbb{1}(M, V^n) \in \mathcal{A}_n | H = 1\right). \tag{4.11}$$

Lemma 4.1. *Any error-exponent that is achievable is also achievable by a deterministic detector of the form given in* (4.9) *for some* $\mathcal{A}_n \subseteq [e^{nR}] \times \mathcal{V}^n$*, where* $\mathcal{A}_n$ *and* $\mathcal{A}_n^c$ *denote the acceptance regions for* H_0 *and* H_1*, respectively.*

The proof of Lemma 4.1 is given in Appendix C.1 for completeness. Due to Lemma 4.1, henceforth we restrict our attention to deterministic $g^{(n)}$. The next result shows

that without loss of generality (w.l.o.g), it is also sufficient to consider $g_r^{(n)}$ (in Definition 4.2) to be a deterministic function of the form

$$g_r^{(n)} = \{\bar{\phi}_i(m, v^n)\}_{i=1}^n, \tag{4.12}$$

for the minimization in (4.5), where $\bar{\phi}_i : \mathcal{M} \times \mathcal{V}^n \mapsto \hat{\mathcal{S}}$, $i \in [n]$, denotes an arbitrary deterministic function.

Lemma 4.2. *The infimum in* (4.5) *is achieved by a deterministic function* $g_r^{(n)}$ *as given in* (4.12)*, and hence it is sufficient to restrict our attention to such deterministic* $g_r^{(n)}$*.*

The proof of Lemma 4.2 is given in Appendix C.2. Next, we state some lemmas that will be handy for upper bounding the amount of privacy leakage in the proofs of the main results stated below. The following one is a well known result proved in [20] that upper bounds the difference in entropy of two r.v.'s (with a common support) in terms of the total variation distance between their probability distributions.

Definition 4.3. The total variation between probability distributions P_X and Q_X defined on the same support $\mathcal{X}$ is defined as

$$||P_X - Q_X|| = \frac{1}{2} \sum_{x \in \mathcal{X}} |P_X(x) - Q_X(x)|.$$

Lemma 4.3. *[20, Lemma 2.7] Let* P_X *and* Q_X *be distributions defined on a common support* $\mathcal{X}$ *and let* $\rho := ||P_X - Q_X||$*. Then,*

$$|H_{P_X} - H_{Q_X}| \leq -2\rho \log\left(\frac{2\rho}{|\mathcal{X}|}\right).$$

Next, we state some properties of total variation distance that will be used later.

Property 4.4.1. [61] Let P, Q and Φ be probability distributions defined on the same alphabet and sigma algebra $(\mathcal{X}, \Sigma)$.

(a) Let $\delta > 0$ and let $f(x)$ be a function with bounded range of width $b > 0$. Then $\|P - Q\| < \delta \Rightarrow |\mathbb{E}_P f(X) - \mathbb{E}_Q f(X)| < \delta b$.

(b) $\|P - Q\| \leq \|P - \Phi\| + \|\Phi - Q\|$.

(c) Let $P_X P_{Y|X}$ and $Q_X P_{Y|X}$ be two joint distributions induced by input distributions P_X and Q_X by passing through a common channel $P_{Y|X}$. Then $\|P_X P_{Y|X} - Q_X P_{Y|X}\| = \|P_X - Q_X\|$.

(d) Let P_X and Q_X be marginal distributions of P_{XY} and Q_{XY}. Then,

$\|P_X - Q_X\| \leq \|P_{XY} - Q_{XY}\|$.

The next lemma will be handy in proving Theorem 4.4, Theorem 4.5, Proposition 4.10 and the counter-example for strong converse presented in Section 4.7.

Lemma 4.4. *Let (X^n, Y^n) denote n i.i.d. copies of r.v.'s (X, Y), and $P_{X^nY^n} = \prod_{i=1}^n P_{XY}$ and $Q_{X^nY^n} = \prod_{i=1}^n Q_{XY}$ denote two joint probability distributions on (X^n, Y^n). For $\delta > 0$, define*

$$I_X(x^n, \delta) := \mathbb{1}\left(x^n \notin T^n_{[P_X]_\delta}\right). \tag{4.13}$$

If $P_X \neq Q_X$, then for $\delta > 0$ sufficiently small, there exists $\bar{\delta} > 0$ and $n_0(\delta, |\mathcal{X}|, |\mathcal{Y}|) \in \mathbb{Z}^+$ such that for all $n \geq n_0(\delta, |\mathcal{X}|, |\mathcal{Y}|)$,

$$\|Q_{Y^n} - Q_{Y^n|I_X(X^n,\delta)=1}\| \leq e^{-n\bar{\delta}}. \tag{4.14}$$

If $P_X = Q_X$, then for any $\delta > 0$, there exists $\bar{\delta} > 0$ and $n_0(\delta, |\mathcal{X}|, |\mathcal{Y}|) \in \mathbb{Z}^+$ such that for all $n \geq n_0(\delta, |\mathcal{X}|, |\mathcal{Y}|)$,

$$\|Q_{Y^n} - Q_{Y^n|I_X(X^n,\delta)=0}\| \leq e^{-n\bar{\delta}}, \tag{4.15}$$

Also, for any $\delta > 0$, there exists $\bar{\delta} > 0$ and $n_0(\delta, |\mathcal{X}|, |\mathcal{Y}|) \in \mathbb{Z}^+$ such that for all $n \geq n_0(\delta, |\mathcal{X}|, |\mathcal{Y}|)$,

$$\|P_{Y^n} - P_{Y^n|I_X(X^n,\delta)=0}\| \leq e^{-n\bar{\delta}}. \tag{4.16}$$

Proof. The proof is presented in Appendix C.3. □

In the next section, we establish an inner bound on $\mathcal{R}_e(\epsilon)$ and $\mathcal{R}_d(\epsilon)$ defined in Definitions 4.2 and 4.1, respectively.

4.5 Main Results

The following two theorems are the main results of this chapter providing inner bounds for $\mathcal{R}_e(\epsilon)$ and $\mathcal{R}_d(\epsilon)$, respectively.

Theorem 4.4. *For $\epsilon \in (0,1)$, $(R, \kappa, \Lambda_0, \Lambda_1) \in \mathcal{R}_e(\epsilon)$ if there exists an auxiliary r.v. W, such that $(V,S) - U - W$, and*

$$R \geq I_P(W;U|V), \tag{4.17}$$

$$\kappa \leq \kappa^*(P_{W|U}, R), \tag{4.18}$$

$$\Lambda_0 \leq H_P(S|W,V), \tag{4.19}$$

$$\Lambda_1 \leq \mathbb{1}\left(P_U = Q_U\right)\ H_Q(S|W,V) + \mathbb{1}\left(P_U \neq Q_U\right)\ H_Q(S|V), \tag{4.20}$$

where

$$\begin{aligned} \kappa^*(P_{W|U}, R) &:= \min\left\{E_1(P_{W|U}),\ E_2(R, P_{W|U})\right\}, \\ E_1(P_{W|U}) &:= \min_{P_{\tilde{U}\tilde{V}\tilde{W}} \in \mathcal{T}_1(P_{UW}, P_{VW})} D(P_{\tilde{U}\tilde{V}\tilde{W}} || Q_{UV} P_{W|U}), \end{aligned} \tag{4.21}$$

$$E_2(R, P_{W|U}) \tag{4.22}$$

$$:= \begin{cases} \min\limits_{P_{\tilde{U}\tilde{V}\tilde{W}} \in \mathcal{T}_2(P_{UW}, P_V)} D(P_{\tilde{U}\tilde{V}\tilde{W}} || Q_{UV} P_{W|U}) & \\ \qquad\qquad + (R - I_P(U;W|V)), & \text{if } I_P(U;W) > R, \\ \infty, & \text{otherwise,} \end{cases} \tag{4.23}$$

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

$$\begin{aligned} \mathcal{T}_2(P_{UW}, P_V) := \{&P_{\tilde{U}\tilde{V}\tilde{W}} \in \mathcal{T}(\mathcal{U} \times \mathcal{V} \times \mathcal{W}) : P_{\tilde{U}\tilde{W}} = P_{UW},\ P_{\tilde{V}} = P_V, \\ &H_P(W|V) \leq H(\tilde{W}|\tilde{V})\}, \end{aligned}$$

$$P_{SUVW} := P_{SUV} P_{W|U} \quad \text{and } Q_{SUVW} := Q_{SUV} P_{W|U}.$$

The proof of Theorem 4.4 is given in Apppendix C.4 along with the proof of Theorem 4.5 below. Notice that the rate-error exponent trade-off derived in [11] is recovered when the privacy constraint in Theorem 4.4 is relaxed. To provide some intuition to the factors $E_1(\cdot)$ and $E_2(\cdot)$ appearing in the characterization of the error-exponent above, $E_1(\cdot)$ corresponds to the event when there is no error at the encoder or decoder, while $E_2(\cdot)$ corresponds to the event when there is a *binning* error at the decoder.

We next state an inner bound for $\mathcal{R}_d(\epsilon)$.

Theorem 4.5. *For a given bounded additive distortion measure $d(\cdot,\cdot)$ and $\epsilon \in (0,1)$, $(R,\kappa,\Delta_0,\Delta_1) \in \mathcal{R}_d(\epsilon)$ if there exist an auxiliary r.v. W and deterministic functions $\phi : \mathcal{W}\times\mathcal{V} \mapsto \hat{S}$ and $\phi' : \mathcal{V} \mapsto \hat{S}$, such that $(V,S) - U - W$ and (4.17), (4.18),*

$$\Delta_0 \leq \min_{\phi(\cdot,\cdot)} \mathbb{E}_P \left[d\left(S, \phi(W,V)\right)\right], \tag{4.24}$$

and $\Delta_1 \leq \mathbb{1}\left(P_U = Q_U\right) \min_{\phi(\cdot,\cdot)} \mathbb{E}_Q \left[d\left(S, \phi(W,V)\right)\right]$

$$+ \mathbb{1}\left(P_U \neq Q_U\right) \min_{\phi'(\cdot)} \mathbb{E}_Q \left[d\left(S, \phi'(V)\right)\right], \tag{4.25}$$

are satisfied, where P_{SUVW} and Q_{SUVW} are as defined in Theorem 4.4.

While the rate-error exponent trade-off in Theorem 4.4 and Theorem 4.5 is the same as that achieved by the Shimokawa-Han-Amari (SHA) scheme [11], the coding strategy is different due to the requirement of the privacy constraint. As mentioned above, in order to obtain a single-letter lower bound for the achievable distortion (and achievable equivocation) of the private part at the detector, it is required that the aposteriori probability distribution of S^n given the observations (M, V^n) at the detector is close in some sense to a desired "target" memoryless distribution. For this purpose, we use stochastic encoding to induce the necessary randomness for S^n at the detector. The coding scheme achieving this is inspired from [67] and to our knowledge has not been used before in the context of distributed HT. The analysis of the type I and type II error probabilities and the privacy achieved by our scheme is novel and involves the application of the well-known *channel resolvability* or *soft-covering lemma* [68–70]. Properties of the total variation distance between probability distributions mentioned in Property 4.4.1 play a key role in this analysis. The analysis

also reveals the interesting fact that the coding schemes in Theorem 4.4 and Theorem 4.5, although quite different from the SHA scheme, achieves the same lower bound on the error-exponent.

Theorems 4.4 and 4.5 provide single-letter inner bounds on $\mathcal{R}_d(\epsilon)$ and $\mathcal{R}_e(\epsilon)$, respectively. A complete computable characterization of these regions would require a matching converse. This is a hard problem, since such a characterization is not available even for the distributed HT problem in general, without a privacy constraint (see [4]). However, it is known that a single-letter characterization of the rate-error exponent region exists for the special case of TACI [12]. In the next section, we show that TACI with a privacy constraint also admits a single-letter characterization, in addition to other optimality results.

4.6 Optimality Results for Special cases

4.6.1 TACI with a Privacy Constraint

Assume that the detector observes two discrete memoryless sources Y^n and Z^n, i.e., $V^n = (Y^n, Z^n)$. Recall that in TACI, the detector tests for the conditional independence of U and Y, given Z. Thus, the joint distribution of the r.v.'s under the null and alternate hypothesis are given by

$$H_0 : \ P_{SUYZ} := P_{S|UYZ} P_{U|Z} P_{Y|UZ} P_Z, \tag{4.26a}$$

and

$$H_1 : \ Q_{SUYZ} := Q_{S|UYZ} P_{U|Z} P_{Y|Z} P_Z, \tag{4.26b}$$

respectively.

Let $\mathcal{R}_e$ and $\mathcal{R}_d$ denote the rate-error exponent-equivocation and rate-error exponent-distortion regions, respectively, for the case of vanishing type I error probability constraint, i.e.,

$$\mathcal{R}_e := \lim_{\epsilon \to 0} \mathcal{R}_e(\epsilon) \text{ and } \mathcal{R}_d := \lim_{\epsilon \to 0} \mathcal{R}_d(\epsilon).$$

Assume that the privacy constraint under the alternate hypothesis is inactive. Thus, we are interested in characterizing the set of all tuples $(R, \kappa, \Lambda_0, \Lambda_1) \in \mathcal{R}_e$ and $(R, \kappa, \Delta_0, \Delta_1) \in \mathcal{R}_d$, where

$$\begin{aligned} \Lambda_1 \leq \Lambda_{min} &:= H_Q(S|U, Y, Z), \\ \text{and } \Delta_1 \leq \Delta_{min} &:= \min_{\phi(u,y,z)} \mathbb{E}_Q \left[d\left(S, \phi(U, Y, Z)\right) \right]. \end{aligned} \tag{4.27}$$

Note that Λ_{min} and Δ_{min} correspond to the equivocation and average distortion of S^n at the detector, respectively, when U^n is available directly at the detector under the alternate hypothesis.

The above assumption is motivated by scenarios, in which, the encoder is more eager to protect S^n when there is a correlation between its own observation and that of the decoder. Consider the following example of user privacy in the context of online shopping, in which the encoder and detector correspond to a consumer and an online shopping portal, respectively. A consumer would like to share some information about his/her shopping behaviour, e.g., shopping history and preferences, with the shopping portal in order to get better deals and recommendations on relevant products. The shopping portal would like to determine whether the consumer belongs to its target age group (e.g., below 30 years old) before sending special offers to this customer. Assuming that the shopping patterns of the users within and outside the target age groups are independent, the shopping portal performs an independence test to check if the consumer's shared data is correlated with the data of its own customers. If the consumer is indeed within the target age group, the shopping portal would like to gather more information about this potential customer, particular interests, more accurate age estimation, etc.; while the user is reluctant to provide any further information. In this example, U^n, S^n and Y^n corresponds to shopping behaviour, more information about the customer, and customers data available to the shopping portal, respectively.

For the above mentioned case, we have the following results.

Proposition 4.6. For the HT given in (4.26), $(R, \kappa, \Lambda_0, \Lambda_{min}) \in \mathcal{R}_e$ if and only if there exists an auxiliary r.v. W, $|\mathcal{W}| \leq |\mathcal{U}| + 2$, such that $(Z, Y, S) - U - W$, and

$$\kappa \leq I_P(W; Y|Z), \tag{4.28}$$

$$R \geq I_P(W; U|Z), \tag{4.29}$$

$$\Lambda_0 \leq H_P(S|W, Z, Y), \tag{4.30}$$

for some joint distribution of the form $P_{SUYZW} := P_{SUYZ} P_{W|U}$.

Proof. For TACI, the inner bound in Theorem 4.4 yields that for $\epsilon \in (0, 1)$, $(R, \kappa, \Lambda_0, \Lambda_1) \in \mathcal{R}_e(\epsilon)$ if there exists an auxiliary r.v. W, such that $(Y, Z, S) - U - W$, and

$$R \geq I_P(W; U|Y, Z), \tag{4.31}$$

$$\kappa \leq \kappa^*(P_{W|U}, R), \tag{4.32}$$

$$\Lambda_0 \leq H_P(S|W, Y, Z), \tag{4.33}$$

$$\Lambda_1 \leq H_Q(S|W, Y, Z), \tag{4.34}$$

where

$$\begin{aligned} \kappa^*(P_{W|U}, R) &:= \min\left\{E_1(P_{W|U}),\ E_2(R, P_{W|U})\right\}, \\ E_1(P_{W|U}) &:= \min_{P_{\tilde{U}\tilde{Y}\tilde{Z}\tilde{W}} \in \mathcal{T}_1(P_{UW}, P_{YZW})} D(P_{\tilde{U}\tilde{Y}\tilde{Z}\tilde{W}} || Q_{UYZ} P_{W|U}), \end{aligned} \tag{4.35}$$

$$E_2(R, P_{W|U}) := \begin{cases} \min\limits_{P_{\tilde{U}\tilde{Y}\tilde{Z}\tilde{W}} \in \mathcal{T}_2(P_{UW}, P_{YZ})} D(P_{\tilde{U}\tilde{Y}\tilde{Z}\tilde{W}} || Q_{UYZ} P_{W|U}) & \\ \qquad + (R - I_P(U; W|Y, Z)), & \text{if } I_P(U; W) > R, \\ \infty, & \text{otherwise,} \end{cases} \tag{4.36}$$

$$\mathcal{T}_1(P_{UW}, P_{YZW}) := \{P_{\tilde{U}\tilde{Y}\tilde{Z}\tilde{W}} \in \mathcal{T}(\mathcal{U} \times \mathcal{Y} \times \mathcal{Z} \times \mathcal{W}) : P_{\tilde{U}\tilde{W}} = P_{UW},\ P_{\tilde{Y}\tilde{Z}\tilde{W}} = P_{YZW}\},$$

$$\begin{aligned} \mathcal{T}_2(P_{UW}, P_{YZ}) := \{&P_{\tilde{U}\tilde{Y}\tilde{Z}\tilde{W}} \in \mathcal{T}(\mathcal{U} \times \mathcal{Y} \times \mathcal{Z} \times \mathcal{W}) : P_{\tilde{U}\tilde{W}} = P_{UW}, \\ &P_{\tilde{Y}\tilde{Z}} = P_{YZ},\ H_P(W|Y, Z) \leq H(\tilde{W}|\tilde{Y}\tilde{Z})\}, \end{aligned}$$

$$P_{SUYZW} := P_{SUYZ} P_{W|U},\ Q_{SUYZW} := Q_{S|YZ} P_{U|Z} P_{Y|Z} P_Z P_{W|U}.$$

Note that since $(Y, Z, S) - U - W$, we have

$$I_P(W;U) \geq I_P(W;U|Z) \geq I_P(W;U|Y,Z). \tag{4.37}$$

Let $\mathcal{B}' := \{P_{W|U} : I_P(U;W|Z) \leq R\}$. Then, for $P_{W|U} \in \mathcal{B}'$, we have,

$$E_1(R, P_{W|U}) = \min_{P_{\tilde{U}\tilde{Y}\tilde{Z}\tilde{W}} \in \mathcal{T}_1(P_{UW}, P_{YZW})} D(P_{\tilde{U}\tilde{Y}\tilde{Z}\tilde{W}} || Q_{UYZ} P_{W|U}) = I_P(Y;W|Z),$$
$$E_2(R, P_{W|U}) \geq I_P(U;W|Z) - I_P(U;W|Y,Z) = I_P(Y;W|Z).$$

Hence,

$$\kappa^*(P_{W|U}, R) \geq I_P(Y;W|Z). \tag{4.38}$$

By noting that $\Lambda_{min} \leq H_Q(S|W,Y,Z)$ (by the data processing inequality), we have shown that for $\Lambda_1 \leq \Lambda_{min}$, $(R, \kappa, \Lambda_0, \Lambda_1) \in \mathcal{R}_e$ if (4.28)-(4.30) are satisfied. This completes the proof of achievability.

Converse: Let $(R, \kappa, \Lambda_0, \Lambda_1) \in \mathcal{R}_e$. Let T be a r.v. uniformly distributed over $[n]$ and independent of all the other r.v.'s (U^n, Y^n, Z^n, M). Define an auxiliary r.v. $W := (W_T, T)$, where $W_i := (M, Y^{i-1}, Z^{i-1}, Z_{i+1}^n)$, $i \in [n]$. Then, for any $\gamma' > 0$ and sufficiently large n, we have

$$\begin{aligned}
n(R+\gamma') &\geq H_P(M) \geq H_P(M|Z^n) \geq I_P(M;U^n|Z^n) \\
&= \sum_{i=1}^n I_P(M;U_i|U^{i-1}, Z^n) \\
&= \sum_{i=1}^n I_P(M, U^{i-1}, Z^{i-1}, Z_{i+1}^n; U_i|Z_i) && (4.39) \\
&= \sum_{i=1}^n I_P(M, U^{i-1}, Z^{i-1}, Z_{i+1}^n, Y^{i-1}; U_i|Z_i) && (4.40) \\
&\geq \sum_{i=1}^n I_P(M, Z^{i-1}, Z_{i+1}^n, Y^{i-1}; U_i|Z_i) \\
&= \sum_{i=1}^n I_P(W_i; U_i|Z_i) = nI_P(W_T; U_T|Z_T, T) \\
&= nI_P(W_T, T; U_T|Z_T) && (4.41) \\
&= nI_P(W;U|Z). && (4.42)
\end{aligned}$$

Here, (4.39) follows since the sequences (U^n, Z^n) are memoryless; (4.40) follows since

$Y^{i-1} - (M, U^{i-1}, Z^{i-1}, Z_{i+1}^n) - U_i$; (4.41) follows from the fact that T is independent of all the other r.v.'s.

The equivocation of source S^n under the null hypothesis can be bounded as follows.

$$\begin{aligned} H(S^n|M, Y^n, Z^n, H=0) &= \sum_{i=1}^n H(S_i|M, S^{i-1}, Y^n, Z^n, H=0) \\ &\leq \sum_{i=1}^n H(S_i|M, Y^{i-1}, Y_i, Z^{i-1}, Z_i, Z_{i+1}^n, H=0) \\ &= \sum_{i=1}^n H(S_i|W_i, Y_i, Z_i, H=0) \\ &= nH(S_T|W_T, Y_T, Z_T, T, H=0) \\ &= nH_P(S|W, Y, Z), \end{aligned} \tag{4.43}$$

where $P_{SUYZW} = P_{SUYZ}P_{W|U}$ for some conditional distribution $P_{W|U}$.

Finally, we prove the upper bound on κ. For any encoding function $f^{(n)}$ and decision region $\mathcal{A}_n \subseteq \mathcal{M} \times \mathcal{Y}^n \times \mathcal{Z}^n$ for H_0 such that $\epsilon_n \to 0$, we have,

$$\begin{aligned} &D\left(P_{MY^nZ^n}||Q_{MY^nZ^n}\right) \\ &\geq P_{MY^nZ^n}(\mathcal{A}_n)\log\left(\frac{P_{MY^nZ^n}(\mathcal{A}_n)}{Q_{MY^nZ^n}(\mathcal{A}_n)}\right) + P_{MY^nZ^n}(\mathcal{A}_n^c)\log\left(\frac{P_{MY^nZ^n}(\mathcal{A}_n^c)}{Q_{MY^nZ^n}(\mathcal{A}_n^c)}\right) \qquad (4.44) \\ &\geq -H(\epsilon_n) - (1-\epsilon_n)\log\left(\beta\left(f^{(n)}, \epsilon_n\right)\right). \end{aligned}$$

Here, (4.44) follows from the log-sum inequality [20]. Thus,

$$\begin{aligned} \limsup_{n\to\infty} \frac{-\log\left(\beta\left(f^n, 0\right)\right)}{n} &\leq \limsup_{n\to\infty} \frac{1}{n} D\left(P_{MY^nZ^n}||Q_{MY^nZ^n}\right) \\ &= \limsup_{n\to\infty} \frac{1}{n} I_P(M; Y^n|Z^n) \qquad (4.45) \\ &= H_P(Y|Z) - \liminf_{n\to\infty} \frac{1}{n} H_P(Y^n|M, Z^n), \qquad (4.46) \end{aligned}$$

where (4.45) follows since $Q_{MY^nZ^n} = P_{MZ^n}P_{Y^n|Z^n}$. The last term can be single-letterized as follows:

$$\begin{aligned} H_P(Y^n|M, Z^n) &= \sum_{i=1}^n H_P(Y_i|Y^{i-1}, M, Z^n) \\ &= \sum_{i=1}^n H_P(Y_i|Z_i, W_i) \\ &= nH_P(Y_T|Z_T, W_T, T) \end{aligned}$$

$$= nH_P(Y|Z,W). \tag{4.47}$$

Substituting (4.47) in (4.46), we obtain

$$\kappa \leq I_P(Y;W|Z). \tag{4.48}$$

Also, note that $(Z,Y) - U - W$ holds. This completes the proof of the converse. The proof of the cardinality bound on W follows using standard arguments based on the Fenchel-Eggleston-Carathéodory's theorem and is given in Appendix C.5. This completes the proof of the theorem. □

Next, we state the result for TACI with a distortion privacy constraint, where the distortion is measured using an arbitrary distortion measure $d(\cdot,\cdot)$. Let $\Delta_{min} := \min_{\phi(u,y,z)} \mathbb{E}_Q\left[d\left(S,\phi(U,Y,Z)\right)\right]$.

Proposition 4.7. For the HT given in (4.26), $(R,\kappa,\Delta_0,\Delta_{min}) \in \mathcal{R}_d$ if and only if there exist an auxiliary r.v. W, $|\mathcal{W}| \leq |\mathcal{U}|+2$, and a deterministic function $\phi : \mathcal{W}\times\mathcal{Y}\times\mathcal{Z} \mapsto \hat{\mathcal{S}}$ such that

$$R \geq I_P(W;U|Z), \tag{4.49}$$

$$\kappa \leq I_P(W;Y|Z), \tag{4.50}$$

$$\Delta_0 \leq \min_{\phi(\cdot,\cdot,\cdot)} \mathbb{E}_P\left[d\left(S,\phi(W,Y,Z)\right)\right], \tag{4.51}$$

for some P_{SUYZW} as defined in Proposition 4.6.

Proof. The proof of achievability follows from Theorem 4.5, similarly to the way Proposition 4.6 is obtained from Theorem 4.4. Hence, only differences will be highlighted. Similar to the inequality $\Lambda_{min} \leq H_Q(S|U,Y,Z)$ in the proof of Proposition 4.6, we need to prove the inequality $\Delta_{min} \leq \mathbb{E}_Q\left[d\left(S,\phi(W,Y,Z)\right)\right]$, where $Q_{SUYZW} := Q_{SUYZ}P_{W|U}$

for some conditional distribution $P_{W|U}$. This can be shown as follows.

$$
\begin{aligned}
&\min_{\phi(\cdot,\cdot,\cdot)} \mathbb{E}_Q\left[d\left(S,\phi(W,Y,Z)\right)\right] \\
&= \sum_{u,y,z} Q_{UYZ}(u,y,z) \sum_{w} P_{W|U}(w|u) \min_{\phi(w,y,z)} \sum_{s} Q_{S|UYZ}(s|u,y,z)\; d(s,\phi(w,y,z)) \\
&\geq \sum_{u,y,z} Q_{UYZ}(u,y,z) \sum_{w,s} P_{W|U}(w|u)\; Q_{S|UYZ}(s|u,y,z)\; d(s,\phi^*(u,y,z)) \qquad (4.52) \\
&\geq \sum_{u,y,z} Q_{UYZ}(u,y,z) \min_{\phi(u,y,z)} \sum_{w,s} P_{W|U}(w|u)\; Q_{S|UYZ}(s|u,y,z)\; d(s,\phi(u,y,z)) \\
&= \sum_{u,y,z} Q_{UYZ}(u,y,z) \min_{\phi(u,y,z)} \sum_{s} Q_{S|UYZ}(s|u,y,z)\; d(s,\phi(u,y,z)) \\
&= \min_{\phi(\cdot,\cdot,\cdot)} \mathbb{E}_Q\left[d\left(S,\phi(U,Y,Z)\right)\right] := \Delta_{min},
\end{aligned}
$$

where, in (4.52), $\phi^*(u,y,z)$ is chosen such that

$$
\phi^*(u,y,z) := \operatorname*{arg\,min}_{\phi(w,y,z), w\in\mathcal{W}} \sum_{s} Q_{S|UYZ}(s|u,y,z)\; d(s,\phi(w,y,z)).
$$

Converse: Let $W = (W_T, T)$ denote the auxiliary r.v. defined in the converse of Proposition 4.6. Inequalities (4.49) and (4.50) follow similarly as obtained in Proposition 4.6. We prove (4.51). Defining $\tilde{\phi}(M, Y^n, Z^n, i) := \bar{\phi}_i(M, Y^n, Z^n)$, we have that,

$$
\begin{aligned}
&\min_{g_r^{(n)}} \mathbb{E}\left[d\left(S^n,\hat{S}^n\right)\Big|H=0\right] \\
&= \min_{\{\tilde{\phi}(m,y^n,z^n,i)\}_{i=1}^n} \mathbb{E}\left[\sum_{i=1}^n d\left(S_i,\tilde{\phi}(M,Y^n,Z^n,i)\right)\Big|H=0\right] \qquad (4.53) \\
&= \min_{\{\tilde{\phi}(\cdot,\cdot,\cdot,\cdot,\cdot)\}_{i=1}^n} \mathbb{E}\left[\sum_{i=1}^n d\left(S_i,\tilde{\phi}(W_i,Z_i,Y_i,Y_{i+1}^n,i)\right)\Big|H=0\right] \\
&\leq \min_{\{\phi(w_i,z_i,y_i,i)\}} \mathbb{E}\left[\sum_{i=1}^n d\left(S_i,\phi(W_i,Z_i,Y_i,i)\right)\Big|H=0\right] \\
&= n \min_{\{\phi(\cdot,\cdot,\cdot,\cdot)\}} \mathbb{E}\left[\mathbb{E}\left[d\left(S_T,\phi(W_T,Z_T,Y_T,T)\right)|T\right]\;\Big|H=0\right] \\
&= n \min_{\{\phi(\cdot,\cdot,\cdot,\cdot)\}} \mathbb{E}\left[d\left(S_T,\phi(W_T,Z_T,Y_T,T)\right)\Big|H=0\right] \\
&= n \min_{\{\phi(w,z,y)\}} \mathbb{E}\left[d\left(S,\phi(W,Z,Y)\right)\Big|H=0\right].
\end{aligned}
$$

In (4.53), we used the fact that (C.1) holds for any arbitrary joint distribution on the r.v.'s (S^n, U^n, M, Y^n, Z^n) in place of $\tilde{P}^{(0)}_{S^n U^n M Y^n Z^n}$. Hence, any Δ_0 satisfying (4.5) satisfies

$$\Delta_0 \leq \min_{\{\phi(w,z,y)\}} \mathbb{E}_P \left[d\left(S, \phi(W, Z, Y)\right) \right].$$

This completes the proof of the converse. The proof of the cardinality bound on W follows using standard arguments based on the Fenchel-Eggleston-Carathéodory's theorem and is given in Appendix C.5. This completes the proof of the theorem. □

A more general version of Proposition 4.6 and Proposition 4.7 is claimed in [71] as Theorem 7 and Theorem 8, respectively, in which a privacy constraint under the alternate hypothesis is also imposed. However, we have identified a mistake in the converse proof; and hence, a single-letter characterization for this general problem remains open.

Remark 4.8. When $Q_{S|UYZ} = Q_{S|YZ}$, a tight single-letter characterization of $\mathcal{R}_e$ and $\mathcal{R}_d$ exists even if the privacy constraint is active under the alternate hypothesis. This is due to the fact that given Y^n and Z^n, M is independent of S^n under the alternate hypothesis. In this case, $(R, \kappa, \Lambda_0, \Lambda_1) \in \mathcal{R}_e$ if and only if there exists an auxiliary r.v. W, such that $(Z, Y, S) - U - W$, and

$$\kappa \leq I_P(W; Y|Z), \tag{4.54}$$

$$R \geq I_P(W; U|Z), \tag{4.55}$$

$$\Lambda_0 \leq H_P(S|W, Z, Y), \tag{4.56}$$

$$\Lambda_1 \leq H_Q(S|Z, Y), \tag{4.57}$$

for some P_{SUYZW} as in Proposition 4.6. Similarly, we have that $(R, \kappa, \Delta_0, \Delta_1) \in \mathcal{R}_d$ if and only if there exist an auxiliary r.v. W and a deterministic function $\phi : \mathcal{W} \times \mathcal{Y} \times \mathcal{Z} \mapsto \hat{\mathcal{S}}$ such that (4.54), (4.55),

$$\Delta_0 \leq \min_{\phi(\cdot,\cdot,\cdot)} \mathbb{E}_P \left[d\left(S, \phi(W, Y, Z)\right) \right], \tag{4.58}$$

$$\Delta_1 \leq \min_{\phi(\cdot,\cdot,\cdot)} \mathbb{E}_Q \left[d\left(S, \phi(Y, Z)\right) \right], \tag{4.59}$$

are satisfied for some P_{SUYZW} as in Proposition 4.6.

The computation of the trade-off given in Proposition 4.6 is challenging in spite of the cardinality bound on the auxiliary r.v. W as closed form solutions do not exist in general. Below, we provide an example where such a solution does exist.

Example 4.9. Let $\mathcal{V} = \mathcal{U} = \mathcal{S} = \{0,1\}$, $V = Y$, $Z =$ constant, $V - S - U$, $P_U(0) = Q_U(0) = 0.5$, $P_{S|U}(0|0) = P_{S|U}(1|1) = Q_{S|U}(0|0) = Q_{S|U}(1|1) = 1-q$, $P_{V|S}(0|0) = P_{V|S}(1|1) = 1-p$ and $Q_{V|S}(0|0) = Q_{V|S}(1|1) = 0.5$. Then, $(R, \kappa, \Lambda_0, 0) \in \mathcal{R}_e$ if there exists $r \in [0, 0.5]$ such that

$$R \geq 1 - h_b(r), \tag{4.60}$$

$$\kappa \leq 1 - h_b((r * q) * p), \tag{4.61}$$

$$\Lambda_0 \leq h_b(p) + h_b(q * r) - h_b(p * (q * r)), \tag{4.62}$$

where, for $a, b \in \mathbb{R}$, $a * b := (1-a) \cdot b + (1-b) \cdot a$, and $h_b : [0,1] \mapsto [0,1]$ is the binary entropy function given by

$$h_b(t) = -(1-t)\log(1-t) - t\log(t).$$

The above characterization[1] is exact for $q = 0$, i.e., $(R, \kappa, \Lambda_0, 0) \in \mathcal{R}_e$ only if there exists $r \in [0, 0.5]$ such that (4.60)-(4.62) are satisfied.

Proof. Achievability: Taking $\mathcal{W} = \{0,1\}$, and $P_{W|U}(0|0) = P_{W|U}(1|1) = 1-r$, the constraints defining the trade-off given in Proposition 4.6 simplifies to

$$\begin{aligned}
I(U;W) &= 1 - h_b(r), \\
I(V;W) &= 1 - h_b((r * q) * p) \\
H(S|V,W) &= H(S|W) - I(S;V|W) \\
&= H(S|W) + H(V|S) - H(V|W) \\
&= h_b(r * q) + h_b(p) - h_b(p * (q * r)).
\end{aligned}$$

[1] Numerical computation shows that the characterization given in (4.60)-(4.62) is exact even when $q \in (0,1)$.

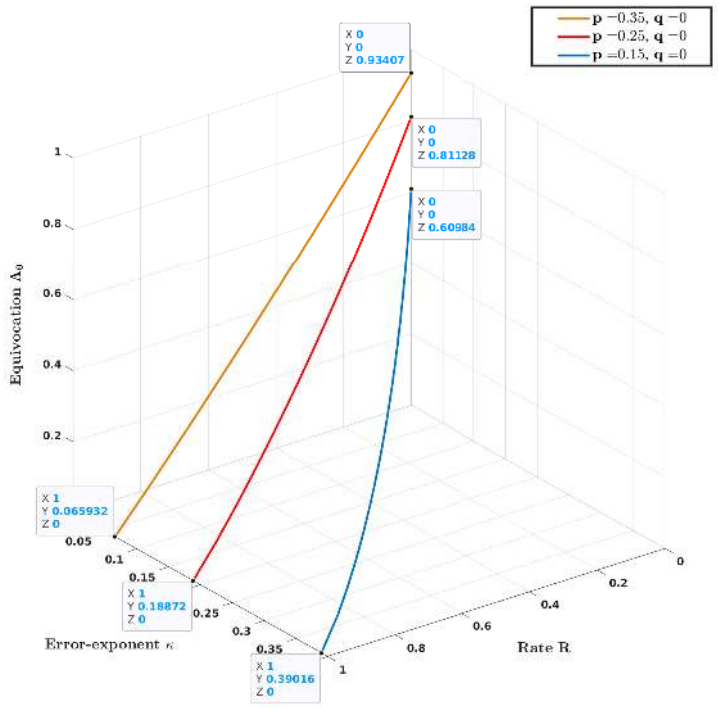


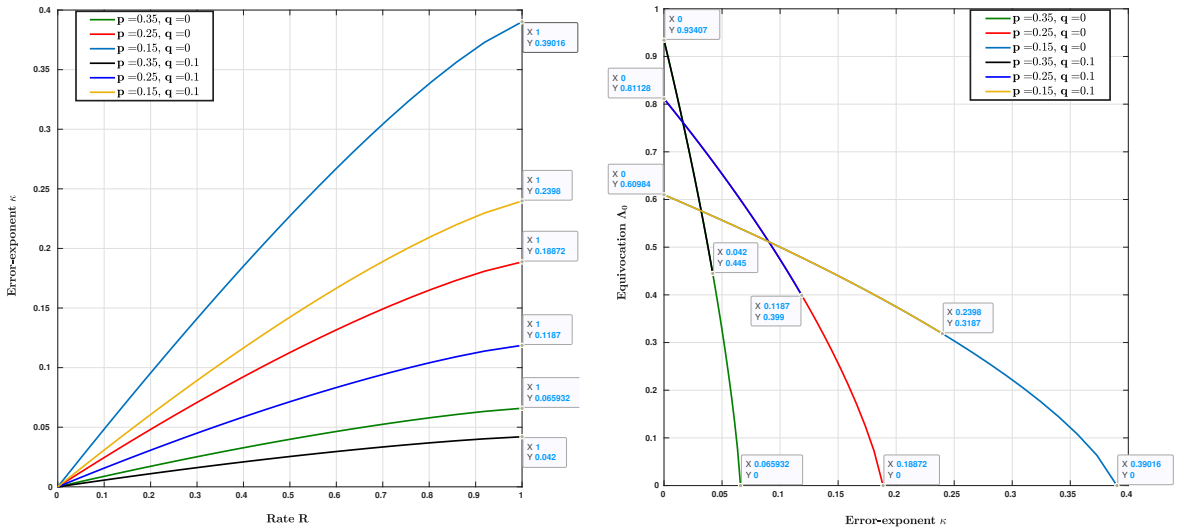
FIGURE 4.2: (R, κ, Λ_0) trade-off at the boundary of $\mathcal{R}_e$ in Example 4.9 (Axes units are in bits)

Converse: On the other hand, if $q = 0$, note that $S = U$. Hence, the same constraints can be bounded as follows:

$$\begin{aligned} I(U;W) &= 1 - H(U|W), \\ I(V;W) &= 1 - H(V|W) \leq 1 - h_b\left(h_b^{-1}(H(U|W)) * p\right) \qquad (4.63) \\ H(U|V,W) &= H(U|W) + H(V|U) - H(V|W) \\ &\leq h_b(p) + H(U|W) - h_b\left(h_b^{-1}(H(U|W)) * p\right), \qquad (4.64) \end{aligned}$$

where, $h_b^{-1} : [0,1] \mapsto [0,0.5]$ is the inverse of the binary entropy function. Here, the inequality in (4.63) and (4.64) follows by an application of Mrs. Gerber's lemma [72], since $V = U \oplus N_p$ under the null hypothesis and $N_p \sim Ber(p)$ is independent of U and W. Also, $\Lambda_{min} = 0$ since $S = U$. Noting that $H(U|W) \in [0,1]$, and defining $r := h_b^{-1}(H(U|W)) \in [0,0.5]$, the result follows. □

Fig. 4.2 depicts the curve $\Big(1 - h_b(r), 1 - h_b(p * (q * r)), h_b(p) + h_b(r * q) - h_b(p * (r * q))\Big)$ for $q = 0$ and $p \in \{0.15, 0.25, 0.35\}$, as r is varied in the range $[0, 0.5]$.



(A) Rate-Error exponent trade-off. (B) Error exponent-Equivocation trade-off.

FIGURE 4.3: Projection of Fig. 4.2 in the $R-\kappa$ plane and $\kappa-\Lambda_0$ plane (Axes units are in bits)

The projection of this curve on the $R-\kappa$ and $\kappa-\Lambda_0$ plane is shown in Fig. 4.3a and Fig. 4.3b, respectively, for $q \in \{0, 0.1\}$ and the same values of p. As expected, the error-exponent κ increases with rate R while the equivocation Λ_0 decreases with κ at the boundary of $\mathcal{R}_e$.

Proposition 4.6 (resp. Proposition 4.7) provide a characterization of $\mathcal{R}_e$ (resp. $\mathcal{R}_d$) under the condition of vanishing type I error probability constraint. Consequently, the converse part of these results are *weak converse* results in the context of HT. In the next subsection, we establish the optimal error exponent-privacy trade-off for the special case of zero-rate compression. This trade-off is independent of the type I error probability constraint $\epsilon \in (0, 1)$, and hence a *strong converse* result.

4.6.2 Zero-rate compression

Assume the following zero-rate constraint on the communication between the observer and the detector,

$$\lim_{n\to\infty} \frac{\log(|\mathcal{M}|)}{n} = 0. \tag{4.65}$$

Note that (4.65) does not imply that $|\mathcal{M}| = 0$, i.e., nothing can be transmitted, but that the message set cardinality can grow at most sub-exponentially in n. Such a scenario is motivated practically by low power or low bandwidth constrained applications in which communication is costly. Considering the HT given in (2.1), Proposition 4.10 and Proposition 4.11 stated below provide an optimal single-letter characterization of $\mathcal{R}_d(\epsilon)$ and $\mathcal{R}_e(\epsilon)$ in this case. While the coding schemes in the achievability part of these results are inspired from that in [9], the analysis of privacy achieved at the detector is new. Lemma 4.4 serves as a crucial tool for this purpose. We next state the results. Let

$$\Delta_0^{max} := \min_{\phi'(\cdot)} \mathbb{E}_P \left[d\left(S, \phi'(V)\right) \right], \tag{4.66a}$$

$$\text{and } \Delta_1^{max} := \min_{\phi'(\cdot)} \mathbb{E}_Q \left[d\left(S, \phi'(V)\right) \right], \tag{4.66b}$$

where $\phi' : \mathcal{V} \mapsto \hat{S}$ denotes a deterministic function.

Proposition 4.10. For $\epsilon \in (0,1)$, $(0, \kappa, \Delta_0, \Delta_1) \in \mathcal{R}_d(\epsilon)$ if and only if it satisfies,

$$\kappa \leq \min_{P_{\tilde{U}\tilde{V}} \in \mathcal{T}_1(P_U, P_V)} D(P_{\tilde{U}\tilde{V}} || Q_{UV}), \tag{4.67}$$

$$\Delta_0 \leq \Delta_0^{max}, \tag{4.68}$$

$$\Delta_1 \leq \Delta_1^{max}, \tag{4.69}$$

where

$$\mathcal{T}_1(P_U, P_V) = \{P_{\tilde{U}\tilde{V}} \in \mathcal{T}(\mathcal{U} \times \mathcal{V}) : P_{\tilde{U}} = P_U,\ P_{\tilde{V}} = P_V\}.$$

Proof. First, we prove that $(0, \kappa, \Delta_0, \Delta_1)$ satisfying (4.67)-(4.69) is achievable. While the encoding and decoding scheme is the same as that in [9], we mention it for the sake of completeness.

Encoding: The encoder sends the message $M = 1$ if $U^n \in T^n_{[P_U]_\delta}$, $\delta > 0$, and $M = 0$ otherwise.

Decoding: The detector declares $\hat{H} = 0$ if $M = 1$ and $V^n \in T^n_{[P_V]_\delta}$, $\delta > 0$. Otherwise, $\hat{H} = 1$ is declared.

We analyze the type I and type II error probabilities for the above scheme. Note that for any $\delta > 0$, the weak law of large numbers implies that

$$\mathbb{P}(U^n \in T^n_{[P_U]_\delta} \cap V^n \in T^n_{[P_V]_\delta})|H=0) = \mathbb{P}(M=1 \cap V^n \in T^n_{[P_V]_\delta})|H=0) \xrightarrow{(n)} 1.$$

Hence, the type I error probability tends to zero, asymptotically. The type II error can be written as follows.

$$\beta(f^{(n)}, g^{(n)}) = \mathbb{P}(U^n \in T^n_{[P_U]_\delta} \cap V^n \in T^n_{[P_V]_\delta})|H=1)$$

$$= \sum_{\substack{u^n \in T^n_{[P_U]_\delta}, \\ v^n \in T_{[P_V]_\delta}}} Q_{U^nV^n}(u^n, v^n) \leq (n+1)^{|\mathcal{U}||\mathcal{V}|} e^{-n(\kappa^* - O(\delta))} = e^{-n\left(\kappa^* - \frac{|\mathcal{U}||\mathcal{V}|\log(n+1)}{n} - O(\delta)\right)},$$

where

$$\kappa^* = \min_{P_{\tilde{U}\tilde{V}} \in \mathcal{T}_1(P_U, P_V)} D(P_{\tilde{U}\tilde{V}} || Q_{UV}).$$

Next, we lower bound the average distortion for S^n achieved by this scheme at the detector. Defining

$$I_U(U^n, \delta) := \mathbb{1}\left(U^n \notin T^n_{[P_U]_\delta}\right), \tag{4.70}$$

$$\rho_0^{(n)}(\delta) := \|P_{S^nV^n} - P_{S^nV^n|I_U(U^n,\delta)=0}\|, \tag{4.71}$$

$$\text{and } \rho_1^{(n)}(\delta) := \|Q_{S^nV^n} - Q_{S^nV^n|I_U(U^n,\delta)=1}\|, \tag{4.72}$$

we can write

$$\begin{aligned}
&\Big| \min_{\{\bar{\phi}_i(m,v^n)\}_{i=1}^n} \mathbb{E}\left[d\left(S^n, \hat{S}^n\right)|H=0\right] - n \min_{\phi'(v)} \mathbb{E}_P\left[d\left(S, \phi'(V)\right)\right] \Big| \\
&= \Big| \min_{\{\bar{\phi}_i(m,v^n)\}_{i=1}^n} \mathbb{E}\left[d\left(S^n, \hat{S}^n\right)|H=0\right] - \min_{\{\phi'_i(v^n)\}_{i=1}^n} \mathbb{E}\left[d\left(S^n, \hat{S}^n\right)|H=0\right] \Big| \\
&\leq \Big| \min_{\{\bar{\phi}_i(m,v^n)\}_{i=1}^n} \mathbb{E}\left[d\left(S^n, \hat{S}^n\right)|H=0\right] - \mathbb{P}\left(M=1|H=0\right)
\end{aligned}$$

$$
\begin{aligned}
&\min_{\{\phi_i'(v^n)\}_{i=1}^n} \mathbb{E}\left[d\left(S^n,\hat{S}^n\right)|M=1,H=0\right]\Big| \\
&\quad + \mathbb{P}\left(M=0|H=0\right) \min_{\{\phi_i'(v^n)\}_{i=1}^n} \mathbb{E}\left[d\left(S^n,\hat{S}^n\right)|M=0,H=0\right] \\
&\leq \Big| \min_{\{\bar{\phi}_i(m,v^n)\}_{i=1}^n} \mathbb{E}\left[d\left(S^n,\hat{S}^n\right)|H=0\right] - \min_{\{\phi_i'(v^n)\}_{i=1}^n} \mathbb{E}\left[d\left(S^n,\hat{S}^n\right)|M=1,H=0\right]\Big| \\
&\quad + \mathbb{P}\left(M=0|H=0\right)\Bigg[\min_{\{\phi_i'(v^n)\}_{i=1}^n} \mathbb{E}\left[d\left(S^n,\hat{S}^n\right)|M=1,H=0\right] \\
&\quad + \min_{\{\phi_i'(v^n)\}_{i=1}^n} \mathbb{E}\left[d\left(S^n,\hat{S}^n\right)|M=0,H=0\right]\Bigg] \\
&\leq \Big| \min_{\{\bar{\phi}_i(m,v^n)\}_{i=1}^n} \mathbb{E}\left[d\left(S^n,\hat{S}^n\right)|H=0\right] - \min_{\{\phi_i'(v^n)\}_{i=1}^n} \mathbb{E}\left[d\left(S^n,\hat{S}^n\right)|I_U(U^n,\delta)=0,H=0\right]\Big| \\
&\quad + \mathbb{P}\left(I_U(U^n,\delta)=1|H=0\right)\Bigg[\min_{\{\phi_i'(v^n)\}_{i=1}^n} \mathbb{E}\left[d\left(S^n,\hat{S}^n\right)|M=1,H=0\right] + \\
&\qquad\qquad \min_{\{\phi_i'(v^n)\}_{i=1}^n} \mathbb{E}\left[d\left(S^n,\hat{S}^n\right)|M=0,H=0\right]\Bigg] \\
&\leq nD_m\rho_0^{(n)}(\delta) + 2\ e^{-n\Omega(\delta)}nD_m \qquad (4.73) \\
&\xrightarrow{(n)} 0, \qquad (4.74)
\end{aligned}
$$

where, (4.73) follows from Property 4.4.1(a), and (4.74) follows from (4.16). Similarly, it can be shown using (4.15) that if $Q_U = P_U$, then

$$
\begin{aligned}
&\Big| \min_{\{\bar{\phi}_i(m,v^n)\}_{i=1}^n} \mathbb{E}\left[d\left(S^n,\hat{S}^n\right)|H=1\right] - \min_{\{\phi_i'(v^n)\}_{i=1}^n} \mathbb{E}\left[d\left(S^n,\hat{S}^n\right)|H=1\right]\Big| \\
&\xrightarrow{(n)} 0. \qquad (4.75)
\end{aligned}
$$

On the other hand, if $Q_U \neq P_U$ and δ is small enough, we have

$$
\mathbb{P}\left(M=0|H=1\right) = \mathbb{P}\left(I_U(U^n,\delta)=1|H=1\right) \geq 1 - e^{-n(D(P_U\|Q_U)-O(\delta))} \xrightarrow{(n)} 1. \quad (4.76)
$$

Hence, we can write for δ small enough,

$$
\begin{aligned}
&\Big| \min_{\{\bar{\phi}_i(m,v^n)\}_{i=1}^n} \mathbb{E}\left[d\left(S^n,\hat{S}^n\right)|H=1\right] - n\min_{\phi'(v)} \mathbb{E}_Q\left[d\left(S,\phi'(V)\right)\right]\Big| \\
&= \Big| \min_{\{\bar{\phi}_i(m,v^n)\}_{i=1}^n} \mathbb{E}\left[d\left(S^n,\hat{S}^n\right)|H=1\right] - \min_{\{\phi_i'(v^n)\}_{i=1}^n} \mathbb{E}\left[d\left(S^n,\hat{S}^n\right)|H=1\right]\Big| \\
&\leq \Big| \min_{\{\bar{\phi}_i(m,v^n)\}_{i=1}^n} \mathbb{E}\left[d\left(S^n,\hat{S}^n\right)|H=1\right] - \mathbb{P}\left(M=0|H=0\right)
\end{aligned}
$$

$$
\begin{aligned}
&\min_{\{\phi_i'(v^n)\}_{i=1}^n} \mathbb{E}\left[d\left(S^n,\hat{S}^n\right)|M=0,H=1\right]\Big| \\
&\quad+\mathbb{P}\left(M=1|H=1\right)\min_{\{\phi_i'(v^n)\}_{i=1}^n} \mathbb{E}\left[d\left(S^n,\hat{S}^n\right)|M=1,H=1\right] \\
&\leq\Big|\min_{\{\bar{\phi}_i(m,v^n)\}_{i=1}^n} \mathbb{E}\left[d\left(S^n,\hat{S}^n\right)|H=1\right]-\min_{\{\phi_i'(v^n)\}_{i=1}^n} \mathbb{E}\left[d\left(S^n,\hat{S}^n\right)|M=0,H=1\right]\Big| \\
&\quad+\mathbb{P}\left(M=1|H=1\right)\Bigg[\min_{\{\phi_i'(v^n)\}_{i=1}^n} \mathbb{E}\left[d\left(S^n,\hat{S}^n\right)|M=1,H=1\right]+ \\
&\qquad\qquad\min_{\{\phi_i'(v^n)\}_{i=1}^n} \mathbb{E}\left[d\left(S^n,\hat{S}^n\right)|M=0,H=1\right]\Bigg] \\
&\leq\Big|\min_{\{\bar{\phi}_i(m,v^n)\}_{i=1}^n} \mathbb{E}\left[d\left(S^n,\hat{S}^n\right)|H=1\right]-\min_{\{\phi_i'(v^n)\}_{i=1}^n} \mathbb{E}\left[d\left(S^n,\hat{S}^n\right)|I_U(U^n,\delta)=1,H=1\right]\Big| \\
&\quad+\mathbb{P}\left(I_U(U^n,\delta)=0|H=1\right)\Bigg[\min_{\{\phi_i'(v^n)\}_{i=1}^n} \mathbb{E}\left[d\left(S^n,\hat{S}^n\right)|M=1,H=1\right] \\
&\qquad\qquad+\min_{\{\phi_i'(v^n)\}_{i=1}^n} \mathbb{E}\left[d\left(S^n,\hat{S}^n\right)|M=0,H=1\right]\Bigg] \\
&\leq nD_m\rho_1^{(n)}(\delta)+2\ e^{-n(D(P_U||Q_U)-O(\delta))}nD_m \qquad (4.77)\\
&\xrightarrow{(n)} 0, \qquad (4.78)
\end{aligned}
$$

This completes the proof of the achievability.

We next prove the converse. Note that by the strong converse result in [10], the R.H.S of (4.67) is an upper bound on the achievable error-exponent for all $\epsilon\in(0,1)$ even without a privacy constraint (hence, also with a privacy constraint). Also,

$$
\begin{aligned}
\min_{g_r^{(n)}} \mathbb{E}\left[d\left(S^n,\hat{S}^n\right)|H=0\right] &\leq \min_{\{\phi_i'(v^n)\}_{i=1}^n}\sum_{i=1}^n \mathbb{E}_{P_{S^nV^n}}\left[d\left(S_i,\phi_i'(V^n)\right)\right] \qquad (4.79)\\
&= \min_{\{\phi'(v)\}} \mathbb{E}_P\left[d(S,\phi'(V))\right].
\end{aligned}
$$

Here, (4.79) follows from the fact that the detector can always reconstruct $\hat{S}_i$ as a function of V^n. Similarly,

$$
\min_{g_r^{(n)}} \mathbb{E}\left[d\left(S^n,\hat{S}^n\right)|H=1\right] \leq \min_{\{\phi'(v)\}} \mathbb{E}_Q\left[d(S,\phi'(V))\right].
$$

Hence, any achievable Λ_0 and Λ_1 has to satisfy (4.68) and (4.69), respectively. This completes the proof. □

The following Proposition is the analogous result to Proposition 4.10 when the privacy measure is equivocation.

Proposition 4.11. For $\epsilon \in (0,1)$, $(0,\kappa,\Lambda_0,\Lambda_1) \in \mathcal{R}_e(\epsilon)$ if and only if it satisfies (4.67) and

$$\Lambda_0 \leq H_P(S|V), \tag{4.80}$$

$$\Lambda_1 \leq H_Q(S|V). \tag{4.81}$$

Proof. For proving the achievablity part, the encoding and decoding scheme is the same as in Proposition 4.10. Hence, the analysis of the error-exponent given in Proposition 4.10 holds. To lower bound the equivocation of S^n at the detector, defining $I_U(U^n,\delta)$, $\rho_0^{(n)}(\delta)$ and $\rho_1^{(n)}(\delta)$ as in (4.70)-(4.72), we can write

$$\begin{aligned}
&|nH_P(S|V) - H(S^n|M,V^n,H=0)| \\
&= |H(S^n|V^n,H=0) - H(S^n|M,V^n,H=0)| \\
&\leq |H(S^n,V^n|H=0) - H(S^n,V^n|M,H=0)| \\
&\leq |H(S^n,V^n|H=0) - \mathbb{P}\left(M=1|H=0\right)\ H(S^n,V^n|M=1,H=0)| \\
&\qquad + \mathbb{P}\left(M=0|H=0\right) H(S^n,V^n|M=0,H=0) \\
&\leq |H(S^n,V^n|H=0) - H(S^n,V^n|M=1,H=0)| \\
&\qquad + \mathbb{P}\left(M=0|H=0\right) \left(H(S^n,V^n|M=1,H=0) + H(S^n,V^n|M=0,H=0)\right) \\
&\leq |H(S^n,V^n|H=0) - H(S^n,V^n|I_U(U^n,\delta)=0,H=0)| \\
&\quad + \mathbb{P}\left(I_U(U^n,\delta)=1|H=0\right) \left(H(S^n,V^n|M=1,H=0) + H(S^n,V^n|M=0,H=0)\right) \\
&\overset{(n)}{\leq} -2\rho_0^{(n)}(\delta) \log\left(\frac{\rho_0^{(n)}(\delta)}{|\mathcal{S}|^n|\mathcal{V}|^n}\right) + 2\ e^{-n\Omega(\delta)} \log\left(|\mathcal{S}|^n|\mathcal{V}|^n\right) \qquad (4.82) \\
&\xrightarrow{(n)} 0, \qquad (4.83)
\end{aligned}$$

where, (4.82) follows due to Lemma 4.3, [20, Lemma 2.12] and the fact that entropy of a r.v. is bounded by the logarithm of cardinality of its support; and, (4.83) follows from (4.16) in Lemma 4.4 since $\delta > 0$. In a similar way, it can be shown using (4.15)

that if $Q_U = P_U$, then

$$|H(S^n|V^n, H=1) - H(S^n|M, V^n, H=1)| \xrightarrow{(n)} 0. \tag{4.84}$$

On the other hand, if $Q_U \neq P_U$ and δ is small enough, we can write,

$$\begin{aligned}
&|nH_Q(S|V) - H(S^n|M, V^n, H=1)| \\
&= |H(S^n|V^n, H=1) - H(S^n|M, V^n, H=1)| \\
&\leq |H(S^n, V^n|H=1) - H(S^n, V^n|M, H=1)| \\
&\leq |H(S^n, V^n|H=1) - H(S^n, V^n|M=0, H=1)| \\
&+ \mathbb{P}\left(I_U(U^n, \delta) = 0|H=1\right)\left(H(S^n, V^n|M=0, H=1) + H(S^n, V^n|M=1, H=1)\right) \\
&\leq -2\rho_1^{(n)}(\delta)\log\left(\frac{\rho_1^{(n)}(\delta)}{|\mathcal{S}|^n|\mathcal{V}|^n}\right) + 2\ e^{-n(D(P_U||Q_U)-O(\delta))}\log\left(|\mathcal{S}|^n|\mathcal{V}|^n\right), \qquad (4.85)
\end{aligned}$$

where (4.85) follows from Lemma 4.3 and (4.76). It follows from (4.14) in Lemma 4.4 that for $\delta > 0$ sufficiently small, $\rho_1^{(n)}(\delta) \leq e^{-n\bar{\delta}}$ for some $\bar{\delta} > 0$, thus implying that the R.H.S. of (4.85) tends to zero. This completes the proof of achievability.

The converse trivially follows from the results in [9] and [10] that the R.H.S of (4.67) is the optimal error-exponent achievable for all values of $\epsilon \in (0, 1)$ even when there is no privacy constraint, and the following inequality

$$H(S^n|M, V^n, H=j) \leq H(S^n|V^n, H=j),\ j=0,1. \tag{4.86}$$

This concludes the proof of the Proposition. □

In Section 4.4.2, we mentioned that it is possible to achieve a positive error-exponent with perfect privacy in our model. Here, we provide an example of TAI with an equivocation privacy constraint under both hypothesis, and show that perfect privacy is possible. Recall that TAI is a special case of TACI, in which, Z = constant, and hence, the null and alternate hypothesis are given by

$$H_0 : (U^n, Y^n) \sim \prod_{i=1}^{n} P_{UY},$$

$$\text{and } H_1 : (U^n, Y^n) \sim \prod_{i=1}^{n} P_U P_Y.$$

Example 4.12. Let $\mathcal{S} = \mathcal{U} = \{0, 1, 2, 3\}$, $\mathcal{Y} = \{0, 1\}$,

$$P_{SU} = 0.125 \cdot \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}, \quad P_{Y|U} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix},$$

$P_{SUY} := P_{SU} P_{Y|U}$ and $Q_{SUY} := P_{SU} P_Y$, where $P_Y = \sum_{u \in \mathcal{U}} P_U(u) P_{Y|U}(y|u)$. Then, we have $H_Q(S|Y) = H_P(S) = H_P(U) = 2$ bits. Also, noting that under the null hypothesis, $Y = U \; mod \; 2$, $H_P(S|Y) = 2$ bits. It follows from the inner bound given by equations (4.31)-(4.34), and, (4.37) and (4.38) that $(R, \kappa, \Lambda_0, \Lambda_1) \in \mathcal{R}_e(\epsilon)$, $\epsilon \in (0, 1)$ if

$$\begin{aligned} R &\geq I_P(W; U), \\ \kappa &\leq I_P(W; Y), \\ \Lambda_0 &\leq H_P(S|W, Y), \\ \Lambda_1 &\leq H_Q(S|W, Y) = H_Q(S|W), \end{aligned}$$

where $P_{SUYW} := P_{SUY} P_{W|U}$ and $Q_{SUYW} := Q_{SUY} P_{W|U}$ for some conditional distribution $P_{W|U}$. If we set $W := U \; mod \; 2$, then we have $I_P(U; W) = 1$ bit, $I_P(Y; W) = H_P(Y) = 1$ bit, $H_P(S|W, Y) = H_P(S|Y) = 2$ bits, and $H_Q(S|W) = H_P(S|Y) = 2$ bits. Thus, by revealing only W to the detector, it is possible to achieve a positive error-exponent while ensuring maximum privacy under both the null and alternate hypothesis, i.e., the tuple $(1, 1, 2, 2) \in \mathcal{R}_e(\epsilon)$, $\forall \; \epsilon \in (0, 1)$.

4.7 A Counterexample to the Strong Converse

Ahlswede and Csiszár obtained a strong converse result for the distributed HT problem without a privacy constraint in [4], where they showed that for any positive rate R, the optimal achievable error-exponent is independent of the type I error probability constraint ϵ. Here, we explore whether a similar result holds in our model, in which,

an additional privacy constraint is imposed. We will show through a counterexample that this is not the case in general. The basic idea used in the counterexample is a "time-sharing" argument which is used to construct from a given coding scheme that achieves the optimal rate-error-exponent-equivocation trade-off under a vanishing type I error probability constraint, a new coding scheme that satisfies the given type I error probability constraint ϵ^* and the same error-exponent as before, yet achieves a higher equivocation for S^n at the detector. This concept has been used previously in other contexts, e.g., in the characterization of the first-order maximal channel coding rate of additive white gaussian noise (AWGN) channel in the finite block-length regime [73], and subsequently in the characterization of the second order maximal coding rate in the same setting [74]. However, we will provide a self-contained proof of the counterexample by utilizing Lemma 4.4 for this purpose.

Assume that the joint distribution P_{SUV} is such that $H_P(S|U,V) < H_P(S|V)$. Proving the strong converse amounts to showing that any $(R, \kappa, \Lambda_0, \Lambda_1) \in \mathcal{R}_e(\epsilon)$ for some $\epsilon \in (0,1)$ also belongs to $\mathcal{R}_e$. Consider TAI problem with an equivocation privacy constraint, in which, $R \geq H_P(U)$ and $\Lambda_1 \leq \Lambda_{min}$. Then, from the optimal single-letter characterization of $\mathcal{R}_e$ given in Proposition 4.6, it follows by taking $W = U$ that $(H_P(U), I_P(V;U), H_P(S|V,U), \Lambda_{min}) \in \mathcal{R}_e$. Note that $I_P(V;U)$ is the maximum error-exponent achievable for any type I error probability constraint $\epsilon \in (0,1)$, even when U^n is observed directly at the detector. Thus, for vanishing type I error probability constraint $\epsilon \to 0$ and $\kappa = I_P(V;U)$, the term $H_P(S|V,U)$ denotes the maximum achievable equivocation for S^n under the null hypothesis. From the proof of Proposition 4.6, it follows that the coding scheme for achieving this tuple is as follows.

1. Quantize U^n to codewords in $\mathcal{C}_U = \{u^n(j),\ j \in [2^{n(H_P(U)+\delta')}],\ u^n(j) \in T^n_{[P_U]_\delta}\}$, i.e., if $U^n = u^n \in T^n_{[P_U]_\delta}$, send $M = j$, where j is the index of u^n in $\mathcal{C}_U$, else, send $M = 0$.

2. At the detector, if $M = 0$, declare $\hat{H} = 1$. Else, if $M \neq 0$, declare $\hat{H} = 0$ if $(U^n(M), V^n) \in T^n_{[P_{UV}]_{\delta'}}$, $\delta' > \delta$ and $\hat{H} = 1$, otherwise.

The type I error probability of the above scheme tends to zero asymptotically with n. Now, for a fixed $\epsilon^* > 0$, consider a modification of this coding scheme as follows.

1. If $U^n = u^n \in T^n_{[P_U]_\delta}$, send $M = j$ with probability $1 - \epsilon^*$, where j is the index of u^n in $\mathcal{C}_U$, and with probability ϵ^*, send $M = 0$. If $U^n = u^n \notin T^n_{[P_U]_\delta}$, send $M = 0$.

2. At the detector, if $M = 0$, declare $\hat{H} = 1$. Else, if $M \neq 0$, declare $\hat{H} = 0$ if $(U^n(M), V^n) \in T^n_{[P_{UV}]_{\delta'}}$, $\delta' > \delta$ and $\hat{H} = 1$, otherwise.

It is easy to see that for this modified coding scheme, the type I error probability is asymptotically equal to ϵ^*, while the error-exponent remains the same as $I(V;U)$ since the probability of declaring $\hat{H} = 0$ is decreased. Recalling that $I_U(u^n, \delta) := \mathbb{1}\left(u^n \notin T^n_{[P_U]_\delta}\right)$, we also have

$$
\begin{aligned}
&\frac{1}{n}H(S^n|M, V^n, H = 0) \\
&= (1-\gamma_n)(1-\epsilon^*)\frac{1}{n}H\left(S^n|U^n, V^n, I_U(U^n,\delta) = 0, H = 0\right) \\
&\quad + (1-\gamma_n)\ \epsilon^*\ \frac{1}{n}H\left(S^n|M = 0, V^n, I_U(U^n,\delta) = 0, H = 0\right) \\
&\quad + \gamma_n\ \frac{1}{n}H(S^n|M = 0, V^n, I_U(U^n,\delta) = 1, H = 0) \\
&\geq (1-\gamma_n)(1-\epsilon^*)\ \left(H_P\left(S|U,V\right) - \gamma_n''\right) \\
&\quad + (1-\gamma_n)\ \epsilon^*\ \frac{1}{n}H\left(S^n|M = 0, V^n, I_U(U^n,\delta) = 0, H = 0\right) \\
&\quad + \gamma_n\ \frac{1}{n}H(S^n|M = 0, V^n, I_U(U^n,\delta) = 1, H = 0) \qquad (4.87) \\
&> (1-\gamma_n)(1-\epsilon^*)\ \left(H_P\left(S|U,V\right) - \gamma_n''\right) + (1-\gamma_n)\ \epsilon^*\ \left(H_P(S|U,V) - \frac{\gamma_n'}{n}\right) \\
&\quad + \gamma_n\ \frac{1}{n}H(S^n|M = 0, V^n, H = 0, I_U(U^n,\delta) = 1) \qquad (4.88) \\
&= (1-\gamma_n)(1-\epsilon^*)\ \left(H_P\left(S|U,V\right) - \gamma_n''\right) + (1-\gamma_n)\ \epsilon^*\left(H_P\left(S|U,V\right) - \frac{\gamma_n'}{n}\right) \\
&\quad + \gamma_n''' \qquad (4.89) \\
&= (1-\gamma_n)H_P\left(S|U,V\right) - \bar{\gamma}_n, \qquad (4.90)
\end{aligned}
$$

where, $\{\gamma_n''\}_{n\in\mathbb{Z}^+}$ denotes some sequence of positive numbers such that $\gamma_n'' \xrightarrow{(n)} 0$,

$$
\begin{aligned}
\gamma_n &:= \mathbb{P}\left(U^n \notin T^n_{P_{[U]_\delta}}\right) \leq e^{-n\Omega(\delta)} \xrightarrow{(n)} 0, \qquad (4.91) \\
\gamma_n' &:= -2\rho_n^* \log\left(\frac{2\rho_n^*}{|\mathcal{S}|^n}\right),
\end{aligned}
$$

$$\rho_n^* := \|P_{S^nV^n|I_U(U^n,\delta)=0,M=0} - P_{S^nV^n}\| = \|P_{S^nV^n|I_U(U^n,\delta)=0} - P_{S^nV^n}\|,$$
$$\gamma_n''' := \frac{\gamma_n}{n} H(S^n|M=0,V^n,,H=0,I_U(U^n,\delta)=1) \xrightarrow{(n)} 0, \tag{4.92}$$
$$\text{and } \bar{\gamma}_n := (1-\gamma_n)(1-\epsilon^*)\gamma_n'' + (1-\gamma_n)\,\epsilon^*\frac{\gamma_n'}{n} - \gamma_n'''.$$

Equation (4.87) follows similarly to the proof of Theorem 1 in [75]. Equation (4.88) is obtained as follows.

$$\frac{1}{n}H\left(S^n|M=0,V^n,H=0,I_U(U^n,\delta=0)\right)$$
$$\geq \frac{1}{n}H\left(S^n|V^n,H=0\right) - \frac{\gamma_n'}{n} \tag{4.93}$$
$$> H_P(S|U,V) - \frac{\gamma_n'}{n}. \tag{4.94}$$

Here, (4.93) is obtained by an application of Lemma 4.3; and (4.94) is due to the assumption that $H_P(S|U,V) < H_P(S|V)$.

It follows from Lemma 4.4 that $\rho_n^* \xrightarrow{(n)} 0$, which in turn implies that

$$\frac{\gamma_n'}{n} \xrightarrow{(n)} 0. \tag{4.95}$$

From (4.91), (4.92) and (4.95), we have that $\bar{\gamma}_n \xrightarrow{(n)} 0$. Hence, equation (4.90) implies that $(H_P(U), I_P(V;U), \Lambda_0^*, \Lambda_{min}) \in \mathcal{R}_e(\epsilon^*)$ for some $\Lambda_0^* > H_P(S|U,V)$. Since $(H_P(U), I_P(V;U), \Lambda_0^*, \Lambda_{min}) \notin \mathcal{R}_e$, this implies that in general, the strong converse does not hold for HT with an equivocation privacy constraint. The same counterexample can be used in a similar manner to show that the strong converse does not hold for HT with an average distortion privacy constraint either.

4.8 Conclusions

In this chapter, we studied a distributed HT problem under a privacy constraint, with equivocation and average distortion as the measures of privacy. We established a single-letter inner bound on the rate-error exponent-equivocation and rate-error exponent-distortion trade-offs. We also obtained an exact single-letter characterization of the

optimal rate-error exponent-equivocation and rate-error exponent-distortion trade-offs for two special cases, when the communication rate over the channel is zero, and for TACI under a privacy constraint. It is interesting to note that the strong converse for distributed HT does not hold when there is an additional privacy constraint in the system. Extending these results to the case where the communication channel between the observer and detector is noisy is an interesting avenue for future research.

Chapter 5

Distributed HT under a Security Constraint

5.1 Overview

In this chapter, we consider a distributed HT problem under security constraints involving three parties, an *observer*, a *detector* and an *eavesdropper*. The observer communicates its observations to the detector which performs TACI of its observations from that of the observer, given some additional correlated side information. Two different scenarios are explored depending on the communication channel between the three parties; (i) *noiseless channel* setting, in which the communication channel between the observer and the detector is a rate-limited noiseless channel, which the eavesdropper also has perfect access to, (ii) *noisy channel* setting, in which the communication channel from the observer to the detector and eavesdropper is a discrete memoryless noisy broadcast channel. Additionally, the eavesdropper has access to side-information different from the detector in general. The objective is to maximize the type II error-exponent (or error-exponent) in the Stein's regime, while keeping the observer's observations as secure from the eavesdropper as possible. With average distortion between the observer's observation and eavesdropper's reconstruction as a measure of secrecy, the trade-off between error-exponent and distortion at the eavesdropper is explored. In the noiseless channel setting, a single-letter inner bound on the rate-error exponent-distortion trade-off is obtained, that is tight when the *less noisy* condition on side-information at the detector holds. In the noisy channel setting, a single-letter inner bound on the above trade-off is established using a hybrid coding scheme.

5.2 Introduction

Data inference and security (or secrecy) are often contradicting objectives in todays interconnected world, neither of which can be compromised. In a distributed setting such as that considered in Chapter 2, it is obvious that the utility gained depends on the communication between the observer and the detector. In practice, communication usually happens over a public channel such as a wireless network. This makes the data vulnerable to external third party attacks such as eavesdropping and adversarial jamming. Clearly, there is a trade-off between the two conflicting requirements of utility and secrecy. In this chapter, we study the trade-off between HT performance and security in the distributed setting studied in Chapter 2 with an additional eavesdropper, which we refer to as the *distributed HT under a security constraint* problem.

5.3 Previous Work and Our Contributions

The information theoretic study of secrecy aspects in communication systems dates back to the works of Shannon, where he introduced the *Shannon Cipher System* (SCS) [76]. It comprises of a data source which is to be communicated reliably (at an arbitrary small probability of error) by the transmitter to a legitimate receiver, such that the data is also kept secure against an eavesdropper. The legitimate parties involved in the communication, i.e., the transmitter and the receiver, share a private resource called a key that is used for encryption at the transmitter and decryption at the receiver. Shannon showed that for the SCS, the minimum amount of secret key rate R_K necessary for reliable communication while ensuring perfect secrecy, i.e., keeping the source samples completely oblivious to the eavesdropper, is equal to the entropy of the source. An important point to note here is that the eavesdropper is assumed to be all powerful, i.e., it is unlimited in resources like time, space and computational power, and that it knows the statistical properties of the system and also the codes used for the communication, except for the realization of the key and the source samples. Thus, the eavesdropper is free to choose its strategy based on its knowledge of the system and its own observations, and the system should ensure security under the worst case situation

when the eavesdropper always chooses the optimal strategy given the information it has.

Several measures have been used to quantify data security. The usage of equivocation of the message as a measure of secrecy in a communication setting first appears in Wyner's landmark chapter on wiretap channels [77]. Equivocation is also used as a secrecy measure in subsequent works, e.g., for the secure transmission of messages over a broadcast channel [78], secure lossy compression of a source with a private and public part [79], secure lossless and lossy compression with side-information [47] [80] [81], and secure transmission of a source over a noisy channel with side information at the receiver [82]. A rate-distortion approach to secrecy is first explored in the work of Yamamoto for the case of a noiseless channel with rate constraint R, where in addition to a distortion constraint D at the legitimate receiver, it is also desired to enforce a certain amount of distortion Δ at the eavesdropper [59]. Yamamoto provided a complete characterization of the set of all admissible (R, D, Δ) tuples when the distortion at the legitimate receiver and eavesdropper are measured using arbitrary additive distortion measures. This approach is extended to the SCS with rate-limited channels in [83], where inner and outer bounds are established for the set of achievable (R, R_K, D, Δ) tuples, in addition to a complete characterization of the (R, R_K, D, Λ) region for general noisy channels, where Λ denotes the equivocation of the source at the eavesdropper and R_K denotes the secret key rate. Recently, the complete characterization of the (R, R_K, D, Δ) trade-off is established for the SCS with rate-limited noiseless channel in [61]. Moreover, it is also shown that equivocation is a special case of the distortion based approach to secrecy with log-loss distortion measure and *causal disclosure* of the source to the eavesdropper (see [61] and [84] for more details). As mentioned in Chapter 4, other models apart from the SCS have also been considered in the information-theoretic security literature such as the availability of different side-information at the legitimate receiver and the eavesdropper. A single-letter characterization of the (R, D, Λ) (rate-distortion-equivocation) trade-off in such a setting where communication between the source and the legitimate receiver happens over a rate-limited noiseless channel is obtained in [81]. On the other hand, when the communication channel from the transmitter to the receiver and eavesdropper

is a noisy broadcast channel, inner bounds on this trade-off are established in [82]. Analogous results with average distortion as the secrecy measure are obtained in [67] and [85], respectively. As already mentioned in Chapter 4, distributed HT under an equivocation secrecy constraint has also been considered previously in the noiseless channel setting [49], where an inner bound on the rate-error exponent-equivocation trade-off is established.

In this chapter, we study distributed HT under security constraints in the setting shown in Fig. 5.1. We focus on TACI with the type II error-exponent (or error-exponent henceforth) in Stein's regime as the performance measure of the HT, and average distortion at the eavesdropper as the secrecy measure. The contributions in this chapter are as follows:

1. We obtain a single-letter inner bound on the rate-error exponent-distortion trade-off in the noiseless channel setting. This inner bound is shown to match with a trivial outer bound under the so-called *less noisy* condition on the side-information at the detector, thus establishing its tightness in that case.

2. In the noisy channel setting, we obtain a single-letter inner bound on the rate-error exponent-distortion trade-off using a hybrid coding scheme for the communication between the observer and the detector.

The organization of the chapter is as follows. In section 5.4, we introduce the detailed system model and definitions along with the supporting lemmas. The main results are stated in Section 5.5. The detailed proofs are deferred to Appendix D. Finally, Section 5.6 concludes the chapter.

5.4 Problem formulation

Consider the setup shown in Fig.5.1, which consists of a single observer connected to the detector via a public channel and an eavesdropper that eavesdrops onto the channel. The i.i.d. observations U^n at the observer are encoded and transmitted to the detector over a public channel, that is also observed by the eavesdropper. Given

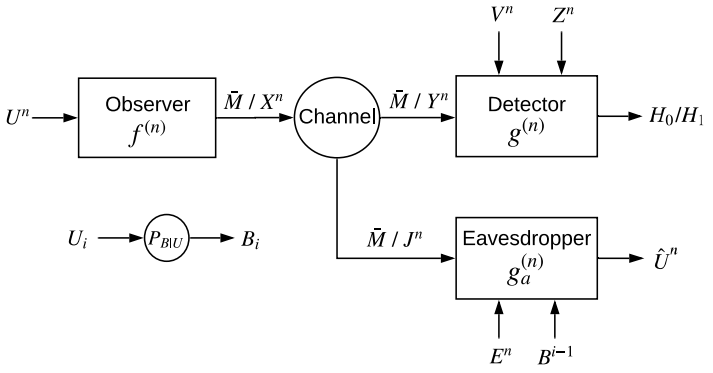


FIGURE 5.1: Distributed HT under security constraints.

its own i.i.d. observations V^n and side-information Z^n, the detector performs TACI with the null and alternate hypothesis given by

$$H_0 : (U^n, V^n, Z^n) \sim \prod_{i=1}^{n} P_{UVZ},$$
$$\text{and } H_1 : (U^n, V^n, Z^n) \sim \prod_{i=1}^{n} P_Z P_{U|Z} P_{V|Z},$$

respectively. The eavesdropper is interested in the reconstruction $\hat{U}^n$, such that the average distortion between U^n and $\hat{U}^n$ is minimized for a given distortion measure $d_a : \mathcal{U} \times \hat{\mathcal{U}} \mapsto [0, D_a]$ with multi-letter distortion given by

$$d_a(u^n, \hat{u}^n) = \sum_{i=1}^{n} d_a(u_i, \hat{u}_i). \tag{5.1}$$

It is assumed that the eavesdropper also observes an i.i.d. side information E^n (correlated with U^n) and has *causal access*[1] to samples B^{i-1} for estimating $\hat{U}_i$, where B^n is the output of a discrete memoryless channel $P_{B|U}$ with input U^n. We consider two distinct scenarios: (i) the public channel is a noiseless channel with communication rate constraint R which the eavesdropper has perfect access to, and (ii) the channel from the observer to the detector and eavesdropper is a noisy discrete memoryless broadcast channel $P_{YJ|X}$ with outputs Y^n and J^n given input X^n, where Y^n is observed by the

[1]This assumption known as causal disclosure makes the notion of distortion as a measure of secrecy more robust [61].

detector and Z^n by the eavesdropper. As mentioned above, the scenarios (i) and (ii) will be referred to as the *noiseless channel* and *noisy channel* setting, respectively.

To summarize the problem formulation, our system model in the noiseless channel setting comprises of:

- A given discrete joint probability distribution of the sources, side-informations and causal disclosure P_{UVZEB} and samples $(U^n, V^n, Z^n, E^n, B^n)$ generated i.i.d. according to $\prod_{i=1}^n P_{UVZEB}$.
- Encoder $f^{(n)} : \mathcal{U}^n \mapsto \bar{\mathcal{M}}$ with output $\bar{M} = f^{(n)}(U^n)$ (possibly stochastic), where $\bar{\mathcal{M}}$ is a set of indices.
- Decoder $g^{(n)} : \bar{\mathcal{M}} \times \mathcal{Z}^n \times \mathcal{V}^n \mapsto \{0, 1\}$, where 0 and 1 indicate H_0 and H_1, respectively.
- Eavesdropper decoder $g_a^{(n)}$, where $g_a^{(n)}$ is chosen from the set of conditional distributions of the form $\{P_{\hat{U}_i|\bar{M},E^n,B^{i-1}}\}_{i=1}^n$.
- Bounded additive distortion measure at the eavesdropper $d_a(\cdot, \cdot)$.

On the other hand, the system model in the noisy channel setting is similar to above, but with the following differences:

- Encoder[2] $f^{(n)} : \mathcal{U}^n \mapsto \mathcal{X}^n$ with output $X^n = f(U^n)$ (possibly stochastic).
- A discrete memoryless broadcast channel $P_{YJ|X}$ with input X^n and outputs (Y^n, Z^n) generated according to $\prod_{i=1}^n P_{YJ|X}$.
- Decoder $g^{(n)} : \mathcal{Y}^n \times \mathcal{Z}^n \times \mathcal{V}^n \mapsto \{0, 1\}$, where 0 and 1 indicate H_0 and H_1, respectively.
- Eavesdropper decoder $g_a^{(n)}$, where $g_a^{(n)}$ is chosen from the set of conditional distributions of the form $\{P_{\hat{U}_i|J^n,E^n,B^{i-1}}\}_{i=1}^n$.

Some more definitions are in order. This is provided below for the noiseless channel setting. The corresponding definitions for the noisy channel setting can be obtained

[2] For simplicity, we take the bandwidth ratio to be unity in the noisy channel setting.

in a straightforward manner by replacing $\bar{M}$ by Y^n or J^n as appropriate. Let $\mathcal{A}_n \subseteq \bar{\mathcal{M}} \times \mathcal{Z}^n \times \mathcal{V}^n$ and $\mathcal{A}_n^c := \bar{\mathcal{M}} \times \mathcal{Z}^n \times \mathcal{V}^n \setminus \mathcal{A}_n$ denote the acceptance region for H_0 and H_1, respectively. The detector is then given by $g^{(n)}(\bar{m}, z^n, v^n) = \mathbb{1}\left((\bar{m}, z^n, v^n) \in \mathcal{A}_n^c\right)$. Let $\bar{\alpha}\left(f^{(n)}, g^{(n)}\right) := P_{\bar{M}Z^nV^n}(\mathcal{A}_n^c)$ and $\bar{\beta}\left(f^{(n)}, g^{(n)}\right) := P_{\bar{M}Z^n} \times P_{V^n|Z^n}(\mathcal{A}_n)$ denote the type I and type II error probabilities for the encoder $f^{(n)}$ and decision rule $g^{(n)}$, respectively. Let $\beta\left(f^{(n)}, \epsilon\right)$ denote the minimum type II error probability achievable for a given type I error probability constraint ϵ as defined in (4.1).

Definition 5.1. For a given type I error probability constraint ϵ, a rate-error exponent-distortion[3] tuple (R, κ, Δ) is *achievable* in the noiseless setting if there exists a sequence of encoding and decoding functions $(f^{(n)}, g^{(n)})$ such that

$$\limsup_{n\to\infty} \frac{\log(|\bar{\mathcal{M}}|)}{n} \leq R, \tag{5.2}$$

$$\liminf_{n\to\infty} \frac{-\log\left(\beta(f^{(n)}, \epsilon)\right)}{n} \geq \kappa, \tag{5.3}$$

and for any $\gamma > 0$, there exists an $n_0 \in \mathbb{Z}^+$ such that

$$\inf_{g_a^{(n)}} \mathbb{E}\left[d_a\left(U^n, \hat{U}^n\right)\right] \geq n\Delta - \gamma, \ \forall\ n \geq n_0. \tag{5.4}$$

The rate-error exponent-distortion region $\bar{\mathcal{R}}_d^*(\epsilon)$ for a given type I error probability constraint ϵ is the closure of the region of all such achievable (R, κ, Δ) tuples.

As already mentioned in Chapter 4, the average distortion constraint given in (5.4) is stronger than a constraint of the form

$$\liminf_{n\to\infty} \inf_{g_a^{(n)}} \mathbb{E}\left[\frac{1}{n} d_a\left(U^n, \hat{U}^n\right) | H = i\right] \geq \Delta. \tag{5.5}$$

The corresponding definition in the noisy channel setting is as follows.

Definition 5.2. An error exponent-distortion pair (κ, Δ) is achievable in the noisy setting if there exists a sequence of encoding and decoding functions $(f^{(n)}, g^{(n)})$ such

[3]It is well known that the equivocation measure of secrecy given by $\frac{1}{n}H(U^n|\bar{M}, E^n)$ in the noiseless setting and by $\frac{1}{n}H(U^n|J^n, E^n)$ in the noisy setting are special cases of the average distortion measure given in (5.5) in the noiseless and noisy setting, respectively, when the distortion measure is log-loss and the source is causally disclosed to the eavesdropper [61]. Accordingly, setting $B = U$ and taking $d_a(u, \hat{u}) = -\log(\hat{u}(u))$ in the above problem formulation results in equivocation secrecy measure.

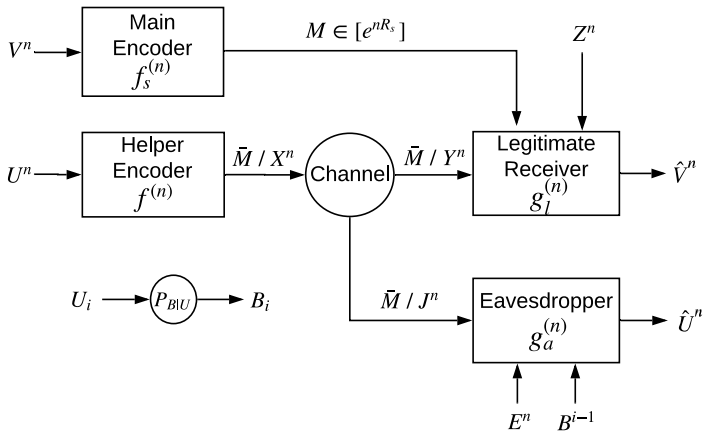


FIGURE 5.2: Equivalent rate-distortion problem in the presence of a helper and eavesdropper.

that (5.3) and (5.4) are satisfied. The error exponent-distortion region $\bar{\mathcal{R}}_d(\epsilon)$ for a given type I error probability constraint ϵ is the closure of the region of all achievable (κ, Δ) tuples.

Our objective is to provide a computable characterization of $\bar{\mathcal{R}}_d^*(\epsilon)$ and $\bar{\mathcal{R}}_d(\epsilon)$ as $\epsilon \to 0$, which we denote by $\bar{\mathcal{R}}_d^*$ and $\bar{\mathcal{R}}_d$, respectively. Using similar methods as in [4], the problem of characterizing $\bar{\mathcal{R}}_d^*$ or $\bar{\mathcal{R}}_d$ can be recast as a lossless source coding problem in the presence of a helper and an eavesdropper, depicted in Fig. 5.2. In here, the helper communicates its observations U^n to the legitimate receiver through a noiseless channel with rate constraint R such that it assists in reducing the compression rate R_s of source V^n to be reconstructed losslessly, while simultaneously also ensuring that the average distortion constraint at the eavesdropper is satisfied. Let $\mathcal{R}_s^*$ denote the closure of the set of all achievable (R, R_s, Δ) tuples. Then, the following equivalence holds:

$$(R, \kappa, \Delta) \in \bar{\mathcal{R}}_d^* \Leftrightarrow (R, H(V|Z) - \kappa, \Delta) \in \mathcal{R}_s^*. \tag{5.6}$$

Let $\mathcal{R}_s$ denote the closure of the set of all achievable (R_s, Δ) pairs for the equivalent lossless source coding problem depicted in Fig. 5.2 with the noiseless channel replaced by the broadcast channel $P_{YJ|X}$. Then, similar to (5.6), we have the equivalence

$$(\kappa,\ \Delta) \in \bar{\mathcal{R}}_d \Leftrightarrow (H(V|Z) - \kappa,\ \Delta) \in \mathcal{R}_s. \tag{5.7}$$

For proving our results, we will in fact consider a more general rate distortion problem in the presence of a helper and an eavesdropper. The problem is to characterize the minimum achievable compression rate R_s in Fig. 5.2 such that the average distortion at the legitimate receiver is below a specified value D while simultaneously ensuring that the average distortion of U at the eavesdropper is above a specified value Δ. More specifically, the system should ensure that (5.2) and (5.4) are satisfied and for sufficiently large n,

$$\mathbb{E}(d_l(V^n, \hat{V}^n)) \leq D, \tag{5.8}$$

where $d_l : \mathcal{V} \times \hat{\mathcal{V}} \mapsto [0, D_l]$ is some additive distortion measure at the legitimate receiver with multi-letter distortion given by

$$d_l(v^n, \hat{v}^n) = \frac{1}{n} \sum_{i=1}^{n} d_l(v_i, \hat{v}_i). \tag{5.9}$$

Denote the set of all such achievable (R, R_s, D, Δ) and (R_s, D, Δ) tuples in the noiseless and noisy channel setting by $\mathcal{R}_g^*$ and $\mathcal{R}_g$, respectively. The lossless source coding problem mentioned above is a special case of this problem with hamming distortion measure $d_l(v, \hat{v}) = \mathbb{1}(v = \hat{v})$ and average distortion constraint $D = 0$.

In section 5.5, we establish single-letter inner bounds to $\bar{\mathcal{R}}_d^*$ and $\bar{\mathcal{R}}_d$ via that of $\mathcal{R}_s^*$ and $\mathcal{R}_s$ using the equivalences in eqns. (5.6) and (5.7). The following lemmas will be useful for our analysis.

Lemma 5.1. *[67, Lemma.2] For a probability distribution P_{YZX} with r.v.'s Y and Z defined on the same alphabet and $0 < \delta < 1$, if $\mathbb{P}(Y \neq Z) \leq \delta$, we have*

$$\|P_{YX} - P_{ZX}\| \leq \delta. \tag{5.10}$$

Lemma 5.2. *[70, Corollary* VII.*8] Given a joint distribution P_{WZXY}, let $\mathcal{C}_W^n$ be a random codebook of sequences $W^n(m_1)$, $m_1 \in [1 : 2^{nR_1}]$ each drawn independently according to $\prod_{i=1}^n P_W$, and $\mathcal{C}_Z^n$ denote the codebook composed of sequences $Z^n(m_1, m_2)$, $m_1 \in [1 : 2^{nR_1}]$, $m_2 \in [1 : 2^{nR_2}]$ each drawn independently according to the distribution*

$\prod_{i=1}^{n} P_{Z|W}(.|W_i(m_1))$. *Let*

$$P_{M_1M_2X^nY^k}(m_1,m_2,x^n,y^k) := \frac{1}{2^{n(R_1+R_2)}} \prod_{i=1}^{n} P_{X|WZ}(x_i|W_i(m_1),Z_i(m_1,m_2)) \prod_{j=1}^{k} P_{Y|XWZ}(y_j|x_j,W_j(m_1),Z_j(m_1,m_2)),$$

$$P'_{M_1X^nY^k}(m_1,x^n,y^k) := \frac{1}{2^{nR_1}} \prod_{i=1}^{n} P_{X|W}(x_i|W_i(m_1)) \prod_{j=1}^{k} P_{Y|XW}(y_j|x_j,W_j(m_1)),$$

$$\text{and } P''_{X^nY^k}(x^n,y^k) := \prod_{i=1}^{n} P_X(x_i) \prod_{j=1}^{k} P_{Y|X}(y_j|x_j).$$

If

$$R_1 > I(W;X), \tag{5.11}$$

$$R_2 > I(W,Z;X) - H(W), \tag{5.12}$$

$$\text{and } R_1 + R_2 > I(W,Z;X), \tag{5.13}$$

then there exists $\gamma_2 > 0$ *such that*

$$\mathbb{E}\left[\|P_{X^nY^k} - P''_{X^nY^k}\|\right] \leq \exp(-\gamma_2 n) \xrightarrow{(n)} 0, \tag{5.14}$$

for any $\zeta' < \min\left(\frac{R_2-I(Z;X|W)}{I(Y;Z|X,W)}, \frac{R_1-I(X;W)}{I(Y;W|X)}\right)$ *and* $k \leq \zeta' n$.

In particular, if $R_2 > I(X;Z|W)$ *(and* $R_1 > 0$*), then there exists* $\gamma_1 > 0$ *such that,*

$$\mathbb{E}_{\mathcal{C}_Z^n}\left[\|P_{M_1X^nY^k} - P'_{M_1X^nY^k}\|\right] \leq \exp(-\gamma_1 n) \xrightarrow{(n)} 0, \tag{5.15}$$

for any $\zeta < \frac{R_2-I(Z;X|W)}{I(Y;Z|X,W)}$ *and* $k \leq \zeta n$.

Next, we present the results.

5.5 Main Results

Theorem 5.3. $(R, \kappa, \Delta) \in \bar{\mathcal{R}}_d^*$ *if there exists auxiliary r.v.'s* W_1 *and* W_2 *such that*

$$R \geq I(W_2; U|Z), \tag{5.16}$$

$$\kappa \leq I(W_2; V|Z), \tag{5.17}$$

$$\begin{aligned}\Delta \leq \min\{\zeta_s, \zeta_p\} \min_{\phi''(\cdot)} \mathbb{E}\left(d_a\left(U, \phi(E)\right)\right) + (\zeta_s - \min\{\zeta_s, \zeta_p\}) \min_{\phi'(\cdot,\cdot)} \mathbb{E}\left(d_a\left(U, \phi(E, W_1)\right)\right), \\ + (1 - \zeta_s) \min_{\phi(\cdot,\cdot)} \mathbb{E}\left(d_a\left(U, \phi(E, W_2)\right)\right),\end{aligned} \tag{5.18}$$

where

$$\zeta_p = \min\left(\frac{[I(W_1; Z) - I(W_1; E)]^+}{I(W_1; B|E)}, 1\right), \tag{5.19}$$

$$\text{and } \zeta_s = \min\left(\frac{[I(W_2; Z|W_1) - I(W_2; E|W_1)]^+}{I(B; W_2|W_1, E)}, 1\right), \tag{5.20}$$

for some joint distribution $P_U P_{W_2|U} P_{W_1|W_2} P_{ZEVB|U}$*. Here,* $\phi : \mathcal{E} \times \mathcal{W}_2 \mapsto \hat{\mathcal{U}}$*,* $\phi' : \mathcal{E} \times \mathcal{W}_1 \mapsto \hat{\mathcal{U}}$ *and* $\phi'' : \mathcal{E} \mapsto \hat{\mathcal{U}}$ *denotes arbitrary deterministic functions.*

Proof. The proof is given in Appendix D.1. □

We also have the following trivial outer-bound for $\bar{\mathcal{R}}_d^*$ when $B =$ constant.

Proposition 5.4. $(R, \kappa, \Delta) \in \bar{\mathcal{R}}_d^*$ only if there exists an auxiliary r.v. W_2 such that (5.16) and (5.17) are satisfied, and

$$\Delta \leq \min_{\phi''(\cdot)} \mathbb{E}\left(d_a\left(U, \phi(E)\right)\right), \tag{5.21}$$

for some joint distribution $P_U P_{W_2|U} P_{ZEV|U}$.

Proof. Equations (5.16) and (5.17) follows from the converse of Proposition 2.8, when the noisy communication channel between the observer and the detector is replaced by a noiseless channel of rate constraint R. Eqn. (5.21) follows by noting that the distortion at the eavesdropper cannot be higher than that obtained by a symbol by

symbol reconstruction $\hat{U}_i = \phi''(E_i)$ using only E^n and ignoring the message $\bar{M}$ from the helper. □

Definition 5.5. Side-information Z is said to be strictly less noisy than E if for all r.v.'s W satisfying the Markov chain $W - U - (Z, E)$, $I(Z;W) > I(E;W)$.

Corollary 5.6. For strictly less noisy side-information Z compared to E at the legitimate receiver and B = constant, $(R, \kappa, \Delta) \in \bar{\mathcal{R}}_d^*$ if and only if there exists an auxiliary r.v. W_2 such that (5.16), (5.17) and (5.21) are satisfied for some distribution $P_U P_{W_2|U} P_{ZEV|U}$.

Proof. For achievability, consider Theorem 5.3 and note that for strictly less noisy case with B =constant, $I(Z;W_1) > I(E;W_1)$, $I(Z;W_2|W_1) > I(E;W_2|W_1)$, $I(W_1;B|E) = 0$ and $I(W_2;B|W_1,E) = 0$. This implies that $\zeta_p = 1$, $\zeta_s = 1$, thus proving the achievability. The converse follows trivially from Proposition 5.4. □

The counterpart of the above results in the noisy channel setting is given below.

Theorem 5.7. $(\kappa, \Delta) \in \bar{\mathcal{R}}_d(\epsilon)$, $\epsilon \in (0,1]$ *if there exists auxiliary r.v.'s* W_1 *and* W_2 *such that*

$$I(W_2;U) \leq I(W_1, W_2; Y, Z), \tag{5.22}$$

$$\kappa \leq I(V; W_2, Y|Z), \tag{5.23}$$

$$\begin{aligned}\Delta \leq \min\{\zeta_s', \zeta_p'\} \min_{\phi''(\cdot,\cdot)} \mathbb{E}\left(d_a\left(U, \phi(J,E)\right)\right) + \left(\zeta_s' - \min\{\zeta_s', \zeta_p'\}\right) \\ \min_{\phi(\cdot,\cdot,\cdot)} \mathbb{E}\left(d_a\left(U, \phi(J,E,W_1)\right)\right) + (1-\zeta_s') \min_{\phi(\cdot,\cdot,\cdot)} \mathbb{E}\left(d_a\left(U, \phi(J,E,W_2)\right)\right),\end{aligned} \tag{5.24}$$

where

$$\zeta_p' = \min\left(\frac{[I(W_1;Y,Z) - I(W_1;J,E)]^+}{I(W_1;B|J,E)}, 1\right), \tag{5.25}$$

$$\zeta_s' = \min\left(\frac{[r_s' - I(W_2;J,E|W_1)]^+}{I(B;W_2|W_1,J,E)}, 1\right), \tag{5.26}$$

$$\text{and } r_s' = \min\left(I(W_2;Y,Z|W_1), I(W_1,W_2;Y,Z) - I(W_1;U)\right), \tag{5.27}$$

for some distribution $P_U P_{W_2|U} P_{W_1|W_2} P_{X|W_1W_2U} P_{YJ|X} P_{ZEVB|U}$, *and* $\phi : \mathcal{J} \times \mathcal{E} \times \mathcal{W}_2 \mapsto \hat{U}$, $\phi' : \mathcal{J} \times \mathcal{E} \times \mathcal{W}_1 \mapsto \hat{U}$ *and* $\phi'' : \mathcal{J} \times \mathcal{E} \mapsto \hat{U}$ *are arbitrary deterministic functions.*

The proof of Theorem 5.7 is provided in Appendix D.2.

5.6 Conclusions

In this chapter, we studied TACI under security constraints. In the noiseless channel setting with average distortion as the measure of secrecy at the eavesdropper, we obtained an inner bound on the rate-error exponent-distortion trade-off using a coding scheme that is a combination of superposition coding and binning. This bound is tight when the side information at the legitimate receiver is less noisy compared to that of the eavesdropper. In the noisy channel setting, we established inner bounds on the error exponent-distortion trade-off by using superposition hybrid coding.

Chapter 6

Conclusions

Distributed HT is a central topic of study at the intersection of information theory and statistics. Understanding the fundamental limits of performance of distributed HT under communication constraints is not merely of theoretical interest, but has far reaching practical implications in several areas such as remote sensing, radar communications, military communications, etc. On the other hand, it is important to take into account the aspects of data security and privacy in distributed settings as computational power is becoming increasingly common and affordable to a potential adversary. Understanding the trade-off between the performance in distributed HT, data privacy and security is the ultimate goal of this dissertation.

In this dissertation, we have focused on the simplest case of binary HT to explore the above mentioned trade-offs. Our basic system model is an extension of the one studied in [4] and [9], in which the rate-limited noiseless channel between the observer and the detector is replaced by a noisy DMC. In Chapter 2, we studied the asymmetric scenario of maximizing the type II error-exponent in the Stein's regime, while in Chapter 3, we studied the trade-off between the type I and type II error-exponents in the Chernoff's regime. Optimal single-letter characterization of the type II error-exponent were obtained in the Stein's regime for the special case of TACI, which revealed the interesting fact that the optimal type II error-exponent depends on the DMC only via its capacity. This is surprising as one would expect that the noisy channel degrades the performance of the HT compared to that of a noiseless channel of the same capacity, and that the reliability function of the channel would play a role in the characterization of the error-exponents in HT. We also obtained single-letter inner bounds on the optimal type II error-exponent for the case of general HT, one using the SHTCC that performs independent HT and channel coding, and the other using the JHTCC scheme that uses hybrid coding for the communication between the observer and the detector.

In Chapter 3, we extended the SHTCC and JHTCC schemes to obtain inner bounds on the trade-off between both the type I and type II error-exponents. An interesting question for future research would be to investigate if computable characterizations can be obtained for some special cases of HT other than those considered in this thesis. Towards this goal, it would be worthwhile to explore if tighter converse bounds can be obtained using novel tools.

In Chapter 4, we introduced the aspect of privacy in the distributed HT problem studied in Chapters 2 and 3, when the channel is rate-limited and noiseless. With equivocation and average distortion as privacy measures, we established a single-letter characterization of the optimal trade-off between communication rate, type II error-exponent and privacy for TACI and zero-rate communication scenarios, and established a single-letter inner bound on this trade-off in the general case. The above mentioned instances of HT where single-letter characterization is obtained are inspired from the analogous results for distributed HT over rate-limited channels without a privacy constraint. The privacy constraints we imposed are slightly stronger than that usually encountered in the literature based on normalized equivocation or normalized average distortion. Although our general problem formulation considered privacy constraints under both the null and alternate hypotheses, the optimal single-letter characterization of the rate-error exponent-privacy trade-off for TACI is obtained when the privacy constraint under the alternate hypothesis is inactive. This is due to the fact that in the converse part of the proof of this result, the auxiliary r.v. identification in the single-letterization step for upper bounding the equivocation (or average distortion) do not match under both hypotheses. We also showed via a counterexample that the strong converse which holds for distributed HT without a privacy constraint, does not hold when a privacy constraint is imposed. This is an interesting observation as we are not aware of any other instance in the distributed HT literature where a strong converse does not hold.

In Chapter 5, we studied distributed HT over a rate-limited noiseless channel under security constraints with average distortion as the security measure. The goal therein is to maximize the type II error-exponent in the Stein's regime such that the observer's observations are protected against an eavesdropper that has either perfect

or noisy access to the message received by the detector, in addition to correlated side-information. We also made the assumption of noisy causal disclosure of the observer's observations to the eavesdropper, which is known to make the average distortion as a measure of security more robust. The results in this chapter generalize analogous results with equivocation as a security measure, as it is well-known that normalized equivocation is a special case of average distortion with the log-loss distortion measure under the assumption of perfect causal disclosure of the observer's data samples to the eavesdropper.

Research Challenges

In this thesis, we studied a distributed HT problem under certain constraints like privacy and security. For these problems, we were able to answer some interesting questions. However, there are still a large number of open questions worth investigating both from a theoretical or practical implementation point of view among the topics studied in this dissertation.

One important question that is of theoretical interest is to investigate whether the *strong converse* holds for the distributed HT over a noisy channel problem considered in Chapter 2. In [23], we were able to show that the strong converse does indeed hold for a special case, when there is no side-information at the detector, and the hypothesis test is on the marginal distribution of the observer's observations. The key tool used for proving this result is the blowing-up lemma of Ahlswede, Gács and Korner [86], which is also a key component in the proof of the strong converse for distributed HT over a rate-limited noiseless channel given in [4]. However, the blowing-up lemma does not appear to be sufficient for proving the strong converse when the channel is noisy. The *change of measure* technique proposed in [87] and the *hypercontractivity* based method proposed in [88] are some promising techniques that could shed further light in this direction.

Another interesting direction of research is to investigate better achievable schemes and novel tools that could result in tighter converse bounds in distributed HT problems.

Any breakthrough or improvement in this direction is hugely significant as it can lead to a better understanding of the trade-off's in the distributed settings studied so far.

It is well known that joint source-channel coding strictly outperforms separation based schemes in multi-terminal communication problems, such as the transmission of correlated sources over a multiple access channel [25]. Also, it is well known that for the transmission of an i.i.d. data source over a DMC, the error exponent achieved by a joint source channel coding scheme is strictly better in general than that achieved by a scheme which performs separate source compression and channel coding [24]. It is an open question whether joint schemes can strictly outperform separation based schemes with regard to the error-exponents trade-off. For TACI, it follows from the achievability scheme in Proposition 2.8 that a separation based scheme achieves the optimal type II error-exponent (Stein's regime). Hence, joint schemes do not offer any advantage compared to separation based schemes in this case. However, in general, we conjecture that joint schemes can achieve a strictly better error-exponent than separation based schemes, and it would be worth investigating specific toy examples for which this claim can be shown to hold.

For the various settings considered in this thesis, the alphabets of the r.v.'s are assumed to be finite, and hence the proof of most of the results rely heavily on the method of types [20]. An interesting direction of research is to consider more general alphabets that have a countable or uncountable support. Of special interest is the Gaussian setting in which the distributions of the r.v.'s involved are jointly Gaussian. Some of the results in this thesis, particularly those related to TACI, are expected to hold in the Gaussian setting via the standard discretization argument (see Remark 3.8 [72]) with the discretization interval not decreasing too fast with the number of samples. However, establishing this and extending the results beyond the Gaussian setting is yet another interesting avenue for further research.

In conclusion, we hope that the research work presented in this dissertation has contributed towards a better understanding of distributed statistical inference problems under communication and privacy/security constraints, and has posed some interesting questions for future research in this area.

Bibliography

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

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

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

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

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

[6] Y. Polyanskiy and Y. Wu. (2017) Lecture notes on information theory. [Online]. Available: http://people.lids.mit.edu/yp/homepage/data/itlectures_v5.pdf

[7] I. Wagner and D. Eckhoff, "Technical privacy metrics: a systematic survey," *arXiv:1512.00327v1 [cs.CR].*

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

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

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

[11] H. Shimokawa, T. S. Han, and S. Amari, "Error bound of hypothesis testing with data compression," in *IEEE Int. Symp. Inf. Theory*, Trondheim, Norway, 1994.

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

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

[14] W. Zhao and L. Lai, "Distributed testing against independence with multiple terminals," in *52nd Annual Allerton Conf.*, IL, USA, Oct. 2014.

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

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

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

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

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

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

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

[22] S. Borade, B. Nakiboğlu, and L. Zheng, "Unequal error protection: An information-theoretic perspective," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5511–5539, Dec 2009.

[23] S. Sreekumar and D. Gündüz, "Hypothesis testing over a noisy channel," in *Proc. IEEE Int. Symp. Inf. Theory*, Paris, France, Jul. 2019.

[24] I. Csiszár, "Joint source-channel error exponent," *Prob. of Control and Inf. Theory*, vol. 9, no. 5, pp. 315–328, Oct. 1980.

[25] T. Cover, A. E. Gamal, and M. Salehi, "Multiple access channels with arbitrarily correlated sources," *IEEE Trans. Inf. Theory*, vol. 26, no. 6, pp. 648–657, Nov. 1980.

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

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

[28] N. Weinberger and Y. Kochman, "On the reliability function of distributed hypothesis testing under optimal detection," *IEEE Trans. Inf. Theory*, vol. 65, no. 8, pp. 4940–4965, Apr. 2019.

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

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

[31] S. Watanabe, "Neyman-Pearson test for zero-rate multiterminal hypothesis testing," *IEEE Trans. Inf. Theory*, vol. 64, no. 7, pp. 4923–4939, Jul. 2018.

[32] N. Weinberger, Y. Kochman, and M. Wigger, "Exponent trade-off for hypothesis testing over noisy channels," in *Proc. IEEE Int. Symp. Inf. Theory*, Paris, France, 2019.

[33] I. Csiszár, "On the error exponent of source-channel transmission with a distortion threshold," *IEEE Trans. Inf. Theory*, vol. 28, no. 6, pp. 823–828, Nov. 1982.

[34] A. Appari and E. Johnson, "Information security and privacy in healthcare: current state of research," *Int. Journ. Internet and Enterprise Management*, vol. 6, no. 4, pp. 279–314, 2010.

[35] R. Gross and A. Acquisti, "Information revelation and privacy in online social networks," in *ACM workshop on Privacy in Electronic Society*, Alexandria, VA, USA, Nov. 2005.

[36] A. Miyazaki and A. Fernandez, "Consumer perceptions of privacy and security risks for online shopping," *Journ. of Consumer Affairs*, vol. 35, no. 1, pp. 27–44, 2001.

[37] G. Giaconi, D. Gündüz, and H. V. Poor, "Privacy-aware smart metering: Progress and challenges," *IEEE Signal Processing Magazine*, vol. 35, no. 6, pp. 59–78, Nov. 2018.

[38] R. Bayardo and R. Agrawal, "Data privacy through optimal k-anonymization," in *Int. Conf. on Data Engineering*, Tokyo, Japan, Apr. 2005.

[39] R. Agrawal and R. Srikant, "Privacy-preserving data mining," in *ACM SIGMOD Int. Conf. on Management of data*, Dallas, USA, May. 2000.

[40] E. Bertino, "Big data-security and privacy," in *IEEE Int. Congress on BigData*, New York, USA, 2015.

[41] Y. Gertner, Y. Ishai, E. Kushilevitz, and T. Malkin, "Protecting data privacy in private information retrieval schemes," *Journ. of Computer and System Sciences*, vol. 60, no. 3, pp. 592–629, Jun 2000.

[42] M. Hay, G. Miklau, D. Jensen, D. Towsley, and P. Weis, "Resisting structural re-identification in anonymized social networks," *Journ. Proc. of the VLDB Endowment*, vol. 1, no. 1, pp. 102–114, Aug. 2008.

[43] A. Narayanan and V. Shmatikov, "De-anonymizing social networks," in *IEEE Symp. on Security and Privacy*, Berkeley, USA, 2009.

[44] J. Liao, L. Sankar, V. Tan, and F. Calmon, "Hypothesis testing under mutual information privacy constraints in the high privacy regime," *IEEE Trans. Inf. Forensics and Security*, vol. 13, no. 4, pp. 1058–1071, Apr. 2018.

[45] J. Liao, L. Sankar, F. Calmon, and V. Tan, "Hypothesis testing under maximal leakage privacy constraints," in *IEEE Int. Symp. Inf. Theory*, Aachen, Germany, Jun. 2017.

[46] A. Gilani, S. B. Amor, S. Salehkalaibar, and V. Y. F. Tan, "Distributed hypothesis testing with privacy constraints," *arXiv:1806.02015*.

[47] D. Gündüz, E. Erkip, and H. V. Poor, "Secure lossless compression with side information," in *IEEE Inf. Theory Workshop*, Porto, Portugal, May. 2008.

[48] ——, "Lossless compression with security constraints," in *IEEE Int. Symp. Inf. Theory*, Jul. 2008.

[49] M. Mhanna and P. Piantanida, "On secure distributed hypothesis testing," in *Proc. IEEE Int. Symp. Inf. Theory*, Hong Kong, China, 2015.

[50] L. Sweeney, "K-anonymity: a model for protecting privacy," *Int. Journ. on Uncertainty, Fuzziness and Knowledge based Systems*, 2002.

[51] C. Dwork, "Differential privacy," *Automata, Languages and Programming. Springer*, vol. 4052, pp. 1–12, 2006.

[52] F. Calmon and N. Fawaz, "Privacy against statistical inference," in *50th Annual Allerton Conf.*, IL, USA, Oct.2012.

[53] A. Makhdoumi, S. Salamatian, N. Fawaz, and M. Medard, "From the information bottleneck to the privacy funnel," in *IEEE Inf. Theory Workshop*, Hobart, Australia, Nov. 2014.

[54] F. Calmon, A. Makhdoumi, and M. Medard, "Fundamental limits of perfect privacy," in *IEEE Int. Symp. Inf. Theory*, Hong Kong, China, Jun. 2015.

[55] I. Issa, S. Kamath, and A. B. Wagner, "An operational measure of information leakage," in *Annual Conf. on Inf. Science and Systems*, Princeton, USA, Mar. 2016.

[56] B. Rassouli and D. Gündüz, "Optimal utility-privacy trade-off with total variation distance as a privacy measure," *IEEE Trans. Inf. Forensics and Security*, to appear, 2019.

[57] S. Goldwasser and S. Micali, "Probabilistic encryption," *Journ. of Computer and System Sciences*, vol. 28, no. 2, pp. 270–299, Apr. 1984.

[58] M. Bellare, S. Tessaro, and A. Vardy, "Semantic security for the wiretap channel," in *Advances in Cryptology-CRYPTO 2012*, Heidelberg, Germany, 2012.

[59] H. Yamamoto, "A rate-distortion problem for a communication system with a secondary decoder to be hindered," *IEEE Trans. Inf. Theory*, vol. 34, no. 4, pp. 835–842, Jul. 1988.

[60] R. Tandon, L. Sankar, and H. V. Poor, "Discriminatory lossy source coding: Side information privacy," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5665–5677, Sep. 2013.

[61] C. Schieler and P. Cuff, "Rate-distortion theory for secrecy systems," *IEEE Trans. Inf. Theory*, vol. 60, no. 12, pp. 7584–7605, Dec. 2014.

[62] G. K. Agarwal, "On information theoretic and distortion-based security," *PhD Thesis-UCLA [Available online] https://escholarship.org/uc/item/7qs7z91g*, 2019.

[63] N. Merhav, "On random coding error exponents of watermarking systems," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 420–430, Mar. 2000.

[64] A. Somekh-Baruch and N. Merhav, "On the error exponent and capacity games of private watermarking systems," *IEEE Trans. Inf. Theory*, vol. 49, no. 3, pp. 537–562, Mar. 2003.

[65] Z. Li, T. Oechtering, and D. Gündüz, "Privacy against a hypothesis testing adversary," *IEEE Trans. Inf. Forensics and Security*, vol. 14, no. 6, pp. 1567–1581, Jun. 2019.

[66] Y. Wang, Y. O. Basciftci, and P. Ishwar, "Privacy-utility tradeoffs under constrained data release mechanisms," *arXiv:1710.09295*.

[67] E. C. Song, P. Cuff, and H. V. Poor, "The likelihood encoder for lossy compression," *IEEE Trans. Inf. Theory*, vol. 62, no. 4, pp. 1836–1849, Apr. 2016.

[68] A. D. Wyner, "The common information of two dependent random variables," *IEEE Trans. Inf. Theory*, vol. 21, no. 2, pp. 163–179, Mar. 1975.

[69] T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 752–772, May. 1993.

[70] P. Cuff, "Distributed channel synthesis," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7071–7096, Nov. 2013.

[71] S. Sreekumar, D. Gündüz, and A. Cohen, "Distributed hypothesis testing under privacy constraints," in *IEEE Inf. Theory Workshop*, Guangzhou, China, Nov. 2018.

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

[73] Y. Polyanskiy, *Channel coding: non-asymptotic fundamental limits*. PhD Thesis, Princeton University, 2010.

[74] W. Yang, G. Caire, G. Durisi, and Y. Polyanskiy, "Optimum power control at finite blocklength," *IEEE Trans. Inf. Theory*, vol. 61, no. 9, pp. 4598–4615, Sep. 2015.

[75] J. Villard and P. Piantanida, "Secure multiterminal source coding with side information at the eavesdropper," *IEEE Trans. Inf. Theory*, vol. 59, no. 6, pp. 3668–3692, Jun. 2013.

[76] C. E. Shannon, "Communication theory of secrecy systems," *Bell System Technical Journal*, vol. 28, no. 4, pp. 656–715, Oct. 1949.

[77] A. D. Wyner, "The wire-tap channel," *Bell System Technical Journal*, vol. 54, no. 8, pp. 1355–1387, Oct. 1975.

[78] I. Csiszár and J. Körner, "Broadcast channels with confidential messages," *IEEE Trans. Inf. Theory*, vol. 24, no. 3, pp. 339–348, May. 1978.

[79] H. Yamamoto, "A source coding problem for sources with additional outputs to keep secret from the receiver or wiretappers," *IEEE Trans. Inf. Theory*, vol. 29, no. 6, pp. 918–923, Nov. 1983.

[80] R. Tandon, S. Ulukus, and K. Ramchandran, "Secure source coding with a helper," *IEEE Trans. Inf. Theory*, vol. 59, no. 4, pp. 2178–2187, Apr. 2013.

[81] J. Villard and P. Piantanida, "Secure lossy source coding with side information at the decoders," in *48th Annual Allerton Conf.*, Illinois, USA, Sept-Oct. 2010.

[82] J. Villard, P. Piantanida, and S. Shamai, "Secure transmission of sources over noisy channels with side information at the receivers," *IEEE Trans. Inf. Theory*, vol. 60, no. 1, pp. 713–739, Jan. 2014.

[83] H. Yamamoto, "Rate-distortion theory for the shannon cipher system," *IEEE Trans. Inf. Theory*, vol. 43, no. 3, pp. 827–835, May. 1997.

[84] T. A. Courtade and T. Weissman, "Multiterminal source coding under logarithmic loss," *IEEE Trans. Inf. Theory*, vol. 60, no. 1, pp. 740–761, Jan. 2014.

[85] E. C. Song, P. Cuff, and H. V. Poor, "Joint source-channel secrecy using hybrid coding," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Hong Kong, China, Jun. 2015.

[86] R. Ahlswede, P. Gács, and J. Körner, "Bounds on conditional probabilities with applications in multi-user communication," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, vol. 34, no. 2, pp. 157–177, 1976.

[87] H. Tyagi and S. Watanabe, "Strong converse using change of measure arguments," *arXiv:1805.04625*.

[88] J. Liu, R. van Handel, and S. Verdú, "Beyond the blowing-up lemma: Sharp converses via reverse hypercontractivity," in *Proc. IEEE Int. Symp. Inf. Theory*, Aachen, Germany, 2017.

[89] R. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Trans. Inf. Theory*, vol. 11, no. 1, pp. 3–18, Jan. 1965.

Appendix A

Proofs for Chapter 2

A.1 Proof of Theorem 2.2

The proof outline is as follows. We first describe the encoding and decoding operations of the SHTCC scheme. The random coding method is used to analyze the type I and type II error probabilities achieved by this scheme, averaged over the ensemble of randomly generated codebooks. By the standard expurgation technique [21] (e.g., removing "worst" codebooks in the ensemble with the highest type I error probability such that the total probability of the removed codebooks lies in the interval $(0.5, 1)$), this guarantees the existence of at least one deterministic codebook that achieves type I and type II error probabilities of the same order, i.e., within a constant multiplicative factor. Since, in our scheme below, the type I error probability averaged over the random code ensemble vanishes asymptotically with the the number of samples k, the same holds for the codebook obtained after expurgation. Moreover, the error-exponent is not affected by a constant multiplicative factor on the type II error probability, and thus, this codebook asymptotically achieves the same type I error probability and error-exponent as the average.

For brevity, in the proof below, we denote the information theoretic quantities like $I_P(U;W)$, $T^k_{[P_{UW}]_\delta}$, etc., that are computed with respect to joint distribution P_{UVWSXY} given in (A.1) below by $I(U;W)$, $T^k_{[UW]_\delta}$, etc.

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

$$P_{UVWSXY}(P_{W|U}, P_{SX}) := P_{UV} P_{W|U} P_{SX} P_{Y|X}. \tag{A.1}$$

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

$$I(U;W|V) + 2\mu \leq R \leq \tau I(X;Y|S) - \mu. \tag{A.2}$$

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

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

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

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

Decoding: At the decoder, $g_c^{(k,n)}$ outputs $\hat{M} = 0$ if for some $1 \leq i \leq |\mathcal{S}|$, the channel outputs corresponding to the i^{th} block does not belong to $T^n_{[P_{Y|S=s_i}]_\delta}$. Otherwise, $\hat{M}$

is set as the index of the codeword corresponding to the maximum-likelihood candidate among the ordinary codewords. If $\hat{M} = 0$, H_1 is declared. Else, given the side information sequence $V^k = v^k$ and estimated bin-index $\hat{M} = \hat{m}$, $g_s^{(k,n)}$ searches for a typical sequence $\hat{w}^k = w^k(\hat{j}) \in T^k_{[W]_{\hat{\delta}}}$, in codebook $\mathcal{C}$ such that

$$\hat{j} = \underset{\substack{l:\ f_B(l)=\hat{m},\\ w^k(l)\in T^k_{[W]_{\hat{\delta}}}}}{\arg\min}\ H_e(w^k(l)|v^k),\ \text{if } I(U;W)+\mu > R,$$

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

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

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

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

$$\mathcal{E}_{TE} = \left\{(U^k, V^k) \notin T^k_{[UV]_{\bar{\delta}}}\right\}$$
$$\mathcal{E}_{EE} = \left\{\nexists\ j \in \left[M'_k\right] : (U^k, W^k(j)) \in T^k_{[UW]_{\delta}}\right\}$$
$$\mathcal{E}_{ME} = \left\{(V^k, W^k(J)) \notin T^k_{[VW]_{\bar{\delta}}}\right\}$$
$$\mathcal{E}_{DE} = \Bigg\{\exists\ l \in \left[M'_k\right],\ l \neq J : f_B(l) = f_B(J),\ W^k(l) \in T^k_{[W]_{\hat{\delta}}},\ H_e(W^k(l)|V^k) \leq H_e(W^k(J)|V^k)\Bigg\}$$
$$\mathcal{E}_{CD} = \left\{g_c^{(k,n)}(Y^n) \neq M\right\}$$

$\mathbb{P}(\mathcal{E}_{TE}|H=0)$ tends to 0 asymptotically by the weak law of large numbers. Conditioned on $\mathcal{E}^c_{TE}$, $U^k \in T_{[U]_{\delta'}}$ and by the covering lemma [20, Lemma 9.1], it is well known that for $\mu = O(\delta)$ chosen appropriately, $\mathbb{P}(\mathcal{E}_{EE}|\mathcal{E}^c_{TE})$ tends to 0 doubly exponentially with k. Given $\mathcal{E}^c_{EE} \cap \mathcal{E}^c_{TE}$ holds, it follows from the Markov chain relation $V - U - W$ and the Markov lemma [72], that $\mathbb{P}(\mathcal{E}_{ME}|\mathcal{E}^c_{TE} \cap \mathcal{E}^c_{EE})$ tends to zero as $k \to \infty$. Next, we consider $\mathbb{P}(\mathcal{E}_{DE})$. Given that $\mathcal{E}^c_{ME} \cap \mathcal{E}^c_{EE} \cap \mathcal{E}^c_{TE}$ holds, note that for k sufficiently large,

$H_e(W^k(J)|V^k) \leq H(W|V) + O(\delta)$. Thus, we have (for sufficiently large k)

$$\begin{aligned}
&\mathbb{P}(\mathcal{E}_{DE}|\ \ V^k = v^k, W^k(J) = w^k, \mathcal{E}_{ME}^c \cap \mathcal{E}_{EE}^c \cap \mathcal{E}_{TE}^c) \\
&\leq \sum_{\substack{l=1,\\ l\neq J}}^{M_k'} \sum_{\substack{\tilde{w}^k \in T^k_{[W]_{\hat{\delta}}}: \\ H_e(\tilde{w}^k|v^k) \\ \leq H_e(w^k|v^k)}} \mathbb{P}\Big(f_B(l) = f_B(J),\ \ W^k(l) = \tilde{w}^k|\ V^k = v^k, W^k(J) = w^k, \\
&\qquad\qquad \mathcal{E}_{ME}^c \cap \mathcal{E}_{EE}^c \cap \mathcal{E}_{TE}^c\Big) \\
&= \sum_{\substack{l=1,\\ l\neq J}}^{M_k'} \sum_{\substack{\tilde{w}^k \in T^k_{[W]_{\hat{\delta}}}: \\ H_e(\tilde{w}^k|v^k) \leq H_e(w^k|v^k)}} \mathbb{P}(W^k(l) = \tilde{w}^k|\ V^k = v^k, W^k(J) = w^k, \mathcal{E}_{ME}^c \cap \mathcal{E}_{EE}^c \cap \mathcal{E}_{TE}^c)\ \frac{1}{e^{kR}} \\
&\leq \sum_{\substack{l=1,\\ l\neq J}}^{M_k'} \sum_{\substack{\tilde{w}^k \in T^k_{[W]_{\hat{\delta}}}: \\ H_e(\tilde{w}^k|v^k) \leq H_e(w^k|v^k)}} 2 \cdot e^{-kR} e^{-k(H(W) - O(\delta))} \qquad \text{(A.3)} \\
&\leq \sum_{\substack{l=1,\\ l\neq J}}^{M_k'} (k+1)^{|\mathcal{V}||\mathcal{W}|}\ e^{k(H(W|V) + O(\delta))} \cdot 2 \cdot e^{-kR} e^{-k(H(W) - O(\delta))} \qquad \text{(A.4)} \\
&\leq e^{-k(R - I(U;W|V) - \delta_1^{(k)})}, \qquad \text{(A.5)}
\end{aligned}$$

where

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

To obtain (A.3), we used the fact that

$$\mathbb{P}(W^k(l) = \tilde{w}^k|\ \mathcal{E}_{ME}^c \cap \mathcal{E}_{EE}^c \cap \mathcal{E}_{TE}^c, W^k(J) = w^k, V^k = v^k) \leq 2 \cdot \mathbb{P}(W^k(l) = \tilde{w}^k). \qquad \text{(A.6)}$$

This follows similarly to (A.28), which is discussed in the type II error analysis section below. In order to obtain the expression in (A.4), we first summed over the types $P_{\tilde{W}}$ of sequences within the typical set $T^k_{[W]_\delta}$ that have empirical entropy less than $H_e(w^k|v^k)$; and used the facts that the number of sequences within such a type is upper bounded by $e^{k(H(W|V) + \gamma_1(k))}$, and the total number of types is upper bounded by $(k+1)^{|\mathcal{V}||\mathcal{W}|}$ [20]. Summing over all $(w^k, v^k) \in T^k_{[VW]_{\tilde{\delta}}}$, we obtain (for sufficiently

large k) that

$$\begin{aligned}
&\mathbb{P}(\mathcal{E}_{DE}|\mathcal{E}_{ME}^c \cap \mathcal{E}_{EE}^c \cap \mathcal{E}_{TE}^c) \\
&\leq \sum_{(w^k,v^k)\in T^k_{[WV]_{\tilde{\delta}}}} \mathbb{P}(W^k(J)=w^k, V^k=v^k|\mathcal{E}_{ME}^c \cap \mathcal{E}_{EE}^c \cap \mathcal{E}_{TE}^c)\; e^{-k(R-I(U;W|V)-\delta_1^{(k)})} \\
&\leq e^{-k(R-I(U;W|V)-\delta_1^{(k)})} \leq e^{-k\frac{\mu}{2}}, \qquad \text{(A.7)}
\end{aligned}$$

where, (A.7) follows from (A.2) by choosing $\mu = O(\delta)$ appropriately.

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

$$\mathbb{P}(\mathcal{E}_{CT}|\mathcal{E}_{EE}^c) \xrightarrow{(k)} 0. \qquad \text{(A.8)}$$

Also, it follows from [20, Exercise 10.18, 10.24] that for sufficiently large n (depending on μ, τ, $|\mathcal{X}|$ and $|\mathcal{Y}|$),

$$\mathbb{P}\left(\mathcal{E}_{CD}|\mathcal{E}_{EE}^c \cap \mathcal{E}_{CT}^c\right) \leq e^{-nE_x(\frac{R}{\tau}+\frac{\mu}{2\tau},P_{SX})}. \qquad \text{(A.9)}$$

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

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

$$\mathcal{E}_0 = \left\{U^k \notin T^k_{[U]_{\delta'}}\right\}. \qquad \text{(A.10)}$$

Then, the type II error probability can be written as

$$\beta\left(k, n, f^{(k,n)}, g^{(k,n)}\right)$$

$$= \sum_{(u^k,v^k)\in\mathcal{U}^k\times\mathcal{V}^k} \mathbb{P}(U^k=u^k,V^k=v^k|H=1)\ \mathbb{P}(\mathcal{D}_0|U^k=u^k,V^k=v^k). \tag{A.11}$$

Let $\mathcal{E}_{NE} := \mathcal{E}_{EE}^c \cap \mathcal{E}_0^c$. The last term in (A.11) can be upper bounded as follows.

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

Thus, we have

$$\begin{aligned}
&\beta\left(k,n,f^{(k,n)},g^{(k,n)}\right)\\
&\le \sum_{\substack{(u^k,v^k)\\ \in\,\mathcal{U}^k\times\mathcal{V}^k}} \mathbb{P}(U^k=u^k,V^k=v^k|H=1)\Big[\mathbb{P}(\mathcal{D}_0|U^k=u^k,V^k=v^k,\mathcal{E}_{NE})\\
&\qquad\qquad + \mathbb{P}(\mathcal{D}_0|U^k=u^k,V^k=v^k,\mathcal{E}_{NE}^c)\Big].
\end{aligned} \tag{A.12}$$

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

$$\begin{aligned}
&\mathbb{P}(\mathcal{D}_0|\ U^k=u^k,V^k=v^k,\mathcal{E}_{NE})\\
&= \sum_{j=1}^{M_k'}\sum_{m=1}^{e^{kR}} \mathbb{P}(J=j,f_B(J)=m|\ U^k=u^k,V^k=v^k,\mathcal{E}_{NE})\\
&\qquad\qquad \mathbb{P}(\mathcal{D}_0|U^k=u^k,V^k=v^k,J=j,f_B(J)=m,\ \mathcal{E}_{NE}).
\end{aligned} \tag{A.13}$$

By the symmetry of the codebook generation, encoding and decoding procedure, the term $\mathbb{P}(\mathcal{D}_0|U^k=u^k,V^k=v^k,J=j,f_B(J)=m,\ \mathcal{E}_{NE})$ in (A.13) is independent of the value of J and $f_B(J)$. Hence, w.l.o.g. assuming $J=1$ and $f_B(J)=1$, we can write

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

$$= \mathbb{P}(\mathcal{D}_0|U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1,\ \mathcal{E}_{NE})$$
$$= \sum_{w^k \in \mathcal{W}^k} \mathbb{P}(W^k(1) = w^k|U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1,\ \mathcal{E}_{NE})$$
$$\mathbb{P}(\mathcal{D}_0|U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, W^k(1) = w^k,\ \mathcal{E}_{NE}). \quad \text{(A.14)}$$

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

$$\mathcal{E}_{CE} = \{\mathcal{E}_{CT}^c \text{ and } \hat{M} = g_c^{(k,n)}(Y^n) \neq M\}, \quad \text{(A.15)}$$

and

$$\mathcal{E}_{BE} = \left\{\exists\ l \in \left[M_k'\right],\ l \neq J,\ f_B(l) = \hat{M},\ W^k(l)) \in T^k_{[W]_\delta}, (V^k, W^k(l)) \in T^k_{[VW]_{\tilde{\delta}}}\right\}. \quad \text{(A.16)}$$

Define the following events

$$\mathcal{F} = \{U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, W^k(1) = w^k,\ \mathcal{E}_{NE}\}, \quad \text{(A.17)}$$
$$\mathcal{F}_1 = \{U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, W^k(1) = w^k,\ \mathcal{E}_{NE},\ \mathcal{E}_{CE}\}, \quad \text{(A.18)}$$
$$\mathcal{F}_2 = \{U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, W^k(1) = w^k,\ \mathcal{E}_{NE},\ \mathcal{E}_{CE}^c\}, \quad \text{(A.19)}$$
$$\mathcal{F}_{21} = \{U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, W^k(1) = w^k,\ \mathcal{E}_{NE},\ \mathcal{E}_{CE}^c,\ \mathcal{E}_{BE}\}, \quad \text{(A.20)}$$
$$\mathcal{F}_{22} = \{U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, W^k(1) = w^k,\ \mathcal{E}_{NE},\ \mathcal{E}_{CE}^c,\ \mathcal{E}_{BE}^c\}. \quad \text{(A.21)}$$

The last term in (A.14) can be expressed as follows.

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

where

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

It follows from (A.9) that for sufficiently large k,

$$\mathbb{P}(\mathcal{E}_{CE}|\mathcal{F}) \leq e^{-nE_x(\frac{R}{\tau}+\frac{\mu}{2\tau},P_{SX})} = e^{-k\tau E_x(\frac{R}{\tau}+\frac{\mu}{2\tau},P_{SX})}. \tag{A.23}$$

Next, consider the type II error event that happens when an error occurs at the channel decoder. We need to consider two separate cases: $I(U;W)+\mu > R$ and $I(U;W)+\mu \leq R$. Note that in the former case, binning is performed and type II error happens at the decoder only if a sequence $W^k(l)$ exists in the wrong bin $\hat{M} \neq M = f_B(J)$ such that $(V^k, W^k(l)) \in T^k_{[VW]_{\tilde{\delta}}}$. As noted in [26], the calculation of the probability of this event does not follow from the standard random coding argument usually encountered in achievability proofs due to the fact that the chosen codeword $W^k(J)$ depends on the entire codebook. Following steps similar to those in [26], we analyze the probability of this event (averaged over codebooks $\mathcal{C}$ and random binning) as follows. We first consider the case when $I(U;W)+\mu > R$.

$$\begin{aligned}
&\mathbb{P}(\mathcal{D}_0|\mathcal{F}_1) \leq \mathbb{P}(\ \exists\ W^k(l):\ f_B(l) = \hat{M} \neq 1,\ (W^k(l), v^k) \in T^k_{[WV]_{\tilde{\delta}}}|\mathcal{F}_1) \\
&\leq \sum_{l=2}^{M'_k} \sum_{\hat{m}\neq 1} \mathbb{P}(\hat{M} = \hat{m}|\mathcal{F}_1)\ \mathbb{P}((W^k(l), v^k) \in T^k_{[WV]_{\tilde{\delta}}}:\ f_B(l) = \hat{m}|\mathcal{F}_1) \\
&= \sum_{l=2}^{M'_k} \sum_{\hat{m}\neq 1} \mathbb{P}(\hat{M} = \hat{m}|\mathcal{F}_1) \sum_{\substack{\tilde{w}^k: \\ (\tilde{w}^k, v^k)\in T^k_{[WV]_{\tilde{\delta}}}}} \mathbb{P}(W^k(l) = \tilde{w}^k:\ f_B(l) = \hat{m}|\mathcal{F}_1) \\
&= \sum_{l=2}^{M'_k} \sum_{\hat{m}\neq 1} \mathbb{P}(\hat{M} = \hat{m}|\mathcal{F}_1) \sum_{\substack{\tilde{w}^k: \\ (\tilde{w}^k, v^k)\in T^k_{[WV]_{\tilde{\delta}}}}} \mathbb{P}(W^k(l) = \tilde{w}^k|\mathcal{F}_1)\ \frac{1}{e^{kR}} \qquad \text{(A.24)} \\
&= \sum_{l=2}^{M'_k} \sum_{\substack{\tilde{w}^k: \\ (\tilde{w}^k, v^k)\in T^k_{[WV]_{\tilde{\delta}}}}} \mathbb{P}(W^k(l) = \tilde{w}^k|\mathcal{F}_1)\ \frac{1}{e^{kR}}. \qquad \text{(A.25)}
\end{aligned}$$

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

$$\mathbb{P}(W^k(l) = \tilde{w}^k|\mathcal{F}_1) = \sum_{\mathcal{C}_{1,l}^- = c} \mathbb{P}(\mathcal{C}_{1,l}^- = c|\mathcal{F}_1)\mathbb{P}(W^k(l) = \tilde{w}^k|\mathcal{F}_1, \mathcal{C}_{1,l}^- = c). \tag{A.26}$$

The term in (A.26) can be upper bounded as follows:

$$\begin{aligned}
&\mathbb{P}(W^k(l) = \tilde{w}^k|\mathcal{F}_1,\ \mathcal{C}_{1,l}^- = c)\\
&= \mathbb{P}(W^k(l) = \tilde{w}^k|U^k = u^k, V^k = v^k,\ \mathcal{C}_{1,l}^- = c)\\
&\quad \frac{\mathbb{P}(W^k(1) = w^k|W^k(l) = \tilde{w}^k, U^k = u^k, V^k = v^k, \mathcal{C}_{1,l}^- = c)}{\mathbb{P}(W^k(1) = w^k|U^k = u^k, V^k = v^k, \mathcal{C}_{1,l}^- = c)}\\
&\quad \frac{\mathbb{P}(J = 1|W^k(1) = w^k, W^k(l) = \tilde{w}^k, U^k = u^k, V^k = v^k, \mathcal{C}_{1,l}^- = c)}{\mathbb{P}(J = 1|W^k(1) = w^k, U^k = u^k, V^k = v^k, \mathcal{C}_{1,l}^- = c)}\\
&\quad \frac{\mathbb{P}(f_B(J) = 1|J = 1, W^k(1) = w^k, W^k(l) = \tilde{w}^k, U^k = u^k, V^k = v^k,\ \mathcal{C}_{1,l}^- = c)}{\mathbb{P}(f_B(J) = 1|J = 1, W^k(1) = w^k, U^k = u^k, V^k = v^k, \mathcal{C}_{1,l}^- = c)}\\
&\quad \frac{\mathbb{P}(\mathcal{E}_{NE},\ \mathcal{E}_{CE}|f_B(J) = 1, J = 1, W^k(1) = w^k, W^k(l) = \tilde{w}^k, U^k = u^k, V^k = v^k,\ \mathcal{C}_{1,l}^- = c)}{\mathbb{P}(\mathcal{E}_{NE},\ \mathcal{E}_{CE}|f_B(J) = 1, J = 1, W^k(1) = w^k, U^k = u^k, V^k = v^k, \mathcal{C}_{1,l}^- = c)}.
\end{aligned} \tag{A.27}$$

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

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

and

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

Also, note that

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

Next, consider the term in (A.27). Let $N(u^k, \mathcal{C}_{1,l}^-) = |\{w^k(l') \in \mathcal{C}_{1,l}^- : l' \neq 1,\ l' \neq l,\ (w^k(l'), u^k) \in T_{[WU]_\delta}^k\}|$. Recall that if there are multiple sequences in codebook $\mathcal{C}$ that are jointly typical with the observed sequence U^k, then the encoder selects one of them uniformly at random. Also, note that given $\mathcal{F}_1$, $(w^k, u^k) \in T_{[WU]_\delta}^k$. Thus, if $(\tilde{w}^k, u^k) \in T_{[WU]_\delta}^k$, then

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

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

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

Hence, the term in (A.26) can be upper bounded as

$$\begin{aligned}
&\mathbb{P}(W^k(l)=\tilde{w}^k|\mathcal{F}_1) \\
&\leq \sum_{\mathcal{C}_{1,l}^-=c} \mathbb{P}(\mathcal{C}_{1,l}^-=c|\mathcal{F}_1)\ 2\ \mathbb{P}(W^k(l)=\tilde{w}^k|U^k=u^k, V^k=v^k,\ \mathcal{C}_{1,l}^-=c) \\
&= 2\ \mathbb{P}(W^k(l)=\tilde{w}^k|U^k=u^k, V^k=v^k) = 2\ \mathbb{P}(W^k(l)=\tilde{w}^k). \qquad \text{(A.28)}
\end{aligned}$$

Substituting (A.28) in (A.25), we obtain

$$\mathbb{P}(\mathcal{D}_0|\mathcal{F}_1) \leq \sum_{l=1}^{M_k'} \sum_{\substack{\tilde{w}^k: \\ (\tilde{w}^k, v^k) \in T_{[WV]_{\tilde{\delta}}}^k}} 2\ \mathbb{P}(W^k(l)=\tilde{w}^k)\ \frac{1}{e^{kR}}$$

$$= \sum_{l=1}^{M'_k} \sum_{\substack{\tilde{w}^k: \\ (\tilde{w}^k, v^k) \in T^k_{[WV]_{\tilde{\delta}}}}} 2 \cdot e^{-k(H(W)-O(\hat{\delta}))} \frac{1}{e^{kR}}$$

$$= 2\ M'_k\ e^{k(H(W|V)+\delta)}\ e^{-k(H(W)-O(\hat{\delta}))}\ \frac{1}{e^{kR}}$$

$$\leq e^{-k(R-I(U;W|V)-\delta_2^{(k)})}, \tag{A.29}$$

where $\delta_2^{(k)} := O(\delta) + \frac{\log(2)}{k}$.

For the case $I(U;W)+\mu \leq R$ (when binning is not done), the terms can be bounded similarly using (A.28) as follows.

$$\begin{aligned}
\mathbb{P}(\mathcal{D}_0|\mathcal{F}_1) &= \sum_{\hat{m}\neq 1} \mathbb{P}(\hat{M}=\hat{m}|\mathcal{F}_1)\ \mathbb{P}((W^k(\hat{m}), v^k) \in T^k_{[WV]_{\tilde{\delta}}}|\mathcal{F}_1) \\
&\leq \sum_{\hat{m}\neq 1} \mathbb{P}(\hat{M}=\hat{m}|\mathcal{F}_1) \sum_{\substack{\tilde{w}^k: \\ (\tilde{w}^k, v^k) \in T^k_{[WV]_{\tilde{\delta}}}}} 2\ \mathbb{P}(W^k(\hat{m})=\tilde{w}^k) \\
&\leq e^{-k(I(V;W)-\delta_2^{(k)})}.
\end{aligned} \tag{A.30}$$

Next, consider the event when there are no encoding or channel errors, i.e., $\mathcal{E}_{NE} \cap \mathcal{E}_{CE}^c$. For the case $I(U;W)+\mu > R$, the binning error event denoted by $\mathcal{E}_{BE}$ happens when a wrong codeword $W^k(l),\ l \neq J$, is retrieved from the bin with index M by the empirical entropy decoder such that $(W^k(l), V^k) \in T^k_{[WV]_\delta}$. Let $P_{\tilde{U}\tilde{V}\tilde{W}}$ denote the type of $P_{U^kV^kW^k(J)}$. Note that $P_{\tilde{U}\tilde{W}} \in \mathcal{T}^k_{[UW]_\delta}$ when $\mathcal{E}_{NE}$ holds. If $H(\tilde{W}|\tilde{V}) < H(W|V)$, then in the bin with index M, there exists a codeword with empirical entropy strictly less than $H(W|V)$. Hence, the decoded codeword $\hat{W}^k$ is such that $(\hat{W}^k, V^k) \notin T^k_{[WV]_{\tilde{\delta}}}$(asymptotically) since $(\hat{W}^k,\ V^k) \in T^k_{[WV]_{\tilde{\delta}}}$ necessarily implies that $H_e(\hat{W}^k|V^k) \geq H(W|V) - O(\delta)$ (for δ small enough). Consequently, a type II error can happen under the event $\mathcal{E}_{BE}$ only when $H(\tilde{W}|\tilde{V}) \geq H(W|V) - O(\delta)$. The probability of the event $\mathcal{E}_{BE}$ can be upper bounded under this condition as follows:

$$\begin{aligned}
&\mathbb{P}(\mathcal{E}_{BE}|\mathcal{F}_2) \\
&\leq \mathbb{P}\left(\exists\ l \neq 1,\ l \in [M'_k]:\ f_B(l)=1 \text{ and } (W^k(l), v^k) \in T^k_{[WV]_{\tilde{\delta}}}|\mathcal{F}_2\right)
\end{aligned}$$

$$\leq \sum_{l=2}^{M'_k} \mathbb{P}\left((W^k(l), v^k) \in T^k_{[WV]_{\tilde{\delta}}} | \mathcal{F}_2\right) \; \mathbb{P}\left(f_B(l) = 1 | \mathcal{F}_2, (W^k(l), v^k) \in T^k_{[WV]_{\tilde{\delta}}}\right)$$

$$= \sum_{l=2}^{M'_k} \mathbb{P}\left((W^k(l), v^k) \in T^k_{[WV]_{\tilde{\delta}}} | \mathcal{F}_2\right) \; e^{-kR}$$

$$\leq \sum_{l=2}^{M'_k} \sum_{\substack{\tilde{w}^k: \\ (\tilde{w}^k, v^k) \in T^k_{[WV]_{\tilde{\delta}}}}} 2 \; \mathbb{P}(W^k(l) = \tilde{w}^k) \; e^{-kR} \tag{A.31}$$

$$= e^{-k(R - I(U;W|V) - \delta_2^{(k)})}. \tag{A.32}$$

In (A.31), we used the fact that

$$\mathbb{P}\left(W^k(l) = \tilde{w}^k | \mathcal{F}_2\right) \leq 2 \; \mathbb{P}(W^k(l) = \tilde{w}^k), \tag{A.33}$$

which follows in a similar way as (A.28). Also, note that, by definition, $\mathbb{P}(\mathcal{D}_0 | \mathcal{F}_{21}) = 1$.

We proceed to analyze the R.H.S of (A.12) which upper bounds the type II error probability. Towards this end, we first focus on the the case when $\mathcal{E}_{NE}$ holds. From (A.14), it follows that

$$\sum_{(u^k, v^k) \in \mathcal{U}^k \times \mathcal{V}^k} \mathbb{P}(U^k = u^k, V^k = v^k | H = 1) \; \mathbb{P}(\mathcal{D}_0 | U^k = u^k, V^k = v^k, \mathcal{E}_{NE}) \tag{A.34}$$

$$= \sum_{(u^k, v^k) \in \mathcal{U}^k \times \mathcal{V}^k} \mathbb{P}(U^k = u^k, V^k = v^k | H = 1)$$

$$\mathbb{P}(\mathcal{D}_0 | U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, \mathcal{E}_{NE}). \tag{A.35}$$

Rewriting the summation in (A.35) as the sum over the types and sequences within a type, we obtain

$$\mathbb{P}(\mathcal{D}_0 | \; \mathcal{E}_{NE}, H = 1)$$

$$= \sum_{\substack{P_{\tilde{U}\tilde{V}\tilde{W}} \\ \in \mathcal{T}^k_{\mathcal{U}\mathcal{V}\mathcal{W}}}} \sum_{\substack{(u^k, v^k, w^k) \\ \in T_{P_{\tilde{U}\tilde{V}\tilde{W}}}}} \Big[\mathbb{P}(U^k = u^k, V^k = v^k | H = 1) \; \mathbb{P}(\mathcal{D}_0 | \mathcal{F})$$

$$\mathbb{P}(W^k(1) = w^k | U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, \mathcal{E}_{NE})\Big]. \tag{A.36}$$

We also have

$$\begin{aligned}
&\mathbb{P}(U^k = u^k, V^k = v^k | H = 1)\ \mathbb{P}(W^k(1) = w^k | U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, \mathcal{E}_{NE}) \\
&= \left[\prod_{i=1}^{k} Q_{UV}(u_i, v_i)\right] \mathbb{P}(W^k(1) = w^k | U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, \mathcal{E}_{NE}) \\
&\leq \left[\prod_{i=1}^{k} Q_{UV}(u_i, v_i)\right] \frac{1}{|T_{P_{\tilde{W}|\tilde{U}}}|} \leq e^{-k(H(\tilde{U}\tilde{V}) + D(P_{\tilde{U}\tilde{V}}||Q_{UV}) + H(\tilde{W}|\tilde{U}) - \frac{1}{k}|\mathcal{U}||\mathcal{W}|\log(k+1))},
\end{aligned} \tag{A.37}$$

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

With (A.23), (A.29), (A.30), (A.32) and (A.37), we have the necessary machinery to analyze (A.36). First, consider that the event $\mathcal{E}_{NE} \cap \mathcal{E}_{CE}^c \cap \mathcal{E}_{BE}^c$ holds. In this case,

$$\begin{aligned}
\mathbb{P}(\mathcal{D}_0|\mathcal{F}_{22}) &= \mathbb{P}(\mathcal{D}_0 | U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, W^k(1) = w^k, \mathcal{E}_{NE}, \mathcal{E}_{CE}^c, \mathcal{E}_{BE}^c) \\
&= \begin{cases} 1, \text{ if } P_{u^k w^k} \in T^k_{[UW]_\delta} \text{ and } P_{v^k w^k} \in T^k_{[VW]_\delta}, \\ 0, \text{ otherwise.} \end{cases}
\end{aligned} \tag{A.38}$$

Thus, the following terms in (A.36) can be simplified (for sufficiently large k) as follows:

$$\begin{aligned}
&\sum_{\substack{P_{\tilde{U}\tilde{V}\tilde{W}} \\ \in \mathcal{T}^k_{\mathcal{U}\mathcal{V}\mathcal{W}}}} \sum_{\substack{(u^k, v^k, w^k) \\ \in T_{P_{\tilde{U}\tilde{V}\tilde{W}}}}} \Big[\mathbb{P}(U^k = u^k, V^k = v^k | H = 1)\ \mathbb{P}(\mathcal{E}_{CE}^c|\mathcal{F})\ \mathbb{P}(\mathcal{E}_{BE}^c|\mathcal{F}_2)\ \mathbb{P}(\mathcal{D}_0|\mathcal{F}_{22}) \\
&\qquad\qquad \mathbb{P}(W^k(1) = w^k | U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, \mathcal{E}_{NE})\Big] \\
&\leq \sum_{\substack{P_{\tilde{U}\tilde{V}\tilde{W}} \\ \in \mathcal{T}^k_{\mathcal{U}\mathcal{V}\mathcal{W}}}} \sum_{\substack{(u^k, v^k, w^k) \\ \in T_{P_{\tilde{U}\tilde{V}\tilde{W}}}}} \Big[\mathbb{P}(U^k = u^k, V^k = v^k | H = 1)\ \mathbb{P}(\mathcal{D}_0|\mathcal{F}_{22}) \\
&\qquad\qquad \mathbb{P}(W^k(1) = w^k | U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, \mathcal{E}_{NE})\Big]
\end{aligned}$$

$$\begin{aligned}
&\leq (k+1)^{|\mathcal{U}||\mathcal{V}||\mathcal{W}|} \max_{\substack{P_{\tilde{U}\tilde{V}\tilde{W}} \in \\ \hat{\mathcal{T}}_1^{(k)}(P_{UW}, P_{VW})}} e^{kH(\tilde{U}\tilde{V}\tilde{W})} e^{-k(H(\tilde{U}\tilde{V}) + D(P_{\tilde{U}\tilde{V}}||Q_{UV}) + H(\tilde{W}|\tilde{U}) - \frac{1}{k}|\mathcal{U}||\mathcal{W}|\log(k+1))} \\
&= e^{-k\tilde{E}_{1k}},
\end{aligned} \tag{A.39}$$

where,

$$\hat{\mathcal{T}}_1^{(k)}(P_{UW}, P_{VW}) := \{P_{\tilde{U}\tilde{V}\tilde{W}} : P_{\tilde{U}\tilde{W}} \in T^k_{[UW]_\delta} \text{ and } P_{\tilde{V}\tilde{W}} \in T^k_{[VW]_{\tilde{\delta}}}\}, \tag{A.40}$$

$$\text{and } \tilde{E}_{1k} := \min_{\substack{P_{\tilde{U}\tilde{V}\tilde{W}} \in \\ \hat{\mathcal{T}}_1^{(k)}(P_{UW}, P_{VW})}} H(\tilde{U}\tilde{V}) + D(P_{\tilde{U}\tilde{V}}||Q_{UV}) + H(\tilde{W}|\tilde{U}) - H(\tilde{U}\tilde{V}\tilde{W}) - \frac{1}{k}|\mathcal{U}||\mathcal{V}||\mathcal{W}|\log(k+1) - \frac{1}{k}|\mathcal{U}||\mathcal{W}|\log(k+1). \tag{A.41}$$

To obtain (A.39), we used (A.37) and (A.38). Note that for δ small enough,

$$\begin{aligned}\tilde{E}_{1k} &\overset{(k)}{\geq} \min_{\substack{P_{\tilde{U}\tilde{V}\tilde{W}} \in \\ \mathcal{T}_1(P_{UW}, P_{VW})}} \sum P_{\tilde{U}\tilde{V}\tilde{W}} \log\left(\frac{P_{\tilde{U}\tilde{V}}}{Q_{UV}} \frac{1}{P_{\tilde{U}\tilde{V}}} \frac{P_{\tilde{U}}}{P_{\tilde{U}\tilde{W}}} P_{\tilde{U}\tilde{V}\tilde{W}}\right) - O(\delta) \\ &= \min_{\substack{P_{\tilde{U}\tilde{V}\tilde{W}} \in \\ \mathcal{T}_1(P_{UW}, P_{VW})}} D(P_{\tilde{U}\tilde{V}\tilde{W}}||Q_{UVW}) - O(\delta) = E_1(P_{W|U}) - O(\delta), \end{aligned} \tag{A.42}$$

Next, consider the terms corresponding to the event $\mathcal{E}_{NE} \cap \mathcal{E}_{CE}^c \cap \mathcal{E}_{BE}$ in (A.36). Note that given the event $\mathcal{F}_{21} = \{U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, W^k(1) = w^k, \mathcal{E}_{NE}, \mathcal{E}_{CE}^c, \mathcal{E}_{BE}\}$ occurs, $P_{u^k w^k} \in T^k_{[UW]_\delta}$. Also, $\mathcal{D}_0$ can happen only if $H_e(w^k|v^k) \geq H(W|V) - O(\tilde{\delta})$, and $P_{v^k} \in T^k_{[V]_{\tilde{\delta}}}$. Using these facts to simplify the terms corresponding to the event $\mathcal{E}_{NE} \cap \mathcal{E}_{CE}^c \cap \mathcal{E}_{BE}$ in (A.36), we obtain

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

$$\begin{aligned}\leq &\max_{\substack{P_{\tilde{U}\tilde{V}\tilde{W}} \in \\ \hat{\mathcal{T}}_2^{(k)}(P_{UW}, P_V)}} e^{kH(\tilde{U}\tilde{V}\tilde{W})} e^{-k\left(H(\tilde{U}\tilde{V}) + D(P_{\tilde{U}\tilde{V}}||Q_{UV}) + H(\tilde{W}|\tilde{U}) + R - I(U;W|V) - O(\delta)\right)} \\ &\qquad e^{(|\mathcal{U}||\mathcal{V}||\mathcal{W}|\log(k+1) + |\mathcal{U}||\mathcal{W}|\log(k+1))} \\ = &\ e^{-k\tilde{E}_{2k}}, \end{aligned} \tag{A.43}$$

where,

$$\hat{\mathcal{T}}_2^{(k)}(P_{UW}, P_V) := \Big\{ P_{\tilde{U}\tilde{V}\tilde{W}} : P_{\tilde{U}\tilde{W}} \in T^k_{[UW]_\delta}, P_{\tilde{V}} \in T^k_{[V]_{\tilde{\delta}}} \text{ and } H(\tilde{W}|\tilde{V}) \geq H(W|V) - O(\delta) \Big\}, \tag{A.44}$$

and

$$\begin{aligned}\tilde{E}_{2k} :=& \min_{\substack{P_{\tilde{U}\tilde{V}\tilde{W}} \in \\ \mathcal{T}_2(P_{UW}, P_V)}} H(\tilde{U}\tilde{V}) + D(P_{\tilde{U}\tilde{V}}||Q_{UV}) + H(\tilde{W}|\tilde{U}) + R - I(U;W|V) \\ &- \frac{1}{k}|\mathcal{U}||\mathcal{V}||\mathcal{W}|\log(k+1) - \frac{1}{k}|\mathcal{U}||\mathcal{W}|\log(k+1) - O(\delta) \\ \overset{(k)}{\geq}& E_2(P_{W|U}, P_{SX}, R) - O(\delta). \end{aligned} \tag{A.45}$$

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

Next, consider that the event $\mathcal{E}_{NE} \cap \mathcal{E}_{CE}$ holds. As in the case above, note that given $\mathcal{F}_1 = \{U^k = u^k, V^k = v^k, J = 1, f_B(J) = 1, W^k(1) = w^k,\ \mathcal{E}_{NE},\ \mathcal{E}_{CE}\}$, $P_{u^k w^k} \in T^k_{[UW]_\delta}$ and $\mathcal{D}_0$ occurs only if $P_{v^k} \in T^k_{[V]_{\tilde{\delta}}}$. Using these facts and eqns. (A.29), (A.30) and (A.23), it can be shown that the terms corresponding to this event in (A.36) results in the factor $E_3(P_{W|U}, P_{SX}, R, \tau) - O(\delta)$ in the error-exponent.

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

$$\begin{aligned}\mathbb{P}(\mathcal{D}_0|\ U^k = u^k, V^k = v^k,\ \mathcal{E}_{NE}^c) =& \mathbb{P}(\mathcal{E}_{CE}|\ U^k = u^k, V^k = v^k,\ \mathcal{E}_{NE}^c) \\ & \mathbb{P}(\mathcal{D}_0|\ U^k = u^k, V^k = v^k,\ \mathcal{E}_{NE}^c \cap\ \mathcal{E}_{CE}). \end{aligned} \tag{A.46}$$

It follows from Borade et al.'s coding scheme [22] that asymptotically,

$$\mathbb{P}(\mathcal{E}_{CE}|\ U^k = u^k, V^k = v^k,\ \mathcal{E}_{NE}^c) \leq e^{-n(E_m(P_{SX}) - O(\delta))} = e^{-k\tau(E_m(P_{SX}) - O(\delta))}. \tag{A.47}$$

When binning is performed at the encoder, $\mathcal{D}_0$ occurs only if there exists a sequence $\hat{W}^k$ in the bin $\hat{M} \neq 0$ such that $(\hat{W}^k, V^k) \in T^k_{[WV]_{\tilde{\delta}}}$. Also, recalling that the encoder sends the error message $M = 0$ independent of the source codebook $\mathcal{C}$, it can be shown

using standard arguments that for such $v^k \in T^k_{[V]_{\tilde{\delta}}}$,

$$\mathbb{P}(\mathcal{D}_0|\ U^k = u^k, V^k = v^k,\ \mathcal{E}_{NE}^c \cap\ \mathcal{E}_{CE}) \leq e^{-k(R-I(U;W|V)-O(\delta))}. \tag{A.48}$$

Thus, from (A.46), (A.47) and (A.48), we obtain (asymptotically) that,

$$\begin{aligned}&\sum_{u^k,v^k} \mathbb{P}(U^k = u^k, V^k = v^k|H=1)\ \mathbb{P}(\mathcal{D}_0|\ U^k = u^k, V^k = v^k,\ \mathcal{E}_{NE}^c \cap\ \mathcal{E}_{CE})\\ &\leq e^{-k(R-I(U;W|V)+D(P_V||Q_V)+\tau E_m(P_{SX})-O(\delta))}.\end{aligned} \tag{A.49}$$

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

$$\begin{aligned}&\sum_{u^k,v^k} \mathbb{P}(U^k = u^k, V^k = v^k|H=1)\ \mathbb{P}(\mathcal{D}_0|\ U^k = u^k, V^k = v^k,\ \mathcal{E}_{NE}^c \cap\ \mathcal{E}_{CE})\\ &\leq e^{-k(I(V;W)+D(P_V||Q_V)+\tau E_m(P_{SX})-O(\delta))}.\end{aligned} \tag{A.50}$$

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

A.2 Proof of Theorem 2.6

We only give a sketch of the proof as the intermediate steps follow similarly to those in the proof of Theorem 2.2. We will use the random coding method combined with the expurgation technique as explained in the proof of Theorem 2.2, to guarantee the existence of at least one deterministic codebook that achieves the type I error probability and error-exponent claimed in Theorem 2.6. For brevity, we will denote information theoretic quantities like $I_{\hat{P}}(U, S; \bar{W})$, $T^n_{[\hat{P}_{US\bar{W}}]_\delta}$, etc., that are computed with respect to joint distribution $\hat{P}_{UVS\bar{W}X'XY}$ given below in (A.51) by $I(U, S; \bar{W})$, $T^n_{[US\bar{W}]_\delta}$, etc.

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

$$\hat{P}_{UVS\bar{W}X'XY}(P_S, P_{\bar{W}|US}, P_{X'|US}, P_{X|US\bar{W}}) := P_{UV}P_S P_{\bar{W}|US}P_{X'|US}P_{X|US\bar{W}}P_{Y|X}. \tag{A.51}$$

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

$$\hat{P}_{\bar{W}} = \sum_{(u,s)\in\mathcal{U}\times\mathcal{S}} P_U(u)P_S(s)P_{\bar{W}|US}(\bar{w}|u,s).$$

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

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

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

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

Analysis of Type I error:

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

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

By the weak law of large numbers, $\tilde{\mathcal{E}}_{TE}$ tends to 0 asymptotically with n. The covering lemma guarantees that $\tilde{\mathcal{E}}_{EE} \cap \tilde{\mathcal{E}}^c_{TE}$ tends to 0 doubly exponentially if $\mu = O(\delta)$ is chosen appropriately. Given $\tilde{\mathcal{E}}^c_{EE} \cap \tilde{\mathcal{E}}^c_{TE}$ holds, it follows from the Markov lemma and the weak law of large numbers, respectively, that $\mathbb{P}(\tilde{\mathcal{E}}_{ME})$ and $\mathbb{P}(\tilde{\mathcal{E}}_{CE})$ tends to zero asymptotically. Next, we consider the probability of the event $\tilde{\mathcal{E}}_{DE}$. Given that $\tilde{\mathcal{E}}^c_{CE} \cap \tilde{\mathcal{E}}^c_{ME} \cap \tilde{\mathcal{E}}^c_{EE} \cap \tilde{\mathcal{E}}^c_{TE}$ holds, note that $H_e(\bar{W}^n(J)|V^n, S^n, Y^n) \overset{(n)}{\geq} H(\bar{W}|V, S, Y) - O(\delta)$. Hence, similarly to (A.5) in Appendix A.1, it can be shown that

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

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

Analysis of Type II error: The analysis of the error-exponent is very similar to that of the SHTCC scheme given in Appendix A.1. Hence, only a sketch of the proof is provided, with the differences from the proof of the SHTCC scheme highlighted.

Let

$$\bar{\mathcal{E}}_0 := \{(U^n, S^n) \notin T^n_{[US]_{\delta'}}\}. \tag{A.52}$$

Then, the type 2 error probability can be written as

$$\begin{aligned}
&\beta\left(n, n, f^{(n,n)}, g^{(n,n)}\right) \\
&\leq \sum_{(u^n,v^n)\in\mathcal{U}^n\times\mathcal{V}^n} \mathbb{P}(U^n=u^n, V^n=v^n|H=1)\Big[\mathbb{P}(\tilde{\mathcal{E}}_{EE}\cap\bar{\mathcal{E}}_0^c|U^n=u^n, V^n=v^n) \\
&\qquad + \mathbb{P}(\mathcal{D}_0|U^n=u^n, V^n=v^n, \tilde{\mathcal{E}}_{NE}) + \mathbb{P}(\mathcal{D}_0|U^n=u^n, V^n=v^n, \bar{\mathcal{E}}_0)\Big], \qquad \text{(A.53)}
\end{aligned}$$

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

$$\begin{aligned}
\tilde{\mathcal{E}}_{T2} = \Big\{\exists\ l\in\Big[e^{n(I(U,S;\bar{W})+\mu)}\Big],\ l\neq J,\ \bar{W}^n(l)\in T^n_{[\bar{W}]_{\tilde{\delta}}}, \\
(V^n, \bar{W}^n(l), S^n, Y^n)\in T^n_{[VS\bar{W}Y]_{\tilde{\delta}}}\Big\}, \qquad \text{(A.54)}
\end{aligned}$$

$$\tilde{\mathcal{F}} = \{U^n=u^n, V^n=v^n, J=1, \bar{W}^n(1)=\bar{w}^n, S^n=s^n, Y^n=y^n, \tilde{\mathcal{E}}_{NE}\}, \qquad \text{(A.55)}$$

$$\tilde{\mathcal{F}}_1 = \{U^n=u^n, V^n=v^n, J=1, \bar{W}^n(1)=\bar{w}^n, S^n=s^n, Y^n=y^n, \tilde{\mathcal{E}}_{NE}, \tilde{\mathcal{E}}_{T2}^c\}, \qquad \text{(A.56)}$$

$$\tilde{\mathcal{F}}_2 = \{U^n=u^n, V^n=v^n, J=1, \bar{W}^n(1)=\bar{w}^n, S^n=s^n, Y^n=y^n, \tilde{\mathcal{E}}_{NE}, \tilde{\mathcal{E}}_{T2}\}. \qquad \text{(A.57)}$$

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

$$\begin{aligned}
&\mathbb{P}(\mathcal{D}_0|\ U^n=u^n, V^n=v^n, \tilde{\mathcal{E}}_{NE}) \\
&= \sum_{j=1}^{e^{n(I(U,S;\bar{W})+\mu)}} \mathbb{P}(J=j|\ U^n=u^n, V^n=v^n, \tilde{\mathcal{E}}_{NE}) \\
&\qquad\qquad \mathbb{P}(\mathcal{D}_0|U^n=u^n, V^n=v^n, J=1,\ \tilde{\mathcal{E}}_{NE}) \\
&= \mathbb{P}(\mathcal{D}_0|U^n=u^n, V^n=v^n, J=1,\ \tilde{\mathcal{E}}_{NE}) \\
&= \sum_{\substack{(\bar{w}^n,s^n,y^n)\\ \in \bar{\mathcal{W}}^n\times\mathcal{S}^n\times\mathcal{Y}^n}} \mathbb{P}(\bar{W}^n(1)=\bar{w}^n, S^n=s^n, Y^n=y^n|U^n=u^n, V^n=v^n, J=1,\ \tilde{\mathcal{E}}_{NE}) \\
&\qquad \mathbb{P}(\mathcal{D}_0|U^n=u^n, V^n=v^n, J=1, \bar{W}^n(1)=\bar{w}^n, S^n=s^n, Y^n=y^n,\ \tilde{\mathcal{E}}_{NE}) \\
&= \sum_{\substack{(\bar{w}^n,s^n,y^n)\\ \in \bar{\mathcal{W}}^n\times\mathcal{S}^n\times\mathcal{Y}^n}} \mathbb{P}(\bar{W}^n(1)=\bar{w}^n, S^n=s^n, Y^n=y^n|U^n=u^n, V^n=v^n, J=1,\ \tilde{\mathcal{E}}_{NE}) \\
&\qquad\qquad \mathbb{P}(\mathcal{D}_0|\ \tilde{\mathcal{F}}). \qquad \text{(A.58)}
\end{aligned}$$

The last term in (A.58) can be upper bounded using the events in (A.55)-(A.57) as follows.

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

We next analyze the R.H.S of (A.53), which upper bounds the type 2 error probability. We can write,

$$\mathbb{P}(\mathcal{D}_0|\tilde{\mathcal{F}}_1) = \begin{cases} 1, \text{ if } P_{u^n s^n \bar{w}^n} \in T^n_{[US\bar{W}]_\delta} \text{ and } P_{v^n \bar{w}^n s^n y^n} \in T^k_{[VS\bar{W}Y]_{\tilde{\delta}}}, \\ 0, \text{ otherwise.} \end{cases} \tag{A.59}$$

Hence, the terms corresponding to the event $\tilde{\mathcal{F}}_1$ in (A.53) can be upper bounded (in the limit $\delta, \tilde{\delta} \to 0$) as

$$\begin{aligned}
&\sum_{\substack{(u^n,v^n,\bar{w}^n,s^n,y^n) \\ \in\, \mathcal{U}^n\times\mathcal{V}^n\times\bar{\mathcal{W}}^n\times\mathcal{S}^n\times\mathcal{Y}^n}} \Big[\mathbb{P}(U^n = u^n, V^n = v^n|H=1)\ \mathbb{P}(\mathcal{D}_0|\tilde{\mathcal{F}}_1) \\
&\qquad\qquad \mathbb{P}(\bar{W}^n(1) = \bar{w}^n, S^n = s^n, Y^n = y^n|U^n = u^n, V^n = v^n, J=1,\ \tilde{\mathcal{E}}_{NE})\Big] \\
&\leq \sum_{\substack{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} \\ \in \mathcal{T}^n_{\mathcal{U}\mathcal{V}\bar{\mathcal{W}}\mathcal{S}\mathcal{Y}}}} \sum_{\substack{(u^n,v^n,\bar{w}^n,s^n,y^n) \\ \in T_{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}}}}} \Big[\mathbb{P}(U^n = u^n, V^n = v^n|H=1)\ \mathbb{P}(\mathcal{D}_0|\tilde{\mathcal{F}}_1) \\
&\qquad\qquad \mathbb{P}(S^n = s^n, \bar{W}^n(1) = \bar{w}^n|U^n = u^n, J=1,\ \tilde{\mathcal{E}}_{NE}) \\
&\qquad\qquad \mathbb{P}(Y^n = y^n|U^n = u^n, S^n = s^n, J=1, \bar{W}^n(1) = \bar{w}^n,\ \tilde{\mathcal{E}}_{NE})\Big] \\
&\leq \sum_{\substack{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} \\ \in \mathcal{T}^n_{\mathcal{U}\mathcal{V}\bar{\mathcal{W}}\mathcal{S}\mathcal{Y}}}} \sum_{\substack{(u^n,v^n,\bar{w}^n,s^n,y^n) \\ \in T_{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}}}}} \Big[\mathbb{P}(\mathcal{D}_0|\tilde{\mathcal{F}}_1)\ e^{-n\left(H(\tilde{U}\tilde{V}) + D(P_{\tilde{U}\tilde{V}}||Q_{UV})\right)} \\
&\qquad e^{-n\left(H(\tilde{S}\tilde{W}|\tilde{U}) - \frac{1}{n}|\mathcal{U}||\bar{\mathcal{W}}||\mathcal{S}|\log(n+1)\right)}\ e^{-n\Big(H(\tilde{Y}|\tilde{U}\tilde{S}\tilde{W}) + D(P_{\tilde{Y}|\tilde{U}\tilde{S}\tilde{W}}||\hat{P}_{Y|US\bar{W}}|P_{\tilde{U}\tilde{S}\tilde{W}})\Big)}\Big] \\
&\leq \max_{\substack{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} \in \\ \mathcal{T}_1'^{(n)}(\hat{P}_{US\bar{W}}, \hat{P}_{VS\bar{W}Y})}} \Big[e^{-n\left(H(\tilde{U}\tilde{V}) + D(P_{\tilde{U}\tilde{V}}||Q_{UV})\right)}\ e^{-n\left(H(\tilde{S}\tilde{W}|\tilde{U}) - \frac{1}{n}|\mathcal{U}||\bar{\mathcal{W}}||\mathcal{S}|\log(n+1)\right)} \\
&\qquad e^{-n\Big(H(\tilde{Y}|\tilde{U}\tilde{S}\tilde{W}) + D(P_{\tilde{Y}|\tilde{U}\tilde{S}\tilde{W}}||\hat{P}_{Y|US\bar{W}}|P_{\tilde{U}\tilde{S}\tilde{W}})\Big)}\, e^{n\left(H(\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}) - \frac{1}{n}||\mathcal{U}||\mathcal{V}||\bar{\mathcal{W}}||\mathcal{S}||\mathcal{Y}|\log(n+1)\right)}\Big] \\
&= e^{-nE^*_{1n}},
\end{aligned} \tag{A.60}$$

where,

$$\mathcal{T}_1'^{(n)}(\hat{P}_{US\bar{W}}, \hat{P}_{VS\bar{W}Y}) := \{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} \in \mathcal{T}_{\mathcal{U}\mathcal{V}\mathcal{S}\mathcal{W}\mathcal{Y}} : P_{\tilde{U}\tilde{S}\tilde{W}} \in T^n_{[US\bar{W}]_\delta},\ P_{\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} \in T^n_{[VS\bar{W}Y]_{\tilde{\delta}}}\},$$

and

$$\begin{aligned}
&E_{1n}^* \\
&:= \min_{\substack{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} \in \\ \mathcal{T}_1'(\hat{P}_{US\bar{W}}, \hat{P}_{VS\bar{W}Y})}} \Bigg[H(\tilde{U}\tilde{V}) + D(P_{\tilde{U}\tilde{V}} || Q_{UV}) + H(\tilde{S}\tilde{W}|\tilde{U}) + H(\tilde{Y}|\tilde{U}\tilde{S}\tilde{W}) - H(\tilde{U}\tilde{V}\tilde{W}\tilde{S}\tilde{Y}) \\
&\qquad + D(P_{\tilde{Y}|\tilde{U}\tilde{S}\tilde{W}} || \hat{P}_{Y|US\bar{W}} | P_{\tilde{U}\tilde{S}\tilde{W}}) - \frac{1}{n}(|\mathcal{U}||\bar{\mathcal{W}}| + |\mathcal{U}||\mathcal{V}||\bar{\mathcal{W}}||\mathcal{S}||\mathcal{Y}|)\log(n+1) \Bigg] \\
&\overset{(n)}{\geq} \min_{\substack{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} \in \\ \mathcal{T}_1'(\hat{P}_{US\bar{W}}, \hat{P}_{VS\bar{W}Y})}} \Bigg[\sum_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} \log \left(\frac{1}{P_{\tilde{U}\tilde{V}}} \frac{P_{\tilde{U}\tilde{V}}}{Q_{UV}} \frac{P_{\tilde{U}}}{P_{\tilde{U}\tilde{S}\tilde{W}}} \frac{1}{P_{\tilde{Y}|\tilde{U}\tilde{S}\tilde{W}}} \frac{P_{\tilde{Y}|\tilde{U}\tilde{S}\tilde{W}}}{\hat{P}_{Y|US\bar{W}}} P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} \right) \\
&\qquad - O(\delta) \Bigg] \\
&= \min_{\substack{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} \in \\ \mathcal{T}_1'(\hat{P}_{US\bar{W}}, \hat{P}_{VS\bar{W}Y})}} \Bigg[D(P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} | Q_{UV} P_{\tilde{S}\tilde{W}|\tilde{U}} \hat{P}_{Y|US\bar{W}}) - O(\delta) \Bigg] \\
&= E_1'(P_S, P_{\bar{W}|US}, P_{X|US\bar{W}}) - O(\delta). \qquad \text{(A.61)}
\end{aligned}$$

Here, (A.61) follows from the fact that $P_{\tilde{S}\tilde{W}|\tilde{U}} \to P_{S\bar{W}|U}$ given $\tilde{\mathcal{E}}_{NE}$, as $\delta \to 0$.

Next, consider the terms corresponding to the event $\tilde{\mathcal{F}}_2$ in (A.53). Given $\tilde{\mathcal{F}}_2$, $P_{\tilde{U}\tilde{S}\tilde{W}} \in T^n_{[US\bar{W}]_\delta}$ and $\mathcal{D}_0$ occurs only if $(V^n, S^n, Y^n) \in T^n_{[VSY]_{\delta''}}$, $\delta'' = |\bar{\mathcal{W}}|\tilde{\delta}$, and $H(\tilde{W}|\tilde{V}, \tilde{S}, \tilde{Y}) \geq H(\bar{W}|V, S, Y) - O(\tilde{\delta})$. Thus, we have,

$$\begin{aligned}
&\sum_{\substack{(u^n, v^n, \bar{w}^n, s^n, y^n) \\ \in \mathcal{U}^n \times \mathcal{V}^n \times \bar{\mathcal{W}}^n \times \mathcal{S}^n \times \mathcal{Y}^n}} \Big[\mathbb{P}(U^n = u^n, V^n = v^n | H = 1)\ \mathbb{P}(\mathcal{D}_0 | \tilde{\mathcal{F}}_2)\ \mathbb{P}(\tilde{\mathcal{E}}_{T2} | \tilde{\mathcal{F}}) \\
&\qquad \mathbb{P}(\bar{W}^n(1) = \bar{w}^n, S^n = s^n, Y^n = y^n | U^n = u^n, V^n = v^n, J = 1,\ \tilde{\mathcal{E}}_{NE}) \Big] \\
&\leq \sum_{\substack{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} \in \\ \mathcal{T}^n(\mathcal{U}\times\mathcal{V}\times\bar{\mathcal{W}}\times\mathcal{S}\times\mathcal{Y})}} \sum_{\substack{(u^n, v^n, \bar{w}^n, s^n, y^n) \\ \in T_{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}}}}} \Big[\mathbb{P}(U^n = u^n, V^n = v^n | H = 1)\ \mathbb{P}(\mathcal{D}_0 | \tilde{\mathcal{F}}_2)\ \mathbb{P}(\tilde{\mathcal{E}}_{T2} | \tilde{\mathcal{F}}) \\
&\qquad \mathbb{P}(S^n = s^n, \bar{W}^n(1) = \bar{w}^n | U^n = u^n, J = 1,\ \tilde{\mathcal{E}}_{NE}) \\
&\qquad \mathbb{P}(Y^n = y^n | U^n = u^n, S^n = s^n, J = 1, \bar{W}^n(1) = \bar{w}^n,\ \tilde{\mathcal{E}}_{NE}) \Big]
\end{aligned}$$

$$
\begin{aligned}
&\leq \sum_{\substack{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}}\in \\ \mathcal{T}^n(\mathcal{U}\times\mathcal{V}\times\mathcal{W}\times\mathcal{S}\times\mathcal{Y})}} \sum_{\substack{(u^n,v^n,\bar{w}^n,s^n,y^n) \\ \in T_{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}}}}} \Big[e^{-n\left(H(\tilde{U}\tilde{V})+D(P_{\tilde{U}\tilde{V}}||Q_{UV})\right)} \mathbb{P}(\mathcal{D}_0|\tilde{\mathcal{F}}_2)\cdot 2 \\
&\qquad e^{-n\left(I(\bar{W};V,S,Y)-I(U,S;\bar{W})-O(\delta)\right)} e^{-n\left(H(\tilde{S}\tilde{W}|\tilde{U})-\frac{1}{n}|\mathcal{U}||\bar{\mathcal{W}}||\mathcal{S}|\log(n+1)\right)} \\
&\qquad e^{-n\left(H(\tilde{Y}|\tilde{U}\tilde{S}\tilde{W})+D(P_{\tilde{Y}|\tilde{U}\tilde{S}\tilde{W}}||\hat{P}_{Y|USW}|P_{\tilde{U}\tilde{S}\tilde{W}})\right)} \Big] \qquad \text{(A.62)} \\
&\leq \max_{\substack{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}}\in \\ \mathcal{T}_2'^{(n)}(\hat{P}_{UW},\hat{P}_{VSWY})}} \Big[e^{-n\left(H(\tilde{U}\tilde{V})+D(P_{\tilde{U}\tilde{V}}||Q_{UV})\right)}\ e^{-n\left(H(\tilde{S}\tilde{W}|\tilde{U})-\frac{1}{n}|\mathcal{U}||\bar{\mathcal{W}}||\mathcal{S}|\log(n+1)\right)} \\
&\quad e^{-n\left(I(\bar{W};V,S,Y)-I(U,S;\bar{W})-O(\delta)-\frac{1}{n}\right)}\ e^{-n\left(H(\tilde{Y}|\tilde{U}\tilde{S}\tilde{W})+D(P_{\tilde{Y}|\tilde{U}\tilde{S}\tilde{W}}||\hat{P}_{Y|USW}|P_{\tilde{U}\tilde{S}\tilde{W}})\right)} \\
&\quad e^{n\left(H(\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y})-\frac{1}{n}|\mathcal{U}||\mathcal{V}||\bar{\mathcal{W}}||\mathcal{S}||\mathcal{Y}|\log(n+1)\right)} \Big] \\
&= e^{-nE_{2n}^*}, \qquad \text{(A.63)}
\end{aligned}
$$

where,

$$
\begin{aligned}
&\mathcal{T}_2'^{(n)}(\hat{P}_{USW},\hat{P}_{VSWY}) \\
&:= \{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} \in \mathcal{T}_{\mathcal{U}\mathcal{V}\mathcal{S}\mathcal{W}\mathcal{Y}} : P_{\tilde{U}\tilde{S}\tilde{W}} \in T^n_{[USW]_\delta},\ P_{\tilde{V}\tilde{S}\tilde{W}\tilde{Y}} \in T^n_{[VSWY]_{\tilde{\delta}}} \\
&\qquad \text{and } H(\tilde{W}|\tilde{V},\tilde{S},\tilde{Y}) \geq H(\bar{W}|V,S,Y) - O(\delta)\},
\end{aligned}
$$

and

$$
\begin{aligned}
E_{2n}^* &\overset{(n)}{\geq} \min_{\substack{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}}\in \\ \mathcal{T}_2'(\hat{P}_{USW},\hat{P}_{VSWY})}} \Big[D(P_{\tilde{U}\tilde{V}\tilde{S}\tilde{W}\tilde{Y}}|Q_{UV}P_{\tilde{S}\tilde{W}|\tilde{U}}\hat{P}_{Y|USW}) + I(\bar{W};V,Y|S) \\
&\qquad\qquad - I(U;\bar{W}|S) - O(\delta) \Big] \\
&= E_2'(P_S, P_{\bar{W}|US}, P_{X|USW}) - O(\delta). \qquad \text{(A.64)}
\end{aligned}
$$

In (A.62), we used the fact that

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

which follows from

$$
\mathbb{P}\left(\bar{W}^n(l)=\tilde{w}^n|\tilde{\mathcal{F}}\right) \leq 2\ \mathbb{P}(\bar{W}^n(l)=\tilde{w}^n). \qquad \text{(A.65)}
$$

Eqn. (A.65) can be proved similarly to (A.28).

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

$$\begin{aligned}
&\sum_{u^n,v^n} \mathbb{P}(U^n=u^n,V^n=v^n|H=1)\ \mathbb{P}(\mathcal{D}_0|\ U^n=u^n,V^n=v^n,\ \bar{\mathcal{E}}_0)\\
&=\sum_{u^n,v^n} \mathbb{P}(U^n=u^n,V^n=v^n|H=1)\\
&\qquad \sum_{s^n,y^n} \mathbb{P}(S^n=s^n,Y^n=y^n,\mathcal{D}_0|\ U^n=u^n,V^n=v^n,\ \bar{\mathcal{E}}_0)\\
&=\sum_{u^n,v^n} \mathbb{P}(U^n=u^n,V^n=v^n|H=1)\\
&\qquad \Big[\ \sum_{s^n,y^n} \mathbb{P}(S^n=s^n,Y^n=y^n|\ U^n=u^n,V^n=v^n,\ \bar{\mathcal{E}}_0)\\
&\qquad \mathbb{P}(\mathcal{D}_0|\ U^n=u^n,V^n=v^n,S^n=s^n,Y^n=y^n,\ \bar{\mathcal{E}}_0)\Big]\\
&=\sum_{u^n,v^n} \mathbb{P}(U^n=u^n,V^n=v^n|H=1)\Big[\ \sum_{s^n,y^n} \mathbb{P}(S^n=s^n,Y^n=y^n|\ U^n=u^n,\ \bar{\mathcal{E}}_0)\\
&\qquad \mathbb{P}(\mathcal{D}_0|\ U^n=u^n,\ V^n=v^n,S^n=s^n,Y^n=y^n,\ \bar{\mathcal{E}}_0)\Big] \qquad \text{(A.66)}
\end{aligned}$$

The event $\mathcal{D}_0$ occurs only if there exists a sequence $(\bar{W}^n(l),V^n,S^n,Y^n)\in T^n_{[\bar{W}VSY]_{\tilde{\delta}}}$ for some $l\in[e^{n(I(U,S;\bar{W})+\mu)}]$. Noting that the quantization codebook is independent of the (V^n,S^n,Y^n) given that $\bar{\mathcal{E}}_0$ holds, it can be shown using standard arguments that

$$\mathbb{P}(\mathcal{D}_0|\ V^n=v^n,S^n=s^n,Y^n=y^n,\ \bar{\mathcal{E}}_0)\leq e^{-n(I(\bar{W};V,Y|S)-I(U;\bar{W}|S)-O(\delta))}. \qquad \text{(A.67)}$$

Also,

$$\mathbb{P}(S^n=s^n,Y^n=y^n|\ U^n=u^n,\bar{\mathcal{E}}_0)\leq e^{-n(H(\tilde{S}\tilde{Y}|\tilde{U})+D(P_{\tilde{S}\tilde{Y}|\tilde{U}}||\check{Q}_{SY|U}|P_{\tilde{U}}))}. \qquad \text{(A.68)}$$

Hence, using (A.67) and (A.68) in (A.66), we obtain

$$\begin{aligned}
&\sum_{u^n,v^n} \mathbb{P}(U^n=u^n,V^n=v^n|H=1)\ \mathbb{P}(\mathcal{D}_0|\ U^n=u^n,V^n=v^n,\ \bar{\mathcal{E}}_0)\\
&\leq (n+1)^{|\mathcal{U}||\mathcal{V}||\mathcal{S}||\mathcal{Y}|} \max_{\substack{P_{\tilde{U}\tilde{V}\tilde{S}\tilde{Y}}:\\ P_{\tilde{V}\tilde{S}\tilde{Y}}=\hat{P}_{VSY}}} e^{nH(\tilde{U}\tilde{V}\tilde{S}\tilde{Y})}\ e^{-n(H(\tilde{U}\tilde{V})+D(P_{\tilde{U}\tilde{V}}||Q_{UV}))}\ e^{-nH(\tilde{S}\tilde{Y}|\tilde{U})}\\
&\qquad\qquad e^{-nD(P_{\tilde{S}\tilde{Y}|\tilde{U}}||\check{Q}_{SY|U}|P_{\tilde{U}})}\ e^{-n(I(\bar{W};V,Y|S)-I(U;\bar{W}|S)-O(\delta))}
\end{aligned}$$

$= e^{-nE^*_{3n}}$,

where,

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

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

A.3 Optimal single-letter characterization of error-exponent when $C(P_{Y|X}) = 0$

The achievability follows from Proposition 2.4 which states that for $\tau \geq 0$, $\kappa(\tau,\epsilon) \geq \kappa_0(\tau)$, $\forall\ \epsilon \in (0,1]$. Now, it is well-known (see [20]) that $C(P_{Y|X}) = 0$ only if

$$P^*_Y := P_{Y|X=x} = P_{Y|X=x'},\ \forall\ x,x' \in \mathcal{X}. \tag{A.69}$$

From (A.69), it follows that $E_c(P_{Y|X}) = 0$. Also,

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

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

Converse: We first show the weak converse, i.e., $\kappa(\tau) \leq D(P_V||Q_V)$, where $\kappa(\tau)$ is as defined in (2.31). For any sequence of encoding functions $f^{(k,n_k)}$ and acceptance regions $\mathcal{A}_{(k,n_k)}$ for H_0 that satisfy $n_k \leq \tau k$ and (2.43), it follows similarly to (2.44),

that

$$\limsup_{k\to\infty} \frac{-1}{k} \log\left(\beta\left(k, n_k, f^{(k,n_k)}, g^{(k,n_k)}\right)\right) \leq \limsup_{k\to\infty} \frac{1}{k} D\left(P_{Y^{n_k}V^k} || Q_{Y^{n_k}V^k}\right). \quad \text{(A.70)}$$

The terms in the R.H.S. of (A.70) can be expanded as

$$\begin{aligned}
&\frac{1}{k} D\left(P_{Y^{n_k}V^k} || Q_{Y^{n_k}V^k}\right) \\
&= D(P_V||Q_V) + \frac{1}{k} \sum_{\substack{(v^k, y^{n_k}) \\ \in \mathcal{V}^k \times \mathcal{Y}^{n_k}}} P_{V^k Y^{n_k}}(v^k, y^{n_k}) \log\left(\frac{P_{Y^{n_k}|V^k}(y^{n_k}|v^k)}{Q_{Y^{n_k}|V^k}(y^{n_k}|v^k)}\right). \quad \text{(A.71)}
\end{aligned}$$

Now, note that

$$\begin{aligned}
P_{Y^{n_k}|V^k}(y^{n_k}|v^k) &= \sum_{\substack{(u^k, x^{n_k}) \\ \in\, \mathcal{U}^k \times \mathcal{X}^{n_k}}} P_{U^k|V^k}(u^k|v^k) P_{X^{n_k}|U^k}(x^{n_k}|u^k) P_{Y^{n_k}|X^{n_k}}(y^{n_k}|x^{n_k}) \\
&= \left(\prod_{i=1}^{n_k} P_Y^*(y_i)\right) \sum_{\substack{(u^k, x^{n_k}) \\ \in\, \mathcal{U}^k \times \mathcal{X}^{n_k}}} P_{U^k|V^k}(u^k|v^k) P_{X^{n_k}|U^k}(x^{n_k}|u^k) \quad \text{(A.72)} \\
&= \prod_{i=1}^{n_k} P_Y^*(y_i), \quad \text{(A.73)}
\end{aligned}$$

where, (A.72) follows from (2.3) and (A.69). Similarly, it follows that

$$Q_{Y^{n_k}|V^k}(y^{n_k}|v^k) = \prod_{i=1}^{n_k} P_Y^*(y_i). \quad \text{(A.74)}$$

From (A.70), (A.71), (A.73) and (A.74), we obtain that

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

This completes the proof of the weak converse.

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

$$\limsup_{k\to\infty} \alpha\left(k, n_k, f^{(k,n_k)}, g^{(k,n_k)}\right) \leq \epsilon. \quad \text{(A.75)}$$

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

$$P_{V^kY^{n_k}}(v^k, y^{n_k}) = \left(\prod_{i=1}^{k} P_V(v_i)\right)\left(\prod_{j=1}^{n_k} P_Y^*(y_j)\right), \tag{A.76}$$

$$\text{and } Q_{V^kY^{n_k}}(v^k, y^{n_k}) = \left(\prod_{i=1}^{k} Q_V(v_i)\right)\left(\prod_{j=1}^{n_k} P_Y^*(y_j)\right), \tag{A.77}$$

respectively. By the weak law of large numbers, for any $\delta > 0$, (A.76) implies that

$$\lim_{k\to\infty} P_{V^kY^{n_k}}\left(T^k_{[P_V]_\delta} \times T^{n_k}_{[P_Y^*]_\delta}\right) = 1. \tag{A.78}$$

Also, from (A.75), we have

$$\liminf_{k\to\infty} P_{V^kY^{n_k}}\left(\mathcal{A}_{(k,n_k)}\right) \geq (1-\epsilon). \tag{A.79}$$

From (A.78) and (A.79), it follows that

$$P_{V^kY^{n_k}}\left(\mathcal{A}_{(k,n_k)} \cap T^k_{[P_V]_\delta} \times T^{n_k}_{[P_Y^*]_\delta}\right) \geq 1-\epsilon', \tag{A.80}$$

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

$$\mathcal{A}(v^k, \delta) := \left\{y^{n_k} : (v^k, y^{n_k}) \in \mathcal{A}_{(k,n_k)} \cap T^k_{[P_V]_\delta} \times T^{n_k}_{[P_Y^*]_\delta}\right\}, \tag{A.81}$$

$$\text{and } \mathcal{D}(\eta, \delta) := \left\{v^k \in T^k_{[P_V]_\delta} : P_{Y^{n_k}}(\mathcal{A}(v^k, \delta)) \geq \eta\right\}. \tag{A.82}$$

Fix $0 < \eta' < 1 - \epsilon'$. Then, we have from (A.80) that for any $\delta > 0$ and sufficiently large k,

$$P_{V^k}\left(\mathcal{D}(\eta', \delta)\right) \geq \frac{1-\epsilon'-\eta'}{1-\eta'}. \tag{A.83}$$

From [20, Lemma 2.14], (A.83) implies that $\mathcal{D}(\eta', \delta)$ should contain atleast $\frac{1-\epsilon'-\eta'}{1-\eta'}$ fraction (approx.) of sequences in $T^k_{[P_V]_\delta}$ and for each $v^k \in \mathcal{D}(\eta', \delta)$, (A.82) implies that

$\mathcal{A}(v^k, \delta)$ should contain atleast η' fraction (approx.) of sequences in $T^{n_k}_{[P^*_Y]_\delta}$, asymptotically. Hence, for sufficiently large k, we have

$$Q_{V^kY^{n_k}}\left(\mathcal{A}_{(k,n_k)}\right) \geq \sum_{v^k \in \mathcal{D}(\eta',\delta)} Q_{V^k}(v^k) \sum_{y^{n_k} \in \mathcal{A}(v^k,\delta)} P_{Y^n}(y^{n_k}) \tag{A.84}$$

$$\geq e^{-k\left(D(P_V||Q_V) - \frac{\log\left(\frac{1-\epsilon'-\eta'}{1-\eta'}\right)}{k} - \frac{\log(\eta')}{k} - O(\delta)\right)}. \tag{A.85}$$

Here, (A.85) follows from [20, Lemma 2.6].

Let $\mathcal{A}'_{(k,n_k)} := T^k_{[P_V]_\delta} \times T^{n_k}_{[P^*_Y]_\delta}$. Then, for sufficiently large k,

$$P_{V^kY^{n_k}}\left(\mathcal{A}'_{(k,n_k)}\right) \xrightarrow{(k)} 1, \tag{A.86}$$

$$\text{and } Q_{V^kY^{n_k}}\left(\mathcal{A}'_{(k,n_k)}\right) \leq e^{-k(D(P_V||Q_V)-O(\delta))}, \tag{A.87}$$

where, (A.86) and (A.87) follows from weak law of large numbers and [20, Lemma 2.6], respectively. Together (A.85), (A.86) and (A.87) implies that

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

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

Appendix B

Proofs for Chapter 3

B.1 Proof of Proposition 3.6

First, we show the proof of achievability, i.e., for $-D(P_{Y_{X_0}}||P_{Y_{X_1}}|P_{X_0X_1}) < \theta \leq D(P_{Y_{X_1}}||P_{Y_{X_0}}|P_{X_0X_1})$,

$$\kappa\left(\mathbb{E}_{P_{X_0X_1}}\left[\psi^*_{Y_{X_0},\tilde{h}_{X_0,X_1}}(\theta)\right], x_0^n, x_1^n\right) \geq \mathbb{E}_{P_{X_0X_1}}\left[\psi^*_{Y_{X_0},\tilde{h}_{X_0,X_1}}(\theta)\right] - \theta.$$

Let $\tilde{h}'_{\mathcal{Y}^n} : \mathcal{Y}^n \mapsto \mathbb{R}$ defined as

$$\tilde{h}'_{\mathcal{Y}^n}(y^n) := \log\left(\frac{P_{Y^n|X^n=x_1^n}(y^n)}{P_{Y^n|X^n=x_0^n}(y^n)}\right).$$

For the decision rule $g^{(n)}_{\theta,\mathcal{Y}}$ defined in (3.13), the type I error probability can be upper bounded for $\theta > -D(P_{Y_{X_0}}||P_{Y_{X_1}}|P_{X_0X_1})$ as follows:

$$\begin{aligned}
\alpha\left(n, g^{(n)}_{\theta,\mathcal{Y}}, x_0^n, x_1^n\right) &= P_{Y^n|X^n=x_0^n}\left(\log\left(\frac{P_{Y^n|X^n=x_1^n}(Y^n)}{P_{Y^n|X^n=x_0^n}(Y^n)}\right) \geq n\theta\right) \\
&\leq e^{-\sup\limits_{\lambda\geq 0}\left(n\theta\lambda - \psi_{P_{Y^n|X^n=x_0^n},\tilde{h}'_{\mathcal{Y}^n}}(\lambda)\right)} \qquad \text{(B.1)} \\
&= e^{-\sup\limits_{\lambda\in\mathbb{R}}\left(n\left(\theta\lambda - \frac{1}{n}\psi_{P_{Y^n|X^n=x_0^n},\tilde{h}'_{\mathcal{Y}^n}}(\lambda)\right)\right)}. \qquad \text{(B.2)}
\end{aligned}$$

Here, (B.1) follows using the standard Chernoff bound. Eqn. (B.2) follows due to the fact that for $\theta > -D(P_{Y_{X_0}}||P_{Y_{X_1}}|P_{X_0X_1})$, the supremum in (B.1) is always achieved at $\lambda \geq 0$, which in turn follows from Lemma 3.1 (i) and (ii).

Simplifying the term within the exponent in (B.2), we obtain

$$\begin{aligned}
\frac{1}{n}\psi_{P_{Y^n|X^n=x_0^n},\tilde{h}'_{\mathcal{Y}^n}}(\lambda) &:= \frac{1}{n}\log\left(\mathbb{E}_{P_{Y^n|X^n=x_0^n}}\left(\frac{P^{\lambda}_{Y^n|X^n=x_1^n}(Y^n)}{P^{\lambda}_{Y^n|X^n=x_0^n}(Y^n)}\right)\right)\\
&= \frac{1}{n}\log\left(\mathbb{E}_{P_{Y^n|X^n=x_0^n}}\left(\prod_{i=1}^{n}\frac{P^{\lambda}_{Y_i|X_i=x_{1i}}(Y_i)}{P^{\lambda}_{Y_i|X_i=x_{0i}}(Y_i)}\right)\right)\\
&= \frac{1}{n}\log\left(\prod_{i=1}^{n}\mathbb{E}_{P_{Y_i|X_i=x_{0i}}}\left(\frac{P^{\lambda}_{Y_i|X_i=x_{1i}}(Y_i)}{P^{\lambda}_{Y_i|X_i=x_{0i}}(Y_i)}\right)\right)\\
&= \frac{1}{n}\sum_{i=1}^{n}\log\left(\mathbb{E}_{P_{Y_i|X_i=x_{0i}}}\left(\frac{P^{\lambda}_{Y_i|X_i=x_{1i}}(Y_i)}{P^{\lambda}_{Y_i|X_i=x_{0i}}(Y_i)}\right)\right)\\
&= \sum_{x,x'}P_{x_0^n x_1^n}(x,x')\log\left(\mathbb{E}_{P_{Y_x}}\left(\frac{P^{\lambda}_{Y_{x'}}(Y)}{P^{\lambda}_{Y_x}(Y)}\right)\right) \qquad \text{(B.3)}\\
&\xrightarrow{(n)} \mathbb{E}_{P_{X_0X_1}}\left[\log\left(\mathbb{E}_{P_{Y_{X_0}}}\left(e^{\lambda\tilde{h}_{X_0,X_1}(Y)}\right)\right)\right], \qquad \text{(B.4)}
\end{aligned}$$

where, (B.4) follows from (3.9) and Assumption 3.2. Substituting (B.4) in (B.2) and using (3.6), we obtain for arbitrarily small but fixed $\delta > 0$ and sufficiently large n that

$$\begin{aligned}
\alpha\left(n, g^{(n)}_{\theta,\mathcal{Y}}, x_0^n, x_1^n\right) &\leq e^{-\sup\limits_{\lambda\in\mathbb{R}}\left(n\left(\theta\lambda-\mathbb{E}_{P_{X_0X_1}}\left[\log\left(\mathbb{E}_{P_{Y_{X_0}}}\left(e^{\lambda\tilde{h}_{X_0,X_1}(Y)}\right)\right)\right]-\delta\right)\right)}\\
&= e^{-n\left(\mathbb{E}_{P_{X_0X_1}}\left[\sup\limits_{\lambda\in\mathbb{R}}\left(\theta\lambda-\mathbb{E}_{P_{Y_{X_0}}}\left(e^{\lambda\tilde{h}_{X_0,X_1}(Y)}\right)\right)\right]-\delta\right)}\\
&= e^{-n\left(\mathbb{E}_{P_{X_0X_1}}\left[\psi^*_{Y_{X_0},\tilde{h}_{X_0,X_1}}(\theta)\right]-\delta\right)}. \qquad \text{(B.5)}
\end{aligned}$$

Similarly, we can show that for $\theta \leq D(P_{Y_{X_1}}||P_{Y_{X_0}}|P_{X_0X_1})$,

$$\beta\left(n, g^{(n)}_{\theta,\mathcal{Y}}, x_0^n, x_1^n\right) \leq e^{-n\left(\mathbb{E}_{P_{X_0X_1}}\left[\psi^*_{Y_{X_1},\tilde{h}_{X_0,X_1}}(\theta)\right]-\delta\right)}. \qquad \text{(B.6)}$$

We also have

$$\psi_{Y_{x'},\tilde{h}_{x,x'}}(\lambda) = \sum_{y\in\mathcal{Y}}P_{Y_{x'}}\frac{P^{\lambda}_{Y_{x'}}}{P^{\lambda}_{Y_x}} = \sum_{y\in\mathcal{Y}}P_{Y_x}\frac{P^{\lambda+1}_{Y_{x'}}}{P^{\lambda+1}_{Y_x}} = \psi_{Y_x,\tilde{h}_{x,x'}}(\lambda+1).$$

It follows that

$$\psi^*_{Y_{x'},\tilde{h}_{x,x'}}(\theta) := \sup_{\lambda\in\mathbb{R}}\left(\lambda\theta - \psi_{Y_{x'},\tilde{h}_{x,x'}}(\lambda)\right)$$

$$= \sup_{\lambda \in \mathbb{R}} \left(\lambda\theta - \psi_{Y_x, \tilde{h}_{x,x'}}(\lambda + 1) \right) = \psi^*_{Y_x, \tilde{h}_{x,x'}}(\theta) - \theta.$$

Hence,

$$\mathbb{E}_{P_{X_0X_1}} \left[\psi^*_{Y_{X_1}, \tilde{h}_{X_0,X_1}}(\theta) \right] = \mathbb{E}_{P_{X_0X_1}} \left[\psi^*_{Y_{X_0}, \tilde{h}_{X_0,X_1}}(\theta) \right] - \theta. \tag{B.7}$$

From (B.5), (B.6) and (B.7), it follows that for $-D(P_{Y_{X_0}} || P_{Y_{X_1}} | P_{X_0X_1}) < \theta \leq D(P_{Y_{X_1}} || P_{Y_{X_0}} | P_{X_0X_1})$,

$$\bar{\kappa} \left(\mathbb{E}_{P_{X_0X_1}} \left[\psi^*_{Y_{X_0}, \tilde{h}_{X_0,X_1}}(\theta) \right] - \delta, P_{X_0X_1} \right) \geq \mathbb{E}_{P_{X_0X_1}} \left[\psi^*_{Y_{X_0}, \tilde{h}_{X_0,X_1}}(\theta) \right] - \theta - \delta.$$

Noting that $\delta > 0$ is arbitrary and $\bar{\kappa}(\kappa_\alpha, P_{X_0X_1})$ is a continuous function of κ_α for a fixed $P_{X_0X_1}$, the proof of achievability is complete.

Next, we prove the converse. Let

$$\mu_n(x, x') := P_{x_0^n x_1^n}(x, x'),$$
$$\text{and } \mathcal{I}_n(x, x') := \{i \in [n] \text{ s.t. } x_{0i} = x \text{ and } x_{1i} = x'\}.$$

Denoting $\alpha\left(n, g^{(k,n)}, x_0^n, x_1^n\right)$ and $\beta\left(n, g^{(k,n)}, x_0^n, x_1^n\right)$ by α_n and β_n, respectively, we obtain that for any $\theta \in \mathbb{R}$,

$$\begin{aligned}
&\alpha_n + e^{-n\theta}\beta_n \\
&\geq P_{Y^n|X^n=x_0^n} \left(\log \left(\frac{P_{Y^n|X^n=x_1^n}(Y^n)}{P_{Y^n|X^n=x_0^n}(Y^n)} \right) \geq n\theta \right) \qquad \text{(B.8)} \\
&= P_{Y^n|X^n=x_0^n} \left(\sum_{i=1}^n \log \left(\frac{P_{Y_i|X=x_{1i}}(Y_i)}{P_{Y_i|X=x_{0i}}(Y_i)} \right) \geq n\theta \right) \\
&= P_{Y^n|X^n=x_0^n} \left(\sum_{x,x'} \sum_{i \in \mathcal{I}_n(x,x')} \log \left(\frac{P_{Y_i|X=x_{1i}}(Y_i)}{P_{Y_i|X=x_{0i}}(Y_i)} \right) \geq n\theta \right) \\
&= P_{Y^n|X^n=x_0^n} \left(\sum_{x,x'} \sum_{i \in \mathcal{I}_n(x,x')} \log \left(\frac{P_{Y_i|X=x_{1i}}(Y_i)}{P_{Y_i|X=x_{0i}}(Y_i)} \right) \geq \sum_{(x,x') \in \mathcal{X} \times \mathcal{X}} n\mu_n(x, x')\theta \right) \\
&\geq P_{Y^n|X^n=x_0^n} \left(\bigcap_{x,x'} \left(\sum_{i \in \mathcal{I}_n(x,x')} \log \left(\frac{P_{Y_i|X=x_{1i}}(Y_i)}{P_{Y_i|X=x_{0i}}(Y_i)} \right) \geq n\mu_n(x, x')\theta \right) \right)
\end{aligned}$$

$$= \prod_{(x,x')\in\mathcal{X}\times\mathcal{X}} P_{Y^n|X^n=x_0^n}\left(\sum_{i\in\mathcal{I}_n(x,x')} \log\left(\frac{P_{Y_i|X=x_{1i}}(Y_i)}{P_{Y_i|X=x_{0i}}(Y_i)}\right) \geq n\mu_n(x,x')\theta\right).$$

Here, (B.8) follows by applying Theorem 3.4. Then, for arbitrary $\delta > 0$ and sufficiently large n, we can write

$$\alpha_n + e^{-n\theta}\beta_n \geq \prod_{(x,x')\in\mathcal{X}\times\mathcal{X}} e^{-n\mu_n(x,x')\left(\left(\min\limits_{\tilde{Q}_x:\mathbb{E}_{\tilde{Q}_x}(Y)\geq\theta} D\left(\tilde{Q}_x||P_{Y_x}\right)\right)+\delta\right)} \tag{B.9}$$

$$\geq \prod_{(x,x')\in\mathcal{X}\times\mathcal{X}} e^{-n\mu_n(x,x')\left(\psi^*_{Y_x,\tilde{h}_{x,x'}}(\theta)+\delta\right)} \tag{B.10}$$

$$= e^{-n\left(\mathbb{E}_{P_{X_0X_1}}\left(\psi^*_{Y_{X_0},\tilde{h}_{X_0,X_1}}(\theta)\right)+\delta'\right)}, \tag{B.11}$$

where, $\delta' > \delta$ is arbitrary. Here, (B.9) follows from [6, Theorem 14.1]; (B.10) follows from [6, Theorem 13.3] and [6, Theorem 14.3]; and (B.11) follows from (3.9). Note that (B.11) holds even if $\psi^*_{Y_x,\tilde{h}_{x,x'}}(\theta) = \infty$ for some $x, x' \in \mathcal{X}\times\mathcal{X}$ and $\theta > 0$ since in this case, both (B.10) and (B.11) equal 0. Equation (B.11) implies that

$$\limsup_{n\to\infty} \min\left(-\frac{\log(\alpha_n)}{n}, -\frac{\log(\beta_n)}{n} + \theta\right) \leq \mathbb{E}_{P_{X_0X_1}}\left(\psi^*_{Y_{X_0},\tilde{h}_{X_0,X_1}}(\theta)\right) + \delta'. \tag{B.12}$$

Hence, if it holds that for all sufficiently large n,

$$\alpha_n < e^{-n\left(\mathbb{E}_{P_{X_0X_1}}\left(\psi^*_{Y_{X_0},\tilde{h}_{X_0,X_1}}(\theta)\right)+\delta'\right)}, \tag{B.13}$$

then

$$\limsup_{n\to\infty} -\frac{\log(\beta_n)}{n} \leq \mathbb{E}_{P_{X_0X_1}}\left(\psi^*_{Y_{X_0},\tilde{h}_{X_0,X_1}}(\theta)\right) - \theta + \delta'. \tag{B.14}$$

Since δ (and δ') is arbitrary, this implies via the continuity of $\bar{\kappa}(\kappa_\alpha, P_{X_0X_1})$ in κ_α that

$$\bar{\kappa}\left(\mathbb{E}_{P_{X_0X_1}}\left[\psi^*_{Y_{X_0},\tilde{h}_{X_0,X_1}}(\theta)\right], P_{X_0X_1}\right) \leq \mathbb{E}_{P_{X_0X_1}}\left[\psi^*_{Y_{X_0},\tilde{h}_{X_0,X_1}}(\theta)\right] - \theta.$$

To complete the proof, we need to show that θ can be restricted to lie in $\mathcal{I}(P_{X_0X_1}, P_{Y|X})$. To prove this, it suffices to show the following:

(i) $\mathbb{E}_{P_{X_0X_1}}\left(\psi^*_{Y_{X_0},\tilde{h}_{X_0,X_1}}\left(-D\left(P_{Y_{X_0}}||P_{Y_{X_1}}|P_{X_0X_1}\right)\right)\right)=0.$

(ii) $\mathbb{E}_{P_{X_0X_1}}\left(\psi^*_{Y_{X_0},\tilde{h}_{X_0,X_1}}\left(D\left(P_{Y_{X_1}}||P_{Y_{X_0}}|P_{X_0X_1}\right)\right)\right)=D\left(P_{Y_{X_1}}||P_{Y_{X_0}}|P_{X_0X_1}\right).$

(iii) $\mathbb{E}_{P_{X_0X_1}}\left(\psi^*_{Y_{X_0},\tilde{h}_{X_0,X_1}}(\theta)\right)$ and $\mathbb{E}_{P_{X_0X_1}}\left(\psi^*_{Y_{X_0},\tilde{h}_{X_0,X_1}}(\theta)\right)-\theta$ are convex functions of θ.

We have,

$$\begin{aligned}
&\mathbb{E}_{P_{X_0X_1}}\left(\psi^*_{Y_{X_0},\tilde{h}_{X_0,X_1}}\left(-D\left(P_{Y_{X_0}}||P_{Y_{X_1}}|P_{X_0X_1}\right)\right)\right)\\
&=\sup_{\lambda\in\mathbb{R}}\left[-\lambda\ D\left(P_{Y_{X_0}}||P_{Y_{X_1}}|P_{X_0X_1}\right)-\mathbb{E}_{P_{X_0X_1}}\left(\psi_{Y_{X_0},\tilde{h}_{X_0,X_1}}(\lambda)\right)\right]\\
&\leq\sum_{x_0,x_1}P_{X_0X_1}(x_0,x_1)\ \left[\sup_{\lambda\in\mathbb{R}}-\lambda D\left(P_{Y_{X_0}}||P_{Y_{X_1}}\right)-\psi_{Y_{X_0},\tilde{h}_{x_0,x_1}}(\lambda)\right] &\text{(B.15)}\\
&=0, &\text{(B.16)}
\end{aligned}$$

where, (B.16) follows since each term inside the square braces in (B.15) is zero, which in turn follows from Lemma 3.1 (iii). Also,

$$\begin{aligned}
&\mathbb{E}_{P_{X_0X_1}}\left(\psi^*_{Y_{X_0},\tilde{h}_{X_0,X_1}}\left(-D\left(P_{Y_{X_0}}||P_{Y_{X_1}}|P_{X_0X_1}\right)\right)\right)\\
&=\sum_{x_0,x_1}P_{X_0X_1}(x_0,x_1)\ \psi^*_{Y_{X_0},\tilde{h}_{X_0,X_1}}\left(-D\left(P_{Y_{X_0}}||P_{Y_{X_1}}|P_{X_0X_1}\right)\right)\\
&\geq 0, &\text{(B.17)}
\end{aligned}$$

where, (B.17) again follows from Lemma 3.1 (iii). Combining (B.16) and (B.17) proves (i). We also have that

$$\begin{aligned}
&\mathbb{E}_{P_{X_0X_1}}\left(\psi^*_{Y_{X_0},\tilde{h}_{X_0,X_1}}\left(D\left(P_{Y_{X_1}}||P_{Y_{X_0}}|P_{X_0X_1}\right)\right)\right)-D\left(P_{Y_{X_1}}||P_{Y_{X_0}}|P_{X_0X_1}\right)\\
&=\mathbb{E}_{P_{X_0X_1}}\left(\psi^*_{Y_{X_1},\tilde{h}_{X_0,X_1}}\left(D\left(P_{Y_{X_1}}||P_{Y_{X_0}}|P_{X_0X_1}\right)\right)\right)\\
&=0, &\text{(B.18)}
\end{aligned}$$

where, (B.18) follows similarly to the proof of (i). This proves (ii). Finally, (iii) follows from Lemma 3.1 (iii) and the fact that a weighted sum of convex functions is convex provided the weights are non-negative. This completes the proof.

B.2 Proof of Theorem 3.9

Fix $\kappa_\alpha > 0$ and $(\omega, R, P_{SX}, \theta) \in \mathcal{L}(\kappa_\alpha, \tau)$. Let $\eta > 0$ be a small number, and let $R' \geq 0$ and $R \geq 0$ be defined as

$$R' := \zeta_q(\kappa_\alpha, \omega), \tag{B.19}$$

$$\text{and } \zeta_q(\kappa_\alpha, \omega) - \rho(\kappa_\alpha, \omega) \leq R < \tau I(X;Y|S). \tag{B.20}$$

Encoding:

The encoder is composed of two stages, a source encoder followed by a channel encoder. The source encoding comprises of a quantization scheme followed by binning (if necessary). The details are as follows:

Quantization scheme: Let

$$\mathcal{D}_k^U(\eta) := \{\hat{U} \in \mathcal{T}_k(\mathcal{U}) : D(\hat{U}||U) \leq \kappa_\alpha + \eta\}. \tag{B.21}$$

Consider some ordering on the types in $\mathcal{D}_k^U(\eta)$ and denote the elements as $\hat{U}_1$, $\hat{U}_2$, ..., etc. For each type variable $\hat{U}_i \in \mathcal{D}_k^U(\eta)$, $1 \leq i \leq |\mathcal{D}_k^U(\eta)|$, choose a joint type variable $\hat{U}_i\hat{W}_i$, $\hat{W}_i \in \mathcal{T}_k(\mathcal{W})$, such that

$$D\left(\hat{W}_i|\hat{U}_i||W_i|U|\hat{U}_i\right) \leq \frac{\eta}{3}, \tag{B.22}$$

$$I(\hat{U}_i;\hat{W}_i) \leq R' + \frac{\eta}{3}, \tag{B.23}$$

where $P_{W_i|U} = \omega(P_{\hat{U}_i})$. This is always possible for k large enough due to (B.19) and the continuity of ω (see [27]). Let

$$\mathcal{D}_k^{UW}(\eta) := \{\hat{U}_i\hat{W}_i : 1 \leq i \leq |\mathcal{D}_k^U(\eta)|\}, \tag{B.24}$$

$$\text{and } R_i' := I(\hat{U}_i;\hat{W}_i) + \frac{\eta}{3}, 1 \leq i \leq |\mathcal{D}_k^U(\eta)|. \tag{B.25}$$

Let

$$\mathcal{C}_k = \left\{ w^k(j), j \in \left[1 : \sum_{i=1}^{|\mathcal{D}_k^U(\eta)|} e^{kR_i'}\right]\right\},$$

denote a quantization codebook such that each codeword $w^k(j)$, $j \in \mathcal{M}'_i := [1 + \sum_{m=1}^{i-1} e^{kR'_m} : \sum_{m=1}^{i} e^{kR'_m}]$, $1 \le i \le |\mathcal{D}_k^U(\eta)|$, belongs to the set $\mathcal{T}_k(\hat{W}_i)$. For $u^k \in \mathcal{T}_k(\hat{U}_i)$ such that $\hat{U}_i \in \mathcal{D}_k^U(\eta)$ for some $1 \le i \le |\mathcal{D}_k^U(\eta)|$, let

$$\mu(u^k, \mathcal{C}_k) := \{j \in \mathcal{M}'_i : \ w^k(j) \in \mathcal{C}_k \text{ and } (u^k, w^k(j)) \in \mathcal{T}_k(\hat{U}_i\hat{W}_i),\ \hat{U}_i\hat{W}_i \in \mathcal{D}_k^{UW}(\eta)\}.$$

If $|\mu(u^k, \mathcal{C}_k)| \ge 1$, let $M'(u^k, \mathcal{C}_k)$ denote an index selected uniformly at random from the set $\mu(u^k, \mathcal{C}_k)$, otherwise, set $M'(u^k, \mathcal{C}_k) = 0$. Given $\mathcal{C}_k$ and $u^k \in \mathcal{U}^k$, the quantizer outputs $M' = M'(u^k, \mathcal{C}_k)$, where the support of M' is given by

$$\mathcal{M}' := \left[0 : \sum_{i=1}^{|\mathcal{D}_k^U(\eta)|} e^{kR'_i}\right].$$

Note that for sufficiently large k,

$$\begin{aligned} |\mathcal{M}'| &\le 1 + \sum_{i=1}^{|\mathcal{D}_k^U(\eta)|} e^{kR'_i} \le 1 + |\mathcal{D}_k^U(\eta)| e^{k\left(\max\limits_{\hat{U}\hat{W} \in \mathcal{D}_k^{UW}(\eta)} I(\hat{U};\hat{W}) + \frac{\eta}{3}\right)} \\ &\le 1 + |\mathcal{D}_k^U(\eta)| e^{k\left(R' + \frac{2\eta}{3}\right)} \le e^{k(R'+\eta)}, \end{aligned} \tag{B.26}$$

where, in (B.26), we used the fact that $|\mathcal{D}_k^U(\eta)| \le (k+1)^{|\mathcal{U}|}$.

Let

$$\begin{aligned} R_k &:= \log\left(\frac{e^{kR}}{|\mathcal{D}_k^U(\eta)|}\right), \\ \mathcal{M}_i &:= [1 + (i-1)R_k : iR_k],\ 1 \le i \le |\mathcal{D}_k^U(\eta)|, \\ \text{and } \mathcal{M} &:= \{0\} \bigcup \bigcup_{i=1}^{|\mathcal{D}_k^U(\eta)|} \mathcal{M}_i. \end{aligned}$$

Note that

$$e^{kR_k} \ge e^{k\left(R - \frac{|\mathcal{U}|}{k}\log(k+1)\right)}. \tag{B.27}$$

Let $f_b : \mathcal{M}' \mapsto \mathcal{M}$ denote a function such that $f_b(j) = 0$ iff $j = 0$, and for each index $j \in \mathcal{M}'_i$, $f_b(j) \in \mathcal{M}_i$, $1 \le i \le |\mathcal{D}_k^U(\eta)|$. Given f_b, the source encoder outputs

$M = f_b(M')$. If $R' + \eta \leq R$, then f_b is taken to be the identity map, and in this case, $M = M'$.

Channel Encoding: Let $n = \lfloor \tau k \rfloor$. Each index in $\mathcal{M}$ is mapped to a codeword in the *channel codebook* $\mathcal{C}_X^n := \{X^n(j), j \in \mathcal{M}\}$, which is generated similar to the codebook used for the unequal error protection of a single message in [22]. Without loss of generality (w.l.o.g.), denote the elements of the set $\mathcal{S} = \mathcal{X}$ by $\{1, \ldots, |\mathcal{X}|\}$. 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 so forth, and the length of the last block is chosen such that the total length is n. The codeword $X^n(0) = s^n$ corresponding to $M = 0$ is obtained by repeating i in block i for $1 \leq i \leq |\mathcal{X}|$. The remaining[1] $\lceil e^{kR} \rceil$ ordinary codewords $X^n(j),\ j \in \lceil e^{kR} \rceil$, are obtained by blockwise i.i.d. random coding, i.e., the symbols in the i^{th} block of each codeword are generated i.i.d. according to $P_{X|S=i}$. The sequence s^n is revealed to the detector.

Decoding:

The decoder consists of two parts, a channel decoder followed by a tester.

Channel decoding: At the detector, the channel decoder first performs a NP test on the channel output Y^n using the decision rule $g_\theta : \mathcal{Y}^n \mapsto \{0, 1\}$, where

$$g_\theta(y^n) := \mathbb{1}\left(\sum_{j=1}^{n} \log \left(\frac{P_{Y|X=s(j))}(y_j)}{P_{Y|S=s(j)}(y_j)} \right) \geq n\theta \right),$$

$$s(j) := i \text{ if } \sum_{l=1}^{i-1} \lceil P_S(l)n \rceil < j \leq \sum_{l=1}^{i} \lceil P_S(l)n \rceil. \tag{B.28}$$

In (B.28), the empty sum is defined to be equal to 0. If $g_\theta(y^n) = 1$, then $\hat{M} = 0$ and $\hat{H} = 1$ is declared. Else, maximum likelihood (ML) decoding is done on the remaining codewords $X^n(j), j \in \lceil e^{kR} \rceil$, and $\hat{M}$ is set equal to the ML estimate. Note that since the i^{th} block of each codeword $X^n(j),\ j \in \lceil e^{kR} \rceil$, is generated independently and i.i.d. according to distribution $P_{X|S=i}$, the channel outputs in the i^{th} block is distributed i.i.d. according to $P_{Y|S=i}$. It then follows similar to Proposition 3.6 that

[1] Actually, the number of codewords generated should be slightly higher (e.g. $e^{k(R+\delta)}$ for a small positive number δ), as an expurgation step is involved later.

for k sufficiently large,

$$\mathbb{P}\left(\hat{M}=0|M\neq 0\right)\leq e^{-k\tau(E_m(P_{SX},\theta)-\eta)} \tag{B.29}$$

and

$$\mathbb{P}\left(\hat{M}\neq 0|M=0\right)\leq e^{-k\tau(E_m(P_{SX},\theta)-\theta-\eta)}. \tag{B.30}$$

Also, given $\hat{M}\neq 0$, it follows from the analysis based on random coding and expurgation (see [20, Exercise 10.18, 10.24] and [89]) that there exists a deterministic codebook $\mathcal{C}_X^n$ such that (B.29) and (B.30) holds, and the ML-decoding described above asymptotically yields

$$\mathbb{P}\left(\hat{M}\neq m|M=m\neq 0,\hat{M}\neq 0\right)\leq e^{-n\left(E_x\left(\frac{R}{\tau},P_{SX}\right)-\eta\right)}. \tag{B.31}$$

This deterministic codebook is used for channel coding.

Testing: The acceptance region for the hypothesis test is the same as that given in [27, Theorem 1]. More specifically, for a given codebook $\mathcal{C}_k$, let $\mathcal{O}_{m'}$ denote the set of u^k such that the source encoder outputs m', $m'\in\mathcal{M}'\backslash\{0\}$. For each $m'\in\mathcal{M}'\backslash\{0\}$ and $u^k\in\mathcal{O}_{m'}$, let

$$\mathcal{B}_{m'}(u^k)=\{v^k\in\mathcal{V}^k:(w_{m'}^k,u^k,v^k)\in\mathcal{J}_k^{\kappa_\alpha+\eta}(W_{m'}UV)\},$$

where $W_{m'}UV$ is uniquely specified by

$$W_{m'}-U-V \text{ and } P_{W_{m'}|U}=\omega(P_{u^k}). \tag{B.32}$$

For $m'\in\mathcal{M}'\backslash\{0\}$, we define

$$\mathcal{B}_{m'}:=\{v^k:v^k\in\mathcal{B}_{m'}(u^k) \text{ for some } u^k\in\mathcal{O}_{m'}\}.$$

Define the acceptance region for H_0 at the detector as

$$\mathcal{A}_k:=\bigcup_{m'\in\mathcal{M}'\backslash 0} m'\times\mathcal{B}_{m'}, \tag{B.33}$$

or equivalently as

$$\mathcal{A}_k^e := \bigcup_{m' \in \mathcal{M}' \setminus 0} \mathcal{O}_{m'} \times \mathcal{B}_{m'}. \tag{B.34}$$

The tester takes $\hat{M}$ as input, decodes for the quantization codeword $w^k(\hat{M}')$ (if required) using the empirical conditional entropy decoder (ECED), and declares the output of the hypothesis test based on $w^k(\hat{M}')$ and V^k. More specifically, if binning is not performed, i.e., if $R' + \eta \leq R$, set $\hat{M}' = \hat{M}$. Otherwise (if $R' + \eta > R$), given $\hat{M} = \hat{m}$ and $V^k = v^k$, set $\hat{M}' = \hat{m}'$, where

$$\hat{m}' := \begin{cases} 0, & \text{if } \hat{M} = 0, \\ \underset{j: f_b(j) = \hat{m}}{\arg\min} H_e(w^k(j)|v^k), & \text{otherwise.} \end{cases}$$

If $\hat{M}' = 0$, $\hat{H} = 1$ is declared. Otherwise, given $\hat{M}' = \hat{m}' \neq 0$ and $V^k = v^k$, $\hat{H} = 0$ or $\hat{H} = 1$ is declared depending on whether $(\hat{m}', v^k) \in \mathcal{A}_k$ or $(\hat{m}', v^k) \notin \mathcal{A}_k$, respectively.

Analysis of the type I and type II error probabilities:

Using the method of random coding, we will analyze the type I and type II error probabilities over an ensemble of randomly generated quantization and binning codebooks. Then, the standard random coding argument followed by an expurgation technique [89] guarantees the existence of a deterministic quantization and binning codebook that achieves the lower bound given in Theorem 3.9. Let each codeword $w^k(j)$, $j \in \mathcal{M}_i'$, $1 \leq i \leq |\mathcal{D}_k^U(\eta)|$, be selected (with replacement) independently and uniformly at random from the set $\mathcal{T}_k(\hat{W}_i)$ (see quantization scheme above). Let f_B denote the random binning function such that for each index $j \in \mathcal{M}_i'$, an index $f_B(j)$ is selected (with replacement) independently and uniformly at random from the set $\mathcal{M}_i$. We proceed to analyze the type I and type II error probabilities averaged over these random codebooks. Note that a type I error can occur only under the following events:

(i) $\mathcal{E}_{EE} := \bigcup_{\hat{U} \in \mathcal{D}_k^U(\eta)} \bigcup_{u^k \in \mathcal{T}_k(\hat{U})} \mathcal{E}_{EE}(u^k)$, where

$$\mathcal{E}_{EE}(u^k) := \Big\{ \nexists\, W^k(j) \in \mathcal{C}_k,\ j \in [1 : |\mathcal{M}'|],\ \text{s.t. } (u^k, W^k(j)) \in \mathcal{T}_k(\hat{U}_i \hat{W}_i), \\ P_{\hat{U}_i} = P_{u^k}, \hat{U}_i \hat{W}_i \in \mathcal{D}_k^{UW}(\eta) \Big\}.$$

(ii) $\hat{M}' = M'$.

(iii) $M' \neq 0$ and $\hat{M} \neq M$.

(iv) $M' = M = 0$ and $\hat{M} \neq M$.

(v) $M' \neq 0$, $\hat{M} = M$ and $\hat{M}' \neq M'$.

Here, (i) corresponds to the event that there does not exist a quantization codeword corresponding to atleast one sequence u^k of type $P_{u^k} \in \mathcal{D}_k^U(\eta)$; (ii) corresponds to the event, in which, there is neither an error at the channel decoder nor at the ECED; (iii) and (iv) corresponds to the case, in which, there is an error at the channel decoder (hence also at the ECED); and, (v) corresponds to the case such that there is an error only at the ECED.

As we show later in (B.72), it follows by a generalization of the *type-covering lemma* [20, Lemma 9.1] that

$$\mathbb{P}(\mathcal{E}_{EE}) \leq e^{-e^{k\Omega(\eta)}}. \tag{B.35}$$

Since $\frac{e^{k\Omega(\eta)}}{k} \xrightarrow{(k)} \infty$ for $\eta > 0$, we may safely ignore this event from the analysis of the exponent of type I and type II error probability. Given $\mathcal{E}_{EE}^c$ and that event (ii) holds, it follows from [27, Equation 4.22] that for any given codebook $\mathcal{C}_k$, the type I error probability is asymptotically upper bounded by $e^{-k\kappa_\alpha}$, since the acceptance region is the same. Hence, it also holds when averaged over the random quantization codebooks such that $\mathcal{E}_{EE}^c$ holds, implying that

$$\mathbb{P}\left(\hat{H} = 1 | \mathcal{E}_{EE}^c, \hat{M}' = M'\right) \leq e^{-k\kappa_\alpha}. \tag{B.36}$$

Next, consider event (iii). By the design of the channel codebook $\mathcal{C}_X^n$, it holds asymptotically that

$$\begin{aligned}
\mathbb{P}\left(M' \neq 0, \hat{M} \neq M | H = 0\right) &= \mathbb{P}\left(M' \neq 0 | H = 0\right) \mathbb{P}\left(\hat{M} \neq M | M \neq 0\right) \\
&\leq \mathbb{P}\left(\hat{M} \neq M | M \neq 0\right) \\
&\leq \mathbb{P}\left(\hat{M} = 0 | M \neq 0\right) + \mathbb{P}\left(\hat{M} \neq M | M \neq 0, \hat{M} \neq 0\right)
\end{aligned}$$

$$\leq e^{-k\tau(E_m(P_{SX},\theta)-\eta)} + e^{-k\tau\left(E_x\left(\frac{R}{\tau},P_{SX}\right)-\eta\right)} \tag{B.37}$$

$$= e^{-k\tau\left(\min\left(E_m(P_{SX},\theta),E_x\left(\frac{R}{\tau},P_{SX}\right)\right)-\eta\right)}, \tag{B.38}$$

where, in (B.37), we used (B.29) and (B.31). Also, note by the definition of $\mathcal{D}_k(\hat{U})$ and (B.35) that the probability of event (iv) can be upper bounded as

$$\mathbb{P}\left(M=0,\hat{M}\neq M|H=0\right) \leq \mathbb{P}\left(M'=0|H=0\right) \leq e^{-k\kappa_\alpha}. \tag{B.39}$$

Next, consider the event (v). Note that this event is impossible when $R'+\eta \leq R$, since there is no binning involved. Hence, assume that $R'+\eta > R$. Since $M=0$ iff $M'=0$, $M'\neq 0$ and $\hat{M}=M$ implies that $\hat{M}\neq 0$. Let

$$\mathcal{D}_k^{VW}(\eta) := \Big\{\hat{V}\hat{W}:\ \exists\,(w^k,u^k,v^k)\in \underset{m'\in\mathcal{M}'\setminus\{0\}}{\cup}\mathcal{J}_k^{\kappa_\alpha+\eta}(W_{m'}UV), W_{m'}UV \text{ satisfies (B.32)}$$
$$\text{and } P_{w^ku^kv^k}=P_{\hat{W}\hat{U}\hat{V}}\Big\}.$$

We can write,

$$\begin{aligned}
&\mathbb{P}\left(M'\neq 0,\hat{M}=M,\hat{M}'\neq M'|H=0\right)\\
&=\mathbb{P}\left(M'\neq 0,\hat{M}=M,\hat{M}'\neq M',(M',V^k)\in\mathcal{A}_k|H=0\right)\\
&\quad+\mathbb{P}\left(M'\neq 0,\hat{M}=M,\hat{M}'\neq M',(M',V^k)\notin\mathcal{A}_k|H=0\right).
\end{aligned} \tag{B.40}$$

The second term in (B.40) can be upper-bounded as

$$\begin{aligned}
&\mathbb{P}\left(M'\neq 0,\hat{M}=M,\hat{M}'\neq M',(M',V^k)\notin\mathcal{A}_k|H=0\right)\\
&\leq \mathbb{P}\left((M',V^k)\notin\mathcal{A}_k,\mathcal{E}_{EE}|H=0\right)+\mathbb{P}\left((M',V^k)\notin\mathcal{A}_k,\mathcal{E}_{EE}^c|H=0\right)\\
&\leq e^{-e^{k\Omega(\eta)}}+\mathbb{P}\left((M',V^k)\notin\mathcal{A}_k|\mathcal{E}_{EE}^c,H=0\right)\\
&\leq e^{-e^{k\Omega(\eta)}}+\mathbb{P}\left((U^k,V^k)\notin\mathcal{A}_k^e\right)\\
&\leq e^{-e^{k\Omega(\eta)}}+e^{-k\kappa_\alpha},
\end{aligned} \tag{B.41}$$

where, the inequality in (B.41) follows from [27, Equation 4.22] for sufficiently large k, since the acceptance region is the same. Let

$$\mathcal{D}_k(V) := \{\hat{V} : \exists\ \hat{W} \text{ s.t. } \hat{V}\hat{W} \in \mathcal{D}_k^{VW}(\eta)\}.$$

The first term in (B.40) can be bounded as shown below:

$$\begin{aligned}
&\mathbb{P}\left(M' \neq 0, \hat{M} = M, \hat{M}' \neq M', (M', V^k) \in \mathcal{A}_k | H = 0\right) \\
&\leq \sum_{\substack{v^k \in \mathcal{T}_k(\hat{V}): \\ \hat{V} \in \mathcal{D}_k(V)}} \mathbb{P}\Big(V^k = v^k, \exists\ j \in f_B^{-1}(M),\ j \neq M' : H_e(W^k(j)|v^k) \leq \\
&\qquad\qquad H_e(W^k(M')|v^k) | M' \neq 0\Big) \\
&= \sum_{\substack{v^k \in \mathcal{T}_k(\hat{V}): \\ \hat{V} \in \mathcal{D}_k(V)}} \mathbb{P}\left(V^k = v^k | M' \neq 0\right) \mathbb{P}\Big(\exists\ j \in f_B^{-1}(M),\ j \neq M' : H_e(W^k(j)|v^k) \leq \\
&\qquad\qquad H_e(W^k(M')|v^k) | V^k = v^k, M' \neq 0\Big)
\end{aligned} \tag{B.42}$$

Defining the events

$$\begin{aligned}
\mathcal{E}_1' &:= \{V^k = v^k, M' \neq 0\}, \\
\mathcal{E}_2' &:= \{V^k = v^k, M' = m' \neq 0, M = m\},
\end{aligned}$$

we can write

$$\begin{aligned}
&\mathbb{P}\left(\exists\ j \in f_B^{-1}(M),\ j \neq M' : H_e(W^k(j)|v^k) \leq H_e(W^k(M')|v^k) | \mathcal{E}_1'\right) \\
&= \sum_{\substack{m' \in \\ \mathcal{M}' \setminus \{0\}}} \sum_{\substack{m \in \\ \mathcal{M} \setminus \{0\}}} \mathbb{P}\left(M' = m', M = m | \mathcal{E}_1'\right)\ \mathbb{P}\Big(\exists\ j \in f_B^{-1}(m),\ j \neq m' : \\
&\qquad\qquad H_e(W^k(j)|v^k) \leq H_e(W^k(m')|v^k) | \mathcal{E}_2'\Big).
\end{aligned} \tag{B.43}$$

Consider the second term in (B.43). Denoting the type of v^k by $\hat{V}$, it follows that

$$\begin{aligned}
&\mathbb{P}\left(\exists\ j \in f_B^{-1}(m),\ j \neq m' : H_e(W^k(j)|v^k) \leq H_e(W^k(m')|v^k) | \mathcal{E}_2'\right) \\
&= \sum_{j \in \mathcal{M}' \setminus \{0, m'\}} \mathbb{P}\left(f_B(j) = m : H_e(W^k(j)|v^k) \leq H_e(W^k(m')|v^k) | \mathcal{E}_2'\right)
\end{aligned}$$

$$\leq \frac{1}{e^{kR_k}} \sum_{j \in \mathcal{M}' \setminus \{0,m'\}} \mathbb{P}\left(H_e(W^k(j)|v^k) \leq H_e(W^k(m')|v^k) | \mathcal{E}_2' \cup \{f_B(j) = m\} \right) \tag{B.44}$$

$$\leq \frac{1}{e^{kR_k}} \sum_{j \in \mathcal{M}' \setminus \{0,m'\}} \sum_{\substack{\hat{W}: \\ \hat{V}\hat{W} \in \mathcal{D}_k^{VW}(\eta)}} \sum_{\substack{w^k: \\ (v^k, w^k) \in \mathcal{T}_k(\hat{V}\hat{W})}} \mathbb{P}\left(W^k(m') = w^k | \mathcal{E}_2' \cup \{f_B(j) = m\} \right)$$

$$\sum_{\substack{\tilde{w}^k \in \mathcal{T}_k(\hat{W}) \\ H_e(\tilde{w}^k|v^k) \leq H(\hat{W}|\hat{V})}} \mathbb{P}\left(W^k(j) = \tilde{w}^k | \mathcal{E}_2' \cup \{f_B(j) = m\} \cup \{W^k(m') = w^k\} \right) \tag{B.45}$$

$$\leq \frac{1}{e^{kR_k}} \sum_{j \in \mathcal{M}' \setminus \{0,m'\}} \sum_{\substack{\hat{W}: \\ \hat{V}\hat{W} \in \mathcal{D}_k^{VW}(\eta)}} \sum_{\substack{w^k: \\ (v^k, w^k) \in \mathcal{T}_k(\hat{V}\hat{W})}} \mathbb{P}\left(W^k(m') = w^k | \mathcal{E}_2' \cup \{f_B(j) = m\} \right)$$

$$\sum_{\substack{\tilde{w}^k \in \mathcal{T}_k(\hat{W}): \\ H_e(\tilde{w}^k|v^k) \leq H(\hat{W}|\hat{V})}} 2\, \mathbb{P}\left(W^k(j) = \tilde{w}^k \right). \tag{B.46}$$

In (B.44), we used the fact that binning is done uniformly at random; in (B.45), we used the following: if $v^k \in \mathcal{T}_k(\hat{V})$ is such that $\hat{V} \in \mathcal{D}_k(V)$, then $M' \neq 0$ implies that $(W^k(M'), v^k) \in \mathcal{T}_k(\hat{V}\hat{W})$ for some $\hat{V}\hat{W} \in \mathcal{D}_k^{VW}(\eta)$. In (B.46), we used

$$\mathbb{P}\left(W^k(j) = \tilde{w}^k | \mathcal{E}_2' \cup \{f_B(j) = m\} \cup \{W^k(m') = w^k\} \right) \leq 2\, \mathbb{P}\left(W^k(j) = \tilde{w}^k \right), \tag{B.47}$$

which will be shown later. Continuing, we can write (for sufficiently large k)

$$\mathbb{P}\left(\exists\, j \in f_B^{-1}(m),\ j \neq m' : H_e(W^k(j)|v^k) \leq H_e(W^k(m')|v^k) | \mathcal{E}_2' \right)$$

$$\leq \frac{1}{e^{kR_k}} \sum_{j \in \mathcal{M}' \setminus \{0,m'\}} \sum_{\substack{\hat{W}: \\ \hat{V}\hat{W} \in \\ \mathcal{D}_k^{VW}(\eta)}} \sum_{\substack{w^k: \\ (v^k, w^k) \in \\ \mathcal{T}_k(\hat{V}\hat{W})}} \mathbb{P}\left(W^k(m') = w^k | \mathcal{E}_2' \cup \{f_B(j) = m\} \right)$$

$$\sum_{\substack{\tilde{w}^k \in T_{\hat{W}}: \\ H_e(\tilde{w}^k|v^k) \\ \leq H(\hat{W}|\hat{V})}} 2\, e^{-k(H(\hat{W}) - \eta)} \tag{B.48}$$

$$\leq \frac{1}{e^{kR_k}} \sum_{j \in \mathcal{M}' \setminus \{0,m'\}} \sum_{\substack{\hat{W}: \\ \hat{V}\hat{W} \in \\ \mathcal{D}_k^{VW}(\eta)}} \sum_{\substack{w^k: \\ (v^k, w^k) \in \\ \mathcal{T}_k(\hat{V}\hat{W})}} \mathbb{P}\left(W^k(m') = w^k | \mathcal{E}_2' \cup \{f_B(j) = m\} \right)$$

$$(k+1)^{|\mathcal{V}||\mathcal{W}|} e^{kH(\hat{W}|\hat{V})} 2\, e^{-k(H(\hat{W}) - \eta)} \tag{B.49}$$

$$\leq \frac{1}{e^{kR_k}} \sum_{j \in \mathcal{M}' \setminus \{0, m'\}} \sum_{\substack{\hat{W}: \\ \hat{V}\hat{W} \in \mathcal{D}_k^{VW}(\eta)}} 2\ (k+1)^{|\mathcal{V}||\mathcal{W}|}\ e^{-k\left(I(\hat{W};\hat{V})-\eta\right)}$$

$$\leq \frac{1}{e^{kR_k}} \sum_{j \in \mathcal{M}' \setminus \{0, m'\}} 2\ (k+1)^{|\mathcal{W}|}\ (k+1)^{|\mathcal{V}||\mathcal{W}|}\ e^{-k\left(\min\limits_{\hat{V}\hat{W} \in \mathcal{D}_k^{VW}(\eta)} I(\hat{W};\hat{V})-\eta\right)} \tag{B.50}$$

$$\leq e^{-k(R-R'+\rho_k-\eta'_k)}, \tag{B.51}$$

where,

$$\rho_k := \min_{\hat{V}\hat{W} \in \mathcal{D}_k^{VW}(\eta)} I(\hat{V};\hat{W})$$
$$\text{and } \eta'_k := 3\eta + \frac{|\mathcal{W}|(|\mathcal{V}|+1)\log(k+1)}{k} + \frac{\log(2)}{k} + \frac{|\mathcal{U}|\log(k+1)}{k}.$$

In (B.48), we used [20, Lemma 2.3] and the fact that codewords are chosen uniformly at random from $\mathcal{T}_k(\hat{W})$; in (B.49), we used that the total number of sequences $\tilde{w}^k \in \mathcal{T}_k(\tilde{W})$ such that $P_{\tilde{w}^k v^k} = P_{\tilde{W}\tilde{V}}$ and $H(\tilde{W}|\tilde{V}) \leq H(\hat{W}|\hat{V})$ is upper bounded by $e^{kH(\hat{W}|\hat{V})}$ and $|\mathcal{T}_n(\mathcal{W} \times \mathcal{V})| \leq (k+1)^{|\mathcal{V}||\mathcal{W}|}$; in (B.50), we used [20, Lemma 2.2]; and, in (B.51), we used (B.19), (B.20), (B.26) and (B.27). Thus, for sufficiently large k, since $\rho_k \to \rho(\kappa_\alpha, \omega) + O(\eta)$, we have from (B.41), (B.42), (B.43) and (B.51) that for sufficiently large k,

$$\mathbb{P}\left(M' \neq 0, \hat{M} = M, \hat{M}' \neq M' | H = 0\right) \leq e^{-k(\min(\kappa_\alpha, R-\zeta_q(\kappa_\alpha,\omega)+\rho(\kappa_\alpha,\omega)-O(\eta)))}. \tag{B.52}$$

By choice of $(\omega, P_{SX}, \theta) \in \mathcal{L}(\kappa_\alpha, \tau)$, it follows from (B.35), (B.36), (B.38), (B.39) and (B.52) that the type I error probability is upper bounded by $e^{-k(\kappa_\alpha - O(\eta))}$, asymptotically.

Next, we analyze the type II error probability averaged over the random codebooks. For a given codebook $\mathcal{C}_k$, let $\tilde{U}, \tilde{V}, \tilde{W}$ and $\tilde{W}_d$ denote the type variable for the realizations of $\bar{U}^k, \bar{V}^k, W^k(M')$ $(M' \neq 0)$ and $W^k(\hat{M}')$ $(\hat{M}' \neq 0)$, respectively. A type II error can occur only under the following events:

(a) $\mathcal{E}_a := \{\hat{M} = M, \hat{M}' = M' \neq 0, (\bar{U}^k, \bar{V}^k, W^k(M')) \in \mathcal{T}_k(\hat{U}\hat{V}\hat{W})$ such that $\hat{U}\hat{W} \in \mathcal{D}_k^{UW}(\eta)$ and $\hat{V}\hat{W} \in \mathcal{D}_k^{VW}(\eta)\}$.

(b)

$$\mathcal{E}_b := \left\{ \begin{array}{l} M' \neq 0, \hat{M} = M, \hat{M}' \neq M', f_B(\hat{M}') = f_B(M'), \Big(\bar{U}^k, \bar{V}^k, W^k(M'), \\ W^k(\hat{M}')\Big) \in \mathcal{T}_k(\hat{U}\hat{V}\hat{W}\hat{W}_d) \text{ s.t. } \hat{U}\hat{W} \in \mathcal{D}_k^{UW}(\eta), \hat{V}\hat{W}_d \in \mathcal{D}_k^{VW}(\eta), \\ \text{and } H_e\left(W^k(\hat{M}')|\bar{V}^k\right) \leq H_e\left(W^k(M')|\bar{V}^k\right) \end{array} \right\}.$$

(c)

$$\mathcal{E}_c := \left\{ \begin{array}{l} M' \neq 0, \hat{M} \neq M \text{ or } 0, (\bar{U}^k, \bar{V}^k, W^k(M'), W^k(\hat{M}')) \in \mathcal{T}_k(\hat{U}\hat{V}\hat{W}\hat{W}_d) \\ \text{such that } \hat{U}\hat{W} \in \mathcal{D}_k^{UW}(\eta) \text{ and } \hat{V}\hat{W}_d \in \mathcal{D}_k^{VW}(\eta) \end{array} \right\}.$$

(d) $\mathcal{E}_d := \{M = M' = 0, \hat{M} \neq M, (\bar{V}^k, W^k(\hat{M}')) \in \mathcal{T}_k(\hat{V}\hat{W}_d) \text{ s.t. } \hat{V}\hat{W}_d \in \mathcal{D}_k^{VW}(\eta)\}$.

Since the exponent of probability of the event $\mathcal{E}_{EE}$ tends to ∞ with k by (B.35), we may assume that $\mathcal{E}_{EE}^c$ holds for the type II error-exponent analysis. It then follows from the analysis in [27, Eq. 4.23-4.27] that for sufficiently large k, we have

$$\mathbb{P}\left(\mathcal{E}_a|\mathcal{E}_{EE}^c\right) \leq e^{-k(E_1(\kappa_\alpha,\omega)-O(\eta))}. \tag{B.53}$$

When $R' + \eta \leq R$, note that $\mathcal{E}_b$ is impossible, and hence, the exponent of this event is ∞. Assume that $R' + \eta > R$. Let

$$\begin{aligned} \mathcal{F}_{2,k}(\eta) :=& \{\tilde{U}\tilde{V}\tilde{W}\tilde{W}_d \in \mathcal{T}_k(\mathcal{U} \times \mathcal{V} \times \mathcal{W} \times \mathcal{W}) : \tilde{U}\tilde{W} \in \mathcal{D}_k^{UW}(\eta), \tilde{V}\tilde{W}_d \in \\ & \mathcal{D}_k^{VW}(\eta) \text{ and } H(\tilde{W}_d|\tilde{V}) \leq H(\tilde{W}|\tilde{V})\}. \end{aligned} \tag{B.54}$$

Then, we can write

$$\begin{aligned} &\mathbb{P}\left(\mathcal{E}_b|H=1\right) \\ &\leq \sum_{\substack{\tilde{U}\tilde{V}\tilde{W}\tilde{W}_d \\ \in \mathcal{F}_{2,k}(\eta)}} \sum_{\substack{(u^k,v^k,w^k,\bar{w}^k) \\ \in \mathcal{T}_k(\tilde{U}\tilde{V}\tilde{W}\tilde{W}_d)}} \sum_{m' \in \mathcal{M}' \setminus \{0\}} \mathbb{P}\Big(\bar{U}^k = u^k, \bar{V}^k = v^k, M' = m', \\ &\qquad\qquad W^k(M') = w^k|H=1\Big) \\ &\Bigg[\sum_{\hat{m}' \in \mathcal{M}' \setminus \{0\}} \mathbb{P}\Big(W^k(\hat{m}') = \bar{w}^k, f_B(m') = f_B(\hat{m}')|\bar{U}^k = u^k, \bar{V}^k = v^k, \end{aligned}$$

$$M' = m', W^k(m') = w^k\Big)\Big] \tag{B.55}$$

The first term in (B.55) can be written as

$$\begin{aligned}
&\mathbb{P}\left(\bar{U}^k = u^k, \bar{V}^k = v^k, M' = m', W^k(M') = w^k | H = 1\right) \\
&= \mathbb{P}\left(\bar{U}^k = u^k, \bar{V}^k = v^k, M' = m' | H = 1\right) \\
&\quad \mathbb{P}\left(W^k(m') = w^k | \bar{U}^k = u^k, \bar{V}^k = v^k, M' = m'\right)
\end{aligned} \tag{B.56}$$

Note that $M' \neq 0$ and $\bar{U}^k = u^k$ implies that $\tilde{U}\tilde{W} \in \mathcal{D}_k(UW)$. Hence, we can bound the second term in (B.55) for sufficiently large k as

$$\begin{aligned}
&\mathbb{P}\left(W^k(m') = w^k | \bar{U}^k = u^k, \bar{V}^k = v^k, M' = m'\right) \\
&\leq \begin{cases} \frac{1}{e^{k\left(H(\tilde{W}|\tilde{U})-\eta\right)}}, & \text{if } w^k \in \mathcal{T}_k(\tilde{W}), \\ 0, & \text{otherwise,} \end{cases}
\end{aligned} \tag{B.57}$$

where we used the fact that given $M' = m'$ and $\bar{U}^k = u^k$, $W^k(m')$ is uniformly distributed in the set $\mathcal{T}_k(P_{\tilde{W}|\tilde{U}}, u^k)$ and that for sufficiently large k,

$$|\mathcal{T}_k(P_{\tilde{W}|\tilde{U}}, u^k)| \geq e^{k\left(H(\tilde{W}|\tilde{U})-\eta\right)}.$$

On the other hand, the second term in (B.55) can be bounded as follows:

$$\begin{aligned}
&\mathbb{P}\left(W^k(\hat{m}') = \bar{w}^k, f_B(m') = f_B(\hat{m}') | \bar{U}^k = u^k, \bar{V}^k = v^k, M' = m', W^k(m') = w^k\right) \\
&\leq \frac{1}{e^{kR_k}} \mathbb{P}\left(W^k(\hat{m}') = \bar{w}^k | \bar{U}^k = u^k, \bar{V}^k = v^k, M' = m', W^k(m') = w^k\right) &\text{(B.58)} \\
&\leq \frac{2}{e^{kR_k}} \mathbb{P}\left(W^k(\hat{m}') = \bar{w}^k\right), &\text{(B.59)}
\end{aligned}$$

where, in (B.58), we used the fact that the binning is uniformly distributed and independent of the codebook generation; in (B.59), we used

$$\begin{aligned}
&\mathbb{P}\left(W^k(\hat{m}') = \bar{w}^k | \bar{U}^k = u^k, \bar{V}^k = v^k, M' = m', W^k(m') = w^k\right) \\
&\leq 2\ \mathbb{P}\left(W^k(\hat{m}') = \bar{w}^k\right).
\end{aligned} \tag{B.60}$$

which will be shown below. Thus, from (B.57) and (B.59), we can bound the term in (B.55) (for sufficiently large k) as

$$
\begin{aligned}
&\mathbb{P}\left(\mathcal{E}_b|H=1\right)\\
&\leq \frac{2}{e^{kR_k}} \sum_{\substack{\tilde{U}\tilde{V}\tilde{W}\tilde{W}_d\\ \in\mathcal{F}_{2,k}(\eta)}} \sum_{\substack{(u^k,v^k,w^k,\bar{w}^k)\\ \in\mathcal{T}_k(\tilde{U}\tilde{V}\tilde{W}\tilde{W}_d)}} \sum_{m'\in\mathcal{M}'\setminus\{0\}} \mathbb{P}\left(\bar{U}^k=u^k,\bar{V}^k=v^k,M'=m'|H=1\right)\\
&\qquad\qquad \frac{1}{e^{k\left(H(\tilde{W}|\tilde{U})-\eta\right)}} \sum_{\hat{m}'\in\mathcal{M}'\setminus\{0\}} \mathbb{P}\left(W^k(\hat{m}')=\bar{w}^k\right)\\
&\leq \frac{2}{e^{kR_k}} \sum_{\substack{\tilde{U}\tilde{V}\tilde{W}\tilde{W}_d\\ \in\mathcal{F}_{2,k}(\eta)}} \sum_{\substack{(u^k,v^k,w^k,\bar{w}^k)\\ \in\mathcal{T}_k(\tilde{U}\tilde{V}\tilde{W}\tilde{W}_d)}} e^{-k\left(H(\tilde{U}\tilde{V})+D(\tilde{U}\tilde{V}||\bar{U}\bar{V})\right)} \frac{1}{e^{k\left(H(\tilde{W}|\tilde{U})-\eta\right)}}\\
&\sum_{m'\in\mathcal{M}'\setminus\{0\}} \mathbb{P}\left(M'=m'|\bar{U}^k=u^k,\bar{V}^k=v^k\right) \sum_{\hat{m}'\in\mathcal{M}'\setminus\{0\}} \mathbb{P}\left(W^k(\hat{m}')=\bar{w}^k\right)\\
&\leq \frac{2}{e^{kR_k}} \sum_{\substack{\tilde{U}\tilde{V}\tilde{W}\tilde{W}_d\\ \in\mathcal{F}_{2,k}(\eta)}} \sum_{\substack{(u^k,v^k,w^k,\bar{w}^k)\\ \in\mathcal{T}_k(\tilde{U}\tilde{V}\tilde{W}\tilde{W}_d)}} e^{-k\left(H(\tilde{U}\tilde{V})+D(\tilde{U}\tilde{V}||\bar{U}\bar{V})\right)} \frac{1}{e^{k\left(H(\tilde{W}|\tilde{U})-\eta\right)}}\\
&\qquad\qquad \sum_{\hat{m}'\in\mathcal{M}'\setminus\{0\}} \mathbb{P}\left(W^k(\hat{m}')=\bar{w}^k\right)\\
&\leq \frac{2}{e^{kR_k}} \sum_{\substack{\tilde{U}\tilde{V}\tilde{W}\tilde{W}_d\\ \in\mathcal{F}_{2,k}(\eta)}} \sum_{\substack{(u^k,v^k,w^k,\bar{w}^k)\\ \in\mathcal{T}_k(\tilde{U}\tilde{V}\tilde{W}\tilde{W}_d)}} e^{-k\left(H(\tilde{U}\tilde{V})+D(\tilde{U}\tilde{V}||\bar{U}\bar{V})\right)} \frac{1}{e^{k\left(H(\tilde{W}|\tilde{U})-\eta\right)}} \frac{e^{k(R'+\eta)}}{e^{k\left(H(\tilde{W}_d)-\eta\right)}}\\
&\leq \frac{2}{e^{kR_k}} \sum_{\substack{\tilde{U}\tilde{V}\tilde{W}\tilde{W}_d\\ \in\mathcal{F}_{2,k}(\eta)}} \sum_{\substack{(u^k,v^k,w^k)\\ \in\mathcal{T}_k(\tilde{U}\tilde{V}\tilde{W})}} e^{-k\left(H(\tilde{U}\tilde{V})+D(\tilde{U}\tilde{V}||\bar{U}\bar{V})\right)} \frac{1}{e^{k\left(H(\tilde{W}|\tilde{U})-\eta\right)}} \frac{e^{k(R'+\eta)}}{e^{k\left(H(\tilde{W}_d)-\eta\right)}}\\
&\qquad\qquad e^{kH(\tilde{W}_d|\tilde{V})}\\
&\leq \frac{2}{e^{kR_k}} \sum_{\substack{\tilde{U}\tilde{V}\tilde{W}\tilde{W}_d\\ \in\mathcal{F}_{2,k}(\eta)}} e^{kH(\tilde{U}\tilde{V}\tilde{W})} e^{-k\left(H(\tilde{U}\tilde{V})+D(\tilde{U}\tilde{V}||\bar{U}\bar{V})\right)} \frac{1}{e^{k\left(H(\tilde{W}|\tilde{U})-\eta\right)}} \frac{e^{k(R'+\eta)}}{e^{k\left(H(\tilde{W}_d)-\eta\right)}}\\
&\qquad\qquad e^{kH(\tilde{W}_d|\tilde{V})}\\
&\leq e^{-kE_{2,k}},
\end{aligned}
$$

where

$$
\begin{aligned}
E_{2,k} := \min_{\tilde{U}\tilde{V}\tilde{W}\tilde{W}_d\in\mathcal{F}_{2,k}(\eta)} &-H(\tilde{U}\tilde{V}\tilde{W})+H(\tilde{U}\tilde{V})+D(\tilde{U}\tilde{V}||\bar{U}\bar{V})+H(\tilde{W}|\tilde{U})\\
&+I(\tilde{V};\tilde{W}_d)+R-R'-3\eta-\delta'_k,
\end{aligned}
$$

$$\delta'_k := \frac{|\mathcal{U}||\mathcal{V}||\mathcal{W}|^2}{k}\log(k+1) + \frac{|\mathcal{U}|}{k}\log(k+1) + \frac{\log(2)}{k}. \tag{B.61}$$

Note that since $\tilde{U}\tilde{V}\tilde{W}\tilde{W}_d \in \mathcal{F}_{2,k}(\eta)$ implies that $\tilde{V}\tilde{W}_d \in \mathcal{D}_k^{VW}(\eta)$, we have

$$\begin{aligned} E_{2,k} \geq \min_{\tilde{U}\tilde{V}\tilde{W}\tilde{W}_d \in \mathcal{F}_{2,k}(\eta)} & -H(\tilde{U}\tilde{V}\tilde{W}) + H(\tilde{U}\tilde{V}) + D(\tilde{U}\tilde{V}||\bar{U}\bar{V}) + H(\tilde{W}|\tilde{U}) \\ & + \rho_k + R - R' - 3\eta - \delta'_k. \end{aligned} \tag{B.62}$$

Simplifying the terms in (B.62) and using $\rho_k \xrightarrow{(k)} \rho(\kappa_\alpha, \omega) + O(\eta)$, we obtain by the continuity of KL-divergence that

$$\begin{aligned} & \frac{-1}{k}\log\left(\mathbb{P}\left(\mathcal{E}_b|H=1\right)\right) \\ & \overset{(k)}{\geq} \begin{cases} \min\limits_{\tilde{U}\tilde{V}\tilde{W}\in\mathcal{T}_2(\kappa_\alpha,\omega)} D(\tilde{U}\tilde{V}\tilde{W}||\bar{U}\bar{V}\bar{W}) + E_b(\kappa_\alpha,\omega,R) - O(\eta), & \text{if } R < \zeta_q(\kappa_\alpha,\omega)+\eta, \\ \infty, & \text{otherwise,} \end{cases} \\ & = E_2(\kappa_\alpha,\omega,R) - O(\eta). \end{aligned} \tag{B.63}$$

Next, consider the event $\mathcal{E}_c$. Assume that $R' + \eta > R$ (i.e., binning is required). Let

$$\mathcal{F}_{3,k}(\eta) := \{\tilde{U}\tilde{V}\tilde{W}\tilde{W}_d \in \mathcal{T}_k(\mathcal{U}\times\mathcal{V}\times\mathcal{W}\times\mathcal{W}) : \tilde{U}\tilde{W} \in \mathcal{D}_k^{UW}(\eta) \text{ and } \tilde{V}\tilde{W}_d \in \mathcal{D}_k^{VW}(\eta)\}.$$

Then, we can write (for sufficiently large k) that,

$$\begin{aligned} & \mathbb{P}\left(\mathcal{E}_c|H=1\right) \\ & \leq \sum_{\substack{\tilde{U}\tilde{V}\tilde{W}\tilde{W}_d \\ \in\mathcal{F}_{3,k}(\eta)}} \sum_{\substack{(u^k,v^k,w^k,\bar{w}^k) \\ \in\mathcal{T}_k(\tilde{U}\tilde{V}\tilde{W}\tilde{W}_d)}} \sum_{m'\in\mathcal{M}'\setminus\{0\}} \mathbb{P}\left(\bar{U}^k = u^k, \bar{V}^k = v^k, M' = m', W^k(M') = w^k | H=1\right) \\ & \quad \sum_{\substack{m\neq 0, \hat{m}\neq 0: \\ \hat{m}\neq m}} \mathbb{P}\left(M=m|H=1\right)\mathbb{P}\left(\hat{M}=\hat{m}|M=m\right) \\ & \quad \left[\sum_{\hat{m}'\in\mathcal{M}'\setminus\{0\}} \mathbb{P}\left(W^k(\hat{m}') = \bar{w}^k, f_B(\hat{m}') = \hat{m} | \bar{U}^k = u^k, \bar{V}^k = v^k, M' = m', W^k(m') = w^k\right)\right] \\ & \leq \frac{2}{e^{kR_k}} \sum_{\substack{\tilde{U}\tilde{V}\tilde{W}\tilde{W}_d \\ \in\mathcal{F}_{3,k}(\eta)}} e^{kH(\tilde{U}\tilde{V}\tilde{W})} e^{-k\left(H(\tilde{U}\tilde{V}) + D(\tilde{U}\tilde{V}||\bar{U}\bar{V})\right)} \frac{1}{e^{k\left(H(\tilde{W}|\tilde{U})-\eta\right)}} \frac{e^{k(R'+\eta)}}{e^{k\left(H(\tilde{W}_d)-\eta\right)}} \end{aligned}$$

$$e^{kH(\tilde{W}_d|\tilde{V})}\ e^{-k\tau\left(E_x\left(\frac{R}{\tau},P_{SX}\right)-\eta\right)} \tag{B.64}$$

$$\leq e^{-kE_{3,k}},$$

where,

$$E_{3,k} := \min_{\tilde{U}\tilde{V}\tilde{W}\tilde{W}_d\in\mathcal{F}_{3,k}(\eta)} -H(\tilde{U}\tilde{V}\tilde{W}) + H(\tilde{U}\tilde{V}) + D(\tilde{U}\tilde{V}||\bar{U}\bar{V}) + H(\tilde{W}|\tilde{U}) + \tau E_x\left(\frac{R}{\tau},P_{SX}\right) + \rho_k + R - R' - O(\eta) - \delta'_k,$$

and δ'_k is as defined in (B.61). To obtain (B.64), we used (B.31), (B.57) and (B.59). On the other hand, if $R' + \eta \leq R$, it can be shown similarly that,

$$\mathbb{P}\left(\mathcal{E}_c|H=1\right) \leq e^{-kE'_{3,k}},$$

where

$$E'_{3,k} := \min_{\tilde{U}\tilde{V}\tilde{W}\tilde{W}_d\in\mathcal{F}_{3,k}(\eta)} -H(\tilde{U}\tilde{V}\tilde{W}) + H(\tilde{U}\tilde{V}) + D(\tilde{U}\tilde{V}||\bar{U}\bar{V}) + H(\tilde{W}|\tilde{U}) + \tau E_x\left(\frac{R}{\tau},P_{SX}\right) + \rho_k - O(\eta) - \frac{|\mathcal{U}||\mathcal{V}||\mathcal{W}|^2}{k}\log(k+1) - \frac{\log(2)}{k}.$$

Hence, we obtain

$$\frac{-1}{k}\log\left(\mathbb{P}\left(\mathcal{E}_c|H=1\right)\right)$$

$$\overset{(k)}{\geq} \begin{cases} \min\limits_{\tilde{U}\tilde{V}\tilde{W}\in\mathcal{T}_3(\kappa_\alpha,\omega)} D(\tilde{U}\tilde{V}\tilde{W}||\bar{U}\bar{V}\bar{W}) + E_b(\kappa_\alpha,\omega,R) & \\ \qquad +\tau E_x\left(\frac{R}{\tau},P_{SX}\right) - O(\eta), & \text{if } R < \zeta_q(\kappa_\alpha,\omega)+\eta, \\ \min\limits_{\tilde{U}\tilde{V}\tilde{W}\in\mathcal{T}_3(\kappa_\alpha,\omega)} D(\tilde{U}\tilde{V}\tilde{W}||\bar{U}\bar{V}\bar{W}) + \rho(\kappa_\alpha,\omega) & \\ \qquad +\tau E_x\left(\frac{R}{\tau},P_{SX}\right) - O(\eta), & \text{otherwise,} \end{cases}$$

$$= E_3(\kappa_\alpha,\omega,R,P_{SX},\tau) - O(\eta). \tag{B.65}$$

Finally, we consider the event $\mathcal{E}_d$. Assume that $R' + \eta > R$. We have

$$\mathbb{P}\left(\mathcal{E}_d|H=1\right) = \sum_{\substack{u^k \in \mathcal{T}_k(\tilde{U}):\\ \tilde{U} \in \mathcal{D}_k^U(\eta)}} \mathbb{P}\left(\bar{U}^k = u^k, \mathcal{E}_{EE}, \mathcal{E}_d|H=1\right) + \sum_{\substack{u^k \in \mathcal{T}_k(\tilde{U}):\\ \tilde{U} \notin \mathcal{D}_k^U(\eta)}} \mathbb{P}\left(\bar{U}^k = u^k, \mathcal{E}_d|H=1\right), \tag{B.66}$$

where, (B.66) follows from the fact that if $\tilde{U} \in \mathcal{D}_k^U(\eta)$, then $\mathcal{E}_d$ can occur only if $\mathcal{E}_{EE}$ occurs. From (B.35), for any $u^k \in \mathcal{T}_k(\tilde{U})$ such that $\tilde{U} \in \mathcal{D}_k^U(\eta)$, we have

$$\mathbb{P}\left(\bar{U}^k = u^k, \mathcal{E}_{EE}, \mathcal{E}_d|H=1\right) \leq e^{-e^{k\Omega(\eta)}}.$$

Next, note that if $\tilde{U} \notin \mathcal{D}_k^U(\eta)$, then $M' = 0$ is chosen with probability 1 independent of the codebook $\mathcal{C}_k$. Hence, we can write the second term in (B.66) as follows:

$$\begin{aligned}
&\sum_{\substack{u^k \in \mathcal{T}_k(\tilde{U}):\\ \tilde{U} \notin \mathcal{D}_k^U(\eta)}} \mathbb{P}\left(\bar{U}^k = u^k, \mathcal{E}_d|H=1\right) \\
&\leq \sum_{\substack{u^k \in \mathcal{T}_k(\tilde{U}):\\ \tilde{U} \notin \mathcal{D}_k^U(\eta)}} \sum_{\substack{(v^k,\bar{w}^k) \in \mathcal{T}_k(\tilde{V}\tilde{W}_d):\\ \tilde{V}\tilde{W}_d \in \mathcal{D}_k^{VW}(\eta)}} \sum_{\hat{m} \in \mathcal{M}\setminus\{0\}} \mathbb{P}\left(\bar{U}^k = u^k, \bar{V}^k = v^k, M' = M = 0|H=1\right) \\
&\qquad \mathbb{P}\left(\hat{M} = \hat{m}|M=0\right) \sum_{\hat{m}' \in \mathcal{M}'\setminus\{0\}} \mathbb{P}\left(f_B(\hat{m}') = \hat{m}, W^k(\hat{m}') = \bar{w}^k\right) \\
&\leq \sum_{\substack{u^k \in \mathcal{T}_k(\tilde{U}):\\ \tilde{U} \notin \mathcal{D}_k^U(\eta)}} \sum_{\substack{(v^k,\bar{w}^k) \in \mathcal{T}_k(\tilde{V}\tilde{W}_d):\\ \tilde{V}\tilde{W}_d \in \mathcal{D}_k^{VW}(\eta)}} \mathbb{P}\left(\bar{U}^k = u^k, \bar{V}^k = v^k, M' = M = 0|H=1\right) \\
&\qquad \sum_{\hat{m} \in \mathcal{M}\setminus\{0\}} \mathbb{P}\left(\hat{M} = \hat{m}|M=0\right) \sum_{\hat{m}' \in \mathcal{M}'\setminus\{0\}} \frac{1}{e^{kR_k}} \frac{1}{e^{k\left(H(\tilde{W}_d)-\eta\right)}} \\
&\leq \sum_{\substack{u^k \in \mathcal{T}_k(\tilde{U}):\\ \tilde{U} \notin \mathcal{D}_k^U(\eta)}} \sum_{\substack{(v^k,\bar{w}^k) \in \mathcal{T}_k(\tilde{V}\tilde{W}_d):\\ \tilde{V}\tilde{W}_d \in \mathcal{D}_k^{VW}(\eta)}} \mathbb{P}\left(\bar{U}^k = u^k, \bar{V}^k = v^k, M' = M = 0|H=1\right) \\
&\qquad \sum_{\hat{m} \in \mathcal{M}\setminus\{0\}} \mathbb{P}\left(\hat{M} = \hat{m}|M=0\right) \frac{e^{k(R'+\eta)}}{e^{kR_k}} \frac{1}{e^{k\left(H(\tilde{W}_d)-\eta\right)}} \\
&\leq \sum_{\substack{u^k \in \mathcal{T}_k(\tilde{U}):\\ \tilde{U} \notin \mathcal{D}_k^U(\eta)}} \sum_{\substack{(v^k,\bar{w}^k) \in \mathcal{T}_k(\tilde{V}\tilde{W}_d):\\ \tilde{V}\tilde{W}_d \in \mathcal{D}_k^{VW}(\eta)}} \mathbb{P}\left(\bar{U}^k = u^k, \bar{V}^k = v^k, M' = M = 0|H=1\right)
\end{aligned}$$

$$
\begin{aligned}
& e^{-k\tau(E_m(P_{SX},\theta)-\theta-\eta)}\ \frac{e^{k(R'+\eta)}}{e^{kR_k}}\ \frac{1}{e^{k\left(H(\tilde{W}_d)-\eta\right)}} \qquad \text{(B.67)}\\
\leq \sum_{\substack{\tilde{U}\tilde{V}\tilde{W}_d\\ \in\mathcal{D}_k^U(\eta)^c\times\mathcal{D}_k^{VW}(\eta)}} & e^{kH(\tilde{U}\tilde{V})}e^{-k\left(H(\tilde{U}\tilde{V})+D(\tilde{U}\tilde{V}||\bar{U}\bar{V})\right)}\ e^{-k\tau(E_m(P_{SX},\theta)-\theta-\eta)}\\
& \frac{e^{k(R'+\eta)}}{e^{kR_k}}\ \frac{e^{kH(\tilde{W}_d|\tilde{V})}}{e^{k\left(H(\tilde{W}_d)-\eta\right)}}
\end{aligned}
$$

$$\leq e^{-kE_{4,k}},$$

where,

$$
\begin{aligned}
E_{4,k} := \min_{\substack{\tilde{U}\tilde{V}\tilde{W}_d\\ \in\mathcal{D}_k^U(\eta)^c\times\mathcal{D}_k^{VW}(\eta)}} & D(\tilde{U}\tilde{V}||\bar{U}\bar{V})+\tau\left(E_m(P_{SX},\theta)-\theta\right)+\rho_k+R-R'\\
& -O(\eta)-\frac{|\mathcal{U}||\mathcal{V}||\mathcal{W}|}{k}\log(k+1)\\
\geq \min_{\substack{\tilde{V}:\exists\tilde{W},\\ \tilde{V}\tilde{W}\in\mathcal{D}_k^{VW}(\eta)}} & D(\tilde{V}||\bar{V})+\tau\left(E_m(P_{SX},\theta)-\theta\right)+\rho_k+R-R'\\
& -O(\eta)-\frac{|\mathcal{U}||\mathcal{V}||\mathcal{W}|}{k}\log(k+1).
\end{aligned}
$$

In (B.67), we used (B.30).

If $R'+\eta\leq R$, it can be shown that,

$$\mathbb{P}\left(\mathcal{E}_c|H=1\right)\leq e^{-kE'_{4,k}},$$

where

$$
\begin{aligned}
E'_{4,k}\geq \min_{\substack{\tilde{V}:\exists\tilde{W},\\ \tilde{V}\tilde{W}\in\mathcal{D}_k^{VW}(\eta)}} & D(\tilde{V}||\bar{V})+\tau\left(E_m(P_{SX},\theta)-\theta\right)+\rho_k-O(\eta)\\
& -\frac{|\mathcal{U}||\mathcal{V}||\mathcal{W}|}{k}\log(k+1).
\end{aligned}
$$

Hence, we obtain

$$\frac{-1}{k}\log\left(\mathbb{P}\left(\mathcal{E}_d|H=1\right)\right)$$

$$\overset{(k)}{\geq} \begin{cases} \min\limits_{\substack{\tilde{V}:\exists \tilde{W}, \\ \tilde{V}\tilde{W}\in\mathcal{D}_k^{VW}(\eta)}} D(\tilde{V}||\bar{V}) + E_b(\kappa_\alpha,\omega,R) \\ \qquad\qquad +\tau\left(E_m(P_{SX},\theta)-\theta\right) - O(\eta) \qquad \text{if } R < \zeta_q(\kappa_\alpha,\omega)+\eta, \\ \min\limits_{\substack{\tilde{V}:\exists \tilde{W}, \\ \tilde{V}\tilde{W}\in\mathcal{D}_k^{VW}(\eta)}} D(\tilde{V}||\bar{V}) + \rho(\kappa_\alpha,\omega) \\ \qquad\qquad +\tau\left(E_m(P_{SX},\theta)-\theta\right) - O(\eta), \qquad \text{otherwise,} \end{cases}$$

$$= E_4(\kappa_\alpha,\omega,R,P_{SX},\theta,\tau) - O(\eta). \tag{B.68}$$

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 (B.53), (B.63), (B.65) and (B.68) that for a fixed $(\omega,R,P_{SX},\theta)\in\mathcal{L}(\kappa_\alpha,\tau)$,

$$\begin{aligned}\kappa(\tau,\kappa_\alpha) \geq \min\Big(&E_1(\kappa_\alpha,\omega), E_2(\kappa_\alpha,\omega,R), E_3(\kappa_\alpha,\omega,R,P_{SX},\tau), \\ &E_4(\kappa_\alpha,\omega,R,P_{SX},\theta,\tau)\Big) - O(\eta).\end{aligned} \tag{B.69}$$

To complete the proof, we need to show (B.35), (B.47) and (B.60). Since $W^k(j), j\in\mathcal{M}_i'$, is selected uniformly at random from the set $\mathcal{T}_k(\hat{W}_i)$, we have from [20, Lemma 2.5] that, for any $u^k\in\mathcal{T}_k(\hat{U}_i)$ and sufficiently large k,

$$\mathbb{P}\left((u^k,W^k(j))\notin\mathcal{T}_k(\hat{U}_i\hat{W}_i)\right) \leq \left(1-\frac{e^{k\left(H(\hat{W}_i|\hat{U}_i)-\frac{\eta}{4}\right)}}{e^{kH(\hat{W}_i)}}\right). \tag{B.70}$$

Since the codewords are selected independently, we have by the union bound that

$$\begin{aligned}\mathbb{P}\left(\nexists (u^k,W^k(j))\notin\mathcal{T}_k(\hat{U}_i\hat{W}_i),\ j\in\mathcal{M}_i'\right) &\leq \left(1-\frac{e^{k\left(H(\hat{W}_i|\hat{U}_i)-\frac{\eta}{4}\right)}}{e^{kH(\hat{W}_i)}}\right)^{e^{kR_i'}} \\ &\leq e^{-e^{k\left(R_i'-I(\hat{U}_i;\hat{W}_i)-\frac{\eta}{4}\right)}}.\end{aligned} \tag{B.71}$$

Hence, by the choice of R_i' in (B.25), we have for sufficiently large k that

$$\mathbb{P}(\mathcal{E}_{EE}) = \sum_{i=1}^{|\mathcal{D}_k^U(\eta)|} e^{-e^{k\frac{\eta}{12}}} \leq (k+1)^{|\mathcal{U}|} e^{-e^{k\frac{\eta}{12}}} \leq e^{-e^{k\frac{\eta}{15}}}. \tag{B.72}$$

This completes the proof of (B.35).

Next, we prove (B.47). Note that by the encoding procedure, $M' \neq 0$ and $w^k \in \mathcal{T}_k(\hat{W}_i)$ for some $1 \leq i \leq |\mathcal{D}_k^U(\eta)|$ implies that $U^k \in \mathcal{T}_k(P_{\hat{U}_i|\hat{W}_i}, w^k)$. Hence, we can write for $j \neq m'$, that

$$\begin{aligned}
&\mathbb{P}\left(W^k(j) = \tilde{w}^k | V^k = v^k, M' = m' \neq 0, M = m, f_B(j) = m, W^k(m') = w^k\right) \\
&= \sum_{u^k \in \mathcal{T}_k(P_{\hat{U}_i|\hat{W}_i}, w^k)} \mathbb{P}\Big(U^k = u^k | V^k = v^k, M' = m' \neq 0, M = m, f_B(j) = m, \\
&\qquad\qquad W^k(m') = w^k\Big) \\
&\mathbb{P}\Big(W^k(j) = \tilde{w}^k | U^k = u^k, V^k = v^k, M' = m' \neq 0, M = m, f_B(j) = m, \\
&\qquad\qquad W^k(m') = w^k\Big)
\end{aligned}$$

Let

$$\begin{aligned}
&\mathcal{C}^-_{m',j} := \mathcal{C}_k \backslash \{W^k(m'), W^k(j)\}, \\
&\mathcal{E} := \{U^k = u^k, V^k = v^k, M' = m' \neq 0, M = m, f_B(j) = m, W^k(m') = w^k\}.
\end{aligned}$$

Then, we can write,

$$\mathbb{P}\left(W^k(j) = \tilde{w}^k | \mathcal{E}\right) = \sum_{\mathcal{C}^-_{m',j} = c} \mathbb{P}(\mathcal{C}^-_{m',j} = c | \mathcal{E}) \mathbb{P}(W^k(j) = \tilde{w}^k | \mathcal{E}, \mathcal{C}^-_{m',j} = c). \tag{B.73}$$

We can write the term within the summation in (B.73) as follows:

$$\begin{aligned}
&\mathbb{P}(W^k(j) = \tilde{w}^k | \mathcal{E}, \mathcal{C}^-_{m',j} = c) \\
&= \mathbb{P}(W^k(j) = \tilde{w}^k | U^k = u^k, V^k = v^k,\ \mathcal{C}^-_{m',j} = c) \\
&\frac{\mathbb{P}(M' = m' | W^k(j) = \tilde{w}^k, W^k(m') = w^k, U^k = u^k, V^k = v^k, \mathcal{C}^-_{m',j} = c)}{\mathbb{P}(M' = m' | W^k(m') = w^k, U^k = u^k, V^k = v^k, \mathcal{C}^-_{m',j} = c)} \\
&\frac{\mathbb{P}(M = m, f_B(j) = m | M' = m', W^k(j) = \tilde{w}^k, W^k(m') = w^k, U^k = u^k, V^k = v^k, \mathcal{C}^-_{m',j} = c)}{\mathbb{P}(M = m, f_B(j) = m | M' = m', W^k(m') = w^k, U^k = u^k, V^k = v^k, \mathcal{C}^-_{m',j} = c)}
\end{aligned} \tag{B.74}$$

$$= \mathbb{P}(W^k(j) = \tilde{w}^k) \ \frac{\mathbb{P}(M' = m' | W^k(j) = \tilde{w}^k, W^k(m') = w^k, U^k = u^k, V^k = v^k, \mathcal{C}^-_{m',j} = c)}{\mathbb{P}(M' = m' | W^k(m') = w^k, U^k = u^k, V^k = v^k, \mathcal{C}^-_{m',j} = c)}. \tag{B.75}$$

In (B.75), we used

$$\begin{aligned}
&\mathbb{P}(M = m, f_B(j) = m | M' = m', W^k(j) = \tilde{w}^k, W^k(m') = w^k, U^k = u^k, V^k = v^k, \mathcal{C}^-_{m',j} = c) \\
&= \mathbb{P}(M = m, f_B(j) = m | M' = m', W^k(m') = w^k, U^k = u^k, V^k = v^k, \mathcal{C}^-_{m',j} = c) \\
&= \mathbb{P}(M = m, f_B(j) = m),
\end{aligned}$$

which in turn follows from the fact that the binning is performed independent of the $\mathcal{C}_k$, U^k and V^k. Let

$$N(u^k, \mathcal{C}^-_{m',j}) = |\{w^k(l) \in \mathcal{C}^-_{m',j} : l \neq m', j,\ (u^k, w^k(l)) \in \mathcal{T}_k(\hat{U}_i \hat{W}_i)\}|.$$

Recall that if there are multiple indices l in the codebook $\mathcal{C}_k$ such that $(u^k, w^k(l)) \in \mathcal{T}_k(\hat{U}_i \hat{W}_i)$, then the encoder selects one of them uniformly at random. Also, note that since $M' = m' \neq 0$, $(u^k, w^k(m')) \in \mathcal{T}_k(\hat{U}_i \hat{W}_i)$. Thus, if $(u^k, \tilde{w}^k) \in \mathcal{T}_k(\hat{U}_i \hat{W}_i)$, then

$$\begin{aligned}
&\frac{\mathbb{P}(M' = m' | W^k(j) = \tilde{w}^k, W^k(m') = w^k, U^k = u^k, V^k = v^k, \mathcal{C}^-_{m',j} = c)}{\mathbb{P}(M' = m' | W^k(m') = w^k, U^k = u^k, V^k = v^k, \mathcal{C}^-_{m',j} = c)} \\
&= \left[\frac{1}{N(u^k, \mathcal{C}^-_{m',j}) + 2}\right] \frac{1}{\mathbb{P}(M = m | U^k = u^k, V^k = v^k, \mathcal{C}^-_{m',j} = c)} \\
&\leq \frac{N(u^k, \mathcal{C}^-_{m',j}) + 2}{N(u^k, \mathcal{C}^-_{m',j}) + 2} = 1.
\end{aligned} \tag{B.76}$$

On the other hand, if $(u^k, \tilde{w}^k) \notin \mathcal{T}_k(\hat{U}_i \hat{W}_i)$, then

$$\begin{aligned}
&\frac{\mathbb{P}(M' = m' | W^k(j) = \tilde{w}^k, W^k(m') = w^k, U^k = u^k, V^k = v^k, \mathcal{C}^-_{m',j} = c)}{\mathbb{P}(M' = m' | W^k(m') = w^k, U^k = u^k, V^k = v^k, \mathcal{C}^-_{m',j} = c)} \\
&= \left[\frac{1}{N(u^k, \mathcal{C}^-_{m',j}) + 1}\right] \frac{1}{\mathbb{P}(M = m | U^k = u^k, V^k = v^k, \mathcal{C}^-_{m',j} = c)} \leq \frac{N(u^k, \mathcal{C}^-_{m',j}) + 2}{N(u^k, \mathcal{C}^-_{m',j}) + 1} \leq 2.
\end{aligned} \tag{B.77}$$

Substituting (B.76) and (B.77) in (B.73), we obtain (B.47). The proof of (B.60) is similar to that of (B.47), and hence, omitted.

Thus, we have shown that for a fixed $(\omega, R, P_{SX}, \theta) \in \mathcal{L}(\kappa_\alpha, \tau)$, the probability of type I and type II error probabilities averaged over the ensemble of randomly generated

codebooks and binning functions satisfy

$$\mathbb{P}\left(\hat{H}=1|H=0\right) \leq e^{-k(\kappa_\alpha - O(\eta))}, \tag{B.78}$$

$$\text{and } \mathbb{P}\left(\hat{H}=0|H=1\right) \leq e^{-k(\kappa_s^*(\tau,\kappa_\alpha) - O(\eta))}, \tag{B.79}$$

for all sufficiently large k. By the random coding argument followed by an expurgation step [89], there exists a deterministic codebook $\mathcal{C}_k$ and deterministic binning function f_b such that (B.78) and (B.79) are satisfied. Maximizing over $(\omega, R, P_{SX}, \theta) \in \mathcal{L}(\kappa_\alpha, \tau)$ and noting that $\eta > 0$ is arbitrary completes the proof.

B.3 Proof of Theorem 3.14

Fix $\kappa_\alpha > 0$ and $\left(P_S, \omega'(\cdot, P_S), P_{X|USW'}, P_{X'|US}\right) \in \mathcal{L}_h(\kappa_\alpha)$. Let $\eta > 0$ be a small number, and choose a sequence $s^n \in \mathcal{T}_n(\hat{S}^*)$ which is revealed to both the encoder and the detector, where $\hat{S}^*$ satisfies $D(\hat{S}^*||S) \leq \eta$. Let $R' := \zeta'_q(\kappa_\alpha, \omega', P_{\hat{S}^*})$.

Encoding:

The encoder performs type based quantization followed by channel coding similar to that in hybrid coding [26]. The details are as follows:

Quantization scheme: Let

$$\mathcal{D}_n^U(\eta) := \{\hat{U} \in \mathcal{T}_n(\mathcal{U}) : D(\hat{U}||U) \leq \kappa_\alpha + \eta\}. \tag{B.80}$$

Consider some ordering on the types in $\mathcal{D}_n^U(\eta)$ and denote the elements as $\hat{U}_1$, $\hat{U}_2$, ..., etc. For each joint type variable $\hat{S}^*\hat{U}_i$, $\hat{U}_i \in \mathcal{D}_n^U(\eta)$, $1 \leq i \leq |\mathcal{D}_n^U(\eta)|$, such that $\hat{S}^* \perp \hat{U}_i$, choose a joint type variable $\hat{S}^*\hat{U}_i\hat{W}'_i$, $\hat{W}'_i \in \mathcal{T}_n(\mathcal{W}')$, such that

$$D\left(\hat{W}'_i|\hat{U}_i, \hat{S}^*||W'_i|U, \hat{S}^*\middle|\hat{U}_i, \hat{S}^*\right) \leq \frac{\eta}{3},$$
$$I(\hat{S}^*, \hat{U}_i; \hat{W}'_i) \leq R' + \frac{\eta}{3},$$

where $P_{W'_i|U,S} = \omega'(P_{\hat{U}_i}, P_{\hat{S}^*})$. Let

$$\mathcal{D}_n^{SUW'}(\eta) := \{\hat{S}^*\hat{U}_i\hat{W}'_i : 1 \leq i \leq |\mathcal{D}_n^U(\eta)|\},$$

$$\text{and } R_i' := I(\hat{S}^*, \hat{U}_i; \hat{W}_i') + \frac{\eta}{3}, 1 \leq i \leq |\mathcal{D}_n^U(\eta)|. \tag{B.81}$$

Let

$$\mathcal{C}_n' = \left\{ w'^n(j), j \in \left[1 : \sum_{i=1}^{|\mathcal{D}_n^U(\eta)|} e^{nR_i'} \right] \right\},$$

denote a quantization codebook such that each codeword $w'^n(j)$, $j \in \mathcal{M}_i' := [1 + \sum_{m=1}^{i-1} e^{nR_m'} : \sum_{m=1}^{i} e^{nR_m'}]$, $1 \leq i \leq |\mathcal{D}_n^U(\eta)|$, belongs to the set $\mathcal{T}_n(\hat{W}_i')$. For $u^n \in \mathcal{T}_n(\hat{U}_i)$ such that $\hat{U}_i \in \mathcal{D}_n(U)$ for some $1 \leq i \leq |\mathcal{D}_n^U(\eta)|$, let

$$\begin{aligned}\mu'(u^n, \mathcal{C}_n') :=& \{ j \in \mathcal{M}_i' : \ w'^n(j) \in \mathcal{C}_n' \text{ and } (s^n, u^n, w'^n(j)) \in \mathcal{T}_n(\hat{S}^* \hat{U}_i \hat{W}_i'), \\ & \hat{S}^* \hat{U}_i \hat{W}_i' \in \mathcal{D}_n^{SUW'}(\eta) \}.\end{aligned}$$

If $|\mu'(u^n, \mathcal{C}_n')| \geq 1$, let $M'(u^n, \mathcal{C}_n')$ denote an index selected uniformly at random from the set $\mu'(u^n, \mathcal{C}_n')$, otherwise, set $M'(u^n, \mathcal{C}_n') = 0$. Given $\mathcal{C}_n'$ and $u^n \in \mathcal{U}^n$, the quantizer outputs $M' = M'(u^n, \mathcal{C}_k')$, where the support of M' is given by

$$\mathcal{M}' := \left[0 : \sum_{i=1}^{|\mathcal{D}_n^U(\eta)|} e^{nR_i'} \right].$$

Note that for sufficiently large n, it follows similarly to (B.26) that

$$|\mathcal{M}'| \leq e^{n(R'+\eta)}.$$

If $M' = m' \neq 0$, the encoder transmits X^n over the channel, where $X^n = x^n$ is generated according to the distribution $\prod_{i=1}^n P_{X|USW'}(x_i|u_i, s_i, w_i'(m'))$. If $M' = 0$, the encoder transmits $X'^n = x'^n$ randomly according to $\prod_{i=1}^n P_{X'|US}(x_i'|u_i, s_i)$.

Decoding:

For a given codebook $\mathcal{C}_n'$ and $m' \in \mathcal{M}' \backslash \{0\}$, let $\mathcal{O}_{m'}$ denote the set of u^n such that $M'(u^n, \mathcal{C}_n') = m'$. For each $m' \in \mathcal{M}' \backslash \{0\}$ and $u^n \in \mathcal{O}_{m'}$, let

$$\mathcal{B}_{m'}'(u^n) = \{ (v^n, y^n) \in \mathcal{V}^n \times \mathcal{Y}^n : (s^n, u^n, w_{m'}'^n, v^n, y^n) \in \mathcal{J}_n^{\kappa_\alpha + \eta}(\hat{S}^* U W_{m'}' V Y) \},$$

where $\hat{S}^* U W'_{m'} V Y$ is uniquely specified by $\hat{S}^* \perp (U, V)$

$$W'_{m'} - (U, \hat{S}^*) - V,\ Y - (U, \hat{S}^*, W'_{m'}) - V,\ P_{W'_{m'}|U\hat{S}^*} = \omega'(P_{u^n}, P_{\hat{S}^*}), \tag{B.82}$$

$$P_{Y|U\hat{S}^*W'_{m'}}(y|u,s,w') = \sum\nolimits_{x\in\mathcal{X}} P_{X|U\hat{S}^*W'_{m'}}(x|u,s,w') P_{Y|X}(y|x),$$
$$\forall\ (y,u,s,w') \in \mathcal{Y}\times\mathcal{U}\times\mathcal{S}\times\mathcal{W}'. \tag{B.83}$$

For $m' \in \mathcal{M}'\backslash\{0\}$, we define

$$\mathcal{B}'_{m'} := \{(v^n, y^n) : (v^n, y^n) \in B_{m'}(u^n) \text{ for some } u^n \in \mathcal{O}_{m'}\}.$$

Define the acceptance region for H_0 at the detector as

$$\mathcal{A}_n := \bigcup_{m'\in\mathcal{M}'\backslash 0} s^n \times m' \times \mathcal{B}'_{m'},$$

or equivalently as

$$\mathcal{A}^e_n := \bigcup_{m'\in\mathcal{M}'\backslash 0} s^n \times \mathcal{O}_{m'} \times \mathcal{B}'_{m'}.$$

Given $Y^n = y^n$ and $V^n = v^n$, if $(s^n, v^n, y^n) \in \{s^n\}\times\bigcup_{m'\in\mathcal{M}'\backslash\{0\}} \mathcal{B}'_{m'}$, then set $\hat{M}' = m'$, where

$$\hat{m}' := \arg\min_{j\in\mathcal{M}'\backslash 0} H_e(w'^n(j)|v^n, y^n, s^n).$$

Otherwise, set $\hat{M}' = 0$. If $\hat{M}' = 0$, $\hat{H} = 1$ is declared. Otherwise, $\hat{H} = 0$ or $\hat{H} = 1$ is declared depending on whether $(s^n, \hat{m}', v^n, y^n) \in \mathcal{A}_n$ or $(s^n, \hat{m}', v^n, y^n) \notin \mathcal{A}_n$, respectively.

Analysis of the type I and type II error probabilities:

Similar to Theorem 3.9, we will analyze the average type I and type II error probabilities over an ensemble of randomly generated quantization codebooks. Then, the standard random coding argument followed by the expurgation technique in [89] guarantees the existence of a deterministic quantization codebook that achieves the lower bound given in Theorem 3.14. Let each codeword $w'^n(j)$, $j \in \mathcal{M}'_i$, $1 \le i \le |\mathcal{D}^U_n(\eta)|$, be selected (with replacement) independently and uniformly at random from the set $\mathcal{T}_n(\hat{W}'_i)$ (see

quantization scheme above). We proceed to analyze the type I and type II error probabilities averaged over these random codebooks. Note that a type I error can occur only under the following events:

(i) $\mathcal{E}'_{EE} := \bigcup_{\hat{U}\in\mathcal{D}'_n(SU)} \bigcup_{u^n\in\mathcal{T}_n(\hat{U})} \mathcal{E}'_{EE}(u^n)$, where

$$\mathcal{E}'_{EE}(u^n) := \left\{ \begin{array}{l} \nexists\ W'^n(j) \in \mathcal{C}'_n,\ j \in [1:|\mathcal{M}'|],\ \text{s.t. } (s^n, u^n, W'^n(j)) \in \\ \mathcal{T}_n(\hat{S}^*\hat{U}_i\hat{W}'_i),\ P_{\hat{S}^*\hat{U}_i} = P_{s^n u^n},\ \hat{S}^*\hat{U}_i\hat{W}'_i \in \mathcal{D}_n^{SUW'}(\eta) \end{array} \right\}.$$

(ii) $\hat{M}' = M'$.

(iii) $M' \neq 0$ and $\hat{M}' \neq M'$.

(iv) $M' = 0$ and $\hat{M}' \neq M'$.

Similar to (B.35), we have since R'_i satisfies (B.81), that

$$\mathbb{P}(\mathcal{E}'_{EE}) \leq e^{-e^{n\Omega(\eta)}}. \tag{B.84}$$

Next, consider event (ii). Due to (B.84), we can write

$$\mathbb{P}\left(\hat{H}=1|\hat{M}'=M', H=0\right) \leq e^{-e^{n\Omega(\eta)}} + \mathbb{P}\left(\hat{H}=1|\hat{M}'=M', \mathcal{E}'^c_{EE}, H=0\right). \tag{B.85}$$

The second term in (B.85) can be bounded as

$$\begin{aligned} &\mathbb{P}\left(\hat{H}=1|\hat{M}'=M', \mathcal{E}'^c_{EE}, H=0\right) \\ &= \mathbb{P}\left((s^n, M', V^n, Y^n) \notin \mathcal{A}_n | \mathcal{E}'^c_{EE}, H=0\right) &\text{(B.86)}\\ &= 1 - \mathbb{P}\left((s^n, U^n, V^n, Y^n) \in \mathcal{A}_n^e | \mathcal{E}'^c_{EE}, H=0\right) &\text{(B.87)} \end{aligned}$$

We have similar to [27, Equation 4.17] that for $u^n \in \mathcal{O}_{m'}$ that

$$\begin{aligned} &\mathbb{P}((V^n, Y^n) \in \mathcal{B}'_{m'}(u^n) | U^n = u^n, \mathcal{E}'^c_{EE}) \\ &= \mathbb{P}((V^n, Y^n) \in \mathcal{B}'_{m'}(u^n) | U^n = u^n, W'^n(m') = w'^n_{m'}, \mathcal{E}'^c_{EE}) &\text{(B.88)}\\ &\geq 1 - e^{-n\left(\kappa_\alpha + \frac{\eta}{3} - D(P_{u^n}||P_U)\right)}. &\text{(B.89)} \end{aligned}$$

Then, using (B.80) and (B.89), it follows similarly to [27, Equation 4.22] that

$$\mathbb{P}((s^n, U^n, V^n, Y^n) \in \mathcal{A}_e^n | \mathcal{E}_{EE}'^c) \geq 1 - e^{-n\kappa_\alpha}. \tag{B.90}$$

Substituting (B.90) in (B.87), it follows that

$$\mathbb{P}\left(\hat{H} = 1 | \hat{M}' = M', \mathcal{E}_{EE}'^c, H = 0\right) \leq e^{-n\kappa_\alpha}. \tag{B.91}$$

The probability of event (iii) can be upper bounded as follows:

$$\begin{aligned}
&\mathbb{P}\left(M' \neq 0, \hat{M}' \neq M' | H = 0\right) \\
&\leq \mathbb{P}\left(M' \neq 0, \hat{M}' \neq M', (s^n, M', V^n, Y^n) \in \mathcal{A}_n | H = 0\right) \\
&\quad + \mathbb{P}\left(M' \neq 0, \hat{M}' \neq M', (s^n, M', V^n, Y^n) \notin \mathcal{A}_n | H = 0\right) \\
&\leq \mathbb{P}\left(M' \neq 0, \hat{M}' \neq M', (s^n, M', V^n, Y^n) \in \mathcal{A}_n | H = 0\right) + e^{-e^{n\Omega(\eta)}} + e^{-n\kappa_\alpha} \qquad \text{(B.92)} \\
&\leq \mathbb{P}\left(\hat{M}' \neq M' | M' \neq 0, (s^n, M', V^n, Y^n) \in \mathcal{A}_n, H = 0\right) + e^{-e^{n\Omega(\eta)}} + e^{-n\kappa_\alpha} \\
&\leq e^{-n\left(\rho'(\kappa_\alpha, \omega', P_S, P_{X|USW'}) - \zeta_q'(\kappa_\alpha, \omega', P_{\hat{S}*}) - O(\eta)\right)} + e^{-e^{n\Omega(\eta)}} + e^{-n\kappa_\alpha} \qquad \text{(B.93)}
\end{aligned}$$

where (B.92) follows similar to (B.41) using (B.84) and (B.90), and (B.93) follows similar to (B.51) by noting that $(s^n, M', V^n, Y^n) \in \mathcal{A}_n$ implies that $\hat{M}' \neq 0$. Also, from (B.84) and the definition of $\mathcal{D}_n^U(\eta)$, we can bound the probability of event (iv) as

$$\mathbb{P}\left(M' = 0, \hat{M}' \neq M' | H = 0\right) \leq \mathbb{P}\left(M' = 0 | H = 0\right) \leq e^{-n\kappa_\alpha}. \tag{B.94}$$

From (B.84), (B.91), (B.93) and (B.94), it follows that the type I error probability satisfies $e^{-k(\kappa_\alpha - O(\eta))}$, asymptotically.

Next, we analyze the type II error probability of the above scheme averaged over the random codebooks. For a given codebook $\mathcal{C}_n'$, let $\tilde{U}, \tilde{V}, \tilde{W}, \tilde{Y}$ and $\tilde{W}_d$ denote the type variable for the realizations of $\bar{U}^n, \bar{V}^n, W'^n(M')$ $(M' \neq 0)$, $\bar{Y}^n$ and $W'^n(\hat{M}')$ $(\hat{M}' \neq 0)$, respectively. Let

$\mathcal{D}_n^{SVW'Y}(\eta)$

$$:= \left\{ \begin{array}{l} \hat{S}^*\hat{V}\hat{W}\hat{Y}: \ \exists\ (s^n,u^n,v^n,w^n,y^n) \in \underset{m'\in\mathcal{M}'\setminus\{0\}}{\cup} \mathcal{J}_n^{\kappa_\alpha+\eta}(\hat{S}^*UVW'_{m'}Y), \\ \hat{S}^*UVW'_{m'}Y \text{ satisfies (B.82) and (B.83), and } P_{s^nu^nv^nw^ny^n} = P_{\hat{S}^*\hat{U}\hat{V}\hat{W}\hat{Y}} \end{array} \right\}.$$

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

(a)

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

(b)

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

(c)

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

Let

$$\begin{aligned} \mathcal{F}'_{1,n}(\eta) :=& \{\hat{S}^*\tilde{U}\tilde{V}\tilde{W}\tilde{Y} \in \mathcal{T}_n(\mathcal{S}\times\mathcal{U}\times\mathcal{V}\times\mathcal{W}'\times\mathcal{Y}) : \hat{S}^*\tilde{U}\tilde{W} \in \mathcal{D}_n^{SUW'}(\eta), \\ & \hat{S}^*\tilde{V}\tilde{W}\tilde{Y} \in \mathcal{D}_n^{SVW'Y}(\eta)\}. \end{aligned}$$

Then, we can write

$$\begin{aligned} &\mathbb{P}\left(\mathcal{E}'_a|H=1\right) \\ &\leq \sum_{\substack{\hat{S}^*\tilde{U}\tilde{V}\tilde{W}\tilde{Y} \\ \in\mathcal{F}'_{1,n}(\eta)}} \sum_{\substack{(u^n,v^n,w'^n,y^n): \\ (s^n,u^n,v^n,w'^n,y^n) \\ \in\mathcal{T}_n(\hat{S}^*\tilde{U}\tilde{V}\tilde{W}\tilde{Y})}} \sum_{m'\in\mathcal{M}'\setminus\{0\}} \mathbb{P}\Big(\bar{U}^n = u^n, \bar{V}^n = v^n, M' = m', \\ &\qquad W'^n(m') = w'^n, \bar{Y}^n = y^n | S^n = s^n\Big) \end{aligned}$$

$$
\begin{aligned}
&\leq \sum_{\substack{\hat{S}^*\tilde{U}\tilde{V}\tilde{W}\tilde{Y} \\ \in\mathcal{F}'_{1,n}(\eta)}} \sum_{\substack{(u^n,v^n,w'^n,y^n): \\ (s^n,u^n,v^n,w'^n,y^n) \\ \in\mathcal{T}_n(\hat{S}^*\tilde{U}\tilde{V}\tilde{W}\tilde{Y})}} \sum_{m'\in\mathcal{M}'\setminus\{0\}} \mathbb{P}\left(\bar{U}^n = u^n, \bar{V}^n = v^n, M' = m'|S^n = s^n\right) \\
&\qquad \mathbb{P}\left(W'^n(m') = w'^n|\bar{U}^n = u^n, \bar{V}^n = v^n, M' = m', S^n = s^n\right) \\
&\qquad \mathbb{P}\left(\bar{Y}^n = y^n|\bar{U}^n = u^n, \bar{V}^n = v^n, M' = m', W'^n(m') = w'^n, S^n = s^n\right) \\
&\leq \sum_{\substack{\hat{S}^*\tilde{U}\tilde{V}\tilde{W}\tilde{Y} \\ \in\mathcal{F}'_{1,n}(\eta)}} \sum_{\substack{(u^n,v^n,w'^n,y^n): \\ (s^n,u^n,v^n,w'^n,y^n) \\ \in\mathcal{T}_n(\hat{S}^*\tilde{U}\tilde{V}\tilde{W}\tilde{Y})}} \sum_{\substack{m'\in \\ \mathcal{M}'\setminus\{0\}}} e^{-n\left(H(\tilde{U}\tilde{V})+D(\tilde{U}\tilde{V}||\bar{U}\bar{V})\right)} \\
&\qquad \mathbb{P}\left(M' = m'|\bar{U}^n = u^n, \bar{V}^n = v^n, S^n = s^n\right) \ \frac{1}{e^{n\left(H(\tilde{W}|\hat{S}^*\tilde{U})-\eta\right)}} \\
&\qquad e^{-n\left(H(\tilde{Y}|\tilde{U}\hat{S}^*\tilde{W})+D\left(\tilde{Y}|\tilde{U}\hat{S}^*\tilde{W}||Y|USW'\middle|\tilde{U}\hat{S}^*\tilde{W}\right)\right)} \qquad \text{(B.95)} \\
&\leq \sum_{\substack{\hat{S}^*\tilde{U}\tilde{V}\tilde{W}\tilde{Y} \\ \in\mathcal{F}'_{1,n}(\eta)}} e^{nH(\tilde{U}\tilde{V}\tilde{W}\tilde{Y}|\hat{S}^*)} e^{-n\left(H(\tilde{U}\tilde{V})+D(\tilde{U}\tilde{V}||\bar{U}\bar{V})\right)} \ \frac{1}{e^{n\left(H(\tilde{W}|\hat{S}^*\tilde{U})-\eta\right)}} \\
&\qquad e^{-n\left(H(\tilde{Y}|\tilde{U}\hat{S}^*\tilde{W})+D\left(\tilde{Y}|\tilde{U}\hat{S}^*\tilde{W}||Y|USW'\middle|\tilde{U}\hat{S}^*\tilde{W}\right)\right)} \\
&\leq e^{-nE'_{1,n}}, \qquad \text{(B.96)}
\end{aligned}
$$

where

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

In (B.95), we used the fact that

$$
\begin{aligned}
&\mathbb{P}\left(W'^n(m') = w'^n|\bar{U}^n = u^n, \bar{V}^n = v^n, S^n = s^n, M' = m'\right) \\
&\leq \begin{cases} \frac{1}{e^{n\left(H(\tilde{W}|\hat{S}^*\tilde{U})-\eta\right)}}, & \text{if } w'^n \in \mathcal{T}_n(\tilde{W}), \\ 0, & \text{otherwise,} \end{cases}
\end{aligned}
$$

which in turn follows from the fact that given $M' = m'$ and $\bar{U}^n = u^n$, $W'^n(m')$ is uniformly distributed in the set $\mathcal{T}_n(P_{\tilde{W}|\hat{S}^*\tilde{U}}, (s^n, u^n))$ and that for sufficiently large n,

$$|\mathcal{T}_n(P_{\tilde{W}|\hat{S}^*\tilde{U}}, (s^n, u^n))| \geq e^{n\left(H(\tilde{W}|\hat{S}^*\tilde{U})-\eta\right)}.$$

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

$$\mathcal{F}'_{2,n}(\eta) := \left\{ \begin{array}{l} \hat{S}^*\tilde{U}\tilde{V}\tilde{W}\tilde{Y}\tilde{W}_d \in \mathcal{T}_n(\mathcal{S}\times\mathcal{U}\times\mathcal{V}\times\mathcal{W}'\times\mathcal{Y}\times\mathcal{W}') : \hat{S}^*\tilde{U}\tilde{W} \in \mathcal{D}_n^{SUW'}(\eta), \\ \hat{S}^*\tilde{V}\tilde{W}_d\tilde{Y} \in \mathcal{D}_n^{SVW'Y}(\eta) \text{ and } H\left(\tilde{W}_d|\hat{S}^*\tilde{V}\tilde{Y}\right) \leq H\left(\tilde{W}|\hat{S}^*\tilde{V}\tilde{Y}\right) \end{array} \right\}.$$

Then,

$$\begin{aligned}
&\mathbb{P}\left(\mathcal{E}'_b|H=1\right) \\
&\leq \sum_{\substack{\hat{S}^*\tilde{U}\tilde{V}\tilde{W}\tilde{Y}\tilde{W}_d \\ \in\mathcal{F}'_{2,n}(\eta)}} \sum_{\substack{(u^n,v^n,w'^n,y^n,\bar{w}^n): \\ (s^n,u^n,v^n,w'^n,y^n,\bar{w}^n) \\ \in\mathcal{T}_n(\hat{S}^*\tilde{U}\tilde{V}\tilde{W}\tilde{Y}\tilde{W}_d)}} \sum_{m'\in\mathcal{M}'\setminus\{0\}} \mathbb{P}\Big(\bar{U}^n = u^n, \bar{V}^n = v^n, M' = m', \\
&\qquad\qquad W'^n(m') = w'^n, \bar{Y}^n = y^n | S^n = s^n\Big) \\
&\quad \sum_{\hat{m}'\in\mathcal{M}'\setminus\{0,m'\}} \mathbb{P}\left(W'^n(\hat{m}') = \bar{w}^n | \bar{U}^n = u^n, M' = m', W'^n(m') = w'^n, S^n = s^n\right) \\
&\leq \sum_{\substack{\hat{S}^*\tilde{U}\tilde{V}\tilde{W}\tilde{Y}\tilde{W}_d \\ \in\mathcal{F}'_{2,n}(\eta)}} \sum_{\substack{(u^n,v^n,w'^n,y^n,\bar{w}^n): \\ (s^n,u^n,v^n,w'^n,y^n,\bar{w}^n) \\ \in\mathcal{T}_n(\hat{S}^*\tilde{U}\tilde{V}\tilde{W}\tilde{Y}\tilde{W}_d)}} \sum_{m'\in\mathcal{M}'\setminus\{0\}} \Bigg[e^{-n\left(H(\tilde{U}\tilde{V})+D(\tilde{U}\tilde{V}||\bar{U}\bar{V})\right)} \\
&\mathbb{P}\left(M' = m' | \bar{U}^n = u^n, \bar{V}^n = v^n, S^n = s^n\right) \ \frac{1}{e^{n\left(H(\tilde{W}|\hat{S}^*\tilde{U})-\eta\right)}} \\
&e^{-n\left(H(\tilde{Y}|\tilde{U}\hat{S}^*\tilde{W})+D\left(\tilde{Y}|\tilde{U}\hat{S}^*\tilde{W}||Y|USW'\middle|\tilde{U}\hat{S}^*\tilde{W}\right)\right)} \\
&\quad \sum_{\hat{m}'\in\mathcal{M}'\setminus\{0,m'\}} \mathbb{P}\left(W'^n(\hat{m}') = \bar{w}^n | \bar{U}^n = u^n, M' = m', W'^n(m') = w'^n, S^n = s^n\right) \Bigg] \\
&\leq \sum_{\substack{\hat{S}^*\tilde{U}\tilde{V}\tilde{W}\tilde{Y}\tilde{W}_d \\ \in\mathcal{F}'_{2,n}(\eta)}} \sum_{\substack{(u^n,v^n,w'^n,y^n): \\ (s^n,u^n,v^n,w'^n,y^n) \\ \in\mathcal{T}_n(\hat{S}^*\tilde{U}\tilde{V}\tilde{W}\tilde{Y})}} \sum_{m'\in\mathcal{M}'\setminus\{0\}} \Bigg[e^{-n\left(H(\tilde{U}\tilde{V})+D(\tilde{U}\tilde{V}||\bar{U}\bar{V})\right)} \\
&\mathbb{P}\left(M' = m' | \bar{U}^n = u^n, \bar{V}^n = v^n, S^n = s^n\right) \ \frac{1}{e^{n\left(H(\tilde{W}|\hat{S}^*\tilde{U})-\eta\right)}}
\end{aligned}$$

$$
\begin{aligned}
& e^{-n\left(H(\tilde{Y}|\tilde{U}\hat{S}^*\tilde{W})+D\left(\tilde{Y}|\tilde{U}\hat{S}^*\tilde{W}||Y|USW'\middle|\tilde{U}\hat{S}^*\tilde{W}\right)\right)} \sum_{\hat{m}'\in\mathcal{M}'\setminus\{0,m'\}} \frac{2e^{nH(\tilde{W}_d|\hat{S}^*\tilde{V}\tilde{Y})}}{e^{n\left(H(\tilde{W}_d)-\eta\right)}}\Bigg] \\
& \leq \sum_{\substack{\hat{S}^*\tilde{U}\tilde{V}\tilde{W}\tilde{Y}\tilde{W}_d \\ \in\mathcal{F}'_{2,n}(\eta)}} \sum_{\substack{(u^n,v^n,w'^n,y^n): \\ (s^n,u^n,v^n,w'^n,y^n) \\ \in\mathcal{T}_n(\hat{S}^*\tilde{U}\tilde{V}\tilde{W}\tilde{Y})}} \Bigg[e^{-n\left(H(\tilde{U}\tilde{V})+D(\tilde{U}\tilde{V}||\bar{U}\bar{V})\right)} \frac{1}{e^{n\left(H(\tilde{W}|\hat{S}^*\tilde{U})-\eta\right)}} \\
& e^{-n\left(H(\tilde{Y}|\tilde{U}\hat{S}^*\tilde{W})+D\left(\tilde{Y}|\tilde{U}\hat{S}^*\tilde{W}||Y|USW'\middle|\tilde{U}\hat{S}^*\tilde{W}\right)\right)} e^{n\left(\zeta'_q(\kappa_\alpha,\omega',P_{\hat{S}^*})+\eta\right)} \frac{2e^{nH(\tilde{W}_d|\hat{S}^*\tilde{V}\tilde{Y})}}{e^{n\left(H(\tilde{W}_d)-\eta\right)}}\Bigg] \\
& \leq e^{-nE'_{2,n}},
\end{aligned} \tag{B.97}
$$

where

$$
\begin{aligned}
E'_{2,n} :=& \min_{\hat{S}^*\tilde{U}\tilde{V}\tilde{W}\tilde{Y}\tilde{W}_d\in\mathcal{F}'_{2,n}(\eta)} H(\tilde{U}\tilde{V}) + D(\tilde{U}\tilde{V}||\bar{U}\bar{V}) + H(\tilde{W}|\hat{S}^*\tilde{U}) - 2\eta \\
& + H(\tilde{Y}|\tilde{U}\hat{S}^*\tilde{W}) + D\left(\tilde{Y}|\tilde{U}\hat{S}^*\tilde{W}||Y|USW'\middle|\tilde{U}\hat{S}^*\tilde{W}\right) + I(\tilde{W}_d;\hat{S}^*\tilde{V}\tilde{Y}) \\
& - H(\tilde{U}\tilde{V}\tilde{W}\tilde{Y}|\hat{S}^*) - \zeta'_q(\kappa_\alpha,\omega',P_{\hat{S}^*}) - \frac{\log 2}{n} - \frac{|\mathcal{S}||\mathcal{U}||\mathcal{V}||\mathcal{W}|^2|\mathcal{Y}|\log(n+1)}{n} \\
\geq& \min_{\hat{S}^*\tilde{U}\tilde{V}\tilde{W}\tilde{Y}\tilde{W}_d\in\mathcal{F}'_{2,n}(\eta)} D(\tilde{U}\tilde{V}\tilde{W}\tilde{Y}||\bar{U}\bar{V}\bar{W}'\bar{Y}|S) + I(\tilde{W}_d;\hat{S}^*\tilde{V}\tilde{Y}) \\
& - \zeta'_q(\kappa_\alpha,\omega',P_{\hat{S}^*}) - \frac{\log 2}{n} - \frac{|\mathcal{S}||\mathcal{U}||\mathcal{V}||\mathcal{W}|^2|\mathcal{Y}|\log(n+1)}{n} \\
\overset{(n)}{\geq}& \min_{\substack{\tilde{U}\tilde{V}\tilde{W}\tilde{Y}S\in \\ \mathcal{T}'_2(\kappa_\alpha,\omega',P_S,P_{X|USW'})}} D(\tilde{U}\tilde{V}\tilde{W}\tilde{Y}||\bar{U}\bar{V}\bar{W}'\bar{Y}|S) + \rho'(\kappa_\alpha,\omega',P_S,P_{X|USW'}) \\
& - \zeta'_q(\kappa_\alpha,\omega',P_S) - O(\eta) \\
=& E'_2(\kappa_\alpha,\omega',P_S,P_{X|USW'}) - O(\eta).
\end{aligned}
$$

Similar to (B.66), we can write

$$
\begin{aligned}
\mathbb{P}\left(\mathcal{E}'_c|H=1\right) =& \sum_{\substack{u^n\in\mathcal{T}_n(\tilde{U}): \\ \tilde{U}\in\mathcal{D}_n^U(\eta)}} \mathbb{P}\left(\bar{U}^n=u^n,\mathcal{E}'_{EE},\mathcal{E}'_c|S^n=s^n,H=1\right) \\
& + \sum_{\substack{u^n\in\mathcal{T}_n(\tilde{U}): \\ \tilde{U}\notin\mathcal{D}_n^U(\eta)}} \mathbb{P}\left(\bar{U}^n=u^n,\mathcal{E}'_c|S^n=s^n,H=1\right).
\end{aligned} \tag{B.98}
$$

The first term in (B.98) decays double exponentially as $e^{-e^{n\Omega(\eta)}}$. The second term in (B.98) can be simplified as follows:

$$\sum_{\substack{u^n\in\mathcal{T}_n(\tilde{U}):\\ \tilde{U}\notin\mathcal{D}_n^U(\eta)}} \mathbb{P}\left(\bar{U}^n=u^n,\mathcal{E}_c'|S^n=s^n,H=1\right)$$

$$\leq \sum_{\substack{u^n\in\mathcal{T}_n(\tilde{U}):\\ \tilde{U}\notin\mathcal{D}_n^U(\eta)}} \sum_{\substack{(v^n,y^n,\bar{w}^n):\\ (s^n,v^n,y^n,\bar{w}^n)\in\mathcal{T}_n(\hat{S}^*\tilde{V}\tilde{Y}\tilde{W}_d)\\ \hat{S}^*\tilde{V}\tilde{W}_d\tilde{Y}\in\mathcal{D}_n^{SVW'Y}(\eta)}} \sum_{\hat{m}\in\mathcal{M}\setminus\{0\}} \mathbb{P}\Big(\bar{U}^n=u^n,\bar{V}^n=v^n,M'=0,\bar{Y}^n=y^n|$$

$$S^n=s^n,H=1\Big) \sum_{\hat{m}'\in\mathcal{M}'\setminus\{0\}} \mathbb{P}\left(W^k(\hat{m}')=\bar{w}^k\right)$$

$$\leq \sum_{\substack{u^n\in\mathcal{T}_n(\tilde{U}):\\ \tilde{U}\notin\mathcal{D}_n^U(\eta)}} \sum_{\substack{(v^n,y^n,\bar{w}^n):\\ (s^n,v^n,y^n,\bar{w}^n)\in\mathcal{T}_n(\hat{S}^*\tilde{V}\tilde{Y}\tilde{W}_d)\\ \hat{S}^*\tilde{V}\tilde{W}_d\tilde{Y}\in\mathcal{D}_n^{SVW'Y}(\eta)}} \mathbb{P}\Big(\bar{U}^n=u^n,\bar{V}^n=v^n,M'=0,\bar{Y}^n=y^n|$$

$$S^n=s^n,H=1\Big) \sum_{\hat{m}'\in\mathcal{M}'\setminus\{0\}} \frac{1}{e^{k\left(H(\tilde{W}_d)-\eta\right)}}$$

$$\leq \sum_{\substack{u^n\in\mathcal{T}_n(\tilde{U}):\\ \tilde{U}\notin\mathcal{D}_n^U(\eta)}} \sum_{\substack{(v^n,y^n):\\ (s^n,v^n,y^n)\in\mathcal{T}_n(\hat{S}^*\tilde{V}\tilde{Y})\\ \hat{S}^*\tilde{V}\tilde{W}_d\tilde{Y}\in\mathcal{D}_n^{SVW'Y}(\eta)}} \mathbb{P}\left(\bar{U}^n=u^n,\bar{V}^n=v^n\right)$$

$$\mathbb{P}\left(\bar{Y}^n=y^n|\bar{U}^n=u^n,\bar{V}^n=v^n,M'=0,S^n=s^n,H=1\right)\ \frac{e^{nH(\tilde{W}_d|\hat{S}^*\tilde{V}\tilde{Y})}e^{n(R'+\eta)}}{e^{n\left(H(\tilde{W}_d)-\eta\right)}}$$

$$\leq \sum_{\substack{\tilde{U}\hat{S}^*\tilde{V}\tilde{W}_d\tilde{Y}\\ \in\mathcal{D}_n^U(\eta)^c\times\mathcal{D}_n^{SVW'Y}(\eta)}} e^{nH(\tilde{U}\tilde{V}\tilde{Y}|\hat{S}^*)}e^{-n\left(H(\tilde{U}\tilde{V}\tilde{Y}|\hat{S}^*)+D\left(\tilde{U}\tilde{V}\tilde{Y}||\bar{U}\bar{V}\check{Y}\middle|\hat{S}^*\right)\right)}$$

$$\frac{e^{nH(\tilde{W}_d|\hat{S}^*\tilde{V}\tilde{Y})}e^{n(R'+\eta)}}{e^{n\left(H(\tilde{W}_d)-\eta\right)}}$$

$$\leq e^{-nE'_{3,n}}, \tag{B.99}$$

where,

$$E'_{3,n}:=\min_{\substack{\tilde{U}\hat{S}^*\tilde{V}\tilde{W}_d\tilde{Y}\\ \in\mathcal{D}_n^U(\eta)^c\times\mathcal{D}_n^{SVW'Y}(\eta)}} D\left(\tilde{U}\tilde{V}\tilde{Y}||\bar{U}\bar{V}\check{Y}\middle|\hat{S}^*\right)+I(\tilde{W}_d;\hat{S}^*,\tilde{V},\tilde{Y})-R'-O(\eta)$$

$$-\frac{|\mathcal{U}||\mathcal{V}||\mathcal{W}||\mathcal{Y}||\mathcal{S}|}{n}\log(n+1)$$

$$\overset{(n)}{\geq} \min_{\substack{\hat{V}\hat{Y}S:\hat{U}\hat{V}\hat{W}\hat{Y}S\ \in\\ \hat{\mathcal{L}}_h(\kappa_\alpha,\omega',P_S,P_{X|USW'})}} D\left(\hat{V}\hat{Y}||\bar{V}\check{Y}|S\right)+\rho'(\kappa_\alpha,\omega',P_S,P_{X|USW'})$$

$$
\begin{aligned}
&\quad - \zeta_q'(\kappa_\alpha, \omega', P_S) - O(\eta) \\
&= E_3'(\kappa_\alpha, \omega', P_S, P_{X|USW'}, P_{X'|US}) - O(\eta).
\end{aligned}
$$

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 (B.96), (B.97) and (B.99) that for a fixed $\big(P_S, \omega'(\cdot, P_S), P_{X|USW'}, P_{X'|US}\big) \in \mathcal{L}_h(\kappa_\alpha)$,

$$
\begin{aligned}
\kappa(\tau, \kappa_\alpha) \geq \min \big\{ & E_1'(\kappa_\alpha, \omega'), E_2'(\kappa_\alpha, \omega', P_S, P_{X|USW'}), E_3'(\kappa_\alpha, \omega', P_S, P_{X|USW'}, P_{X'|US}) \big\} \\
& - O(\eta).
\end{aligned}
$$

Maximizing over $\big(P_S, \omega'(\cdot, P_S), P_{X|USW'}, P_{X'|US}\big) \in \mathcal{L}_h(\kappa_\alpha)$ and noting that $\eta > 0$ is arbitrary completes the proof.

Appendix C

Proofs for Chapter 4

C.1 Proof of Lemma 4.1

Note that for a stochastic detector, the type I and type II error probabilities are linear functions of $P_{\hat{H}|M,V^n}$. As a result, for each fixed n and $f^{(n)}$, $\bar{\alpha}\left(f^{(n)}, g^{(n)}\right)$ and $\bar{\beta}\left(f^{(n)}, g^{(n)}\right)$ for a stochastic detector $g^{(n)}$ can be thought of as the type I and type II errors achieved by "time sharing" among a finite number of deterministic detectors. To see this, consider some ordering on the elements of the set $\mathcal{M} \times \mathcal{V}^n$ and let $\nu_i := P_{\hat{H}|M,V^n}(0|i)$, $i \in [1:N]$, where i denotes the i^{th} element of $\mathcal{M}\times\mathcal{V}^n$ and $N = |\mathcal{M}\times\mathcal{V}^n|$. Then, we can write

$$P_{\hat{H}|M,V^n} = \begin{bmatrix} \nu_1 & 1-\nu_1 \\ \nu_2 & 1-\nu_2 \\ \vdots & \vdots \\ \nu_N & 1-\nu_N \end{bmatrix}.$$

Then, it is easy to see that $P_{\hat{H}|M,V^n} = \sum_{i=1}^{n} \nu_i I_i$, where $I_i := [e_i \;\; 1-e_i]$ and e_i is an N length vector with 1 at the i^{th} component and 0 elsewhere. Now, suppose $(\bar{\alpha}_1^{(n)}, \bar{\beta}_1^{(n)})$ and $(\bar{\alpha}_2^{(n)}, \bar{\beta}_2^{(n)})$ denote the pair of type I and type II error probabilities achieved by deterministic detectors $g_1^{(n)}$ and $g_2^{(n)}$, respectively. Let $\mathcal{A}_{1,n}$ and $\mathcal{A}_{2,n}$ denote their corresponding acceptance regions for H_0. Let $g_\theta^{(n)}$ denote the stochastic detector formed by using $g_1^{(n)}$ and $g_2^{(n)}$ with probabilities θ and $1-\theta$, respectively. From the above mentioned linearity property, it follows that $g_\theta^{(n)}$ achieves type I and type II error probabilities of $\bar{\alpha}\left(f^{(n)}, g_\theta^{(n)}\right) = \theta\bar{\alpha}_1^{(n)} + (1-\theta)\bar{\alpha}_2^{(n)}$ and $\bar{\beta}\left(f^{(n)}, g_\theta^{(n)}\right) = \theta\bar{\beta}_1^{(n)} + (1-\theta)\bar{\beta}_2^{(n)}$, respectively. Let $r(\theta) = \min(\theta, 1-\theta)$. Then, for $\theta \in (0,1)$,

$$-\frac{1}{n}\log\left(\bar{\beta}\left(f^{(n)}, g_\theta^{(n)}\right)\right)$$

$$\leq \min\left(-\frac{1}{n}\log\left(\bar{\beta}_1^{(n)}\left(f^{(n)}, g_1^{(n)}\right)\right), -\frac{1}{n}\log\left(\bar{\beta}_2^{(n)}\left(f^{(n)}, g_2^{(n)}\right)\right)\right) - \frac{1}{n}\log(r(\theta)).$$

Hence, either

$$\begin{aligned}
&\bar{\alpha}_1^{(n)} \leq \bar{\alpha}\left(f^{(n)}, g_\theta^{(n)}\right) \\
\text{and } &-\frac{1}{n}\log\left(\bar{\beta}_1^{(n)}\left(f^{(n)}, g_1^{(n)}\right)\right) \geq -\frac{1}{n}\log\left(\bar{\beta}\left(f^{(n)}, g_\theta^{(n)}\right)\right) + \frac{1}{n}\log(r(\theta)),
\end{aligned}$$

or

$$\begin{aligned}
&\bar{\alpha}_2^{(n)} \leq \bar{\alpha}\left(f^{(n)}, g_\theta^{(n)}\right) \\
\text{and } &-\frac{1}{n}\log\left(\bar{\beta}_2^{(n)}\left(f^{(n)}, g_2^{(n)}\right)\right) \geq -\frac{1}{n}\log\left(\bar{\beta}\left(f^{(n)}, g_\theta^{(n)}\right)\right) + \frac{1}{n}\log(r(\theta)).
\end{aligned}$$

Thus, since $\frac{1}{n}\log(r(\theta)) \xrightarrow{(n)} 0$, a stochastic detector does not offer any advantage over deterministic detectors in the trade-off between the error-exponent and the type I error probability.

C.2 Proof of Lemma 4.2

Let $\tilde{P}^{(j)}_{S^n U^n V^n M}$ denote the joint distribution of the r.v.'s (S^n, U^n, V^n, M) under hypothesis H_j, $j = 0, 1$, and $\tilde{P}_{\hat{S}^n|M,V^n}$ denote an arbitrary stochastic function for $g_r^{(n)}$. Then, we have

$$\begin{aligned}
&\min_{g_r^{(n)}} \mathbb{E}\left[d\left(S^n, \hat{S}^n\right)|H = j\right] \\
&= \min_{\tilde{P}_{\hat{S}^n|M,V^n}} \mathbb{E}_{\tilde{P}^{(j)}}\left[d\left(S^n, \hat{S}^n\right)\right] \\
&= \min_{\{\tilde{P}_{\hat{S}_i|M,V^n}\}_{i=1}^n} \frac{1}{n}\sum_{i=1}^n \mathbb{E}_{\tilde{P}^{(j)}}\left[d\left(S_i, \hat{S}_i\right)\right] \\
&= \frac{1}{n}\sum_{i=1}^n \sum_{M=m, V^n=v^n} \tilde{P}^{(j)}_{MV^n}(m, v^n) \min_{\tilde{P}_{\hat{S}_i|M=m,V^n=v^n}} \sum_{\hat{s}_i} \tilde{P}_{\hat{S}_i|M=m,V^n=v^n}(\hat{s}_i) \\
&\qquad\qquad \mathbb{E}_{\tilde{P}^{(j)}_{S_i|M=m,V^n=v^n}}\left[d\left(S_i, \hat{s}_i\right)\right] \\
&= \frac{1}{n}\sum_{i=1}^n \sum_{M=m, V^n=v^n} \tilde{P}^{(j)}_{MV^n}(m, v^n)\, \mathbb{E}_{\tilde{P}^{(j)}_{S_i|M=m,V^n=v^n}}\left[d\left(S_i, \phi_{ij}(m, v^n)\right)\right],
\end{aligned}$$

where,

$$\phi_{ij}(m, v^n) = \arg\min_{\hat{s}\in\hat{\mathcal{S}}} \mathbb{E}_{\tilde{P}^{(j)}_{S_i|M=m,V^n=v^n}} \left[d(S_i, \hat{s})\right].$$

Continuing, we have

$$\begin{aligned}
&\min_{g_r^{(n)}} \mathbb{E}\left[d\left(S^n, \hat{S}^n\right)|H=j\right] \\
&= \frac{1}{n}\sum_{i=1}^{n} \sum_{M=m,V^n=v^n} \tilde{P}^{(j)}_{MV^n}(m, v^n) \min_{\phi_i(m,v^n)} \mathbb{E}_{\tilde{P}^{(j)}_{S_i|M=m,V^n=v^n}} \left[d\left(S_i, \phi_i(m, v^n)\right)\right] \\
&= \min_{\{\phi_i(m,v^n)\}_{i=1}^n} \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_{\tilde{P}^{(j)}} \left[d\left(S_i, \phi_i(M, V^n)\right)\right]. \qquad \text{(C.1)}
\end{aligned}$$

C.3 Proof of Lemma 4.4

We will first prove (4.14). Fix $\delta > 0$. For $\gamma > 0$, define the following sets:

$$\begin{aligned}
\mathcal{B}^{\delta}_{0,\gamma} &:= \{y^n \in T^n_{[P_Y]_\gamma} : P_{Y^n}(y^n) \geq P_{Y^n|I_X(X^n,\delta)=0}(y^n)\}, \\
\mathcal{C}^{\delta}_{0,\gamma} &:= \{y^n \in T^n_{[P_Y]_\gamma} : P_{Y^n}(y^n) < P_{Y^n|I_X(X^n,\delta)=0}(y^n)\}, \\
\mathcal{B}^{\delta}_{1,\gamma} &:= \{y^n \in T^n_{[Q_Y]_\gamma} : Q_{Y^n}(y^n) \geq Q_{Y^n|I_X(X^n,\delta)=0}(y^n)\}, \\
\mathcal{C}^{\delta}_{1,\gamma} &:= \{y^n \in T^n_{[Q_Y]_\gamma} : Q_{Y^n}(y^n) < Q_{Y^n|I_X(X^n,\delta)=0}(y^n)\}, \\
\mathcal{B}^{\delta}_{2,\gamma} &:= \{y^n \in T^n_{[Q_Y]_\gamma} : Q_{Y^n}(y^n) \geq Q_{Y^n|I_X(X^n,\delta)=1}(y^n)\}, \\
\mathcal{C}^{\delta}_{2,\gamma} &:= \{y^n \in T^n_{[Q_Y]_\gamma} : Q_{Y^n}(y^n) < Q_{Y^n|I_X(X^n,\delta)=1}(y^n)\}.
\end{aligned}$$

Then, we can write

$$\begin{aligned}
&\|Q_{Y^n} - Q_{Y^n|I_X(X^n,\delta)=1}\| \\
&= \sum_{y^n} |Q_{Y^n}(y^n) - Q_{Y^n|I_X(X^n,\delta)=1}(y^n)| \\
&= \sum_{y^n\notin T^n_{[Q_Y]_\gamma}} |Q_{Y^n}(y^n) - Q_{Y^n|I_X(X^n,\delta)=1}(y^n)| + \sum_{y^n\in T^n_{[Q_Y]_\gamma}} |Q_{Y^n}(y^n) - Q_{Y^n|I_X(X^n,\delta)=1}(y^n)| \\
&\leq \sum_{y^n\notin T^n_{[Q_Y]_\gamma}} Q_{Y^n}(y^n) + Q_{Y^n|I_X(X^n,\delta)=1}(y^n)
\end{aligned}$$

$$+ \sum_{y^n \in T^n_{[Q_Y]_\gamma}} |Q_{Y^n}(y^n) - Q_{Y^n|I_X(X^n,\delta)=1}(y^n)|. \tag{C.2}$$

Next, note that

$$\begin{aligned} Q_{Y^n|I_X(X^n,\delta)=1}(y^n) &= Q_{Y^n}(y^n)\ \frac{Q_{I_X(X^n,\delta)|Y^n}(1|y^n)}{Q(I_X(X^n,\delta)=1)} \\ &\leq \frac{Q_{Y^n}(y^n)}{Q(I_X(X^n,\delta)=1)} \leq 2Q_{Y^n}(y^n), \end{aligned} \tag{C.3}$$

for sufficiently large n (depending on $|\mathcal{X}|$), since $Q(I_X(X^n,\delta)=1) \xrightarrow{(n)} 1$. Thus, for n large enough,

$$\sum_{y^n \notin T^n_{[Q_Y]_\gamma}} Q_{Y^n}(y^n) + Q_{Y^n|I_X(X^n,\delta)=1}(y^n) \leq 3 \sum_{y^n \notin T^n_{[Q_Y]_\gamma}} Q_{Y^n}(y^n) \leq e^{-n\Omega(\gamma)}. \tag{C.4}$$

We can bound the term in (C.2) as follows.

$$\begin{aligned} &\sum_{y^n \in T^n_{[Q_Y]_\gamma}} |Q_{Y^n}(y^n) - Q_{Y^n|I_X(X^n,\delta)=1}(y^n)| \\ &= \sum_{y^n \in \mathcal{B}^\delta_{2,\gamma}} Q_{Y^n}(y^n) - Q_{Y^n|I_X(X^n,\delta)=1}(y^n) + \sum_{y^n \in \mathcal{C}^\delta_{2,\gamma}} Q_{Y^n|I_X(X^n,\delta)=1}(y^n) - Q_{Y^n}(y^n) \\ &= \sum_{y^n \in \mathcal{B}^\delta_{2,\gamma}} Q_{Y^n}(y^n) - Q_{Y^n|I_X(X^n,\delta)=1}(y^n) + \sum_{y^n \in \mathcal{C}^\delta_{2,\gamma}} Q_{Y^n|I_X(X^n,\delta)=1}(y^n) - Q_{Y^n}(y^n) \\ &= \sum_{y^n \in \mathcal{B}^\delta_{2,\gamma}} Q_{Y^n}(y^n) \left(1 - \frac{Q_{Y^n|I_X(X^n,\delta)=1}(y^n)}{Q_{Y^n}(y^n)}\right) \\ &\qquad + \sum_{y^n \in \mathcal{C}^\delta_{2,\gamma}} Q_{Y^n}(y^n) \left(\frac{Q_{Y^n|I_X(X^n,\delta)=1}(y^n)}{Q_{Y^n}(y^n)} - 1\right) \\ &= \sum_{y^n \in \mathcal{B}^\delta_{2,\gamma}} Q_{Y^n}(y^n) \left(1 - \frac{Q_{I_X(X^n,\delta)|Y^n}(1|y^n)}{Q(I_X(X^n,\delta)=1)}\right) \\ &\qquad + \sum_{y^n \in \mathcal{C}^\delta_{2,\gamma}} Q_{Y^n}(y^n) \left(\frac{Q_{I_X(X^n,\delta)|Y^n}(1|y^n)}{Q(I_X(X^n,\delta)=1)} - 1\right) \\ &\leq \sum_{y^n \in \mathcal{B}^\delta_{2,\gamma}} Q_{Y^n}(y^n) \left(1 - Q_{I_X(X^n,\delta)|Y^n}(1|y^n)\right) \end{aligned}$$

$$+ \sum_{y^n \in \mathcal{C}^{\delta}_{2,\gamma}} Q_{Y^n}(y^n) \left(\frac{1}{Q(I_X(X^n, \delta) = 1)} - 1 \right). \tag{C.5}$$

Let $P_{\tilde{Y}}$ denote the type of y^n and define

$$E^{(n)}(\delta, \gamma) := \min_{P_{\tilde{Y}} \in T^n_{[Q_Y]_\gamma}} \min_{P_{\tilde{X}} \in T^n_{[P_X]_\delta}} D(P_{\tilde{X}|\tilde{Y}} || Q_{X|Y} | P_{\tilde{Y}}).$$

Then, for $y^n \in T^n_{[Q_Y]_\gamma}$, arbitrary $\tilde{\gamma} > 0$ and n sufficiently large (depending on $|\mathcal{X}|, |\mathcal{Y}|, \delta, \gamma$), it follows from [20, Lemma 2.6] that

$$Q_{I_X(X^n,\delta)|Y^n}(1|y^n) \geq 1 - e^{-n\left(E^{(n)}(\delta,\gamma) - \tilde{\gamma}\right)},$$
$$\text{and } Q(I_X(X^n, \delta) = 1) \geq 1 - e^{-n(D(P_X||Q_X) - \tilde{\gamma})}.$$

From (C.2), (C.4) and (C.5), it follows that

$$\|Q_{Y^n} - Q_{Y^n|I_X(X^n,\delta)=1}\| \leq e^{-n\Omega(\gamma)} + e^{-n\left(E^{(n)}(\delta,\gamma) - \tilde{\gamma}\right)} + e^{-n(D(P_X||Q_X) - \tilde{\gamma})}. \tag{C.6}$$

We next show that $E^{(n)}(\delta, \gamma) > 0$ for sufficiently small $\delta > 0$ and $\gamma > 0$. This would imply that the R.H.S of (C.6) converges exponentially to zero (for $\tilde{\gamma}$ small enough) with exponent $\bar{\delta} := \min\left(\Omega(\gamma), E^{(n)}(\delta, \gamma) - \tilde{\gamma}, D(P_X||Q_X) - \tilde{\gamma}\right)$, thus proving (4.14). We can write,

$$E^{(n)}(\delta, \gamma) \geq \min_{P_{\tilde{Y}} \in T^n_{[Q_Y]_\gamma}} \min_{P_{\tilde{X}} \in T^n_{[P_X]_\delta}} D(P_{\tilde{X}} || \hat{Q}_X) \tag{C.7}$$

$$\geq 2 \left[\min_{P_{\tilde{Y}} \in T^n_{[Q_Y]_\gamma}} \min_{P_{\tilde{X}} \in T^n_{[P_X]_\delta}} \|P_{\tilde{X}} - \hat{Q}_X\|^2 \right], \tag{C.8}$$

where

$$\hat{Q}_X(x) := \sum_y P_{\tilde{Y}}(y) Q_{X|Y}(x|y).$$

Here, (C.7) follows due to the convexity of KL divergence (C.8) is due to Pinsker's inequality [20]. We also have from the triangle inequality satisfied by total variation

that,

$$\|P_{\tilde{X}} - \hat{Q}_X\| \geq \|P_X - Q_X\| - \|P_{\tilde{X}} - P_X\| - \|\hat{Q}_X - Q_X\|.$$

For $y^n \in T^n_{[Q_Y]_\gamma}$,

$$\|\hat{Q}_X - Q_X\| \leq \|Q_{X|Y} P_{\tilde{Y}} - Q_{XY}\| \leq \|P_{\tilde{Y}} - Q_Y\| \leq O(\gamma).$$

Also, for $P_{\tilde{X}} \in T^n_{[P_X]_\delta}$,

$$\|P_{\tilde{X}} - P_X\| \leq O(\delta).$$

Hence,

$$E^{(n)}(\delta, \gamma) \geq 2(\|P_X - Q_X\| - O(\gamma) - O(\delta))^2.$$

Since $P_X \neq Q_X$, $E^{(n)}(\delta, \gamma) > 0$ for sufficiently small $\gamma > 0$ and $\delta > 0$. This completes the proof of (4.14).

We next prove (4.16). Similar to (C.2) and (C.3), we have,

$$\begin{aligned}
&\|P_{Y^n} - P_{Y^n|I_X(X^n,\delta)=0}\| \\
&\leq \sum_{y^n \notin T^n_{[P_Y]_\gamma}} \left[P_{Y^n}(y^n) + P_{Y^n|I_X(X^n,\delta)=0}(y^n)\right] \\
&\qquad + \sum_{y^n \in T^n_{[P_Y]_\gamma}} |P_{Y^n}(y^n) - P_{Y^n|I_X(X^n,\delta)=0}(y^n)|, \qquad \text{(C.9)}
\end{aligned}$$

and

$$P_{Y^n|I_X(X^n,\delta)=0}(y^n) \leq 2P_{Y^n}(y^n), \qquad \text{(C.10)}$$

since $P(I_X(X^n, \delta) = 0) \xrightarrow{(n)} 1$.

Also, for $\gamma < \frac{\delta}{|\mathcal{Y}|}$ and sufficiently large n (depending on $\delta, \gamma, |\mathcal{X}|, |\mathcal{Y}|$), we have

$$\begin{aligned}
&\sum_{y^n \in T^n_{[P_Y]_\gamma}} |P_{Y^n}(y^n) - P_{Y^n|I_X(X^n,\delta)=0}(y^n)| \\
&= \sum_{y^n \in \mathcal{B}^\delta_{0,\gamma}} P_{Y^n}(y^n) - P_{Y^n|I_X(X^n,\delta)=0}(y^n) + \sum_{y^n \in \mathcal{C}^\delta_{0,\gamma}} P_{Y^n|I_X(X^n,\delta)=0}(y^n) - P_{Y^n}(y^n) \\
&\leq \sum_{y^n \in \mathcal{B}^\delta_{0,\gamma}} P_{Y^n}(y^n) \left(1 - P_{I_X(X^n,\delta)|Y^n}(0|y^n)\right) \\
&\qquad + \sum_{y^n \in \mathcal{C}^\delta_{0,\gamma}} P_{Y^n}(y^n) \left(\frac{1}{P(I_X(X^n,\delta)=0)} - 1\right) \\
&\leq \sum_{y^n \in \mathcal{B}^\delta_{0,\gamma}} P_{Y^n}(y^n) e^{-n\Omega(\delta-\gamma|\mathcal{Y}|)} + \sum_{y^n \in \mathcal{C}^\delta_{0,\gamma}} P_{Y^n}(y^n) e^{-n\Omega(\delta)} \qquad \text{(C.11)} \\
&\leq e^{-n\Omega(\delta-\gamma|\mathcal{Y}|)}, \qquad \text{(C.12)}
\end{aligned}$$

where, to obtain (C.11), we used

$$P(I_X(X^n,\delta)=0) \geq 1 - e^{-n\Omega(\delta)}, \qquad \text{(C.13)}$$

$$\text{and } P_{I_X(X^n,\delta)|Y^n}(0|y^n) \geq 1 - e^{-n\Omega(\delta-\gamma|\mathcal{Y}|)}, \text{ for } y^n \in \mathcal{B}^\delta_{0,\gamma} \text{ and } \gamma < \frac{\delta}{|\mathcal{Y}|}. \qquad \text{(C.14)}$$

Here, (C.13) follows from Lemma 2.12, and (C.14) follows from Lemma 2.10 and Lemma 2.12, in [20], respectively. Thus, from (C.9), (C.10) and (C.12), we can write that,

$$\|P_{Y^n} - P_{Y^n|I_X(X^n,\delta)=0}\| \leq e^{-n\Omega(\gamma)} + e^{-n\Omega(\delta-\gamma|\mathcal{Y}|)} \xrightarrow{(n)} 0.$$

This completes the proof of (4.16). The proof of (4.15) is exactly the same as (4.16), with the only difference that the sets $\mathcal{B}^\delta_{1,\gamma}$ and $\mathcal{C}^\delta_{1,\gamma}$ are used in place of $\mathcal{B}^\delta_{0,\gamma}$ and $\mathcal{C}^\delta_{0,\gamma}$, respectively.

C.4 Proof of Theorem 4.4 and Theorem 4.5

We first describe the encoding and decoding operations which is the same for both Theorem 4.4 and Theorem 4.5.

Codebook Generation: Fix a finite alphabet $\mathcal{W}$, a conditional distribution $P_{W|U}$, and positive number (small) $\delta > 0$. Let $\mu = O(\delta)$ subject to constraints that will be specified below, and let $\delta' := \frac{\delta}{2}$, $\hat{\delta} := |\mathcal{U}|\delta$, $\tilde{\delta} := 2\delta$, $\bar{\delta} := \frac{\delta'}{|\mathcal{V}|}$, and $M_n' := e^{n(I_P(U:W)+\mu)}$. Generate M_n' independent sequences $w^n(k)$, $k \in [M_n']$ randomly according to the distribution $\prod_{i=1}^n P_W(w_i)$, where

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

Denote this codebook by $\mathcal{C}^n$.

Encoding: For a given codebook $\mathcal{C}^n$, define a conditional probability distribution

$$P_{E_u}(j|u^n, \mathcal{C}^n) := \frac{\prod_{i=1}^n P_{U|W}(u_i|w_i(j))}{\sum_j \prod_{i=1}^n P_{U|W}(u_i|w_i(j)))}. \tag{C.15}$$

If $I_P(U;W) + \mu + \frac{|\mathcal{U}||\mathcal{W}|\log(n+1)}{n} > R$, the encoder performs uniform random binning on the sequences $w^n(k)$, $k \in [M_n']$ in $\mathcal{C}^n$, i.e., for each codeword in $\mathcal{C}^n$, it selects an index uniformly at random from the set $\left[e^{n\left(R-\frac{|\mathcal{U}||\mathcal{W}|\log(n+1)}{n}\right)}\right]$. Denote the bin assignment by $\mathcal{C}_B^n$ and the bin index selected for $w^n(k)$ by $f_B(k)$. If the observed sequence $U^n = u^n$ is typical, i.e., $u^n \in T^n_{[P_U]_{\delta'}}$, then the encoder outputs the message[1] $M := (T, M') = (t, m')$, $m' = f_B(j)$, $M \in [e^{nR}]$ or $M := (T, J) = (t, j)$ depending on whether $I_P(U;W) + \mu + \frac{|\mathcal{U}||\mathcal{W}|\log(n+1)}{n} > R$ or otherwise. Here, $j \in [M_n']$ is selected according to the probability $P_{E_u}(j|u^n, \mathcal{C}^n)$ and t denotes the index of the joint type of $(u^n, w^n(j))$ in the set of types $\mathcal{T}^n(\mathcal{U}\times\mathcal{W})$. If $u^n \notin T^n_{[P_U]_{\delta'}}$, the encoder outputs the *error* message $M = 0$. Note that the encoder $f^{(n)} : \mathcal{U}^n \mapsto \mathcal{M} := [e^{nR}]$ described by the above operations is a stochastic encoder with output M.

Decoding: If $M = 0$ or $t \notin T^n_{[P_{UW}]_\delta}$, $\hat{H} = 1$ is declared. Else, given $M = (t, m')$ and $V^n = v^n$, the detector looks for a typical sequence $\hat{w}^n := w^n(\hat{j}) \in T^n_{[P_W]_{\hat{\delta}}}$, in the codebook $\mathcal{C}^n$ such that

$$\hat{j} = \underset{\substack{l:\ m'=f_B(l),\\ w^n(l)\in T^n_{[P_W]_{\hat{\delta}}}}}{\arg\min} H_e(w^n(l)|v^n), \text{ if } I_P(U;W) + \mu + \frac{1}{n}|\mathcal{U}||\mathcal{W}|\log(n+1) > R,$$

[1]Note that this is valid assignment since the total number of types in $\mathcal{T}^n(\mathcal{U}\times\mathcal{W})$ is upper bounded by $(n+1)^{|\mathcal{U}||\mathcal{W}|}$ [20].

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

Denote the above decoding rule by $P_{ED}(m, v^n)$. The detector declares $\hat{H} = 0$ if $(\hat{w}^n, v^n) \in T^n_{[P_{WV}]_{\tilde{\delta}}}$. Else, $\hat{H} = 1$.

We next analyze the average of the type I and type II error probabilities achieved by the above scheme averaged over the random ensemble of codebooks $\mathcal{C}^n$ and $\mathcal{C}^n_B$.

Analysis of Type I error:

The system induced distribution when $H = 0$ is given by

$$\begin{aligned}
&\tilde{P}^{(0)}(s^n, u^n, v^n, j, w^n, m, \hat{j}, \hat{w}^n) \\
&= \left[\prod_{i=1}^n P_{SUV}(s_i, u_i, v_i, z_i)\right] P_{E_u}(j|u^n, \mathcal{C}^n) \mathbb{1}(W^n(j) = w^n) \\
&\quad \mathbb{1}(f_B(j) = m)\ \mathbb{1}(\hat{j} = P_{ED}(m, v^n))\ \mathbb{1}(W^n(\hat{j}) = \hat{w}^n), \qquad \text{if } u^n \in T^n_{[P_U]_{\delta'}},
\end{aligned} \tag{C.16}$$

and

$$\tilde{P}^{(0)}(s^n, u^n, v^n, m) = \left[\prod_{i=1}^n P_{SUV}(s_i, u_i, v_i)\right] \mathbb{1}(m = 0), \text{ if } u^n \notin T^n_{[P_U]_{\delta'}}. \tag{C.17}$$

Consider two auxiliary distribution $\tilde{\Psi}$ and Ψ defined as

$$\begin{aligned}
&\tilde{\Psi}^{(0)}(s^n, u^n, v^n, j, w^n, m, \hat{j}, \hat{w}^n) \\
&:= \left[\prod_{i=1}^n P_{SUV}(s_i, u_i, v_i)\right] P_{E_u}(j|u^n, \mathcal{C}^n) \mathbb{1}(W^n(j) = w^n) \mathbb{1}(f_B(j) = m) \\
&\qquad \mathbb{1}(\hat{j} = P_{ED}(m, v^n)) \mathbb{1}(W^n(\hat{j}) = \hat{w}^n),
\end{aligned} \tag{C.18}$$

and

$$\begin{aligned}
&\Psi^{(0)}(s^n, u^n, v^n, j, w^n, m, \hat{j}, \hat{w}^n) \\
&:= \frac{1}{M'_n}\ \mathbb{1}(W^n(j) = w^n)\ \left[\prod_{i=1}^n P_{U|W}(u_i|w_i)\right] \left[\prod_{i=1}^n P_{VS|U}(v_i, s_i|u_i)\right] \\
&\quad \mathbb{1}(f_B(j) = m)\ \mathbb{1}(\hat{j} = P_{ED}(m, v^n))\ \mathbb{1}(W^n(\hat{j}) = \hat{w}^n).
\end{aligned} \tag{C.19}$$

Note that the distributions $\tilde{P}^{(0)}, \Psi^{(0)}$ and $\tilde{\Psi}^{(0)}$ defined are r.v.'s, and depend on the codebook realizations $\mathcal{C}^n$ and $\mathcal{C}_B^n$. Also, observe that the stochastic encoder is chosen such that $P_{E_u}(j|u^n, \mathcal{C}^n) = \Psi^{(0)}(j|u^n)$ and hence, the only difference between the joint distribution $\Psi^{(0)}$ and $\tilde{\Psi}^{(0)}$ is the marginal distribution of U^n. By the soft-covering lemma [68] [70], it follows that for some $\gamma_1 > 0$,

$$\mathbb{E}_{\mathcal{C}^n}\left[\|\Psi^{(0)}_{U^n} - \tilde{\Psi}^{(0)}_{U^n}\|\right] \leq e^{-n\gamma_1} \xrightarrow{(n)} 0. \tag{C.20}$$

Hence, from (C.20) and Property 4.4.1(c), we have

$$\mathbb{E}_{\mathcal{C}^n}\left[\|\Psi^{(0)} - \tilde{\Psi}^{(0)}\|\right] \leq e^{-n\gamma_1}, \tag{C.21}$$

where the distributions $\Psi^{(0)}$ and $\tilde{\Psi}^{(0)}$ are over the r.v.'s given in (C.18). Also, note that the only difference between the distributions $\tilde{P}^{(0)}$ and $\tilde{\Psi}^{(0)}$ is P_{E_u} when $u^n \notin T^n_{[P_U]_{\delta'}}$. Since

$$\mathbb{P}(U^n \notin T^n_{[P_U]_{\delta'}}|H = 0) \leq e^{-n\Omega(\delta')}, \tag{C.22}$$

it follows that

$$\mathbb{E}_{\mathcal{C}^n}\left[\|\tilde{P}^{(0)} - \tilde{\Psi}^{(0)}\|\right] \leq e^{-n\Omega(\delta')}. \tag{C.23}$$

Equations (C.21) and (C.23) together imply via Property 4.4.1(b) that

$$\mathbb{E}_{\mathcal{C}^n}\left[\|\tilde{P}^{(0)} - \Psi^{(0)}\|\right] \leq e^{-n\Omega(\delta')} + e^{-\gamma_1 n} \xrightarrow{(n)} 0. \tag{C.24}$$

This means that for large n, the system distribution $\tilde{P}^{(0)}$ induced by encoding and decoding operations (when H_0 is the true hypothesis) can be approximated by that under $\Psi^{(0)}$. Let $\tilde{P}^{(1)}$ and $\tilde{\Psi}^{(1)}$ be defined by the R.H.S. of (C.16), (C.17) and (C.18), with P_{SUV} replaced by Q_{SUV}. Let $\Psi^{(1)}$ denote the R.H.S. of (C.19) with $P_{VS|U}$ replaced by $Q_{VS|U}$. Note that under joint distribution $\Psi^{(l)}$, $l \in \{0, 1\}$,

$$S_i - (W_i(J), V_i) - (M, W^n(J), V^n), \ i \in [n]. \tag{C.25}$$

Also, since $I_P(U;W)+\mu>0$, by the application of soft-covering lemma,

$$\mathbb{E}_{\mathcal{C}^n}\left[\sum_{i=1}^{n}\|P_W-\Psi^{(l)}_{W_i}(J)\|\big|H=l\right]\leq e^{-\gamma_3 n}\xrightarrow{(n)}0,\ l=0,1, \tag{C.26}$$

for some $\gamma_3>0$.

If $Q_U=P_U$, then it again follows from the soft-covering lemma that

$$\mathbb{E}_{\mathcal{C}^n}\left[\|\Psi^{(1)}_{U^n}-\tilde{\Psi}^{(1)}_{U^n}\|\right]\leq e^{-\gamma_1 n}\xrightarrow{(n)}0, \tag{C.27}$$

thereby implying that

$$\mathbb{E}_{\mathcal{C}^n}\left[\|\Psi^{(1)}-\tilde{\Psi}^{(1)}\|\right]\leq e^{-\gamma_1 n}, \tag{C.28}$$

where the distributions $\Psi^{(1)}$ and $\tilde{\Psi}^{(1)}$ are over the r.v.'s given in (C.18). Also, note that the only difference between the distributions $\tilde{P}^{(1)}$ and $\tilde{\Psi}^{(1)}$ is P_{E_u} when $u^n\notin T^n_{[P_U]_{\delta'}}$. Since $Q_U=P_U$ implies $\mathbb{P}(U^n\notin T^n_{[P_U]_{\delta'}}|H=1)\leq e^{-n\Omega(\delta')}$, it follows that

$$\mathbb{E}_{\mathcal{C}^n}\left[\|\tilde{P}^{(1)}-\tilde{\Psi}^{(1)}\|\right]\leq e^{-n\Omega(\delta')}. \tag{C.29}$$

Eqns. (C.28) and (C.29) together imply that

$$\mathbb{E}_{\mathcal{C}^n}\left[\|\tilde{P}^{(1)}-\Psi^{(1)}\|\right]\leq e^{-n\Omega(\delta')}+e^{-\gamma_1 n}\xrightarrow{(n)}0. \tag{C.30}$$

This means that for large n, the system distribution $\tilde{P}^{(1)}$ induced by encoding and decoding operations when H_1 is the true hypothesis can be approximated by that under $\Psi^{(1)}$.

Also, from (C.19), (C.24) and (C.26) and the weak law of large numbers,

$$\mathbb{P}((U^n,W^n(J))\in T^n_{[P_{UW}]_\delta}|H=0)\geq 1-e^{-n\Omega(\delta)}\xrightarrow{(n)}1. \tag{C.31}$$

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

$$\mathcal{E}_{TE}=\left\{(U^n,V^n)\notin T^n_{[P_{UV}]_{\bar{\delta}}}\right\},$$

$$\mathcal{E}_{SE} = \{T \notin T^n_{[P_{UW}]_\delta}\},$$
$$\mathcal{E}_{ME} = \left\{(V^n, W^n(J)) \notin T^n_{[P_{VW}]_{\tilde{\delta}}}\right\},$$
$$\mathcal{E}_{DE} = \Bigg\{\exists\, l \in \left[e^{n(I(U;W)+\delta')}\right],\ l \neq J : f_B(l) = f_B(J),\ W^n(l) \in T^n_{[P_W]_\delta},$$
$$H_e(W^n(l)|V^n) \leq H_e(W^n(J)|V^n)\Bigg\}.$$

Let $\mathcal{E} := \mathcal{E}_{TE} \cup \mathcal{E}_{SE} \cup \mathcal{E}_{ME} \cup \mathcal{E}_{DE}$. Then, the type I error can be upper bounded as

$$\alpha(f^{(n)}) := \inf_{g^{(n)}} \alpha(f^{(n)}, g^{(n)}) \leq \mathbb{P}(\mathcal{E}|H=0).$$

$\mathbb{P}(\mathcal{E}_{TE})$ tends to 0 asymptotically by the weak law of large numbers. From (C.31), $\mathbb{P}(\mathcal{E}_{SE}) \xrightarrow{(n)} 0$. Given $\mathcal{E}^c_{SE}$ and $\mathcal{E}^c_{TE}$ holds, it follows from the Markov chain relation $V - U - W$ and the Markov lemma [72], that $\mathbb{P}(\mathcal{E}_{ME}) \xrightarrow{(n)} 0$. Also, as in the proof of Theorem 2.2, it follows that

$$\mathbb{P}(\mathcal{E}_{DE}|\ V^n = v^n, W^n(J) = w^n, \mathcal{E}^c_{ME} \cap \mathcal{E}^c_{SE} \cap \mathcal{E}^c_{TE}, H=0)$$
$$\leq e^{-n(R - I_P(U;W|V) - \delta_1^{(n)})}, \tag{C.32}$$

where $\delta_1^{(n)} \xrightarrow{(n)} \mu + O(\delta)$. Thus, if $R > I_P(U;W|V)$, it follows by choosing $\mu = O(\delta)$ appropriately, that for $\delta > 0$ small enough, the R.H.S. of (C.32) tends to zero asymptotically. By the union bound, $\alpha(f^{(n)}) \xrightarrow{(n)} 0$.

Analysis of Type II error:

Note that a type II error occurs only if $V^n \in T^n_{[P_V]_{\delta''}}$, $\delta'' = |\mathcal{W}|\tilde{\delta}$, and $M \neq 0$, i.e., $U^n \in T^n_{[P_U]_{\delta'}}$ and $T \in T^n_{[P_{UW}]_\delta}$. Hence, we can restrict the type II error analysis to only such (U^n, V^n). Denote the event that a type II error happens by $\mathcal{D}_0$. The type II error probability can be written as

$$\beta(f^{(n)}, \epsilon) = \sum_{(u^n, v^n) \in\, \mathcal{U}^n \times \mathcal{V}^n} \mathbb{P}(U^n = u^n, V^n = v^n | H=1)$$
$$\mathbb{P}(\mathcal{D}_0 | U^n = u^n, V^n = v^n, H=1). \tag{C.33}$$

Let $\mathcal{E}_{NE} := \mathcal{E}_{SE}^c \cap \{V^n \in T^n_{[V]_{\delta''}}\} \cap \{U^n \in T^n_{[U]_{\delta'}}\}$. The last term in (C.33) can be upper bounded as follows.

$$\begin{aligned}
&\mathbb{P}(\mathcal{D}_0 | U^n = u^n, V^n = v^n, H = 1) \\
&= \mathbb{P}(\mathcal{E}_{NE} | U^n = u^n, V^n = v^n, H = 1)\ \mathbb{P}(\mathcal{D}_0 | U^n = u^n, V^n = v^n, \mathcal{E}_{NE}, H = 1) \\
&\leq \mathbb{P}(\mathcal{D}_0 | U^n = u^n, V^n = v^n, \mathcal{E}_{NE}, H = 1).
\end{aligned}$$

By averaging over all codebooks $\mathcal{C}^n$, $\mathcal{C}_B^n$ and using the symmetry of the codebook generation, encoding and decoding procedure, we can write,

$$\begin{aligned}
&\mathbb{P}(\mathcal{D}_0 |\ U^n = u^n, V^n = v^n, \mathcal{E}_{NE}, H = 1) \\
&= \sum_{w^n \in \mathcal{W}^n} \mathbb{P}(W^n(1) = w^n | U^n = u^n, V^n = v^n, J = 1, f_B(J) = 1,\ \mathcal{E}_{NE}, H = 1) \\
&\mathbb{P}(\mathcal{D}_0 | U^n = u^n, V^n = v^n, J = 1, f_B(J) = 1, W^n(1) = w^n,\ \mathcal{E}_{NE}, H = 1). \qquad \text{(C.34)}
\end{aligned}$$

The first term in (C.34) can upper bounded as

$$\begin{aligned}
&\mathbb{P}(W^n(1) = w^n | U^n = u^n, V^n = v^n, J = 1, f_B(J) = 1,\ \mathcal{E}_{NE}, H = 1) \\
&\leq \frac{1}{|T_{P_{\tilde{W}|\tilde{U}}}|} \leq e^{-n(H(\tilde{W}|\tilde{U}) - \frac{1}{n}|\mathcal{U}||\mathcal{W}| \log(n+1))}. \qquad \text{(C.35)}
\end{aligned}$$

To obtain (C.35), we used the fact that $P_{E_u}(1|u^n, \mathcal{C}^n)$ in (C.15) is invariant to the joint type $T_{P_{\tilde{U}\tilde{W}}}$ of $(U^n, W^n(1)) = (u^n, w^n)$ (keeping all the other codewords fixed), which in turn implies that given $\mathcal{E}_{NE}$ and the type $P_{\tilde{U}}$ of $U^n = u^n$, each sequence in the conditional type $T_{P_{\tilde{W}|\tilde{U}}}$ is equally likely (in the randomness induced by the random codebook generation and stochastic encoding in (C.15)) and its probability is upper bounded by $\frac{1}{|T_{P_{\tilde{W}|\tilde{U}}}|}$. Defining the events

$$\begin{aligned}
\mathcal{E}_{BE} := \Big\{ &\exists\ l \in \left[M_n'\right],\ l \neq J,\ f_B(l) = M,\ W^n(l)) \in T^n_{[P_W]_{\tilde{\delta}}}, \\
&(V^n, W^n(l)) \in T^n_{[P_{VW}]_{\tilde{\delta}}} \Big\}, \qquad \text{(C.36)}
\end{aligned}$$

$$\mathcal{F} := \{U^n = u^n, V^n = v^n, J = 1, f_B(J) = 1, W^n(1) = w^n,\ \mathcal{E}_{NE}\}, \qquad \text{(C.37)}$$

$$\mathcal{F}_1 := \{U^n = u^n, V^n = v^n, J = 1, f_B(J) = 1, W^n(1) = w^n,\ \mathcal{E}_{NE},\ \mathcal{E}_{BE}^c\}, \qquad \text{(C.38)}$$

and

$$\mathcal{F}_2 := \{U^n = u^n, V^n = v^n, J = 1, f_B(J) = 1, W^n(1) = w^n,\ \mathcal{E}_{NE},\ \mathcal{E}_{BE}\}, \tag{C.39}$$

the last term in (C.34) can be written as

$$\begin{aligned}\mathbb{P}(\mathcal{D}_0|\mathcal{F}, H = 1) =& \mathbb{P}(\mathcal{E}_{BE}^c|\mathcal{F}, H = 1)\ \mathbb{P}(\mathcal{D}_0|\mathcal{F}_1, H = 1) \\ &+ \mathbb{P}(\mathcal{E}_{BE}|\mathcal{F}, H = 1)\ \mathbb{P}(\mathcal{D}_0|\mathcal{F}_2, H = 1).\end{aligned} \tag{C.40}$$

The analysis of the terms in (C.40) is essentially similar to that given in the proof of Theorem 2.2, except for a subtle difference that we mention next. In order to bound the *binning* error event $\mathcal{E}_{BE}$, we require a bound similar to

$$\mathbb{P}\left(W^n(l) = \tilde{w}^n|\mathcal{F}\right) \leq 2\ \mathbb{P}(W^n(l) = \tilde{w}^n),\ \forall\ \tilde{w}^n \in \mathcal{W}^n, \tag{C.41}$$

that is used in the proof of Theorem 2.2. Note that the stochastic encoding scheme considered here is different from the one in Theorem 2.2. In place of (C.41), we will show that for $l \neq 1$,

$$\mathbb{P}(W^n(l) = \tilde{w}^n|\ \mathcal{F}) \leq (1 + o(1))\ \mathbb{P}(W^n(l) = \tilde{w}^n).$$

We can write

$$\begin{aligned}&\mathbb{P}(W^n(l) = \tilde{w}^n|\mathcal{F}) \\ &= \mathbb{P}(W^n(l) = \tilde{w}^n|U^n = u^n, V^n = v^n)\frac{\mathbb{P}(W^n(1) = w^n|W^n(l) = \tilde{w}^n, U^n = u^n, V^n = v^n)}{\mathbb{P}(W^n(1) = w^n|U^n = u^n, V^n = v^n)} \\ &\quad \frac{\mathbb{P}(J = 1|W^n(1) = w^n, W^n(l) = \tilde{w}^n, U^n = u^n, V^n = v^n)}{\mathbb{P}(J = 1|W^n(1) = w^n, U^n = u^n, V^n = v^n)} \\ &\quad \frac{\mathbb{P}(f_B(J) = 1|J = 1, W^n(1) = w^n, W^n(l) = \tilde{w}^n, U^n = u^n, V^n = v^n)}{\mathbb{P}(f_B(J) = 1|J = 1, W^n(1) = w^n, U^n = u^n, V^n = v^n)} \\ &\quad \frac{\mathbb{P}(\mathcal{E}_{NE}|f_B(J) = 1, J = 1, W^n(1) = w^n, W^n(l) = \tilde{w}^n, U^n = u^n, V^n = v^n)}{\mathbb{P}(\mathcal{E}_{NE}|f_B(J) = 1, J = 1, W^n(1) = w^n, U^n = u^n, V^n = v^n)}.\end{aligned} \tag{C.42}$$

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

$$\begin{aligned}&\mathbb{P}(W^n(1)=w^n|W^n(l)=\tilde{w}^n,U^n=u^n,V^n=v^n)\\&\quad=\mathbb{P}(W^n(1)=w^n|U^n=u^n,V^n=v^n),\end{aligned}$$

and

$$\begin{aligned}&\mathbb{P}(f_B(J)=1|J=1,W^n(1)=w^n,W^n(l)=\tilde{w}^n,U^n=u^n,V^n=v^n)\\&=\mathbb{P}(f_B(J)=1|J=1,W^n(1)=w^n,U^n=u^n,V^n=v^n).\end{aligned}\tag{C.43}$$

Also, note that

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

Next, consider the term in (C.42). Let

$$\begin{aligned}\mathcal{F}'&:=\{W^n(1)=w^n,U^n=u^n,V^n=v^n\},\\\mathcal{F}''&:=\{W^n(1)=w^n,W^n(l)=\tilde{w}^n,U^n=u^n,V^n=v^n\},\\\mathcal{C}_l^n&:=\mathcal{C}^n\backslash\{W^n(1)\},\end{aligned}$$

and $\mathcal{C}_{1,l}^n:=\mathcal{C}^n\backslash\{W^n(1),W^n(l)\}$.

Then, the numerator and denominator of (C.42) can be written as,

$$\begin{aligned}&\mathbb{P}(J=1|\mathcal{F}'')\\&=\mathbb{E}_{\mathcal{C}_{1,l}^n|\mathcal{F}''}\left[\frac{\prod_{i=1}^n P_{U|W}(u_i|w_i)}{\prod_{i=1}^n P_{U|W}(u_i|w_i)+\prod_{i=1}^n P_{U|W}(u_i|\tilde{w}_i)+\sum_{j\neq 1,l}\prod_{i=1}^n P_{U|W}(u_i|W_i(j))}\right]\\&\leq\mathbb{E}_{\mathcal{C}_{1,l}^n|\mathcal{F}''}\left[\frac{\prod_{i=1}^n P_{U|W}(u_i|w_i)}{\prod_{i=1}^n P_{U|W}(u_i|w_i)+\sum_{j\neq 1,l}\prod_{i=1}^n P_{U|W}(u_i|W_i(j))}\right],\end{aligned}\tag{C.44}$$

and

$$\mathbb{P}(J=1|\mathcal{F}') = \mathbb{E}_{\mathcal{C}_1^n|\mathcal{F}'}\left[\frac{\prod_{i=1}^n P_{U|W}(u_i|w_i)}{\prod_{i=1}^n P_{U|W}(u_i|w_i)+\sum_{j\neq 1}\prod_{i=1}^n P_{U|W}(u_i|W_i(j))}\right], \tag{C.45}$$

respectively. Note that $\mathcal{C}_1^n$ and $\mathcal{C}_{1,l}^n$ consists of codewords that are distributed i.i.d. according to $\prod_{i=1}^n P_W$ given $\mathcal{F}'$ and $\mathcal{F}''$, respectively. The R.H.S. of (C.44) (resp. (C.45)) denote the average probability that $J=1$ is chosen by the stochastic encoder P_{E_u} given $W^n(1)=w^n$, $U^n=u^n$ and $\lceil M_n' \rceil - 2$ (resp. $\lceil M_n' \rceil - 1$) other independent codewords in the codebook. Note that for $W^n(j)=w^n(j)$,

$$\prod_{i=1}^n P_{U|W}(u_i|w_i(j)) = e^{-n\left(H(\tilde{U}|\tilde{W}_j)+D(P_{\tilde{U}|\tilde{W}_j}||P_{U|W}|P_{\tilde{W}_j})\right)}, \tag{C.46}$$

where $P_{\tilde{U}\tilde{W}_j}$ denote the joint type of $(u^n, w^n(j))$. Since each term of the form above is exponentially decreasing in n, it follows that the term inside the square braces in (C.44) and (C.45) differ (significantly) asymptotically only if the event

$$\begin{aligned}\mathcal{E}_{max}^l :=& \left\{\prod_{i=1}^n P_{U|W}(u_i|W_i(l)) \geq \max\left(\left\{\prod_{i=1}^n P_{U|W}(u_i|W_i(j)),\ j\in\lceil M_n'\rceil\backslash\{1\}\right\}\right.\right.\\ &\left.\left.\cup\left\{\prod_{i=1}^n P_{U|W}(u_i|w_i)\right\}\right)\right\},\end{aligned} \tag{C.47}$$

occurs, and being probabilities, the difference is atmost 1. Since the probability of the event $\mathcal{E}_{max}^l$ decreases as $2^{-e^{n(I_P(U;W)+\mu)}}$ with n , we have that

$$\begin{aligned}&\frac{\mathbb{P}(J=1|\mathcal{F}'')}{\mathbb{P}(J=1|\mathcal{F}')}\\ &\leq \frac{\mathbb{E}_{\mathcal{C}_{1,l}^n|\mathcal{F}''}\left[\frac{\prod_{i=1}^n P_{U|W}(u_i|w_i)}{\prod_{i=1}^n P_{U|W}(u_i|w_i)+\sum_{j\neq 1,l}\prod_{i=1}^n P_{U|W}(u_i|W_i(j))}\right]}{\mathbb{E}_{\mathcal{C}_1^n|\mathcal{F}'}\left[\frac{\prod_{i=1}^n P_{U|W}(u_i|w_i)}{\prod_{i=1}^n P_{U|W}(u_i|w_i)+\sum_{j\neq 1}\prod_{i=1}^n P_{U|W}(u_i|W_i(j))}\right]}\\ &\leq \frac{(1+o(1))\mathbb{E}_{\mathcal{C}_{1,l}^n|\mathcal{F}''}\left[\frac{\prod_{i=1}^n P_{U|W}(u_i|w_i)}{\prod_{i=1}^n P_{U|W}(u_i|w_i)+\sum_{j\neq 1,l}\prod_{i=1}^n P_{U|W}(u_i|W_i(j))}\right]}{\mathbb{E}_{\mathcal{C}_{1,l}^n|\mathcal{F}''}\left[\frac{\prod_{i=1}^n P_{U|W}(u_i|w_i)}{\prod_{i=1}^n P_{U|W}(u_i|w_i)+\sum_{j\neq 1,l}\prod_{i=1}^n P_{U|W}(u_i|W_i(j))}\right]-2^{-(e^{n(I_P(U;W)+\mu)}-1)}}\\ &\leq 1+o(1),\end{aligned} \tag{C.48}$$

where the final inequality in (C.48) follows since the term within the expectation which is exponential in order dominates the double exponential term $2^{-e^{n(I_P(U;W)+\mu)}}$.

The analysis of the other terms in (C.40) is the same as in the SHA scheme in [11] and follows similar[2] to Theorem 2.2. This results in the error-exponent (within a additive $O(\delta)$ term) claimed in the Theorem. By the random coding argument followed by the standard expurgation technique [21] (see also proof of Theorem 2.2), there exists a deterministic codebook pair $(\mathcal{C}^n, \mathcal{C}_B^n)$ such that the type I and type II error probabilities are within a constant multiplicative factor of their average values over the random ensemble, and

$$S_i - (w_i(J), V_i) - (M, w^n(J), V^n),\ i \in [n], \tag{C.49}$$

$$\|\tilde{P}^{(0)} - \Psi^{(0)}\| \leq e^{-\gamma_4 n}, \tag{C.50}$$

$$\|\tilde{P}^{(1)} - \Psi^{(1)}\| \leq e^{-\gamma_4 n}, \quad \text{if } Q_U = P_U, \tag{C.51}$$

$$\text{and } \sum_{i=1}^{n} \|P_W - \Psi_{w_i}^{(l)}(J)\| \leq e^{-\gamma_5 n}\ l = 0, 1, \tag{C.52}$$

where γ_4 and γ_5 are some positive numbers. Since the average type I error probability for our scheme tends to zero asymptotically, and the error-exponent is unaffected by a constant multiplicative scaling of the type II error probability, this codebook achieves the same type I error probability and error-exponent as the average over the random ensemble. Using this deterministic codebook for encoding and decoding, we first lower bound the equivocation and average distortion of S^n at the detector as follows:

First consider the equivocation of S^n under the null hypothesis.

$$\begin{aligned}
&H(S^n|M, V^n, H = 0) \\
&\geq \mathbb{P}(M \neq 0|H = 0) H(S^n|M \neq 0, V^n, H = 0) \\
&\geq (1 - e^{-n\Omega(\delta)})\ H(S^n|M \neq 0, V^n, H = 0) && \text{(C.53)} \\
&\geq (1 - e^{-n\Omega(\delta)})\ H(S^n|w^n(J), V^n, H = 0) && \text{(C.54)} \\
&= (1 - e^{-n\Omega(\delta)})\ H_{\tilde{P}^{(0)}}(S^n|w^n(J), V^n) && \text{(C.55)} \\
&\geq (1 - e^{-n\Omega(\delta)})\ H_{\Psi^{(0)}}(S^n|w^n(J), V^n) - 2e^{-\gamma_4 n} \log\left(\frac{|\mathcal{S}|^n |\mathcal{V}|^n}{e^{-\gamma_4 n}}\right) && \text{(C.56)}
\end{aligned}$$

[2] In Theorem 2.2, the communication channel between the observer and the detector is a DMC. However, since the coding scheme is a separation based scheme, the type II error-exponent when the channel is noiseless can be recovered by setting $E_3(\cdot)$ and $E_4(\cdot)$ in Theorem 2.2 to ∞.

$$= \sum_{i=1}^{n} H_{\Psi^{(0)}}(S_i|w_i(J), V_i) - e^{-n\Omega(\delta)} \sum_{i=1}^{n} H_{\Psi^{(0)}}(S_i|w_i(J), V_i) - o(1) \tag{C.57}$$

$$\geq \sum_{i=1}^{n} H_{\Psi^{(0)}}(S_i|w_i(J), V_i) - ne^{-n\Omega(\delta)} H_P(S|V) - o(1) \tag{C.58}$$

$$= \left[\sum_{i=1}^{n} H_{\Psi^{(0)}}(S_i|w_i(J), V_i) \right] - o(1) \tag{C.59}$$

$$= nH_P(S|W, V) - o(1). \tag{C.60}$$

Here, (C.53) follows from (C.22); (C.54) follows since M is a function of $w^n(J)$ for a deterministic codebook; (C.56) follows from (C.50) and Lemma 4.3; (C.57) follows from (C.19); and (C.60) follows from (C.52) and $\Psi^{(0)}_{S_iV_i|w_i} = P^{(0)}_{SV|W},\ i \in [n]$.

If $Q_U = P_U$, it follows similarly to above that

$$H(S^n|M, V^n, H = 1)$$

$$\geq (1 - e^{-n\Omega(\delta)})\ H_{\Psi^{(1)}}(S^n|w^n(J), V^n) - 2e^{-\gamma_4 n} \log \left(\frac{|\mathcal{S}|^n |\mathcal{V}|^n}{e^{-\gamma_4 n}} \right) \tag{C.61}$$

$$= \sum_{i=1}^{n} H_{\Psi^{(1)}}(S_i|w_i(J), V_i) - e^{-n\Omega(\delta)} \sum_{i=1}^{n} H_{\Psi^{(1)}}(S_i|w_i(J), V_i) - o(1) \tag{C.62}$$

$$\geq \sum_{i=1}^{n} H_{\Psi^{(1)}}(S_i|w_i(J), V_i) - ne^{-n\Omega(\delta)} H_Q(S|V) - o(1) \tag{C.63}$$

$$= \left[\sum_{i=1}^{n} H_{\Psi^{(1)}}(S_i|w_i(J), V_i) \right] - o(1) \tag{C.64}$$

$$= nH_Q(S|W, V) - o(1). \tag{C.65}$$

Finally, consider the case $H = 1$ and $Q_U \neq P_U$. We have for δ' small enough that,

$$\begin{aligned} \mathbb{P}\left(M = 0|H = 1\right) &= \mathbb{P}\left(U^n \notin T^n_{[P_U]_{\delta'}}|H = 1\right) \\ &\geq 1 - e^{-n(D(P_U||Q_U) - O(\delta'))} \xrightarrow{(n)} 1. \end{aligned} \tag{C.66}$$

Hence, for δ' small enough, we can write

$$H(S^n|M, V^n, H = 1) \geq H(S^n|M, V^n, I_U(U^n, \delta'), H = 1)$$

$$\geq \left(1 - e^{-n(D(P_U||Q_U) - O(\delta'))}\right)\ H(S^n|M, V^n, I_U(U^n, \delta') = 1, H = 1) \tag{C.67}$$

$$= \left(1 - e^{-n(D(P_U||Q_U)-O(\delta'))}\right) \; H(S^n|V^n, I_U(U^n,\delta') = 1, H = 1) \tag{C.68}$$

$$\geq \left(1 - e^{-n(D(P_U||Q_U)-O(\delta'))}\right) \; (H(S^n|V^n, H = 1) - o(1)) \tag{C.69}$$

$$= nH_Q(S|V) - ne^{-n(D(P_U||Q_U)-O(\delta'))}H_Q(S|V) - o(1) = nH_Q(S|V) - o(1). \tag{C.70}$$

Here, (C.67) follows from (C.66); (C.68) follows since $I_U(U^n,\delta') = 1$ implies $M = 0$; (C.69) follows from Lemma 4.3 and (4.14). Thus, since $\delta > 0$ is arbitrary, we have shown that for $\epsilon \in (0,1)$, $(R,\kappa,\Lambda_0,\Lambda_1) \in \mathcal{R}_e^s(\epsilon)$ if (4.17)-(4.20) holds.

On the other hand, average distortion of S^n at the detector can be lower bounded under $H = 0$ as follows:

$$\min_{g_r^{(n)}} \mathbb{E}\left[d\left(S^n, \hat{S}^n\right)|H = 0\right]$$

$$= \min_{\{\bar{\phi}_i(m,v^n)\}_{i=1}^n} \mathbb{E}_{\tilde{P}^{(0)}}\left[\sum_{i=1}^n d\left(S_i, \bar{\phi}_i(M,V^n)\right)\right] \tag{C.71}$$

$$\geq \min_{\{\bar{\phi}_i(m,v^n)\}_{i=1}^n} \mathbb{E}_{\Psi^{(0)}}\left[\sum_{i=1}^n d(S_i, \bar{\phi}_i(M,V^n))\right] - ne^{-n\gamma_4}D_m \tag{C.72}$$

$$\geq \min_{\{\bar{\phi}_i(\cdot,\cdot)\}_{i=1}^n} \mathbb{E}_{\Psi^{(0)}}\left[\sum_{i=1}^n d(S_i, \bar{\phi}_i(w_i(J),V_i))\right] - ne^{-n\gamma_4}D_m \tag{C.73}$$

$$\geq n \min_{\{\phi(\cdot,\cdot)\}} \mathbb{E}_P\left[d(S,\phi(W,V))\right] - n\left(e^{-n\gamma_4} + e^{-n\gamma_5}\right)D_m. \tag{C.74}$$

$$= n \min_{\{\phi(\cdot,\cdot)\}_{i=1}^n} \mathbb{E}_P\left[d(S,\phi(W,V))\right] - o(1). \tag{C.75}$$

Here, (C.71) follows from Lemma 4.2; (C.72) follows from (C.50) and boundedness of distortion measure; (C.73) follows from (C.49); (C.74) follows from (C.52) and the fact that $\Psi^{(0)}_{S_iV_i|w_i} = P^{(0)}_{SV|W},\ i \in [n]$.

Next, consider that the alternate hypothesis holds and that $Q_U = P_U$. Then, similarly to above, we can write

$$\min_{g_r^{(n)}} \mathbb{E}\left[d\left(S^n, \hat{S}^n\right)|H = 1\right]$$

$$= \min_{\{\bar{\phi}_i(m,v^n)\}_{i=1}^n} \mathbb{E}_{\tilde{P}^{(1)}}\left[\sum_{i=1}^n d\left(S_i, \phi_i(M,V^n)\right)\right]$$

$$\geq \min_{\{\bar{\phi}_i(m,v^n)\}_{i=1}^n} \mathbb{E}_{\Psi^{(1)}}\left[\sum_{i=1}^n d(S_i, \phi_i(M,V^n))\right] - ne^{-n\gamma_4}D_m \tag{C.76}$$

$$\geq \min_{\{\phi_i(\cdot,\cdot)\}_{i=1}^n} \mathbb{E}_{\Psi^{(1)}} \left[\sum_{i=1}^n d(S_i, \phi_i(w_i, V_i)) \right] - ne^{-n\gamma_4} D_m \tag{C.77}$$

$$\geq n \min_{\{\phi(\cdot,\cdot)\}_{i=1}^n} \mathbb{E}_Q \left[d(S, \phi(W, V)) \right] - n(e^{-n\gamma_4} + e^{-n\gamma_5}) D_m. \tag{C.78}$$

$$= n \min_{\{\phi(\cdot,\cdot)\}_{i=1}^n} \mathbb{E}_Q \left[d(S, \phi(W, V)) \right] - o(1). \tag{C.79}$$

If $Q_U \neq P_U$, we have

$$\begin{aligned}
&\min_{g_r^{(n)}} \mathbb{E}\left[d\left(S^n, \hat{S}^n\right) | H = 1 \right] \\
&\geq \mathbb{P}\left(M = 0 | H = 1\right) \min_{\{\bar{\phi}_i(m, v^n)\}_{i=1}^n} \sum_{i=1}^n \mathbb{E}_{\tilde{P}^{(1)}} \left[d\left(S_i, \phi_i(0, V^n)\right) \right] \\
&\geq \mathbb{P}\left(M = 0 | H = 1\right) \left[\min_{\{\phi_i'(v)\}_{i=1}^n} \mathbb{E}_Q \left[\sum_{i=1}^n d(S_i, \phi_i'(V_i)) \right] - D_m o(1) \right] \qquad \text{(C.80)} \\
&= n \min_{\{\phi'(\cdot)\}} \mathbb{E}_Q \left[d(S, \phi'(V)) \right] - o(1). \qquad \text{(C.81)}
\end{aligned}$$

Here, (C.80) follows from (4.14) in Lemma 4.4 and (C.81) follows from (C.66). Thus, since $\delta > 0$ is arbitrary, we have shown that $(R, \kappa, \Delta_0, \Delta_1) \in \mathcal{R}_d(\epsilon)$, $\epsilon \in (0, 1)$, provided that (4.17), (4.18), (4.24) and (4.25) are satisfied. This completes the proof of the theorem.

C.5 Proof of Lemma ??

Consider the $|\mathcal{U}| + 2$ functions of $P_{U|W}$,

$$P_U(u_i) = \sum_{w \in \mathcal{W}} P_W(w) P_{U|W}(u_i|w), i = 1, 2, \ldots, |\mathcal{U}| - 1, \tag{C.82}$$

$$H_P(U|W, Z) = \sum_w P_W(w) g_1(P_{U|W}, w), \tag{C.83}$$

$$H_P(Y|W, Z) = \sum_w P_W(w) g_2(P_{U|W}, w), \tag{C.84}$$

$$H_P(S|W, Y, Z) = \sum_w P_W(w) g_3(P_{U|W}, w), \tag{C.85}$$

where,

$$g_1(P_{U|W}, w) = -\sum_{u,z} P_{U|W}(u|w)P_{Z|U}(z|u)\log\left(\frac{P_{U|W}(u|w)P_{Z|U}(z|u)}{\sum_u P_{U|W}(u|w)P_{Z|U}(z|u)}\right),$$

$$g_2(P_{U|W}, w) = -\sum_{y,z,u} P_{U|W}(u|w)P_{YZ|U}(y,z|u)\log\left(\frac{\sum_u P_{U|W}(u|w)P_{YZ|U}(y,z|u)}{\sum_u P_{U|W}(u|w)P_{Z|U}(z|u)}\right),$$

$$g_3(P_{U|W}, w) = -\sum_{s,y,z,u} P_{U|W}(u|w)P_{SYZ|U}(s,y,z|u) \log\left(\frac{\sum_u P_{U|W}(u|w)P_{SYZ|U}(s,y,z|u)}{\sum_u P_{U|W}(u|w)P_{YZ|U}(y,z|u)}\right).$$

Thus, by the Fenchel-Eggleston-Carathéodory's theorem [72], it is sufficient to have at most $|\mathcal{U}| - 1$ points in the support of W to preserve P_U and three more to preserve $H_P(U|W,Z)$, $H_P(Y|W,Z)$ and $H_P(S|W,Z,Y)$. Noting that $H_P(Y|Z)$ and $H_P(U|Z)$ are automatically preserved since P_U is preserved (and $(Y,Z,S) - U - W$ holds), $|\mathcal{W}| = |\mathcal{U}| + 2$ points are sufficient to preserve the R.H.S. of equations (4.28)-(4.30). This completes the proof for the case of $\mathcal{R}_e$. Similarly, considering the $|\mathcal{U}|+1$ functions of $P_{W|U}$ given in (C.82)-(C.84) and

$$\mathbb{E}_P\left[d\left(S, \phi(W,Y,Z)\right)\right] = \sum_w P_W(w) g_4(w, P_{W|U}), \tag{C.86}$$

$$\text{where, } g_4(w, P_{W|U}) = \sum_{s,u,y,z} P_{U|W}(u|w)P_{YZS|U}(y,z,s|u)\; d(s,\phi(w,y,z)), \tag{C.87}$$

similar result holds also for the case of $\mathcal{R}_d$.

Appendix D

Proofs for Chapter 5

D.1 Proof of Theorem 5.3

By the equivalence in (5.6) and the fact that the region $\bar{\mathcal{R}}_d^*$ (and also $\mathcal{R}_s^*$) is defined as a closed set, it is sufficient to show that $(R, R_s, \Delta) \in \mathcal{R}_s^*$ if,

$$R > I(W_2; U|Z), \tag{D.1}$$

$$R_s > H(V|W_2, Z), \tag{D.2}$$

$$\begin{aligned}\text{and } \Delta < \min\{\zeta_s, \zeta_p\} \min_{\phi''(\cdot)} \mathbb{E}\left(d_a\left(U, \phi(E)\right)\right) + (\zeta_s - \min\{\zeta_s, \zeta_p\}) \min_{\phi'(\cdot,\cdot)} \mathbb{E}\left(d_a\left(U, \phi(E, W_1)\right)\right) \\ + (1 - \zeta_s) \min_{\phi(\cdot,\cdot)} \mathbb{E}\left(d_a\left(U, \phi(E, W_2)\right)\right). \end{aligned} \tag{D.3}$$

To show this, we will in fact consider a more general rate distortion problem described in Section 5.4 and provide an inner bound for the region $\mathcal{R}_g^*$. More specifically, we will show that $(R, R_s, D, \Delta) \in \mathcal{R}_g^*$ if there exist auxiliary r.v.'s W, W_1 and W_2 with joint distribution $P_{UVEZB} P_{W_2|U} P_{W_1|W_2} P_{W|V}$ such that

$$R > I(W_2; U|Z), \tag{D.4}$$

$$R_s > I(V; W|W_2, Z), \tag{D.5}$$

$$\begin{aligned}\Delta < \min\{\zeta_s, \zeta_p\} \min_{\phi''(\cdot)} \mathbb{E}\left(d_a\left(U, \phi(E)\right)\right) + (\zeta_s - \min\{\zeta_s, \zeta_p\}) \min_{\phi'(\cdot,\cdot)} \mathbb{E}\left(d_a\left(U, \phi(E, W_1)\right)\right) \\ + (1 - \zeta_s) \min_{\phi(\cdot,\cdot)} \mathbb{E}\left(d_a\left(U, \phi(E, W_2)\right)\right), \end{aligned} \tag{D.6}$$

$$\text{and } D > \mathbb{E}(d_l(V, \hat{V})), \tag{D.7}$$

where $\hat{V} := \phi_l(W, Z, W_2)$ for some function $\phi_l : \mathcal{W} \times \mathcal{Z} \times \mathcal{W}_2 \mapsto \hat{\mathcal{V}}$. Our coding scheme and its analysis are inspired from [75] and [67] respectively. The proof technique uses the soft covering lemma and the standard random coding argument.

Codebook generation: Fix a joint distribution $P_{UVEZB}P_{W_2|U}P_{W_1|W_2}P_{W|V}$ and $\phi_l : \mathcal{W} \times \mathcal{Z} \times \mathcal{W}_2 \mapsto \hat{\mathcal{V}}$ satisfying (D.6) and (D.7). Assume that

$$I(E;W_1) < I(Z;W_1), \tag{D.8}$$

$$\text{and } I(E;W_2|W_1) < I(Z;W_2|W_1), \tag{D.9}$$

holds, so that ζ_p and ζ_s are strictly positive. The other cases can be handled similarly. Let $R'_s, R_1, R'_1, R_2, R'_2$ be non-negative numbers that will be specified later, and let

$$R := R_1 + R_2. \tag{D.10}$$

Codebook used by encoder of source V: Generate codewords $W^n(m,m'), (m,m') \in \mathcal{M} \times \mathcal{M}' := [2^{nR_s}] \times [2^{nR'_s}]$, each independently drawn according to the distribution $\prod_{i=1}^n P_W$. The index M is transmitted while M' is recovered by the legitimate receiver using the side information Z and the message from the helper. Denote this random codebook by $\mathcal{C}_v^n$.

Codebook used by encoder of source U: Generate codewords $W_1^n(m_1, m'_1),\ m_1 \in [2^{nR_1}],\ m'_1 \in [2^{nR'_1}]$ drawn independently according to the distribution $\prod_{i=1}^n P_{W_1}$. Denote this codebook by $\mathcal{C}_{w_1}^n$. For each (m_1, m'_1), generate codewords $W_2^n(m_1, m'_1, m_2, m'_2)$, $m_2 \in [2^{nR_2}]$, $m'_2 \in [2^{nR'_2}]$ independently drawn according to the distribution $\prod_{i=1}^n P_{W_2|W_1}\left(w_{2i}|W_{1i}(m_1,m'_1)\right)$. Denote this codebook by $\mathcal{C}_{w_2}^n$. The indices (M_1, M_2) are transmitted while (M'_1, M'_2) are not, but can be recovered by the legitimate receiver using its side information Z^n. Denote the two codebooks $\mathcal{C}_{w_1}^n$ and $\mathcal{C}_{w_2}^n$ together by $\mathcal{C}_u^n$. The codebooks $\mathcal{C}_u^n$ and $\mathcal{C}_v^n$ are known to all the parties including the eavesdropper.

Encoder: We use a stochastic encoder that chooses messages according to the following probability:

$$P_{E_u}(\check{m}|u^n) = \frac{\prod_{i=1}^n P_{U|W_2}(u_i|W_{2i}(\check{m}))}{\sum_{\tilde{m}\in\check{\mathcal{M}}} \prod_{i=1}^n P_{U|W_2(u_i|W_{2i}(\tilde{m}))}},$$
$$P_{E_v}(m,m'|v^n) = \frac{\prod_{i=1}^n P_{V|W}(v_i|W_i(m,m'))}{\sum_{\tilde{m}\in\mathcal{M}\times\mathcal{M}'} \prod_{i=1}^n P_{V|W}\left(u_i|W_i(\tilde{m})\right)},$$

where $\check{m} \in \check{\mathcal{M}} := [2^{nR_1}] \times [2^{nR'_1}] \times [2^{nR_2}] \times [2^{nR'_2}]$ and $(m,m') \in \mathcal{M} \times \mathcal{M}'$. Note that

this probability is a random variable that depends on the codebook realization. The messages (M_1, M_2) and M are transmitted over the respective noiseless channels.

Decoder: The decoder first uses a good channel decoding rule $P_{D_1}(\hat{m}_1', \hat{m}_2'|m_1, m_2, z^n)$ to estimate the messages (M_1', M_2') using the side information Z^n. Note that Z^n can be considered as the channel output of discrete memoryless channel $\prod_{i=1}^n P_{Z|W_2}$ using the superposition channel sub-codebook $\{W_2^n(m_1, m_1', m_2, m_2')\}$ for transmission of messages (M_1', M_2'). Next, the message index M' is decoded similarly by using a good channel decoding rule for the transmission of messages M' over a memoryless channel $\prod_{i=1}^n P_{ZW_2|W}$ using the sub-codebook $\{W^n(m, m')\}$. The decoder for the source U can be considered to be composed of two parts, $P_{D_u^M}(\hat{m}_1', \hat{m}_2'|m_1, m_2, z^n)$ which is a good channel decoder and P_{D_u} given by

$$P_{D_u}(\hat{w}_2^n|m_1, \hat{m}_1', m_2, \hat{m}_2') = \mathbb{1}(W_2^n(m_1, \hat{m}_1', m_2, \hat{m}_2') = \hat{w}_2^n).$$

Similarly the decoder for source V is composed of two parts, a good channel decoder $P_{D_v^M}(\hat{m}'|m, \hat{w}_2^n, z^n)$ and a symbol by symbol reconstruction given as

$$P_{D_v}(\hat{v}^n|m, \hat{m}', \hat{w}_2^n, z^n) = \prod_{i=1}^n \mathbb{1}(\phi_l(\hat{w}_i(m, \hat{m}'), \hat{w}_{2i}, z_i) = \hat{v}_i).$$

Analysis: Note that (5.2) is satisfied by definition in the above scheme. The joint distribution induced by the encoding and decoding operations is given as follows:

$$\begin{aligned}
&\tilde{P}(e^n, b^n, z^n, u^n, m_1, m_1', m_2, m_2', w_1^n, w_2^n, v^n, m, m', w^n, \hat{m}_1', \hat{m}_2', \hat{w}_2^n, \hat{m}', \hat{v}^n) \\
&= P_{E^nB^nZ^nU^nV^n}(e^n, b^n, z^n, u^n, v^n) P_{E_u}(m_1, m_1', m_2, m_2'|u^n) P_{W_1^n|M_1M_1'}(w_1^n|m_1, m_1') \\
&\quad P_{W_2^n|W_1^n(M_1,M_1')M_2M_2'}(w_2^n|w_1^n, m_2, m_2')\ P_{E_v}(m, m'|v^n)\ P_{W^n|MM'}(w^n|m, m') \\
&\quad P_{D_u^M}(\hat{m}_1', \hat{m}_2'|m_1, m_2, z^n)\ P_{D_u}(\hat{w}_2^n|m_1, \hat{m}_1', m_2, \hat{m}_2')\ P_{D_v^M}(\hat{m}'|m, \hat{w}_2^n, z^n) \\
&\quad P_{D_v}(\hat{v}^n|m, \hat{m}', \hat{w}_2^n, z^n) \\
&= \left[\prod_{i=1}^n P_{EBZUV}(e_i, b_i, z_i, u_i, v_i)\right] P_{E_u}(m_1, m_1', m_2, m_2'|u^n) \mathbb{1}(W_1^n(m_1, m_1') = w_1^n) \\
&\quad \mathbb{1}(W_2^n(m_1, m_1', m_2, m_2') = w_2^n)\ P_{E_v}(m, m'|v^n) \mathbb{1}(W^n(m, m') = w^n) \\
&\quad P_{D_u^M}(\hat{m}_1', \hat{m}_2'|m_1, m_2, z^n) \mathbb{1}(W_2^n(m_1, \hat{m}_1', m_2, \hat{m}_2 = \hat{w}_2^n) P_{D_v^M}(\hat{m}'|m, \hat{w}_2^n, z^n)
\end{aligned}$$

$$P_{D_v}(\hat{v}^n|m,\hat{m}',\hat{w}_2^n,z^n).$$

Define an auxiliary distribution

$$\begin{aligned}
&Q(e^n,b^n,z^n,u^n,m_1,m_1',m_2,m_2',w_1^n,w_2^n,v^n,m,m',w^n,\hat{m}_1',\hat{m}_2',\hat{w}_2^n,\hat{m}',\hat{v}^n)\\
&:= Q_{M_1M_1'M_2M_2'}(m_1,m_1',m_2,m_2')\ Q_{W_1^n|M_1,M_1'}(w_1^n|m_1,m_1')\\
&\quad Q_{W_2^n|W_1^n(M_1,M_1')M_2M_2'}(w_2^n|w_1^n,m_2,m_2')\ Q_{U^n|W_2^n(M_1,M_1',M_2,M_2')}(u^n|w_2^n(m_1,m_1',m_2,m_2'))\\
&\quad Q_{Z^nV^nE^nB^n|U^n}(z^n,v^n,e^n,b^n|u^n)\ Q_{\hat{M}_1'\hat{M}_2'|Z^nM_1M_2}(\hat{m}_1',\hat{m}_2'|z^n,m_1,m_2)\\
&\quad Q_{\hat{W}_2^n|M_1\hat{M}_1'M_2\hat{M}_2'}(\hat{w}_2^n|m_1,\hat{m}_1',m_2,\hat{m}_2')\ Q_{MM'|V^n}(m,m'|v^n)\ Q_{W^n|MM'}(w^n|m,m')\\
&\quad Q_{\hat{M}'|Z^n\hat{W}_2^nM}(\hat{m}'|z^n,\hat{w}_2^n,m)\ Q_{\hat{V}^n|M\hat{M}'\hat{W}_2^nZ^n}(\hat{v}^n|m,\hat{m}',\hat{w}_2^n,z^n)\\
&:= \frac{1}{2^{n(R_1+R_1'+R_2+R_2')}}\mathbb{1}\left(W_1^n(m_1,m_1')=w_1^n\right)\mathbb{1}\left(W_2^n(m_1,m_1',m_2,m_2')=w_2^n\right)\\
&\quad \left[\prod_{i=1}^n P_{U|W_2}(u_i|w_{2i}(m_1,m_1',m_2,m_2'))\right]\ \left[\prod_{i=1}^n P_{ZVEB|U}(z_i,v_i,e_i,b_i|u_i)\right]\\
&\quad P_{D_u^M}(\hat{m}_1',\hat{m}_2'|z^n,m_1,m_2)\ \mathbb{1}(W_2^n(m_1,\hat{m}_1',m_2,\hat{m}_2')=\hat{w}_2^n)\ \times\\
&\left[P_{E_v}(m,m'|v^n)\ \mathbb{1}(W^n(m,m')=w^n)\ P_{D_v^M}(\hat{m}'|m,\hat{w}_2^n,z^n)\ P_{D_v}(\hat{v}^n|m,\hat{m}',\hat{w}_2^n,z^n)\right]\\
&:= Q^H_{M_1M_1'M_2M_2'W_1^nW_2^nU^nZ^nV^nE^nB^n\hat{M}_1'\hat{M}_2'\hat{W}_2^n}(m_1,m_2,m_1',m_2',w_1^n,w_2^n,u^n,z^n,\\
&\qquad\qquad v^n,e^n,b^n,\hat{m}_1',\hat{m}_2',\hat{w}_2^n)\cdot\\
&\quad Q^S_{MM'W^n\hat{M}'\hat{V}^n|V^nZ^n\hat{W}_2^n}(m,m',w^n,\hat{m}',\hat{v}^n|v^n,z^n,\hat{w}_2^n).
\end{aligned}$$

Define

$$\begin{aligned}
&Q^{(1)}(m_1,m_2,m_1',m_2',w_1^n,w_2^n,u^n,z^n,\hat{m}_1',\hat{m}_2',\hat{w}_2^n,e^n,v^n,b^n,m,m',w^n,\hat{m}',\hat{v}^n)\\
&:= Q_{M_1M_1'M_2M_2'W_1^nW_2^nU^nZ^nV^nE^nB^n\hat{M}_1'\hat{M}_2'}(m_1,m_2,m_1',m_2',w_1^n,w_2^n,u^n,z^n,\\
&\qquad\qquad v^n,e^n,b^n,\hat{m}_1',\hat{m}_2')\\
&\qquad P_{D_u}(\hat{w}_2^n|m_1,m_1',m_2,m_2')\ Q^S(m,m',w^n,\hat{m}',\hat{v}^n|v^n,z^n,\hat{w}_2^n)\\
&:= Q_{M_1M_1'M_2M_2'W_1^nW_2^nU^nZ^nV^nE^nB^n\hat{M}_1'\hat{M}_2'}(m_1,m_2,m_1',m_2',w_1^n,w_2^n,u^n,z^n,\\
&\qquad\qquad v^n,e^n,b^n,\hat{m}_1',\hat{m}_2')\\
&\qquad P_{D_u}(\hat{w}_2^n|m_1,m_1',m_2,m_2')\ P_{E_v}(m,m'|v^n)\ \mathbb{1}(W^n(m,m')=w^n)
\end{aligned}$$

$$P_{D_v^M}(\hat{m}'|m, \hat{w}_2^n, z^n)\ P_{D_v}(\hat{v}^n|m, \hat{m}', \hat{w}_2^n, z^n).$$

Note that the only difference between Q and $Q^{(1)}$ is the term P_{D_u}, the part of the decoder that obtains the reconstruction sequence $\hat{W}_2^n$. In $Q^{(1)}$, the actual messages (m_1', m_2') intended by the encoder is used by P_{D_u} instead of the estimates $\hat{m}_1', \hat{m}_2'$. Taking expectation with respect to the codebook $\mathcal{C}_u$, we obtain

$$\mathbb{E}_{\mathcal{C}_u^n}\left[Q^{(1)}_{U^n Z^n \hat{W}_2^n}\right] = P_{U^n Z^n W_2^n} := \prod_{i=1}^n P_{UZW_2}.$$

Let

$$\begin{aligned}
&\bar{Q}^{(1)}_{U^n Z^n V^n E^n B^n \hat{W}_2^n M M' W^n \hat{V}^n}(u^n, z^n, v^n, e^n, b^n, \hat{w}_2^n, m, m', w^n, \hat{m}', \hat{v}^n) \\
&:= \mathbb{E}_{\mathcal{C}_u^n}\left[Q^{(1)}_{U^n Z^n V^n E^n B^n \hat{W}_2^n M M' W^n \hat{V}^n}(u^n, z^n, v^n, e^n, b^n, \hat{w}_2^n, m, m', w^n, \hat{m}', \hat{v}^n)\right] \\
&= \left[\prod_{i=1}^n P_{UZVEBW_2}(u_i, z_i, v_i, e_i, b_i, \hat{w}_{2i})\right] \\
&\quad Q^S_{M M' W^n \hat{M}' \hat{V}^n | V^n Z^n \hat{W}_2^n}(m, m', w^n, \hat{m}', \hat{v}^n | v^n, z^n, \hat{w}_2^n).
\end{aligned}$$

It is easy to see that the likelihood encoder P_{E_u} is chosen such that

$$P_{E_u} = Q_{M_1 M_1' M_2 M_2' | U^n}. \tag{D.11}$$

Let

$$\begin{aligned}
&Q^H_{M_1 M_1' M_2 M_2' W_1^n W_2^n U^n Z^n V^n E^n B^n \hat{M}_1' \hat{M}_2' \hat{W}_2^n}(m_1, m_1', m_2, m_2', w_1^n, w_2^n, u^n, z^n, v^n, e^n, b^n, \\
&\qquad\qquad \hat{m}_1', \hat{m}_2', \hat{w}_2^n) \\
&= Q_{U^n}(u^n) P_{E_u}(m_1, m_1', m_2, m_2' | u^n)\ P_{W_1^n | M_1 M_1'}(w_1^n | m_1, m_1') \\
&\quad P_{W_2^n | W_1^n (M_1, M_1') M_2 M_2'}(w_2^n | w_1^n, m_2, m_2')\ \left[\prod_{i=1}^n P_{ZVEB|U}(z_i, v_i, e_i, b_i | u_i)\right] \\
&\quad P_{D_u^M}(\hat{m}_1', \hat{m}_2' | m_1, m_2, z^n)\ P_{D_u}(\hat{w}_2^n | m_1, \hat{m}_1', m_2, \hat{m}_2').
\end{aligned}$$

Observe that the only difference between Q^H and $\tilde{P}$ is the marginal distribution of U. By Lemma 5.2, it follows that

$$\mathbb{E}_{\mathcal{C}_u^n}\left[\|Q_U^H - \tilde{P}_U\|\right] \leq e^{-\delta_1 n},$$

provided

$$R_1 + R_1' > I(U;W_1), \tag{D.12}$$

$$\text{and } R_2 + R_2' > I(U;W_2|W_1). \tag{D.13}$$

Applying Property 4.4.1 (c), we get

$$\mathbb{E}_{\mathcal{C}_u^n}\left[\|Q^H_{M_1M_1'M_2M_2'W_1^nW_2^nU^nZ^nV^nE^nB^n\hat{M}_1'\hat{M}_2'\hat{W}_2^n} - \tilde{P}_{M_1M_1'M_2M_2'W_1^nW_2^nU^nZ^nV^nE^nB^n\hat{M}_1'\hat{M}_2'\hat{W}_2^n}\|\right] \leq e^{-\delta_1 n}.$$

Since the messages M_1, M_1', M_2, M_2' are uniformly distributed under the joint distribution Q^H, it is well known that if

$$R_1' < I(W_1;Z), \tag{D.14}$$

$$\text{and } R_2' < I(W_2;Z|W_1), \tag{D.15}$$

then, a maximum likelihood decoder achieves a asymptotically vanishing decoding error probability i.e.,

$$\mathbb{E}_{\mathcal{C}_u^n}\left[\Pr(\hat{M}_1', \hat{M}_2') \neq (M_1', M_2')\right] \leq \epsilon_n' \xrightarrow{(n)} 0,$$

where the probability is evaluated based on the joint distribution Q^H. Hence by Lemma 5.1, we have

$$\mathbb{E}_{\mathcal{C}_u^n}\left[\|Q^{(1)}_{M_1M_1'M_2M_2'W_1^nW_2^nU^nZ^nV^nE^nB^n\hat{M}_1'\hat{M}_2'\hat{W}_2^n} - Q^H_{M_1M_1'M_2M_2'W_1^nW_2^nU^nZ^nV^nE^nB^n\hat{M}_1'\hat{M}_2'\hat{W}_2^n}\|\right] \leq \epsilon_n' \xrightarrow{(n)} 0.$$

Applying Property 4.4.1(c) yields

$$\begin{aligned}
&\mathbb{E}_{\mathcal{C}_u^n}\Big[\|Q^{(1)}_{M_1M_1'M_2M_2'W_1^nW_2^nU^nZ^nV^nE^nB^n\hat{M}_1'\hat{M}_2'\hat{W}_2^nMM'W^n\hat{M}'\hat{V}^n} - \\
&\qquad \tilde{P}_{M_1M_1'M_2M_2'W_1^nW_2^nU^nZ^nV^nE^nB^n\hat{M}_1'\hat{M}_2'\hat{W}_2^nMM'W^n\hat{M}'\hat{V}^n}\|\Big] \\
&= \mathbb{E}_{\mathcal{C}_u^n}\Big[\|Q^{(1)}_{M_1M_1'M_2M_2'W_1^nW_2^nU^nZ^nV^nE^nB^n\hat{M}_1'\hat{M}_2'\hat{W}_2^n} - \\
&\qquad \tilde{P}_{M_1M_1'M_2M_2'W_1^nW_2^nU^nZ^nV^nE^nB^n\hat{M}_1'\hat{M}_2'\hat{W}_2^n}\|\Big] \\
&\le \mathbb{E}_{\mathcal{C}_u^n}\Big[\|Q^{(1)} - Q^H\|\Big] + \mathbb{E}_{\mathcal{C}_u^n}\Big[\|Q^H - \tilde{P}\|\Big] \\
&\le e^{-\delta_1 n} + \epsilon_n' := \delta_{2n} \xrightarrow{(n)} 0,
\end{aligned}$$

and

$$\mathbb{E}_{\mathcal{C}_u^n}\Big[\mathbb{E}_{\tilde{P}}\Big[d_l(V^n,\hat{V}^n)\Big]\Big] \le \Big[\mathbb{E}_{\bar{Q}^{(1)}}\Big[d_l(V^n,\hat{V}^n)\Big]\Big] + D_l\ \delta_{2n}, \tag{D.16}$$

In order to bound the first term in (D.16), we will use the results in Section IV in [67]. As shown there, $\bar{Q}^{(1)}$ is equal to the distribution induced by the maximum likelihood encoder P_{E_v} for the rate distortion problem with the source V and side information (Z, W_2) at the decoder (E and B are irrelevant for distortion analysis at the legitimate receiver).

Let

$$\begin{aligned}
&Q^{(3)}_{U^nZ^nV^nE^nB^n\hat{W}_2^nMM'W^n\hat{M}'\hat{V}^n}(u^n,z^n,v^n,e^n,b^n,\hat{w}_2^n,m,m',w^n,\hat{m}',\hat{v}^n) \\
&:= \frac{1}{2^{n(R_s+R_s')}}\mathbb{1}(W^n(m,m')=w^n)\prod_{i=1}^n P_{V|W}(v_i|w_i)P_{UZEBW_2|V}(u_i,z_i,e_i,b_i,\hat{w}_{2i}|v_i) \\
&\quad P_{D_v^M}(\hat{m}'|m,\hat{w}_2^n,z^n)\ P_{D_v}(\hat{v}^n|m,m',\hat{w}_2^n,z^n).
\end{aligned}$$

Note that P_{D_v} uses the actual messages (m,m') rather than estimates $(m,\hat{m}')$ for forming the reconstruction $\hat{V}$. It follows similar to [67, Equation (79)] that for some $\delta_{3n} \xrightarrow{(n)} 0$, we have

$$\mathbb{E}_{\mathcal{C}_v^n}\Big[\|\bar{Q}^{(1)}_{U^nZ^nV^nE^nB^n\hat{W}_2^nMM'W^n\hat{M}'\hat{V}^n} - Q^{(3)}_{U^nZ^nV^nE^nB^n\hat{W}_2^nMM'W^n\hat{M}'\hat{V}^n}\|\Big] \le \delta_{3n},$$

provided

$$R_s + R'_s > I(V;W), \tag{D.17}$$

$$\text{and } R'_s < I(W;W_2,Z). \tag{D.18}$$

Next, note that

$$\mathbb{E}_{\mathcal{C}_v^n}\left[Q^{(3)}_{U^nZ^nV^nE^nB^n\hat{W}_2^n\hat{V}^n}\right] = \prod_{i=1}^{n} P_{UZVEBW_2\phi_l(W)}.$$

Thus, the average distortion at the legitimate receiver averaged over the random codebook $\mathcal{C}_v^n$ and $\mathcal{C}_u^n$ can be bounded as

$$\begin{aligned}
\mathbb{E}_{\mathcal{C}_v^n}\left[\mathbb{E}_{\mathcal{C}_u^n}\left[\mathbb{E}_{\tilde{P}}\left[d_l(V^n,\hat{V}^n)\right]\right]\right] &= \mathbb{E}_{\mathcal{C}_v^n}\left[\mathbb{E}_{\bar{Q}^{(1)}}\left[d_l(V^n,\hat{V}^n)\right] + D_a\delta_{2n}\right] \\
&\leq \mathbb{E}_{\mathcal{C}_v^n}\left[\mathbb{E}_{Q^{(3)}}\left[d_l(V^n,\hat{V}^n)\right]\right] + D_a\left(\delta_{2n}+\delta_{3n}\right) \\
&= \mathbb{E}_{P}\left[d_l(V^n,\hat{V}^n)\right] + D_a\left(\delta_{2n}+\delta_{3n}\right) \\
&= \mathbb{E}_{P}\left[d_l(V,\hat{V})\right] + D_a\left(\delta_{2n}+\delta_{3n}\right) \\
&< D + D_l\left(\delta_{2n}+\delta_{3n}\right). \qquad \text{(D.19)}
\end{aligned}$$

Analysis of distortion at eavesdropper:

For $k \geq 1$, consider an auxiliary distribution

$$\begin{aligned}
&\hat{Q}^{(k)}_{M_1M_1'M_2M_2'W_1^nE^nB^{k-1}U_k}(m_1,m_1',m_2,m_2',w_1^n,e^n,b^{k-1},u_k) \qquad \text{(D.20)} \\
&:= \frac{1}{2^{n(R_1+R_1'+R_2+R_2')}}\mathbb{1}(W_1^n(m_1,m_1')=w_1^n)\prod_{i=1}^{n}P_{E|W_1}(e_i|w_{1i})\prod_{i=1}^{k-1}P_{B|EW_1}(b_i|e_i,w_{1i}) \\
&\quad P_{U|EW_1}(u_k|e_k,w_{1k}).
\end{aligned}$$

Note that the following Markov relation holds under $\hat{Q}^{(k)}$,

$$U_k - \left(E_k, W_{1k}(M_1,M_1')\right) - \left(M_1,M_1',M_2,M_2',B^{k-1},E^n\right).$$

Also, observe by definition that for fixed $M_2 = m_2$,

$$
\begin{aligned}
&Q_{M_1M_1'M_2'E^nB^{k-1}U_k|M_2=m_2}(m_1, m_1', m_2', e^n, b^{k-1}, u_k) \\
&= \frac{1}{2^{n(R_1+R_1'+R_2')}} \prod_{i=1}^{n} P_{E|W_2}(e_i|W_{2i}(m_1, m_1', m_2, m_2')) \\
&\quad \left[\prod_{i=1}^{k-1} P_{B|EW_2}(b_i|e_i, W_{2i}(m_1, m_1', m_2, m_2'))\right] \; P_{U|EW_2}(u_k|e_k, w_{2k}) \\
&= \frac{1}{2^{n(R_1+R_1'+R_2')}} \prod_{i=1}^{n} P_{E|W_2W_1}(e_i|W_{2i}(m_1, m_1', m_2, m_2'), W_{1i}(m_1, m_1')) \\
&\quad \left[\prod_{i=1}^{k-1} P_{B|EW_2W_1}(b_i|e_i, W_{2i}(m_1, m_1', m_2, m_2'), W_{1i}(m_1, m_1'))\right] \\
&\quad P_{U|EW_2W_1}(u_k|e_k, W_{2k}(m_1, m_1', m_2, m_2'), W_{1k}(m_1, m_1')).
\end{aligned}
$$

Let

$$
k_2 := \frac{(R_2' - I(W_2; E|W_1))n}{I(B; W_2|W_1, E)} + \frac{I(B; W_2|W_1, E) - I(U; W_2|B, W_1, E)}{I(B; W_2|W_1, E)}.
$$

By the application of Lemma 5.2, it follows that for arbitrary fixed $M_2 = m_2$,

$$
\mathbb{E}_{\mathcal{C}_v^n}\left[\|\hat{Q}^{(k)}_{M_1M_1'E^nB^{k-1}U_k} - Q_{M_1M_1'E^nB^{k-1}U_k}\|\right] \leq e^{-\delta_4 n} \xrightarrow{(n)} 0, \; 1 \leq k < k_2,
$$

for some $\delta_4 > 0$, if[1]

$$
R_2' > I(E; W_2|W_1). \tag{D.21}
$$

Averaging over M_2, we obtain

$$
\mathbb{E}_{\mathcal{C}_v^n}\left[\|\hat{Q}^{(k)}_{M_1M_1'M_2E^nB^{k-1}U_k} - Q_{M_1M_1'M_2E^nB^{k-1}U_k}\|\right] \leq e^{-\delta_4 n} \xrightarrow{(n)} 0, \; \forall \; k < k_2.
$$

Hence, for some $\delta_5 > 0$,

$$
\begin{aligned}
\mathbb{E}_{\mathcal{C}_v^n}\left[\sum_{k=1}^{k_2-1} \|\hat{Q}^{(k)}_{M_1M_1'M_2E^nB^{k-1}U_k} - Q_{M_1M_1'M_2E^nB^{k-1}U_k}\|\right] &\leq (k_2 - 1)e^{-\delta_4 n} \\
&\leq e^{-\delta_5 n} \xrightarrow{(n)} 0.
\end{aligned} \tag{D.22}
$$

[1]Note that a choice of R_2' simultaneously satisfying (D.15) and (D.21) is possible due to (D.9).

Also, identifying (W, Z, X, Y, R_1, R_2) with $(\emptyset, W_1, \emptyset, W_1, 0, R_1 + R_1')$ in Lemma 5.2, we obtain that for all $1 \leq i \leq n$ and some $\delta_6 > 0$,

$$\mathbb{E}_{\mathcal{C}_v^n}\left[\|P_{W_1} - \hat{Q}^{(i)}_{W_{1i}}(M_1, M_1')\|\right] \leq e^{-\delta_6 n},$$

if $R_1 + R_1' > I(\emptyset; W_1) = 0$. Hence,

$$\mathbb{E}_{\mathcal{C}_v^n}\left[\sum_{i=1}^{n} \|P_{W_1} - \hat{Q}^{(i)}_{W_{1i}}(M_1, M_1')\|\right] \leq n e^{-\delta_6 n} \xrightarrow{(n)} 0. \tag{D.23}$$

For $k \geq 1$, consider another auxiliary distribution,

$$\begin{aligned}
&\check{Q}^{(k)}_{M_1 M_1' M_2 M_2' E^n B^{k-1} U_k}(m_1, m_1', m_2, m_2', e^n, b^{k-1}, u_k) \\
&:= \frac{1}{2^{n(R_1+R_1'+R_2+R_2')}} \left[\prod_{i=1}^{n} P_E(e_i)\right] \left[\prod_{i=1}^{k-1} P_{B|E}(b_i|e_i)\right] P_{U|E}(u_k|e_k).
\end{aligned}$$

Note that under $\check{Q}^{(k)}$,

$$U_k - E_k - \left(M_1, M_1', M_2, M_2', B^{k-1}, E^n\right).$$

By definition, for fixed $M_2 = m_2$ and $M_1 = m_1$,

$$\begin{aligned}
&Q_{M_1' M_2' E^n B^{k-1} U_k | M_1 = m_1, M_2 = m_2} \\
&= \frac{1}{2^{n(R_1'+R_2')}} \left[\prod_{i=1}^{n} P_{E|W_2}(e_i|W_{2i}(m_1, m_1', m_2, m_2'))\right] \\
&\left[\prod_{i=1}^{k-1} P_{B|EW_2}(b_i|e_i, W_{2i}(m_1, m_1', m_2, m_2'))\right] \; P_{U|EW_2}(u_k|e_k, W_{2k}(m_1, m_1', m_2, m_2')) \\
&= \frac{1}{2^{n(R_1'+R_2')}} \left[\prod_{i=1}^{n} P_{E|W_2W_1}(e_i|W_{2i}(m_1, m_1', m_2, m_2'), W_{1i}(m_1, m_1'))\right] \\
&\quad \left[\prod_{i=1}^{k-1} P_{B|EW_2W_1}(b_i|e_i, W_{2i}(m_1, m_1', m_2, m_2'), W_{1i}(m_1, m_1'))\right] \\
&\quad P_{U|EW_2W_1}(u_k|e_k, W_{2k}(m_1, m_1', m_2, m_2'), W_{1k}(m_1, m_1')).
\end{aligned}$$

Let

$$k_1 := \min\left(k_2,\ \frac{(R_1' - I(W_1;E))n}{I(B;W_1|E)} + \frac{I(B;W_1|E) - I(U;W_1|B,E)}{I(B;W_1|E)}\right).$$

By an application of Lemma 5.2, it follows that for fixed $M_2 = m_2, M_1 = m_1$,

$$\mathbb{E}_{\mathcal{C}_v^n}\left[\|\check{Q}^{(k)}_{E^nB^{k-1}U_k} - Q_{E^nB^{k-1}U_k}\|\right] \leq e^{-\delta_7 n} \xrightarrow{(n)} 0,\ 1 \leq k < k_1,$$

for some $\delta_7 > 0$, if (D.21) holds and

$$R_1' > I(E;W_1). \tag{D.24}$$

Averaging over M_2 and M_1, we obtain

$$\mathbb{E}_{\mathcal{C}_v^n}\left[\|\check{Q}^{(k)}_{M_1M_2E^nB^{k-1}U_k} - Q_{M_1M_2E^nB^{k-1}U_k}\|\right] \leq e^{-\delta_7 n} \xrightarrow{(n)} 0.$$

Hence, for some $\delta_8 > 0$, we have

$$\begin{aligned}\mathbb{E}_{\mathcal{C}_v^n}\left[\sum_{k=1}^{k_1-1}\|\check{Q}^{(k)}_{M_1M_2E^nB^{k-1}U_k} - Q_{M_1M_2E^nB^{k-1}U_k}\|\right] &\leq (k_1-1)e^{-\delta_7 n}\\ &\leq e^{-\delta_8 n} \xrightarrow{(n)} 0.\end{aligned} \tag{D.25}$$

By identifying (W, Z, X, Y, R_1, R_2) in Lemma 5.2 with $(W_1, W_2, \emptyset, W_2, R_1+R_1', R_2+R_2')$, it follows again from an application of Lemma 5.2 that for some $\delta_9 > 0$, we have

$$\mathbb{E}_{\mathcal{C}_v^n}\left[\sum_{i=1}^{n}\|P_{W_2} - Q_{W_{2i}}(M_1, M_1', M_2, M_2')\|\right] \leq e^{-\delta_9 n} \xrightarrow{(n)} 0, \tag{D.26}$$

provided $R_1 + R_1' > 0$ and $R_2 + R_2' > 0$. By the random coding argument combined with the standard expurgation technique, there exists a deterministic codebook $\mathcal{C}_u^n$ and $\mathcal{C}_v^n$ such that (D.19), (D.22), (D.23), (D.25) and (D.26) are satisfied.

Now, the distortion at the eavesdropper can be lower bounded as follows:

$$\min_{\{\phi_i(\cdot,\cdot,\cdot,\cdot)\}}\left[\sum_{i=1}^{n}\mathbb{E}_{\tilde{P}}(d_a(U_i, \phi_i(M_1, M_2, E^n, B^{(i-1)}))\right]$$

$$\geq \min_{\{\phi_i(\cdot,\cdot,\cdot,\cdot)\}} \left[\sum_{i=1}^{n} \mathbb{E}_Q(d_a(U_i, \phi_i(M_1, M_2, E^n, B^{(i-1)}))\right] - nD_a\left(e^{-\delta_1 n}\right) \tag{D.27}$$

Now consider the first term in (D.27). For $1 \leq i < k_1$, we can write

$$\min_{\{\phi_i(m_1,m_2,e^n,b^{i-1})\}} \left[\sum_{i=1}^{k_1-1} \mathbb{E}_Q(d_a(U_i, \phi_i(M_1, M_2, E^n, B^{(i-1)}))\right]$$

$$\geq \min_{\{\phi_i(m_1,m_1',m_2,m_2',e^n,b^{i-1})\}} \left[\sum_{i=1}^{k_1-1} \mathbb{E}_Q(d_a(U_i, \phi_i(M_1, M_1', M_2, M_2', E^n, B^{(i-1)}))\right]$$

$$\geq \min_{\{\phi_i(e_i)\}} \left[\sum_{i=1}^{k_1-1} \mathbb{E}_{\check{Q}^{(k_1)}}(d_a(U_i, \phi_i(E_i))\right] - k_1 e^{-\delta_8 n} D_a \tag{D.28}$$

$$\geq (k_1 - 1) \min_{\{\phi''(e)\}} \mathbb{E}_P(d_a(U, \phi''(E)) - ne^{-\delta_8 n} D_a, \tag{D.29}$$

where (D.28) follows from (D.25).

Similarly,

$$\min_{\{\phi_i(m_1,m_2,e^n,b^{i-1})\}} \left[\sum_{i=k_1}^{k_2-1} \mathbb{E}_Q(d_a(U_i, \phi_i(M_1, M_2, E^n, B^{(i-1)}))\right]$$

$$\geq \min_{\{\phi_i(m_1,m_1',m_2,m_2',e^n,b^{i-1})\}} \left[\sum_{i=k_1}^{k_2-1} \mathbb{E}_Q(d_a(U_i, \phi_i(M_1, M_1', M_2, M_2', E^n, B^{(i-1)}))\right]$$

$$\geq \min_{\{\phi_i(e_i,w_{1i})\}} \left[\sum_{i=k_1}^{k_2-1} \mathbb{E}_{\hat{Q}^{(k_2)}}(d_a(U_i, \phi_i(E_i, W_{1i}(M_1, M_1')))\right] - ne^{-\delta_5 n} D_a \tag{D.30}$$

$$\geq (k_2 - k_1 - 1) \min_{\{\phi'(e,w_1)\}} \mathbb{E}_P(d_a(U, \phi'(E, W_1)) - n\left(e^{-\delta_5 n} + ne^{-\delta_6 n}\right) D_a, \tag{D.31}$$

where (D.30) follows from (D.22); and (D.31) follows from (D.23) by noting that $\hat{Q}^{k_2}_{E_i U_i|W_{1i}} = P_{EU|W_1}$.

The remaining terms inside the summation in (D.27) can be bounded as follows

$$\min_{\{\phi_i(m_1,m_2,e^n,b^{i-1})\}} \left[\sum_{i=k_2}^{n} \mathbb{E}_Q(d_a(U_i, \phi_i(M_1, M_2, E^n, B^{(i-1)}))\right]$$

$$\geq \min_{\{\phi_i(m_1,m_2,e^n,b^{i-1})\}} \left[\sum_{i=k_2}^{n} \mathbb{E}_Q(d_a(U_i, \phi_i(M_1, M_1', M_2, M_2', E^n, B^{(i-1)}))\right]$$

$$\geq \min_{\{\phi_i(e_i,w_{2i})\}} \left[\sum_{i=k_2}^{n} \mathbb{E}_Q \left(d_a(U_i, \phi_i(E_i, W_{2i}(M_1, M_1', M_2, M_2'))) \right) \right]$$

$$= (n - k_2 + 1) \min_{\{\phi(e,w_2)\}} \mathbb{E}_P \left(d_a(U, \phi(E, W_2)) \right), \tag{D.32}$$

where (D.32) follows from (D.26). Next, note that

$$\lim_{n\to\infty} \frac{k_2 - 1}{n} = \frac{R_2' - I(W_2; E|W_1)}{I(B; W_2|W_1, E)},$$
$$\text{and } \lim_{n\to\infty} \frac{k_1 - 1}{n} = \min \left(\frac{R_2' - I(W_2; E|W_1)}{I(B; W_2|W_1, E)}, \frac{R_1' - I(W_1; E)}{I(B; W_1|E)} \right).$$

By maximizing the distortion incurred at the eavesdropper with respect to R_1' and R_2', it follows from (D.6), (D.27), (D.29), (D.31) and (D.32) that for any $\gamma > 0$ and sufficiently large n,

$$\min_{\{\phi_i(m_1,m_2,e^n,b^{i-1})\}} \left[\sum_{i=1}^{n} \mathbb{E}_{\tilde{P}}(d_a(U_i, \phi_i(M_1, M_2, E^n, B^{(i-1)})) \right] > n\Delta - \gamma.$$

This is due to the fact that the supremum with respect to R_1' and R_2' occurs at $R_1' = I(W_1; Z)$ and $R_2' = I(W_2; Z|W_1)$. Setting R_1' and R_2' to these values, it follows from (D.10), (D.12)-(D.15), (D.17), (D.18), (D.21) and (D.24) via the Markov conditions $(Z, W_2) - V - W$ and $Z - U - W_2 - W_1$ that $(R, R_s, D, \Delta) \in \mathcal{R}_g^*$ if (D.4)-(D.7) holds. The cases where the joint distribution $P_{UVEZB} P_{W_2|U} P_{W_1|W_2} P_{W|V}$ is such that (D.8) and/or (D.9) does not hold can be handled similarly. Specializing to the lossless case ($D = 0$) with hamming distortion measure, we obtain the condition $R_s > H(V|W_2, Z)$ given in (D.2) by setting $W = V$. This completes the proof of the theorem.

D.2 Proof of Theorem 5.7

The proof of this theorem is similar to that of Theorem 5.3. In lieu of the equivalence (5.7), we will consider the more general rate distortion problem described in Section 5.4 and provide an inner bound on $\mathcal{R}_g$. We will show that $(R_s, D, \Delta) \in \mathcal{R}_g$ if there exist auxiliary r.v.'s W, W_1 and W_2 with joint distribution

$P_{UVEZB}P_{W_2|U}P_{W_1|W_2}P_{W|V}P_{X|UW_1W_2}P_{YJ|X}$ such that

$$I(W_2;U) < I(W_1,W_2;Y,Z), \tag{D.33}$$

$$R_s > I(V;W|W_2,Z,Y), \tag{D.34}$$

$$\begin{aligned}\Delta < \min\{\zeta_s',\zeta_p'\} \min_{\phi''(\cdot)} \mathbb{E}\left(d_a\left(U,\phi(E)\right)\right) + \left(\zeta_s' - \min\{\zeta_s',\zeta_p'\}\right) \min_{\phi'(\cdot,\cdot)} \mathbb{E}\left(d_a\left(U,\phi(E,W_1)\right)\right) \\ + (1-\zeta_s') \min_{\phi(\cdot,\cdot)} \mathbb{E}\left(d_a\left(U,\phi(E,W_2)\right)\right),\end{aligned} \tag{D.35}$$

$$D > \mathbb{E}(d_l(V,\hat{V})), \tag{D.36}$$

for some function $\phi_l : \mathcal{W} \times \mathcal{Z} \times \mathcal{Y} \times \mathcal{W}_2 \mapsto \hat{\mathcal{V}}$. Specializing the result to the lossless case ($D = 0$) with the hamming distortion metric will then complete the proof. Below, we describe the encoding and decoding operations of our scheme, which is basically a hybrid coding scheme with an embedded superposition code [72] to achieve secrecy.

Codebook generation: Fix a joint distribution $P_{UVEZY}P_{W_2|U}P_{W_1|W_2}P_{W|V}P_{X|UW_1W_2}P_{YJ|X}$ and $\phi_l : \mathcal{W} \times \mathcal{Z} \times \mathcal{W}_2 \times \mathcal{Y} \mapsto \hat{\mathcal{V}}$ satisfying (D.33), (D.35) and (D.36). Choose numbers R_1 and R_2 be non-negative numbers such that

$$I(U;W_1) < R_1 < I(W_1;Y,Z), \text{ if } I(U;W_1) < I(W_1;Y,Z), \tag{D.37}$$

$$I(U;W_1) < R_1 < H(W_1), \text{ otherwise,} \tag{D.38}$$

$$R_2 < I(W_2;Y,Z|W_1), \tag{D.39}$$

$$\text{and } I(W_2;U) < R_1 + R_2 < I(W_1,W_2;Y,Z). \tag{D.40}$$

Codebook used by encoder of source V: Let R_s and R_s' be non-negative numbers. Generate codewords $W^n(m,m'), (m,m') \in \mathcal{M} \times \mathcal{M}' := [2^{nR_s}] \times [2^{nR_s'}]$, each drawn independently according to the distribution $\prod_{i=1}^n P_W$. Denote this random codebook by $\mathcal{C}_v^n$.

Codebook used by encoder of source U: Generate codewords $W_1^n(m_1)$, $m_1 \in [2^{nR_1}]$, drawn independently according to the distribution $\prod_{i=1}^n P_{W_1}$. Denote this codebook by $\mathcal{C}_{w_1}^n$. For each m_1, generate codewords $W_2^n(m_1,m_2)$, $m_2 \in [2^{nR_2}]$, drawn independently according to the distribution $\prod_{i=1}^n P_{W_2|W_1}\left(w_{2i}|W_{1i}(m_1)\right)$. Denote this codebook by $\mathcal{C}_{w_2}^n$, and the two codebooks $\mathcal{C}_{w_1}^n$ and $\mathcal{C}_{w_2}^n$ together by $\mathcal{C}_u^n$. The codebooks $\mathcal{C}_u^n$ and $\mathcal{C}_v^n$ are known to all the parties including the eavesdropper.

Encoder: First, the encoder of source U chooses (M_1, M_2) and the encoder of source V chooses (M, M') stochastically according to distributions

$$P_{E_u}(m_1, m_2|u^n) := \frac{\prod_{i=1}^n P_{U|W_2}(u_i|W_{2i}(m_1, m_2))}{\sum_{\check{m}\in\check{\mathcal{M}}} \prod_{i=1}^n P_{U|W_2(u_i|W_{2i}(\check{m}))}}, \ \forall\ (m_1, m_2) \in \check{\mathcal{M}} := \mathcal{M}_1 \times \mathcal{M}_2,$$

$$\text{and } P_{E_v}(m, m'|v^n) := \frac{\prod_{i=1}^n P_{V|W}(v_i|W_i(m, m'))}{\sum_{\tilde{m}\in\mathcal{M}\times\mathcal{M}'} \prod_{i=1}^n P_{V|W}\left(u_i|W_i(\tilde{m})\right)}, \ \forall\ (m, m') \in \mathcal{M} \times \mathcal{M}',$$

respectively. Note that P_{E_u} and P_{E_v} are r.v.'s that depend on the realization of $\mathcal{C}_u^n$ and $\mathcal{C}_v^n$, respectively. The encoder of source U transmits the codeword X^n over the channel $P_{YJ|X}$, where X^n is randomly generated according to the distribution $\prod_{i=1}^n P_{X|W_1W_2U}(x_i|W_{1i}(m_1), W_{2i}(m_1, m_2), u_i)$, and the encoder of source V transmits M over the noiseless channel.

Decoder: Upon receiving $Y^n = y^n$ and $M = m$ and observing $Z^n = z^n$, the decoder first uses a good channel decoding rule $P_{D_1}(\hat{m}_1, \hat{m}_2|y^n, z^n)$ to estimate the messages (M_1, M_2). Subsequently index M' is recovered by using a good channel decoding rule for the transmission of messages M' over a memoryless channel $\prod_{i=1}^n P_{ZYW_2|W}$ using the sub-codebook $\{W^n(m, m'), m' \in \mathcal{M}'\}$. The decoder for the source U can be considered to be composed of two parts, $P_{D_u^M}(\hat{m}_1, \hat{m}_2|y^n, z^n)$ which is a good channel decoder followed by P_{D_u} given by

$$P_{D_u}(\hat{w}_2^n|\hat{m}_1, \hat{m}_2) := \mathbb{1}(W_2^n(\hat{m}_1, \hat{m}_2) = \hat{w}_2^n).$$

Similarly, the decoder for source V is composed of two parts, a good channel decoder $P_{D_v^M}(\hat{m}'|m, \hat{w}_2^n, z^n, y^n)$ followed by a symbol by symbol reconstruction given by

$$P_{D_v}(\hat{v}^n|m, \hat{m}', \hat{w}_2^n, z^n, y^n) = \prod_{i=1}^n \mathbb{1}(\phi_l(\hat{w}_i(m, \hat{m}'), \hat{w}_{2i}, z_i, y_i) = \hat{v}_i).$$

Analysis:

The joint distribution induced by the encoding and decoding operations is given by

$$\begin{aligned}
&\tilde{P}(e^n, b^n, z^n, u^n, m_1, m_2, w_1^n, w_2^n, v^n, m, m', w^n, x^n, y^n, j^n, \hat{m}_1, \hat{m}_2, \hat{w}_2^n, \hat{m}', \hat{v}^n) \\
&= P_{E^nB^nZ^nU^nV^n}(e^n, b^n, z^n, u^n, v^n) P_{E_u}(m_1, m_2|u^n) P_{W_1^n|M_1}(w_1^n|m_1) \\
&\quad P_{W_2^n|W_1^n(M_1)M_2}(w_2^n|w_1^n, m_2)\ P_{E_v}(m, m'|v^n)\ P_{W^n|MM'}(w^n|m, m')
\end{aligned}$$

$$
\begin{aligned}
&P_{X^n|W_1^nW_2^nU^n}(x^n|w_1^n,w_2^n,u^n)\ P_{Y^nJ^n|X^n}(y^n,j^n|x^n)\ P_{D_u^M}(\hat{m}_1,\hat{m}_2|y^n,z^n)\\
&P_{D_u}(\hat{w}_2^n|\hat{m}_1,\hat{m}_2)\ P_{D_v^M}(\hat{m}'|m,\hat{w}_2^n,z^n,y^n)P_{D_v}(\hat{v}^n|m,\hat{m}',\hat{w}_2^n,z^n,y^n)\\
&=\left[\prod_{i=1}^n P_{EBZUV}(e_i,b_i,z_i,u_i,v_i)\right]P_{E_u}(m_1,m_2|u^n)\mathbb{1}(W_1^n(m_1)=w_1^n)\\
&\mathbb{1}(W_2^n(m_1,m_2)=w_2^n)\ P_{E_v}(m,m'|v^n)\mathbb{1}(W^n(m,m')=w^n)\\
&\left[\prod_{i=1}^n P_{X|W_1W_2U}(x_i|w_{ii},w_{2i},u_i)\right]\left[\prod_{i=1}^n P_{YJ|X}(y_i,j_i|x_i)\right]P_{D_u^M}(\hat{m}_1,\hat{m}_2|y^n,z^n)\\
&\mathbb{1}(\hat{W}_2^n(\hat{m}_1,\hat{m}_2)=\hat{w}_2^n)\ P_{D_v^M}(\hat{m}'|m,\hat{w}_2^n,z^n,y^n)\ P_{D_v}(\hat{v}^n|m,\hat{m}',\hat{w}_2^n,z^n,y^n).
\end{aligned}
$$

Let Q and $Q^{(1)}$ denote auxiliary distributions defined as

$$
\begin{aligned}
&Q(e^n,b^n,z^n,u^n,m_1,m_2,w_1^n,w_2^n,v^n,m,m',w^n,x^n,y^n,j^n,\hat{m}_1,\hat{m}_2,\hat{w}_2^n,\hat{m}',\hat{v}^n)\\
&:=Q_{M_1M_2}(m_1,m_2)\ Q_{W_1^n|M_1}(w_1^n|m_1)\ Q_{W_2^n|W_1^n(M_1)M_2}(w_2^n|w_1^n,m_2)\\
&\quad Q_{U^n|W_2^n(M_1,M_2)}(u^n|w_2^n(m_1,m_2))\ Q_{Z^nV^nE^nB^n|U^n}(z^n,v^n,e^n,b^n|u^n)\\
&\quad Q_{X^n|W_1^nW_2^nU^n}(x^n|w_1^n,w_2^n,u^n)\ Q_{Y^nJ^n|X^n}(y^n,j^n|x^n)\ Q_{\hat{M}_1\hat{M}_2|Z^nY^n}(\hat{m}_1,\hat{m}_2|z^n,y^n)\\
&\quad Q_{\hat{W}_2^n|\hat{M}_1\hat{M}_2}(\hat{w}_2^n|\hat{m}_1,\hat{m}_2)\ Q_{MM'|V^n}(m,m'|v^n)\ Q_{W^n|MM'}(w^n|m,m')\\
&\quad Q_{\hat{M}'|MZ^n\hat{W}_2^nY^n}(\hat{m}'|m,z^n,\hat{w}_2^n,y^n)\ Q_{\hat{V}^n|M\hat{M}'\hat{W}_2^nZ^nY^n}(\hat{v}^n|m,\hat{m}',\hat{w}_2^n,z^n,y^n)\\
&:=\frac{1}{2^{n(R_1+R_2)}}\mathbb{1}\left(W_1^n(m_1)=w_1^n\right)\mathbb{1}\left(W_2^n(m_1,m_2)=w_2^n\right)\ \left[\prod_{i=1}^n P_{U|W_2}(u_i|w_{2i}(m_1,m_2))\right]\\
&\quad\left[\prod_{i=1}^n P_{ZVEB|U}(z_i,v_i,e_i,b_i|u_i)\right]\ \left[\prod_{i=1}^n P_{X|W_1W_2U}(x_i|w_{1i},w_{2i},u_i)\right] \qquad \text{(D.41)}\\
&\quad\left[\prod_{i=1}^n P_{YJ|X}(y_i,j_i|x_i)\right]\ P_{D_u^M}(\hat{m}_1,\hat{m}_2|z^n,y^n)\ \mathbb{1}(\hat{W}_2^n(\hat{m}_1,\hat{m}_2)=\hat{w}_2^n)\\
&\quad\left[P_{E_v}(m,m'|v^n)\ \mathbb{1}(W^n(m,m')=w^n)\ P_{D_v^M}(\hat{m}'|m,\hat{w}_2^n,z^n,y^n)\right]\\
&\quad P_{D_v}(\hat{v}^n|m,\hat{m}',\hat{w}_2^n,z^n,y^n)\\
&:=Q^H_{M_1M_2W_1^nW_2^nU^nZ^nV^nE^nB^nX^nY^nJ^n\hat{M}_1\hat{M}_2\hat{W}_2^n}(m_1,m_2,w_1^n,w_2^n,u^n,z^n,v^n,e^n,b^n,\\
&\qquad\qquad x^n,y^n,j^n,\hat{m}_1,\hat{m}_2,\hat{w}_2^n)\\
&\quad Q^S_{MM'W^n\hat{M}'\hat{V}^n|V^nZ^n\hat{W}_2^nY^n}(m,m',w^n,\hat{m}',\hat{v}^n|v^n,z^n,\hat{w}_2^n,y^n),
\end{aligned}
$$

and

$$
\begin{aligned}
&Q^{(1)}(m_1, m_2, w_1^n, w_2^n, u^n, z^n, v^n, e^n, b^n, x^n, y^n, j^n, \hat{m}_1, \hat{m}_2, \hat{w}_2^n, y^n, m, m', w^n, \hat{m}', \hat{v}^n) \\
&:= Q_{M_1 M_2 W_1^n W_2^n U^n Z^n V^n E^n B^n X^n Y^n J^n \hat{M}_1 \hat{M}_2}(m_1, m_2, m_1', m_2', w_1^n, w_2^n, u^n, z^n, \\
&\qquad\qquad\qquad\qquad\qquad\qquad v^n, e^n, b^n, x^n, y^n, j^n, \hat{m}_1, \hat{m}_2) \\
&\quad P_{D_u}(\hat{w}_2^n | m_1, m_2)\ Q^S(m, m', w^n, \hat{m}', \hat{v}^n | v^n, z^n, \hat{w}_2^n, y^n) \\
&:= Q_{M_1 M_2 W_1^n W_2^n U^n Z^n V^n E^n B^n X^n Y^n J^n \hat{M}_1 \hat{M}_2}(m_1, m_2, m_1', m_2', w_1^n, w_2^n, u^n, z^n, \\
&\qquad\qquad\qquad\qquad\qquad\qquad v^n, e^n, b^n, x^n, y^n, j^n, \hat{m}_1, \hat{m}_2) \\
&\quad P_{D_u}(\hat{w}_2^n | m_1, m_2)\ P_{E_v}(m, m' | v^n)\ \mathbb{1}(W^n(m, m') = w^n)\ P_{D_v^M}(\hat{m}' | m, \hat{w}_2^n, z^n, y^n) \\
&\quad P_{D_v}(\hat{v}^n | m, \hat{m}', z^n, \hat{w}_2^n, y^n),
\end{aligned}
$$

respectively. Note that the only difference between Q and $Q^{(1)}$ is P_{D_u}, the part of the decoder that obtains the reconstruction sequence $\hat{W}_2^n$. In $Q^{(1)}$, the actual messages (m_1, m_2) intended by the encoder is used by P_{D_u} instead of the estimates $\hat{m}_1, \hat{m}_2$. Taking expectation with respect to the codebook $\mathcal{C}_u$, we obtain

$$
\mathbb{E}_{\mathcal{C}_u^n}\left[Q^{(1)}_{U^n Z^n Y^n J^n \hat{W}_2^n}\right] = P_{U^n Z^n Y^n J^n W_2^n} := \prod_{i=1}^n P_{UZYJW_2}.
$$

Let

$$
\begin{aligned}
&\bar{Q}^{(1)}_{U^n Z^n V^n E^n B^n Y^n J^n \hat{W}_2^n M M' W^n \hat{V}^n} \\
&:= \mathbb{E}_{\mathcal{C}_u^n}\left[Q^{(1)}_{U^n Z^n V^n E^n B^n Y^n J^n \hat{W}_2^n M M' W^n \hat{V}^n}(u^n, z^n, v^n, e^n, b^n, y^n, j^n, \hat{w}_2^n, m, m', w^n, \hat{m}', \hat{v}^n)\right] \\
&= \left[\prod_{i=1}^n P_{UZVEBYJW_2}(u_i, z_i, v_i, e_i, b_i, y_i, j_i, \hat{w}_{2i})\right] \\
&\quad Q^S_{M M' W^n \hat{M}' \hat{V}^n | V^n Z^n Y^n \hat{W}_2^n}(m, m', w^n, \hat{m}', \hat{v}^n | v^n, z^n, y^n, \hat{w}_2^n).
\end{aligned}
$$

Observe that the likelihood encoder P_{E_u} is chosen such that

$$
P_{E_u} = Q_{M_1 M_2 | U^n}. \tag{D.42}
$$

Let

$$
\begin{aligned}
&Q^H_{M_1M_2W_1^nW_2^nU^nZ^nV^nE^nB^nX^nY^nJ^n\hat{M}_1\hat{M}_2\hat{W}_2^n} \\
&= Q_{U^n}(u^n)\ P_{E_u}(m_1,m_2|u^n)\ P_{W_1^n|M_1}(w_1^n|m_1)\ P_{W_2^n|W_1^n(M_1)M_2}(w_2^n|w_1^n,m_2) \\
&\quad \left[\prod_{i=1}^n P_{ZVEB|U}(z_i,v_i,e_i,b_i|u_i)\right]\ \left[\prod_{i=1}^n P_{X|W_1W_2U}(x_i|w_{ii},w_{2i},u_i)\right] \\
&\quad \left[\prod_{i=1}^n P_{YJ|X}(y_i,j_i|x_i)\right]\ P_{D_u^M}(\hat{m}_1,\hat{m}_2|z^n)\ P_{D_u}(\hat{w}_2^n|m_1,\hat{m}_1',m_2,\hat{m}_2').
\end{aligned}
$$

Note that the only difference between Q^H and $\tilde{P}$ is the marginal distribution of U. By Lemma 5.2, it follows that

$$\mathbb{E}_{\mathcal{C}_u^n}\left[\|Q_U^H-\tilde{P}_U\|\right]\le e^{-\delta_1 n},$$

since $I(U;W_1)<R_1\le H(W)$ and $R_1+R_2>I(U;W_1,W_2)=I(U;W_2)$ by (D.37)-(D.38) and (D.40), respectively. Note that the above conditions also implies that (5.12) is satisfied, i.e., $R_2>I(U;W_2)-R_1\ge I(U;W_2)-H(W_1)$.

Applying Property 4.4.1 (c), we obtain that for some $\delta_1>0$,

$$
\begin{aligned}
&\mathbb{E}_{\mathcal{C}_u^n}\Big[\|Q^H_{M_1M_2W_1^nW_2^nU^nZ^nV^nE^nB^nX^nY^nJ^n\hat{M}_1\hat{M}_2\hat{W}_2^n} \\
&\qquad\qquad -\tilde{P}_{M_1M_2W_1^nW_2^nU^nZ^nV^nE^nB^nX^nY^nJ^n\hat{M}_1\hat{M}_2\hat{W}_2^n}\|\Big] \\
&\le e^{-\delta_1 n}.
\end{aligned}
$$

Since the messages M_1,M_2 are uniformly distributed under the joint distribution Q^H, and $R_2<I(W_2;Z,Y|W_1)$ and $R_1+R_2<I(W_1,W_2;Z,Y)$ (by (D.39) and (D.40)), it is well known that a maximum likelihood decoder drives the decoding error probability to zero, i.e.,

$$\mathbb{E}_{\mathcal{C}_u^n}\left[\Pr(\hat{M}_1,\hat{M}_2)\ne(M_1,M_2)\right]\le\epsilon_n',$$

for some $\epsilon_n'\xrightarrow{(n)}0$, where the probability is evaluated based on the joint distribution Q^H.

Hence by Lemma 5.1, we have

$$\mathbb{E}_{\mathcal{C}_u^n}\Bigg[\|Q^{(1)}_{M_1M_2W_1^nW_2^nU^nZ^nV^nE^nB^nX^nY^nJ^n\hat{M}_1\hat{M}_2\hat{W}_2^n} - Q^{H}_{M_1M_2W_1^nW_2^nU^nZ^nV^nE^nB^nX^nY^nJ^n\hat{M}_1\hat{M}_2\hat{W}_2^n}\|\Bigg] \leq \epsilon'_n.$$

Applying Property 4.4.1(c) yields

$$\begin{aligned}
&\mathbb{E}_{\mathcal{C}_u^n}\Big[\|Q^{(1)}_{M_1M_2W_1^nW_2^nU^nZ^nV^nE^nB^nX^nY^nJ^n\hat{M}_1\hat{M}_2\hat{W}_2^nMM'W^n\hat{M}'\hat{W}^n} - \\
&\qquad \tilde{P}_{M_1M_2W_1^nW_2^nU^nZ^nV^nE^nB^nX^nY^nJ^n\hat{M}_1\hat{M}_2\hat{W}_2^nMM'W^n\hat{M}'\hat{V}^n}\|\Big] \\
&= \mathbb{E}_{\mathcal{C}_u^n}\Big[\|Q^{(1)}_{M_1M_2W_1^nW_2^nU^nZ^nV^nE^nB^nX^nY^nJ^n\hat{M}_1\hat{M}_2\hat{W}_2^n} \\
&\qquad - \tilde{P}_{M_1M_2W_1^nW_2^nU^nZ^nV^nE^nB^nX^nY^nJ^n\hat{M}_1\hat{M}_2\hat{W}_2^n}\|\Big] \\
&\leq \mathbb{E}_{\mathcal{C}_u^n}\Big[\|Q^{(1)} - Q^H\|\Big] + \mathbb{E}_{\mathcal{C}_u^n}\Big[\|Q^H - \tilde{P}\|\Big] \leq e^{-\delta_1 n} + \epsilon'_n := \delta_{2n} \xrightarrow{(n)} 0,
\end{aligned}$$

and

$$\mathbb{E}_{\mathcal{C}_u^n}\Big[\mathbb{E}_{\tilde{P}}\Big[d_l(V^n,\hat{V}^n)\Big]\Big] \leq \Big[\mathbb{E}_{\bar{Q}^{(1)}}\Big[d_l(V^n,\hat{V}^n)\Big]\Big] + D_l\ \delta_{2n}. \tag{D.43}$$

In order to bound the first term in (D.43), we will use similar results to that used in the Wyner- Ziv section in [67]. As shown there, $\bar{Q}^{(1)}$ is equal to the distribution induced by the maximum likelihood encoder P_{E_v} for the rate distortion problem with the source V and side information (Z, W_2, Y) at the decoder.

Let

$$\begin{aligned}
&Q^{(3)}_{U^nZ^nV^nE^nB^nX^nY^nJ^n\hat{W}_2^nMM'W^n\hat{M}'\hat{V}^n}(u^n,z^n,v^n,e^n,b^n,x^n,y^n,j^n,\hat{w}_2^n,m,m',w^n,\hat{m}',\hat{v}^n) \\
&:= \frac{1}{2^{n(R_s+R'_s)}}\mathbb{1}(W^n(m,m')=w^n)\Big[\prod_{i=1}^n P_{V|W}(v_i|w_i)P_{UZEBXYJW_2|V}(u_i,z_i,e_i,b_i,x_i,y_i, \\
&\quad j_i,\hat{w}_{2i}|v_i)\Big]\ P_{D_v^M}(\hat{m}'|m,\hat{w}_2^n,z^n,y^n)\ P_{D_v}(\hat{v}^n|m,m',\hat{w}_2^n,z^n,y^n).
\end{aligned}$$

Note that P_{D_v} uses the actual messages (m, m') rather than estimates $(m, \hat{m}')$ for forming the reconstruction $\hat{V}$. It follows from the results in [67] that for some $\delta_3 > 0$,

$$\mathbb{E}_{\mathcal{C}_v^n}\left[\|\bar{Q}^{(1)}_{U^nZ^nV^nE^nB^nX^nY^nJ^n\hat{W}_2^nMM'W^n\hat{M}'\hat{V}^n} - Q^{(3)}_{U^nZ^nV^nE^nB^nX^nY^nJ^n\hat{W}_2^nMM'W^n\hat{M}'\hat{V}^n}\|\right] \leq e^{-\delta_3 n} \xrightarrow{(n)} 0,$$

provided that

$$R_s + R_s' > I(W;V), \tag{D.44}$$

$$\text{and } R_s' < I(W;Z,W_2,Y). \tag{D.45}$$

Note that (D.44) and (D.45) together implies via the Markov relation $(Z, W_2, Y) - V - W$ that

$$R_s > I(W;V|W_2,Z,Y). \tag{D.46}$$

Next, note that

$$\mathbb{E}_{\mathcal{C}_v^n}\left[Q^{(3)}_{U^nZ^nV^nE^nB^nX^nY^nJ^n\hat{W}_2^n\hat{V}^n}\right] = \prod_{i=1}^{n} P_{UZVEBXYJW_2\phi_l(W)}.$$

Thus, the average distortion at the legitimate receiver averaged over the random codebook $\mathcal{C}_v^n$ and $\mathcal{C}_u^n$ can be bounded as

$$\begin{aligned}
\mathbb{E}_{\mathcal{C}_v^n}\left[\mathbb{E}_{\mathcal{C}_u^n}\left[\mathbb{E}_{\tilde{P}}\left[d_l(V^n,\hat{V}^n)\right]\right]\right] &= \mathbb{E}_{\mathcal{C}_v^n}\left[\mathbb{E}_{\bar{Q}^{(1)}}\left[d_l(V^n,\hat{V}^n)\right] + D_l\delta_{2n}\right] \\
&\leq \mathbb{E}_{\mathcal{C}_v^n}\left[\mathbb{E}_{Q^{(3)}}\left[d_l(V^n,\hat{V}^n)\right]\right] + D_l\left(\delta_{2n}+\delta_{3n}\right) \\
&= \mathbb{E}_P\left[d_l(V^n,\hat{V}^n)\right] + D_l\left(\delta_{2n}+\delta_{3n}\right) \\
&= \mathbb{E}_P\left[d_l(V,\hat{V})\right] + D_l\left(\delta_{2n}+\delta_{3n}\right) \\
&\leq D + D_l\left(\delta_{2n}+\delta_{3n}\right). \qquad \text{(D.47)}
\end{aligned}$$

Analysis of distortion at eavesdropper:

For $k \geq 1$, consider an auxiliary distribution

$$\begin{aligned}
&\hat{Q}^{(k)}_{M_1 M_2 W_1^n E^n J^n B^{k-1} U_k}(m_1, m_2, w_1^n, e^n, j^n, b^{k-1}, u_k) \\
&:= \frac{1}{2^{n(R_1+R_2)}} \mathbb{1}(W_1^n(m_1) = w_1^n) \left[\prod_{i=1}^{n} P_{EJ|W_1}(e_i, j_i|w_{1i})\right] \\
&\quad \left[\prod_{i=1}^{k-1} P_{B|EW_1J}(b_i|e_i, w_{1i}, j_i)\right] P_{U|EW_1J}(u_k|e_k, w_{1k}, j_k).
\end{aligned} \tag{D.48}$$

Note that the following Markov relation holds under $\hat{Q}^{(k)}$,

$$U_k - (E_k, W_{1k}(M_1), J_k) - \left(B^{k-1}, E^n, J^n, M_1, M_2\right).$$

Also, observe that by definition,

$$\begin{aligned}
&Q_{M_1 M_2 E^n J^n B^{k-1} U_k} \\
&= \frac{1}{2^{n(R_1+R_2)}} \left[\prod_{i=1}^{n} P_{EJ|W_2}(e_i, j_i|W_{2i}(m_1, m_2))\right] \\
&\quad \left[\prod_{i=1}^{k-1} P_{B|EW_2J}(b_i|e_i, W_{2i}(m_1, m_2), j_i)\right] P_{U|EW_2J}(u_k|e_k, w_{2k}, j_k) \\
&= \frac{1}{2^{n(R_1+R_2)}} \left[\prod_{i=1}^{n} P_{EJ|W_2W_1}(e_i, j_i|W_{2i}(m_1, m_2), W_{1i}(m_1))\right] \\
&\quad \left[\prod_{i=1}^{k-1} P_{B|EJW_2W_1}(b_i|e_i, j_i, W_{2i}(m_1, m_2), W_{1i}(m_1))\right] \\
&\quad P_{U|EJW_2W_1}(u_k|e_k, j_k, W_{2k}(m_1, m_2), W_{1k}(m_1)).
\end{aligned}$$

By application of Lemma 5.2, it follows that for some $\delta_4 > 0$,

$$\mathbb{E}_{\mathcal{C}_v^n}\left[\|\hat{Q}^{(k)}_{M_1 E^n J^n B^{k-1} U_k} - Q_{M_1 E^n J^n B^{k-1} U_k}\|\right] \leq e^{-\delta_4 n} \xrightarrow{(n)} 0,$$

for any $k < k_2,\ k \in \mathbb{Z}^+$, where

$$k_2 := \frac{(R_2 - I(W_2; E, J|W_1))n}{I(B; W_2|W_1, E, J)} + \frac{I(B; W_2|W_1, E, J) - I(U; W_2|B, W_1, E, J)}{I(B; W_2|W_1, E, J)}.$$

Hence,

$$\mathbb{E}_{\mathcal{C}_v^n}\left[\sum_{k=1}^{k_2-1}\|\hat{Q}^{(k)}_{M_1E^nJ^nB^{k-1}U_k}-Q_{M_1E^nJ^nB^{k-1}U_k}\|\right]\leq (k_2-1)e^{-\delta_4 n}$$
$$\leq e^{-\delta_5 n}\xrightarrow{(n)}0, \tag{D.49}$$

for some $\delta_5>0$. Also, identifying (W,Z,X,Y,R_1,R_2) in Lemma 5.2 with $(\emptyset,W_1,\emptyset,W_1,0,R_1)$, we obtain that for all $1\leq i\leq n$,

$$\mathbb{E}_{\mathcal{C}_v^n}\left[\|P_{W_1}-\hat{Q}^{(i)}_{W_{1i}}(M_1)\|\right]\leq e^{-\delta_6 n},$$

for some $\delta_6>0$ since $R_1>I(\emptyset;W_1)=0$. Hence,

$$\mathbb{E}_{\mathcal{C}_v^n}\left[\sum_{i=1}^{n}\|P_{W_1}-\hat{Q}^{(i)}_{W_{1i}}(M_1)\|\right]\leq ne^{-\delta_6 n}\leq e^{-\delta_7 n}\xrightarrow{(n)}0. \tag{D.50}$$

For $k\geq 1$, consider another auxiliary distribution,

$$\check{Q}^{(k)}_{M_1M_2E^nJ^nB^{k-1}U_k}(m_1,m_2,e^n,j^n,b^{k-1},u_k)$$
$$:=\frac{1}{2^{n(R_1+R_2)}}\left[\prod_{i=1}^{n}P_{EJ}(e_i,j_i)\right]\left[\prod_{i=1}^{k-1}P_{B|E,J}(b_i|e_i,j_i)\right]P_{U|EJ}(u_k|e_k,j_k)$$

Note that under $\check{Q}^{(k)}$,

$$U_k-(E_k,J_k)-\left(M_1,M_2,B^{k-1},E^n,J^n\right).$$

By an application of Lemma 5.2, it follows that

$$\mathbb{E}_{\mathcal{C}_v^n}\left[\|\check{Q}^{(k)}_{E^nJ^nB^{k-1}U_k}-Q_{E^nJ^nB^{k-1}U_k}\|\right]\leq e^{-\delta_8 n}\xrightarrow{(n)}0,$$

for some $\delta_8>0$ and any $k<k_1$, where

$$k_1:=\min\left(k_2,\ \frac{(R_1-I(W_1;E,J))n}{I(B;W_1|E,J)}+\frac{I(B;W_1|E,J)-I(U;W_1|B,E,J)}{I(B;W_1|E,J)}\right).$$

Hence,

$$\mathbb{E}_{\mathcal{C}_v^n}\left[\sum_{k=1}^{k_1-1}\|\check{Q}^{(k)}_{E^nJ^nB^{k-1}U_k}-Q_{E^nJ^nB^{k-1}U_k}\|\right]\leq(k_1-1)e^{-\delta_8 n}$$
$$\leq e^{-\delta_9 n}\xrightarrow{(n)}0, \tag{D.51}$$

for some $\delta_9>0$. By identifying (W,Z,X,Y,R_1,R_2) in Lemma 5.2 with $(W_1,W_2,\emptyset,W_2,R_1,R_2)$ and noting that $R_1>0$, $R_2>0$, we obtain again by an application of Lemma 5.2 that

$$\mathbb{E}_{\mathcal{C}_v^n}\left[\sum_{i=1}^{n}\|P_{W_2}-Q_{W_{2i}}(M_1,M_2)\|\right]\leq e^{-\delta_{10}n}\xrightarrow{(n)}0, \tag{D.52}$$

for some $\delta_{10}>0$. By the random coding argument, there exists a deterministic codebook $\mathcal{C}_u^n$ and $\mathcal{C}_v^n$ such that (D.47), (D.49), (D.50), (D.51) and (D.52) are satisfied. Now, the distortion at the eavesdropper can be lower bounded as follows:

$$\min_{\{\phi_i(e^n,\ j^n,\ b^{i-1})\}}\left[\sum_{i=1}^{n}\mathbb{E}_{\tilde{P}}\left(d_a\left(U_i,\phi_i\left(E^n,J^n,B^{(i-1)}\right)\right)\right)\right]$$
$$\geq\min_{\{\phi_i(e^n,\ j^n,\ b^{i-1})\}}\left[\sum_{i=1}^{n}\mathbb{E}_{Q}\left(d_a\left(U_i,\phi_i\left(E^n,J^n,B^{(i-1)}\right)\right)\right)\right]-nD_a\left(e^{-\delta_1 n}\right). \tag{D.53}$$

Consider the first term in (D.53). We can write

$$\min_{\{\phi_i(e^n,\ j^n,\ b^{i-1})\}}\left[\sum_{i=1}^{k_1-1}\mathbb{E}_{Q}\left(d_a\left(U_i,\phi_i\left(E^n,J^n,B^{(i-1)}\right)\right)\right)\right]$$
$$\geq\min_{\{\phi_i(e_i)\}}\left[\sum_{i=1}^{k_1-1}\mathbb{E}_{\check{Q}^{(i)}}\left(d_a\left(U_i,\phi_i(E_i)\right)\right)\right]-nD_a\left(e^{-\delta_9 n}\right) \tag{D.54}$$
$$\geq(k_1-1)\min_{\{\phi''(\cdot)\}}\mathbb{E}_P\left(d_a\left(U,\phi(E)\right)\right)\Bigg]-nD_a\left(e^{-\delta_9 n}\right),$$

where (D.54) follows from (D.51).

Similarly,

$$\min_{\{\phi_i(e^n,j^n,b^{i-1})\}}\left[\sum_{i=k_1}^{k_2-1}\mathbb{E}_{Q}\left(d_a\left(U_i,\phi_i\left(E^n,J^n,B^{(i-1)}\right)\right)\right)\right]$$

$$\geq \min_{\{\phi_i(e_i,j_i,w_{1i})\}} \left[\sum_{i=k_1}^{k_2-1} \mathbb{E}_{\hat{Q}^{(i)}} \left(d_a \left(U_i, \phi_i(E_i, J_i, W_{1i}(M_1)) \right) \right) \right] - nD_a \left(e^{-\delta_5 n} \right) \tag{D.55}$$

$$\geq (k_2 - k_1 - 1) \min_{\{\phi(e,y_2,w_1)\}} \mathbb{E}_P \left(d_a \left(U, \phi(E, J, W_1) \right) \right) - nD_a \left(e^{-\delta_5 n} \right), \tag{D.56}$$

where (D.55) follows from (D.49). Equation (D.56) follows from (D.50) by noting that $\hat{Q}^i_{E_i J_i U_i | W_{1i}} = P_{EJU|W_1}$.

The remaining terms inside the summation in (D.53) can be written as

$$\begin{aligned}
&\min_{\{\phi_i(e^n,j^n,b^{i-1})\}} \left[\sum_{i=k_2}^{n} \mathbb{E}_Q \left(d_a \left(U_i, \phi_i \left(E^n, J^n, B^{(i-1)} \right) \right) \right) \right] \\
&\geq \min_{\{\phi_i(e_i,j_i,w_{2i})\}} \left[\sum_{i=k_2}^{n} \mathbb{E}_Q \left(d_a(U_i, \phi_i(E_i, J_i, W_{2i}(M_1, M_2))) \right) \right] \\
&= (n - k_2 + 1) \min_{\{\phi(e,j,w_2)\}} \mathbb{E}_P \left(d_a(U, \phi(E, J, W_2)) \right),
\end{aligned} \tag{D.57}$$

where (D.57) follows from (D.52). Finally, note that

$$\lim_{n\to\infty} \frac{k_2 - 1}{n} = \frac{R_2 - I(W_2; E, J|W_1)}{I(B; W_2|W_1, E, J)}, \tag{D.58}$$

$$\lim_{n\to\infty} \frac{k_1 - 1}{n} = \min \left(\frac{R_2 - I(W_2; E, J|W_1)}{I(B; W_2|W_1, E, J)}, \frac{R_1 - I(W_1; E, J)}{I(B; W_1|E, J)} \right), \tag{D.59}$$

and

$$R_2 < \min \ (I(W_2; Z, Y|W_1), I(W_1, W_2; Y, Z) - R_1) \tag{D.60}$$

$$< \min \ (I(W_2; Z, Y|W_1), I(W_1, W_2; Y, Z) - I(W_1; U)), \tag{D.61}$$

where (D.60) and (D.61) follows from (D.39)-(D.40) and (D.37)-(D.38), respectively.

By maximizing the distortion incurred at the eavesdropper with respect to rate R_1 (the supremum occurs at a value of $R_1 = I(W_1; Z, Y)$), it follows that for any $\gamma > 0$ and sufficiently large n,

$$\min_{\{\phi_i(e^n,j^n,b^{i-1})\}} \left[\sum_{i=1}^{n} \mathbb{E}_{\tilde{P}} \left(d_a \left(U_i, \phi_i \left(E^n, J^n, B^{(i-1)} \right) \right) \right) \right] > n\Delta - \gamma. \tag{D.62}$$

Thus, from (D.46), (D.47) and (D.62), we have shown that $(R_s, D, \Delta) \in \mathcal{R}_g$ provided

(D.33)-(D.36) are satisfied. Specializing to the lossless case ($D = 0$) with hamming distortion measure, we obtain the condition $R_s > H(V|W_2, Z, Y)$ given in (D.34) by setting $W = V$. This completes the proof of the theorem.