Hybrid Digital-Analog Source-Channel Coding and Information Hiding: Information-Theoretic Perspectives

by

Yadong Wang

A thesis submitted to the
Department of Mathematics and Statistics
in conformity with the requirements for
the degree of Doctor of Philosophy

Queen's University
Kingston, Ontario, Canada
September, 2007

Abstract

Joint source-channel coding (JSCC) has been acknowledged to have superior performance over separate source-channel coding in terms of coding efficiency, delay and complexity. In the first part of this thesis, we study a hybrid digital-analog (HDA) JSCC system to transmit a memoryless Gaussian source over a memoryless Gaussian channel under bandwidth compression. Information-theoretic upper bounds on the asymptotically optimal mean squared error distortion of the system are obtained. An allocation scheme for distributing the channel input power between the analog and the digital signals is derived for the HDA system with mismatched channel conditions. A low-complexity and low-delay version of the system is next designed and implemented. We then propose an image communication application demonstrating the effectiveness of HDA coding.

In the second part of this thesis, we consider problems in information hiding. We begin by considering a single-user joint compression and private watermarking (JCPW) problem. For memoryless Gaussian sources and memoryless Gaussian attacks, an exponential upper bound on the probability of error in decoding the watermark is derived. Numerical examples show that the error exponent is positive over a (large) subset of the entire achievable region derived by Karakos and Papamarcou (2003).

We then extend the JCPW problem to a multi-user setting. Two encoders independently embed two secret information messages into two correlated host sources subject to a pair of tolerable distortion levels. The (compressed) outputs are subject to multiple access attacks. The tradeoff between the achievable watermarking rates and the compression rates is studied for discrete memoryless host sources and discrete memoryless multiple access channels. We derive an inner bound and an outer bound with single-letter characterization for the achievable compression and watermarking rate region. We next consider a problem where two correlated sources are separately embedded into a common host source. A single-letter sufficient condition is established under which the sources can be successfully embedded into the host source under multiple access attacks. Finally, we investigate a public two-user information hiding problem under multiple access attacks. Inner and outer bounds for the embedding capacity region are obtained with single-letter characterization.

Acknowledgements

I would like to express my sincere gratitude to my supervisors, Dr. Fady Alajaji and Dr. Tamás Linder, for their trust, expert guidance and invaluable professional support throughout this thesis work. Their confidence, kindness, patience and knowledge have been the driving force throughout my graduate studies at Queen's.

I would also like to thank Yangfan Zhong for many enjoyable and helpful discussions, and his research collaboration on the material of Chapters 6 and 7. I am also thankful to our graduate secretary, Mrs. Jennifer Read, for her generous help and dedication.

I am deeply thankful to all my family members for all the support, love and understanding. My special thanks go to my wife, Lei, for her love and trust.

Statement of Originality

All results presented in this thesis are original, unless otherwise stated. No part of this thesis has been submitted elsewhere for any other degree or qualification. The results quoted from the literature are presented as statements with indicated references.

Contents

List of Figures

List of Tables

Chapter 1

General Introduction

1.1 Motivation

1.1.1 Shannon's Source-Channel Coding

One of the ultimate goals in modern communication systems is to provide highly reliable and efficient transmission of data bearing signals over an inherently noisy medium. Various theories and systems have been developed in order to achieve this goal.

In a typical communication system, data bearing signals, such as text, images, video, speech, or combination of these, are often modeled as discrete-time continuous-amplitude random source sequences. This is reasonable since in practice signals are often low-pass filtered, and the sampling theorem guarantees that any band-limited signal with bandwidth W Hz can be accurately represented by sampling it at a rate of $2W$ times per second. Due to restrictions on bandwidth or storage, source sequences are often compressed using a source encoder to remove its natural redundancy. This procedure is called *source coding.* As a result, an inevitable loss of information occurs due to the source coding operation (except in the case of lossless source coding). On the other hand, this removal of redundancy, in turn, may introduce a greater level of sensitivity to transmission noise. Therefore, a channel encoder may be necessary to add some controlled redundancy into the output of the source encoder, which enables detecting and correcting errors at the channel decoder. This procedure is called *channel coding.* The output of the channel encoder is then modulated and sent over the waveform channel. The output of the channel is subsequently demodulated, decoded via a channel decoder and a source decoder, outputting a replica. This communication system is often called a *tandem* source-channel coding system.

In a tandem system, source and channel coders are designed separately and concatenated to form a complete system. Shannon's lossy source-channel separation principle [65] for single-user communication systems guarantees that splitting the coder into a source coder and a channel coder is optimal for most channels, given unconstrained coding delay and complexity. According to Shannon's source coding theorem, $R(D)$ bits per source sample are necessary and sufficient to represent the source samples with an average distortion not exceeding D when we operate on source blocks with arbitrary long length. Conversely, $D(R)$ is the minimum possible average distortion if R is the given source coding rate. We call $R(D)$ the rate-distortion function and $D(R)$ the distortion-rate function. These functions are inverses of each other and they can be calculated from the statistics of the random source. According to Shannon's channel coding theorem, the statistics of the channel determine a number C, called the channel capacity, such that information can only be reliably transmitted at rates below C. Thus we cannot communicate reliably at rates above channel capacity. Combining these two theorems, it is possible to obtain a reconstruction with fidelity D if $R(D) < C$; conversely, if the source can be sent over the channel and reconstructed with fidelity D, then it must hold that $R(D) \leq C$. Shannon's lossy source-channel separation theorem states that we can independently design a source coder with rate as small as possible (given distortion D)

assuming an error-free channel, and a channel coder which provides maximum protection against channel errors at a rate no larger than the channel capacity, no matter what the source statistics are.

1.1.2 Joint Source-Channel Coding

There are many theoretical results and successful practical systems available today which are based on Shannon's source-channel separation principle. However, there are a number of facts about the separation approach which merit attention. Firstly, near optimal performance can be obtained with large coding block lengths, causing large delay and complexity in practice. Secondly, for a tandem system, source and channel codes are designed independently. More specifically, source codes are usually designed assuming that the channel codes can correct all errors introduced by the channel, but this is not always the case even for the most powerful channel codes. Similarly, channel codes are usually designed to protect all the bits created by the source codes equally, assuming that information are equally distributed in these bits which is not always the case for many applications. Indeed, unequal error protection can result in better performance for this situation. Thirdly, there are situations for which the separation principle does not hold anymore; see [82] for an example of a non-stationary source-channel pair for which the converse of the lossless separation theorem does not hold.

These drawbacks have motivated researchers to investigate the design of source and channel codes jointly; such systems are generally called *joint* source-channel coding (JSCC) systems [28], [90]. By designing source and channel codes jointly, it has been shown that various gains may be obtained in terms of coding efficiency, reconstructed signal quality, coding delay and complexity.

Examples of joint source-channel coding systems are: (a) hierarchical protection, also known as unequal error protection, where the basic idea is to apply different channel codes to protect information according to the level of importance of the source data; (b) optimal quantizer design for noisy channels, such as channel-optimized vector quantization (COVQ) (e.g., [21], [38], [40], [92]); (c) optimal index assignment (e.g., [20], [93]); (d) direct source-channel mapping or direct modulation organization, where the encoder includes the modulator and benefits from the flexibility that is naturally present in a constellation (e.g. [60], [67]); (e) channel codes which are designed to exploit the residual redundancy of the source encoder output (e.g., [2], [68]).

We consider the problem of transmitting continuous-valued random sources over a discrete-time memoryless channel. In applications such as broadcasting and robust communication over wireless channels, there is a large variation in channel conditions depending on the physical landscape, the communication distance, the weather situation, etc. Thus, a communication system designed to perform well for a broad range of channel conditions is highly desired. Although most digital JSCC systems perform fairly well in terms of coding efficiency, coding delay, and have a less severe threshold effect (see Section 2.1 for the definition of "threshold effect") than tandem systems when the channel condition falls below the design parameters (i.e., channel signal-to-noise ratio), they usually fail to enhance performance as the channel condition improves due to the distortion introduced by quantizing the source. This leads us to investigate a special kind of JSCC systems: hybrid digital-analog coding systems. By combining digital and analog (or nearly analog) coding/modulation, we may expect a graceful performance improvement/degradation for a wide range of channel conditions.

1.1.3 Information Hiding

As the rapid development of information technology and internet, the communication of multimedia data becomes increasingly popular. People sell their (digital) works, communicate secret information, and do business via the internet. This includes all kinds of digital data, e.g., documents, photos, audio, video, etc. Such applications however raise many problems involving data protection, such as pirating, ownership identification, illegal copyright, and so on. These problems, which can be categorized into the area of "information hiding", have received considerable attention from both the academic world and industry.

In plain words, information hiding is the means to embed/hide a message (known as secret message or watermark) into a host data (*covertext*), so that the information hider is able to transmit the secret message even though the transmission is subject to manipulation by an attacker who tries to make the hidden information undetectable. In some applications such as copyright protection and fingerprinting, the hidden message carries information about the host data, e.g., ownership information, copyright of the host data, etc. In other applications such as secret communication and steganography, the hidden message can also be unrelated to the host data, or the host data acts as a 'carrier' of the secret message.

Generally, information hiding has two desirable characteristics:

- *Transparency:* The embedding procedure should cause as little degradation to the host data as possible. This is easily understood for most applications such as copyright protection and fingerprinting, since the aim is to protect the host data and also preserve the usability of the host data. For applications in secret communication and steganography, transparency can be interpreted as a characteristic to ensure the security of the communications.

- *Robustness:* The embedded message should resist some signal processing procedures (quantization, D/A conversion, print/scan, etc) and/or some malicious attacks, and be detectable even after degradation introduced by these manipulations.

A large number of practical systems have been developed to achieve the aforementioned characteristics (see, e.g., [35], [56], and the references therein).

Information hiding has also been studied from information-theoretic perspectives (see, e.g., [7], [33], [47], [48], [51], [73], [89] and the references therein). One perspective is to model information hiding as a constrained channel coding problem [13]. Secret messages, assumed to be uniformly distributed over a given message set, are embedded into host data source messages. Since the watermarks should not interfere perceptually with the host data, a distortion constraint is placed between the encoder output (also called *stegotext*) and the host data. One information hiding problem is to find the largest watermarking rate (known as *watermarking capacity*) for which, at the encoder, the distortion between the host data and the stegotext does not exceed a preset threshold (*transparency* constraint), and at the decoder, watermarks can be reproduced with an arbitrarily small probability of error (*robustness* constraint). The problem is called *private* information hiding if the host data (side information) is available to both the encoder and the decoder [7], [48], [51], [73]. If the side information is available to the encoder only, the problem is called *public* information hiding [7], [74]. Some previous works on traditional source/channel coding with side information ([66], [25], [8], [49]) are useful for posing information hiding problems as instances of constrained channel coding problems. In another interesting work [3], the duality between the information-embedding problem and the Wyner-Ziv problem of source coding with side information is studied.

Information hiding has also been modeled as a game played between the information embedder and the attacker. Given a certain objective function, e.g., embedding capacity ([7], [19], [51], [52], [74], [75]), or error exponent ([48], [73]), the embedder wishes to maximize the objective function, while the attacker's task is to minimize the objective function.

The problem of joint compression of host data and embedding/watermarking of secret messages has also drawn attention in the information hiding literature ([31], [32], [33], [34], [46], [47], [89], [50]). This model applies to scenarios where a compressed version of the stegotext is transmitted due to bandwidth constraints. Denoting the *compression rate* by R_c and the *watermarking rate* by R_w, the main goal is to determine the achievable rate pairs (R_c, R_w) under transparency and robustness constraints on the system.

In this thesis, we are interested in the information-theoretic aspects of information hiding. We first study a private information hiding problem with joint watermarking and compression, where an encoder jointly embeds a secret message to a host data and compresses the host data. This system can be seen as a special JSCC system where the watermarking/embedding of information messages can be seen as a (constrained) channel coding problem, and the compression of the host data is actually source coding. Due to popular applications in network communications, we also investigate information hiding problems in the multi-user setting, where two encoders wish to embed independent messages to two correlated data over multiple access channels. Both the public and private scenarios are studied. We also study a private information hiding problem where two correlated sources are separately embedded into a common host with multiple access attacks.

1.2 Thesis Organization and Contributions

The rest of this thesis is organized as follows.

In Chapter 2, we investigate a hybrid digital-analog (HDA) system for the coding of a discrete-time memoryless Gaussian source over a discrete-time memoryless Gaussian channel under bandwidth compression. Information-theoretic upper bounds on the asymptotically optimal mean squared error distortion of the system are obtained under both matched and mismatched channel conditions. An allocation scheme for distributing the channel input power between the analog and the digital signals is also derived for the mismatched HDA system. A low-complexity and low-delay version of the system is next designed and implemented without the use of error correcting codes. The parameters of the system, which employs vector quantization in conjunction with binary phase-shift keying modulation in its digital part, are optimized via an iterative algorithm. Simulation results show that the proposed HDA system performs within 0.3 dB of the mismatch distortion upper bound. The results of Chapter 2 have appeared in part in [84] and [85].

In Chapter 3, an image communication application demonstrating the effectiveness of HDA coding is presented by combining the proposed bandwidth compression system with the bandwidth expansion system of Skoglund *et al.* [69]. The results of this chapter have appeared in part in [83].

In Chapter 4, we study an information hiding system where the encoder jointly compresses a host data and embeds a secret message. In particular, we study joint watermarking and compression of a memoryless Gaussian source under memoryless additive Gaussian attacks in a private scenario. The achievable region involving the watermarking and the compression rate pairs has been established by Karakos and Papamarcou [33]. We refine the analysis of the watermarking decoding error probability for given achievable rate pairs by deriving a computable random coding lower bound to the error exponent. Numerical examples show that the random coding exponent is positive within almost the entire achievable region given in [33]. Chapter 4 has appeared in part in [86].

Chapter 5 and 6 deal with private information hiding in a multi-user scenario. In Chapter 5, we consider a model where two information hiders independently and separately embed two secret messages W_1 and W_2 (at rates R_w^1 and R_w^2 respectively) into two correlated host sources U_1 and U_2 subject to a pair of tolerable distortion levels

(D_1, D_2). The (compressed) outputs (at rates R_c^1 and R_c^2 respectively) are subjected to attacks modeled via a multiple-access channel (MAC) $W_{Y|X_1X_2}$. The tradeoff between the achievable watermarking rates and the compression rates with respect to the distortion constraints is studied. Given distortion level (D_1, D_2), we derived an inner bound and an outer bound with single-letter characterization for all the achievable rate quadruple $(R_w^1, R_w^2, R_c^1, R_c^2)$.

In Chapter 6, we consider an information hiding model where two correlated sources (S_1, S_2) are separately embedded into a common host data U. A sufficient condition in single-letter form under which (S_1, S_2) can be successfully embedded into U under the MAC $W_{Y|X_1X_2}$ is established. Chapter 6 has appeared in part in [87].

In Chapter 7, we investigate a public multi-user information embedding system in which two secret messages are independently embedded into two correlated host sources and undergo multiple access attacks. The tradeoff between the achievable embedding rates and the average distortions for the two embedders is studied. For given distortion levels, inner and outer bounds for the embedding capacity region for the public two-user information embedding system are obtained with single-letter characterization. The bounds are tightened when the host sources are independent.

Finally, the thesis is summarized in Chapter 8.

1.3 Notation

Throughout, random variables (RV's) are denoted by capital letters, e.g., X, specific values are denoted by lower case letters, e.g., x, and their alphabets are denoted by calligraphic letters, e.g., $\mathcal{X}$. The cardinality of a finite set $\mathcal{X}$ is denoted by $|\mathcal{X}|$. Similarly, random vectors are denoted by capital letters superscripted by their lengths, e.g., X^n, their alphabets are denoted by calligraphic letters superscripted by their lengths, e.g., $\mathcal{X}^n$, and their realizations are denoted by boldface lower case letters, e.g., $\mathbf{x} \triangleq (x_1, x_2, ..., x_n)^T \in \mathcal{X}^n$, where T denotes transposition. For any RV X, $P_X(x)$ denotes the probability that $X = x$. For jointly distributed RV's X and U, $P_{X|U}(x|u)$ denotes the conditional probability of $X = x$ given that $U = u$. The probability of an independent and identically distributed (i.i.d.) sequence $\mathbf{x} \in \mathcal{X}^n$ is given by $P_X^{(n)}(\mathbf{x}) \triangleq \prod_{i=1}^n P_X(x_i)$. Similar notation applies to the joint and conditional distributions. $\mathbb{E}(X)$ denotes the expectation of X. $\mathbb{1}\{\cdot\}$ is the indicator function.

Part I

Hybrid Digital-Analog Source-Channel Coding

Chapter 2

Hybrid Digital-Analog Joint Source-Channel Coding for Gaussian Source-Channel Pairs

This chapter is based on a paper submitted to the *IEEE Transactions on Communications*, May 2007 [85], and a paper presented at the *IEEE 23nd Biennial Symposium on Communications* (23rd QBSC), Queen's University, Kingston, ON, Canada, May-June 2006 [84].

2.1 Introduction

We consider the problem of transmitting a discrete-time analog-valued source over a discrete-time memoryless channel. Due to the often lacking channel information at the transmitter, a robust system is desirable for a wide range of channel conditions. In terms of the used modulation techniques, systems can be generally categorized as analog, digital or hybrid digital-analog (HDA) systems as shown in Fig. 2.1.

Since the publication of Shannon's landmark paper in 1948 [65], digital communi-

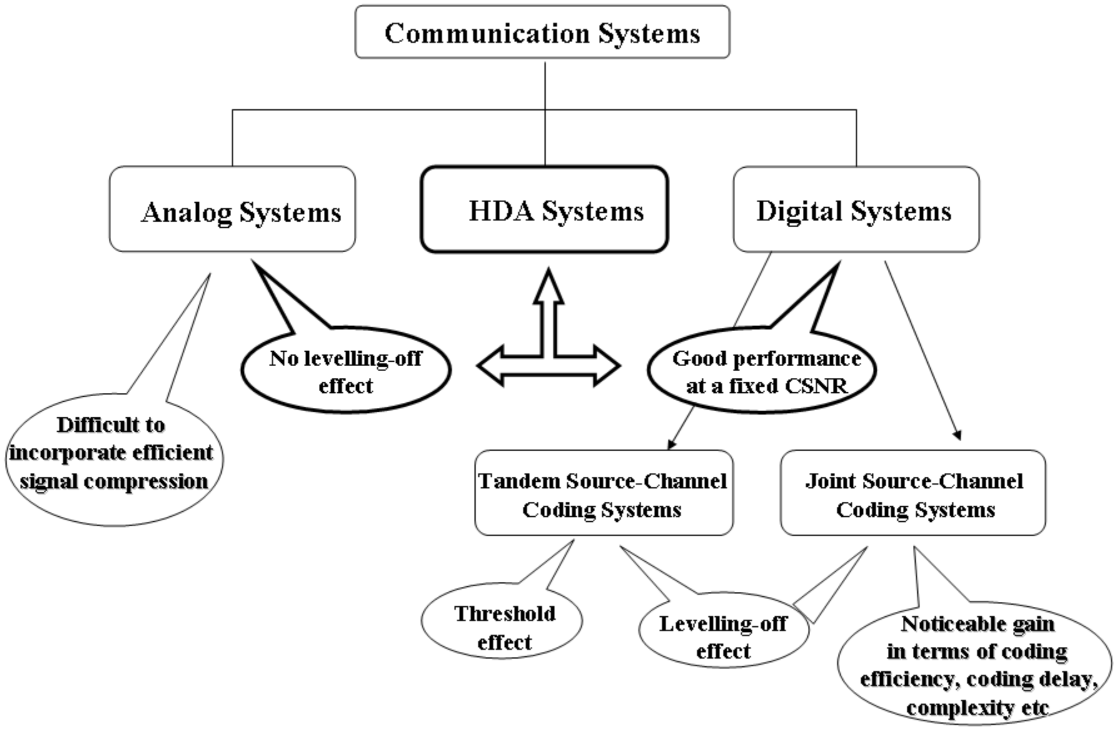

Communication Systems
Analog Systems
HDA Systems
Digital Systems
No levelling-off effect
Difficult to incorporate efficient signal compression
Good performance at a fixed CSNR
Tandem Source-Channel Coding Systems
Joint Source-Channel Coding Systems
Threshold effect
Levelling-off effect
Noticeable gain in terms of coding efficiency, coding delay, complexity etc

Figure 2.1: Characteristics of analog, hybrid, and digital communication systems.

cation has been widely studied. One of the main advantages of digital communication systems is that they can be designed to (asymptotically) achieve the theoretical optimal performance for a fixed channel signal-to-noise ratio (CSNR) via the separate design of optimal source and channel codes [10], [65]. Systems designed based on this principle are often referred to as *tandem source-channel coding* systems. There are, however, two fundamental disadvantages associated with digital tandem systems. One is the *threshold effect*: the system typically performs well at the design CSNR, while its performance degrades drastically when the true CSNR falls below the design CSNR. This effect is due to the quantizer's sensitivity to channel errors and the eventual breakdown of the employed error correcting code at low CSNRs (no matter how powerful it is). The other trait is the *leveling-off effect*: as the CSNR increases, the performance remains constant beyond a certain threshold. This is due to the non-recoverable distortion introduced by the quantizer which limits the system performance at high CSNRs.

The threshold effect can be partly remedied via digital joint source-channel coding (JSCC). By jointly designing the source and channel codes, many results (e.g., [21], [44]) show that noticeable gain can be obtained in terms of coding efficiency, reconstructed signal quality, coding delay and complexity. In particular, JSCC schemes are more robust than tandem systems at low CSNRs. However, such JSCC systems still suffer from the leveling-off effect at high CSNRs, since being digital systems, they employ quantization to "digitize" the source. On the other hand, the leveling-off effect is not a problem for analog systems; actually, their performance can strictly increase as the CSNR increases (we call a system analog if it uses an analog modulation technique such as amplitude modulation). However, it is usually hard to incorporate efficient signal compression schemes in analog systems, particularly when channel bandwidth is valuable and/or the source has memory.

Schemes that exploit the advantage of analog systems are studied by Ramstad and his co-authors in [12], [11], [23], [29] and [43]. These are based on the so-called direct source-channel mapping technique: the output of a source scalar/vector quantizer is mapped directly to a channel symbol using analog (or nearly analog) modulation, i.e., amplitude modulation (AM) or quadrature amplitude modulation (QAM). The direct source-channel codes also enjoy graceful degradation performance at low CSNRs. In [43], a robust image coding system is presented which combines subband coding and QAM. This system allows various compression levels based on block-wise classification. An improved image coding system is proposed in [12]; it utilizes both bandwidth compression and bandwidth expansion mappings, where the bandwidth expansion mapping employs a scalar quantizer and transmits both the quantized value and the quantization error. Recently, a JSCC technique known as the 2:1 Shannon mapping was investigated in [29] and shown to provide very robust performance. It employs the Archimedean spiral to approximately map a point in a plane onto a point on a line. Related works on analog coding methods include [80], [81].

To exploit the advantages of both analog and digital systems, one can allow part of the system to use digital modulation to improve robustness against severe channel conditions, while letting another part of the system use analog signaling to obtain a graceful improvement at high CSNRs. Several recent works have investigated such systems. In [53], a family of HDA systems are introduced and studied theoretically; they are shown to offer better distortion performance than purely digital systems, have a graceful performance improvement, and (asymptotically) achieve the Shannon limit. An HDA system design based on vector quantization (VQ) for bandwidth expansion is investigated in [69], where an algorithm to design optimized codes and performance evaluation are presented. In [70], an HDA system for Gauss-Markov sources with bandwidth compression/expansion is given. It employs the Karhunen-Loève transform to decorrelate the source, Turbo error correcting coding in its digital part to improve the system performance at low CSNRs, and superposition coding of the analog and digital signals. This system allows for both linear and nonlinear mappings in its analog component. In [64], systematic JSCC is studied and is demonstrated to be optimal for a wide class of sources and channels. In [61], an inner distortion bound for broadcasting a single Gaussian source to two listeners over a Gaussian broadcast channel with bandwidth expansion is derived. This bound is obtained based on an HDA coding scheme, which includes one of the HDA systems of [53] and the systematic coding scheme of [64] as two special cases. In [63], systems using an HDA approach, a progressive transmission approach, and a superposition coding approach are compared for a slowly-varying fading additive white Gaussian noise (AWGN) channel. It is shown that the HDA approach has better performance than the other two methods. Most of the gain of this HDA approach is due to the presence of the linear analog part. Other HDA-based techniques are studied in [37], [55] and [58].

In this work, we study the transmission of a memoryless Gaussian source over an AWGN channel with bandwidth compression. We investigate this problem within the HDA coding framework, based on the recent work in [70]. We first obtain an information-theoretical (mean squared) distortion upper bound for the optimal HDA system with a linear analog part. As a direct consequence, we obtain a similar distortion bound for the mismatched HDA system where the encoder does not know the true CSNR. An optimal power allocation formula between the digital and the analog parts is obtained for this mismatched system. A low-complexity and low-delay version of this HDA scheme is next designed and implemented without the use of Turbo error correcting codes (unlike the scheme of [70]) and is shown to be robust over a wide range of CSNRs. These characteristics may be particularly appealing for telemedicine and sensor networks applications where sensitive image data need to be reliably communicated from remote locations irrespective of the channel environment. The digital part of the HDA scheme is formed with a VQ cascaded with a binary phase-shift keying (BPSK) modulated hard-decision decoded AWGN channel. As in [69], the system parameters (in both the digital and analog components) are optimized using an iterative algorithm similar to that for channel-optimized vector quantizer (COVQ) design. Simulations show that this scheme performs within 0.3 dB of the performance bound for the mismatched HDA system for high CSNRs. Comparison are also made with purely analog and purely digital systems, as well as the system in [70]. As an application, an image coding system which combines the bandwidth compression system studied here with the bandwidth expansion system of [69] is presented.

The rest of this chapter is organized as follows. In Section 2.2, a general description of the HDA system is given and information-theoretic bounds on its distortion are derived. A power allocation scheme for distributing the channel input power between the system's analog and digital components is also obtained. In Section 2.3, the HDA system design is examined in detail. Simulation results are given in Section 2.4. Some remarks are given in Section 2.5 for the HDA with non-linear analog and for the Gauss-Markov sources. Finally, conclusions are stated in Section 2.6.

2.2 Information-Theoretic Considerations

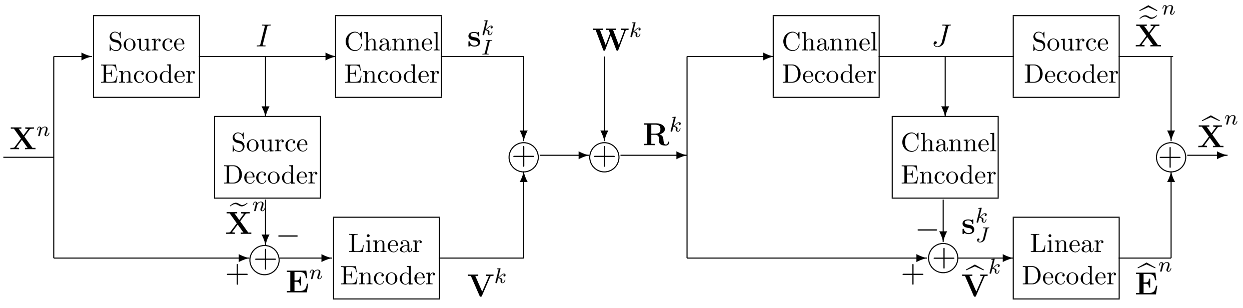


Figure 2.2: HDA coder with bandwidth compression ($k < n$).

The block diagram for the HDA system with bandwidth compression is depicted in Fig.2.2. Samples of a memoryless Gaussian source $\{X_i\}$ with zero mean and variance $\sigma_s^2 > 0$ are grouped into blocks of size n (denoted by $\mathbf{X}^n$) and sent to a source encoder. The discrete output I, which is taken from a finite set of indices, is then fed to a channel encoder/modulator which produces a k-dimensional channel symbol $\mathbf{s}_I^k$, where $k < n$. Here $\mathbf{s}_I^k$ is taken from a finite set of possible symbols and satisfies $\mathbb{E}\|\mathbf{s}_I^k\|^2 \leq k(1-t)P$, where P is the constraint on the total input power per channel use and $t \in [0, 1]$ is the power allocation coefficient for the analog part. The source encoder and the channel encoder/modulator together will often be referred to as tandem source-channel encoder/modulator. The output index I is also sent to a source decoder to form a reconstruction vector $\widetilde{\mathbf{X}}^n$, which is subtracted from $\mathbf{X}^n$ to form an error vector $\mathbf{E}^n$. The first k components of $\mathbf{E}^n$ are further sent to a linear (analog) encoder which performs simple scaling so that the k-dimensional output $\mathbf{V}^k$ satisfies a power constraint $\mathbb{E}\|\mathbf{V}^k\|^2 \leq ktP$. Now $\mathbf{s}_I^k$ and $\mathbf{V}^k$ are superposed and sent over a channel with AWGN $\mathbf{W}^k$ with per symbol noise variance N. The channel output $\mathbf{R}^k$, which is given by $\mathbf{R}^k = \mathbf{s}_I^k + \mathbf{V}^k + \mathbf{W}^k$, is sent to a channel decoder. The discrete output J is sent to the source decoder resulting in vector $\widehat{\widetilde{\mathbf{X}}}^n$. Simultaneously, a channel symbol is chosen according to J, which is subtracted from $\mathbf{R}^k$. The result $\widehat{\mathbf{V}}^k$ is fed to the linear (analog) decoder to form an estimate $\widehat{\mathbf{E}}^k$. The remaining $n-k$ components of the error vector are filled with zeros to produce $\widehat{\mathbf{E}}^n$ which is then added to $\widehat{\widetilde{\mathbf{X}}}^n$ to form an estimate $\widehat{\mathbf{X}}^n$. The overall coding rate of this HDA system is $r = k/n < 1$ channel uses per source sample.

The system normalized mean squared error (MSE) distortion is

$$D_n(N) = \frac{1}{n}\mathbb{E}\left\|\mathbf{X}^n - \widehat{\mathbf{X}}^n\right\|^2. \tag{2.1}$$

For purpose of analysis, we first consider the system's asymptotic distortion, $D(N) = \lim_{n\to\infty} D_n(N)$, as the block length n grows without bound (assuming that the limit exists). The rate-distortion function for the memoryless Gaussian source under the squared-error distortion measure is given by

$$R(D) = \max\left(0, \frac{1}{2}\log_2\frac{\sigma_s^2}{D}\right) \quad \text{(bits/source sample)} \tag{2.2}$$

for any distortion value $D > 0$ [65], [4]. The capacity of the AWGN channel with input power constraint P and noise variance N is given by [10, 65]

$$C(N) = \frac{1}{2}\log_2\left(1 + \frac{P}{N}\right) \quad \text{(bits/channel use).} \tag{2.3}$$

From Shannon's lossy JSCC theorem [10,65] for the memoryless Gaussian source-channel pair, we know that if a code has asymptotic distortion D, then $R(D) \leq rC(N)$ must hold. By letting $R(D) = rC(N)$, a lower bound on the asymptotic distortion of any code can be obtained. This bound is also asymptotically achievable, and is generally referred to as the optimal performance theoretically attainable (OPTA). It is given by

$$D_{opta}(N) \triangleq \frac{\sigma_s^2}{\left(1+\frac{P}{N}\right)^r}. \tag{2.4}$$

By examining the structure of the proposed HDA system in Fig.2.2, we first obtain an upper bound on $D(N)$ for optimally designed HDA systems.

Proposition 2.1 (**Upper bound**) For a memoryless Gaussian source with zero mean and variance σ_s^2 and an AWGN channel with noise variance N, given fixed r, P and t, there exists a sequence of HDA systems with asymptotic distortion $D_{hda}(N)$ given by

$$D_{hda}(N) = r\frac{D_{tan}(N)}{1+\frac{tP}{N}} + (1-r)D_{tan}(N), \tag{2.5}$$

where

$$D_{tan}(N) \triangleq \frac{\sigma_s^2}{\left(1+\frac{(1-t)P}{tP+N}\right)^r}. \tag{2.6}$$

Proof. First we give an informal derivation of the upper bound, and then we provide the outline of a rigorous derivation which uses common randomization at the encoder and the decoder. Some straightforward but tedious details will be omitted. For the source encoder and decoder in the upper "digital" part of the system let $(\varphi_e^{(n)}, \varphi_d^{(n)})$ be a sequence of source codes (vector quantizers) with encoder $\varphi_e^{(n)} : \mathbb{R}^n \to \{1, \ldots, 2^{nR}\}$ and decoder $\varphi_d^{(n)} : \{1, \ldots, 2^{nR}\} \to \mathbb{R}^n$, having rate $R = \frac{r}{2}\log\left(1+\frac{(1-t)P}{tP+N}\right)$ bits per source sample. We choose $(\varphi_e^{(n)}, \varphi_d^{(n)})$ so that it asymptotically achieves the distortion-rate function at rate R of the i.i.d. Gaussian source with zero mean and variance σ_s^2. Thus letting $\widetilde{\mathbf{X}}^n = \varphi_d^{(n)}(\varphi_e^{(n)}(\mathbf{X}^n))$ and $D_n \triangleq \frac{1}{n}\mathbb{E}\|\mathbf{X}^n - \widetilde{\mathbf{X}}^n\|^2$, we have

$$\lim_{n\to\infty} D_n = \sigma_s^2 2^{-2R} = \frac{\sigma_s^2}{\left(1 + \frac{(1-t)P}{tP+N}\right)^r} = D_{tan}(N). \tag{2.7}$$

The output index $I = \varphi_e^{(n)}(\mathbf{X}^n)$ from the source encoder is fed to the channel encoder which operates on blocks of $k = rn$ channel symbols. The sequence of channel codes $(\psi_e^{(k)}, \psi_d^{(k)})$ with encoder $\psi_e^{(k)} : \{1, \ldots, 2^{nR}\} \to \mathbb{R}^k$ and decoder $\psi_d^{(k)} : \mathbb{R}^k \to \{1, \ldots, 2^{nR}\}$ has rate

$$\frac{n}{k}R = \frac{R}{r} = \frac{1}{2}\log\left(1 + \frac{(1-t)P}{tP+N}\right)$$

bits per channel use. This is the capacity of an AWGN channel with noise variance $tP+N$ and input power constraint $(1-t)P$, and we choose the channel code to satisfy this power constraint and such that its error probability is asymptotically (i.e., as $k \to \infty$) zero when it is used on this AWGN channel. Letting $\mathbf{E}^n \triangleq \mathbf{X}^n - \widetilde{\mathbf{X}}^n$, the linear encoder-decoder pair $(\alpha^{(n)}, \beta^{(n)})$ is defined as

$$\mathbf{V}^k \triangleq \alpha^{(n)}(\mathbf{E}^n) = \sqrt{\frac{tP}{D_n}}\,[\mathbf{E}^n]_1^k, \quad \widehat{\mathbf{E}}^n \triangleq \beta^{(n)}(\widehat{\mathbf{V}}^k) = \left(\frac{\sqrt{tPD_n}}{tP+N}(\widehat{\mathbf{V}}^k)^T, (\mathbf{0}^{n-k})^T\right)^T \tag{2.8}$$

where $[\mathbf{E}^n]_1^k$ denotes the first k components of $\mathbf{E}^n$. Since the source code asymptotically achieves the rate-distortion function, one can easily show using a standard information theoretic argument that the normalized relative entropy (Kullback Leibler divergence) [10] between $\mathbf{E}^n$ and an n-dimensional Gaussian random vector with i.i.d. components of zero mean and variance $D_{tan}(N)$ converges to zero as $n \to \infty$. This indicates that the distribution of $\mathbf{E}^n$ is well approximated by that of the Gaussian vector for large n. It is also easy to show that $\mathbf{E}^n$ and $\widetilde{\mathbf{X}}^n$ are (asymptotically) uncorrelated (see, e.g., [53, Lemma 1]). To simplify the informal derivation, let us assume that the following stronger versions of these approximations hold: (i) $\mathbf{E}^n$ is independent of $\widetilde{\mathbf{X}}^n$; (ii) $\mathbf{E}^n$ is Gaussian with independent components of zero mean and equal variance D_n.

Note that since I is a function of $\widetilde{\mathbf{X}}^n$, these assumptions imply that the channel codeword $\mathbf{s}_I^k = \psi_e^{(k)}(I)$ is independent of $\mathbf{V}^k = \sqrt{\frac{tP}{D_n}}[\mathbf{E}^n]_1^k$, and furthermore,

$$\frac{1}{k}\mathbb{E}\|\mathbf{s}_I^k + \mathbf{V}^k\|^2 = \frac{1}{k}\mathbb{E}[\|\mathbf{s}_I^k\|^2] + \frac{1}{k}\mathbb{E}[\|\mathbf{V}^k\|^2] \leq (1-t)P + tP \tag{2.9}$$

so that the total input power constraint P on the channel is met. By assumptions (i) and (ii) the actual channel noise $\mathbf{V}^k + \mathbf{W}^k$ at the channel decoder can be regarded as an AWGN vector with per sample variance $tP + N$ which is independent of the channel encoder input. Under these assumptions the channel code has asymptotically vanishing error probability, i.e.,

$$\lim_{n\to\infty} \Pr\{I \neq J\} = 0. \tag{2.10}$$

It is well known that for the i.i.d. Gaussian source an asymptotically optimal source code can be chosen such that its codevectors lie on a sphere of radius $\sqrt{n(\sigma_s^2 - D_{tan}(N))}$, i.e., we can assume $\frac{1}{n}\|\varphi_d^{(n)}(i)\|^2 = \sigma_s^2 - D_{tan}(N)$ for all i. Using this fact and noting that (2.10) is equivalent to $\lim_{n\to\infty}\Pr\{\widetilde{\mathbf{X}}^n \neq \widehat{\widetilde{\mathbf{X}}}^n\} = 0$, we obtain

$$\lim_{n\to\infty}\frac{1}{n}\mathbb{E}\big\|\widetilde{\mathbf{X}}^n - \widehat{\widetilde{\mathbf{X}}}^n\big\|^2 = 0. \tag{2.11}$$

For simplicity we in fact assume that $\widetilde{\mathbf{X}}^n = \widehat{\widetilde{\mathbf{X}}}^n$ for large n. In this case, the average distortion can be written as

$$\frac{1}{n}\mathbb{E}\big\|\mathbf{X}^n - \widehat{\mathbf{X}}^n\big\|^2 = \frac{1}{n}\mathbb{E}\big\|(\widetilde{\mathbf{X}}^n + \mathbf{E}^n) - (\widehat{\widetilde{\mathbf{X}}}^n + \widehat{\mathbf{E}}^n)\big\|^2 = \frac{1}{n}\mathbb{E}\big\|\mathbf{E}^n - \widehat{\mathbf{E}}^n\big\|^2. \tag{2.12}$$

On the other hand, from (2.8) we have

$$\frac{1}{n}\mathbb{E}\big\|\mathbf{E}^n - \widehat{\mathbf{E}}^n\big\|^2 = \frac{1}{n}\left\|[\mathbf{E}^n]_1^k - \frac{\sqrt{tPD_n}}{tP+N}\widehat{\mathbf{V}}^k\right\|^2 + \frac{1}{n}\big\|[\mathbf{E}^n]_{k+1}^n\big\|^2 \tag{2.13}$$

where $\widehat{\mathbf{V}}^k = \mathbf{V}^k + \mathbf{W}^k + \mathbf{s}_I^k - \mathbf{s}_J^k$. It is well known that the channel codewords can be chosen to lie on a sphere of radius $\sqrt{k(1-t)P}$ (such an equi-energy codebook is often called a Gaussian codebook). Since (2.10) is equivalent to $\lim_{k\to\infty}\Pr\{\mathbf{s}_I^k \neq \mathbf{s}_J^k\} = 0$, we obtain

$$\lim_{k\to\infty}\frac{1}{k}\mathbb{E}\|\mathbf{s}_I^k - \mathbf{s}_J^k\|^2 = 0. \tag{2.14}$$

Again, for simplicity we actually assume $\mathbf{s}_I^k = \mathbf{s}_J^k$, so that $\widehat{\mathbf{V}}^k = \mathbf{V}^k + \mathbf{W}^k$ (for large n). Using (2.8), (2.13), and the assumption in (ii) that the components of $\mathbf{E}^n$ have equal variance D_n, we obtain[1]

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\mathbb{E}\left\|\mathbf{X}^n - \widehat{\mathbf{X}}^n\right\|^2 &= \lim_{n\to\infty}\frac{1}{n}\mathbb{E}\left\|\mathbf{E}^n - \widehat{\mathbf{E}}^n\right\|^2 &(2.15)\\ &= \lim_{n\to\infty}\left(r\frac{D_n}{1+\frac{tP}{N}} + (1-r)D_n\right) &(2.16)\\ &= r\frac{D_{tan}(N)}{1+\frac{tP}{N}} + (1-r)D_{tan}(N) &(2.17)\end{aligned}$$

as desired. The preceding argument in fact forms the basis of a rigorous proof. The crucial point is to prove (2.10), i.e., the existence of a channel code of rate R/r having vanishing error probability which also meets the total power constraint as in (2.9). Indeed, assuming (2.10) holds, we clearly have (2.11) and (2.14). It is then straightforward to show that (2.11) implies (2.15), and that (2.14) implies (2.16) as long as we have

$$\lim_{n\to\infty}\frac{1}{k}\mathbb{E}\left\|[\mathbf{E}_n]_1^k\right\|^2 = D_{tan}(N). \tag{2.18}$$

It is easy to make sure (2.18) holds. Let ℓ be a positive integer which divides n and assume the n-dimensional source code is the n/ℓ-fold product of an ℓ-dimensional vector quantizer $Q^{(\ell)}$ having rate R (i.e., $Q^{(\ell)}$ is used n/ℓ-times when encoding $\mathbf{X}^n$). If $\ell \to \infty$, then the rate-distortion performance (2.7) can be achieved by $Q^{(\ell)}$, and if in addition we have $\ell/n \to 0$, then (2.18) clearly holds.

Thus the entire proof hinges on the existence of channel codes with asymptotically vanishing error probability (2.10) under the power constraint P. In the remainder of the

[1] With these assumptions, $\frac{\sqrt{tPD_n}}{tP+N}\widehat{\mathbf{V}}^k$ becomes the minimum MSE (MMSE) estimate of $[\mathbf{E}^n]_1^k$

proof we show that such codes exist if one allows common randomization at the encoder and decoder. Common randomization, already used in the context of both source and channel coding (see, e.g., [91], [17], [18] and [6]), ensures that the total input power meets the power constraint and also makes the transmitted channel codeword and the "noise" $\mathbf{V}^k + \mathbf{W}^k$ independent. In what follows we first show that the average channel noise $\frac{1}{k}\|\mathbf{V}^k + \mathbf{W}^k\|$ is concentrated near its expectation $tP + N$ with large probability, and then use this fact in showing that the desired channel code exists.

Recall that $D_{tan}(N) = \sigma_s^2 2^{-2R}$ is the distortion-rate function at rate R of a memoryless Gaussian source with variance σ_s^2. It is known (see., e.g., [62] or [42]) that one can choose $Q^{(\ell)}$ so that its codevectors lie on a sphere of radius $\sqrt{\ell(\sigma_s^2 - D_{tan}(N))}$ and it has asymptotically optimal distortion $\lim_{\ell\to\infty} \frac{1}{\ell}\mathbb{E}\|\mathbf{X}^\ell - Q^{(\ell)}(\mathbf{X}^\ell)\|^2 = D_{tan}(N)$, which implies (2.7) since $D_n = \frac{1}{\ell}\mathbb{E}|\mathbf{X}^\ell - Q^{(\ell)}(\mathbf{X}^\ell)\|^2$ by the source code construction.

Since $[\mathbf{E}^n]_1^k$ is the concatenation of $m' = k/\ell$ independent copies of $\mathbf{X}^\ell - Q^{(\ell)}(\mathbf{X}^\ell)$, and $\mathbf{V}^k = \sqrt{\frac{tP}{D_n}}\,[\mathbf{E}^n]_1^k$, we have that $\|\mathbf{V}^k\|^2$ is the sum of $m' = k/\ell$ independent random variables with mean $\frac{tP}{D_n}\mathbb{E}\|\mathbf{X}^\ell - Q^{(\ell)}(\mathbf{X}^\ell)\|^2 = \ell tP$. Thus if ℓ is *fixed*, the weak law of large numbers implies

$$\lim_{k\to\infty} \Pr\left\{\left|\frac{1}{k}\|\mathbf{V}^k\|^2 - tP\right| > \epsilon\right\} = 0 \tag{2.19}$$

for all $\epsilon > 0$. Clearly, we can choose an ℓ sequence such that $\ell \to \infty$, $\ell/k = \ell/(rn) \to 0$ and (2.19) still holds. For the rest of the proof we assume that ℓ increases with n (and k) in this fashion. We have $\frac{1}{k}\|\mathbf{V}^k + \mathbf{W}^k\|^2 = \frac{1}{k}\|\mathbf{V}^k\|^2 + \frac{1}{k}\|\mathbf{W}^k\|^2 + \frac{2}{k}(\mathbf{V}^k)^T\mathbf{W}^k$, where $\frac{1}{k}\|\mathbf{W}^k\|^2$, being the average of k i.i.d. random variables of mean N, converges to N in probability as $k \to \infty$. A direct calculation shows that $\mathbb{E}\left[\left(\frac{1}{k}(\mathbf{V}^k)^T\mathbf{W}^k\right)^2\right] = \frac{N}{k^2}\mathbb{E}\|\mathbf{V}^k\|^2 = \frac{N}{k}tP$, which converges to zero as $k \to \infty$, implying through Chebyshev's inequality that $\Pr\left\{\left|\frac{2}{k}(\mathbf{V}^k)^T\mathbf{W}^k\right| > \epsilon\right\} \to 0$ as $k \to \infty$ for all $\epsilon > 0$. Combining these facts with (2.19) we obtain that for all $\epsilon > 0$,

$$\lim_{k\to\infty} \Pr\left\{ \left| \frac{1}{k}\|\mathbf{V}^k + \mathbf{W}^k\|^2 - (tP+N) \right| > \epsilon \right\} = 0. \tag{2.20}$$

Now consider the fictitious k-dimensional vector channel with input power constraint $k(1-t)P$ and additive noise which is *independent* of the input and has the same distribution as $\mathbf{V}^k + \mathbf{W}^k$. The key point is that (2.20) allows us to use Theorem 1 in [41] which, when applied to our setup, states that given an additive noise channel with power constraint $k(1-t)P$ and input-independent, possibly non-ergodic noise which satisfies (2.20), there exists a sequence of channel codes $(\psi_e^{(k)}, \psi_d^{(k)})$ which has rate $\frac{1}{2}\log\Bigl(1 + \frac{(1-t)P}{tP+N}\Bigr)$ and equi-energy (Gaussian) codebook and whose error probability on this channel approaches zero as $k \to \infty$. (Thus, in effect, a channel code designed for the worst case AWGN noise also works for non-Gaussian channel noise of equal power.)

We will use common randomization to apply $(\psi_e^{(k)}, \psi_d^{(k)})$ to the real system where $\mathbf{V}^k + \mathbf{W}^k$ is not independent of the channel input. Let Π denote a random permutation of the indices $1, \ldots, 2^{nR}$ which is uniformly drawn from the set of all $(2^{nR})!$ permutations and is independent of the source $\mathbf{X}^n$ and the channel noise $\mathbf{W}^k$. Assume that Π is know at both the encoder and the decoder. At the encoder apply Π to the output index I of the source encoder before channel coding, so that the input to the channel encoder is $\Pi(I)$. At the decoder side, if J is the output index at the channel decoder, then $\Pi^{-1}(J)$ is sent to the source decoder, where Π^{-1} denotes the inverse of Π. It is easy to see that the channel with input I and output $\Pi^{-1}(J)$ is statistically equivalent to the discrete channel realized when $(\psi_e^{(k)}, \psi_d^{(k)})$ is used on the fictitious channel with a uniform distribution on its input index set. Since $(\psi_e^{(k)}, \psi_d^{(k)})$ has asymptotically vanishing error probability on the fictitious channel, for the real system we also have $\lim_{k\to\infty} \Pr\{I \neq \Pi^{-1}(J)\} = 0$. It remains to show that the total power input power on the channel does not exceed P.

Since $\mathbf{s}^k_{\Pi(I)} = \psi_e^{(k)}(\Pi(I))$ is independent of $\mathbf{V}^k$,

$$\frac{1}{k}\mathbb{E}\|\mathbf{s}^k_{\Pi(I)} + \mathbf{V}^k\|^2 = \frac{1}{k}\mathbb{E}\|\mathbf{s}^k_{\Pi(I)}\|^2 + \frac{1}{k}\mathbb{E}\|\mathbf{V}^k\|^2 + \frac{2}{k}\mathbb{E}[\mathbf{s}^k_{\Pi(I)}]^T\mathbb{E}[\mathbf{V}^k] \tag{2.21}$$

where $\frac{1}{k}\mathbb{E}\|\mathbf{s}^k_{\Pi(I)}\|^2 = (1-t)P$ and $\frac{1}{k}\mathbb{E}\|\mathbf{V}^k\|^2 = tP$. Let $\mathbf{m}^\ell \triangleq \mathbb{E}[\mathbf{X}^\ell - Q^{(\ell)}(\mathbf{X}^\ell)]$. Then

$$D_n = \frac{1}{n}\mathbb{E}\|\mathbf{X}^\ell - Q^{(\ell)}(\mathbf{X}^\ell)\|^2 = \frac{1}{\ell}\mathbb{E}\|\mathbf{X}^\ell - Q^{(\ell)}(\mathbf{X}^\ell) - \mathbf{m}^\ell\|^2 + \frac{1}{\ell}\|\mathbf{m}^\ell\|^2 \geq D_{tan}(N) + \frac{1}{\ell}\|\mathbf{m}^\ell\|^2$$

where the inequality holds since $Q^{(\ell)}(\mathbf{X}^\ell) + \mathbf{m}^\ell$ is a rate R quantizer for $\mathbf{X}^\ell$. This implies $\lim_{\ell\to\infty}\frac{1}{\ell}\|\mathbf{m}^\ell\|^2 = 0$. Since $\frac{1}{\ell}\|\mathbf{m}^\ell\|^2\frac{tP}{D_n} = \frac{1}{k}\|\mathbb{E}[\mathbf{V}^k]\|^2$, applying Cauchy-Schwarz inequality yields $\lim_{k\to\infty}\frac{1}{k}\mathbb{E}[\mathbf{s}^k_{\Pi(I)}]^T\mathbb{E}[\mathbf{V}^k] = 0$. Substituting this into (2.21) shows that $\lim_{k\to\infty}\frac{1}{k}\mathbb{E}\|\mathbf{s}^k_{\Pi(I)} + \mathbf{V}^k\|^2 = (1-t)P + tP$; thus, the power constraint is (asymptotically) satisfied. □

Remark: It is easy to show that $D_{hda}(N) = D_{opta}(N)$ if and only if $t = 0$.

We next study the realistic situation where the AWGN variance N is not known at the encoder. We assume that the encoder only knows a range of values in which the true noise variance N_{tr} lies; in particular, it chooses the encoding operation for a fixed design noise variance N_{des}. The receiver, on the other hand, has full knowledge of N_{tr} and adapts the decoding accordingly. For this mismatched HDA system, when the true noise variance N_{tr} satisfies $N_{tr} < N_{des}$, the linear decoder can adapt to N_{tr}, resulting in a distortion given by $\frac{D_{tan}(N_{des})}{1+\frac{tP}{N_{tr}}}$. The asymptotic performance of the tandem coder part is still the same. We then obtain the following upper bound on the distortion:

$$D^{mis}_{hda}(N_{tr}, N_{des}) \triangleq r\frac{D_{tan}(N_{des})}{1+\frac{tP}{N_{tr}}} + (1-r)D_{tan}(N_{des}) \tag{2.22}$$

where $D_{tan}(N)$ is given in (2.6).

We now consider the power allocation problem for this mismatched HDA system with the encoder designed for N_{des}, while the true noise variance is N_{tr}. The best power allocation coefficient t that minimizes (2.22) is given by the following proposition.

Proposition 2.2 For $N_{tr} < N_{des}$, P and r, the power allocation coefficient t which minimizes the distortion expression (2.22) at N_{tr} is given by

$$t = \frac{\sqrt{1 + \frac{4(\kappa_{tr} - \kappa_{des})}{(1-r)\kappa_{des}}} - 1}{2\kappa_{tr}}, \tag{2.23}$$

where $\kappa_{tr} = \frac{P}{N_{tr}}$ is the true CSNR and $\kappa_{des} = \frac{P}{N_{des}}$ is the design CSNR.

Proof. Define $\kappa_{tr} = \frac{P}{N_{tr}}$ and $\kappa_{des} = \frac{P}{N_{des}}$, the distortion (2.22) can be rewritten as

$$D_{hda}^{mis}(N_{tr}, N_{des}) = \left[\left(\frac{r}{1 + t\kappa_{tr}} + 1 - r\right)(1 + t\kappa_{des})^r\right] \frac{\sigma_s^2}{(1 + \kappa_{des})^r}. \tag{2.24}$$

Since

$$\begin{aligned} &\frac{d}{dt}\big(D_{hda}^{mis}(N_{tr}, N_{des})\big) \\ = &\left[\frac{(1 + (1-r)t\kappa_{tr})\kappa_{des}}{1 + t\kappa_{des}} - \frac{\kappa_{tr}}{1 + t\kappa_{tr}}\right] \frac{r(1 + t\kappa_{des})^r}{1 + t\kappa_{tr}} \frac{\sigma_s^2}{(1 + \kappa_{des})^r}, \end{aligned} \tag{2.25}$$

setting

$$\frac{d}{dt}\big(D_{hda}^{mis}(N_{tr}, N_{des})\big) = 0 \tag{2.26}$$

yields

$$\frac{(1 + (1-r)t\kappa_{tr})\kappa_{des}}{1 + t\kappa_{des}} - \frac{\kappa_{tr}}{1 + t\kappa_{tr}} = 0, \tag{2.27}$$

or equivalently

$$(1-r)\kappa_{tr}^2\kappa_{des}t^2 + (1-r)\kappa_{tr}\kappa_{des}t + \kappa_{des} - \kappa_{tr} = 0. \tag{2.28}$$

Therefore, we get one stationary point (the negative solution is discarded)

$$t = \frac{\sqrt{1 + \frac{4(\kappa_{tr} - \kappa_{des})}{(1-r)\kappa_{des}}} - 1}{2\kappa_{tr}}. \tag{2.29}$$

As a matter of fact, the above t minimizes $D^{mis}_{hda}(N_{tr}, N_{des})$ since

$$\begin{aligned}
&\frac{d^2}{dt^2}\big(D^{mis}_{hda}(N_{tr}, N_{des})\big) && (2.30)\\
= \;& [2\kappa_{tr}(1+t\kappa_{des}) - (1+r)(1+t\kappa_{tr})\kappa_{des}]\,\frac{(1+t\kappa_{des})^{r-1}}{(1+t\kappa_{tr})^3}\frac{\sigma_s^2}{(1+\kappa_{des})^r}r\kappa_{tr} && (2.31)\\
> \;& [2\kappa_{tr}(1+t\kappa_{des}) - 2(1+t\kappa_{tr})\kappa_{des}]\,\frac{(1+t\kappa_{des})^{r-1}}{(1+t\kappa_{tr})^3}\frac{\sigma_s^2}{(1+\kappa_{des})^r}r\kappa_{tr} && (2.32)\\
= \;& 2(\kappa_{tr} - \kappa_{des})\frac{(1+t\kappa_{des})^{r-1}}{(1+t\kappa_{tr})^3}\frac{\sigma_s^2}{(1+\kappa_{des})^r}r\kappa_{tr} && (2.33)\\
> \;& 0, && (2.34)
\end{aligned}$$

where we have used the fact $r < 1$ in (2.32) and $\kappa_{tr} > \kappa_{des}$ in (2.34). □

Since the optimal t is a function of N_{tr}, it is also unavailable at the encoder. However, via a numerical study (see below) one can choose a value of t which performs well for a large range of CSNRs κ_{tr}. In Fig. 2.3, we plot the optimal t for different system parameters as a function of the true CSNR κ_{tr}. We observe the following.

- It is readily seen that as the true CSNR κ_{tr} increases, t approaches 0. Furthermore, it is also easily seen from (2.23) that the rate of decay of t to 0 is less than that of $1/\kappa_{tr}$. It is easy to see that as $\kappa_{tr} \to \infty$, the distortion performance of the mismatched HDA system (2.22) approaches the constant $(1-r)D_{tan}(N_{des})$. Curves (a), (f), (b) and (c) present the best power allocation for an HDA system of rate 0.5, with design CSNR κ_{des} of 0 dB, 5 dB, 10 dB and 15 dB, respectively. They indicate that, for a system with high design CSNR (which is the case when performance at high CSNRs is the main concern), the best power allocation coefficient at various CSNR pairs $(\kappa_{des}, \kappa_{tr})$ is smaller than that for the low design CSNR case, i.e., the analog part of the HDA system incrementally turns off as κ_{des} increases without bound.

- As κ_{tr} approaches κ_{des}, t approaches 0. Thus the optimal performance at the design

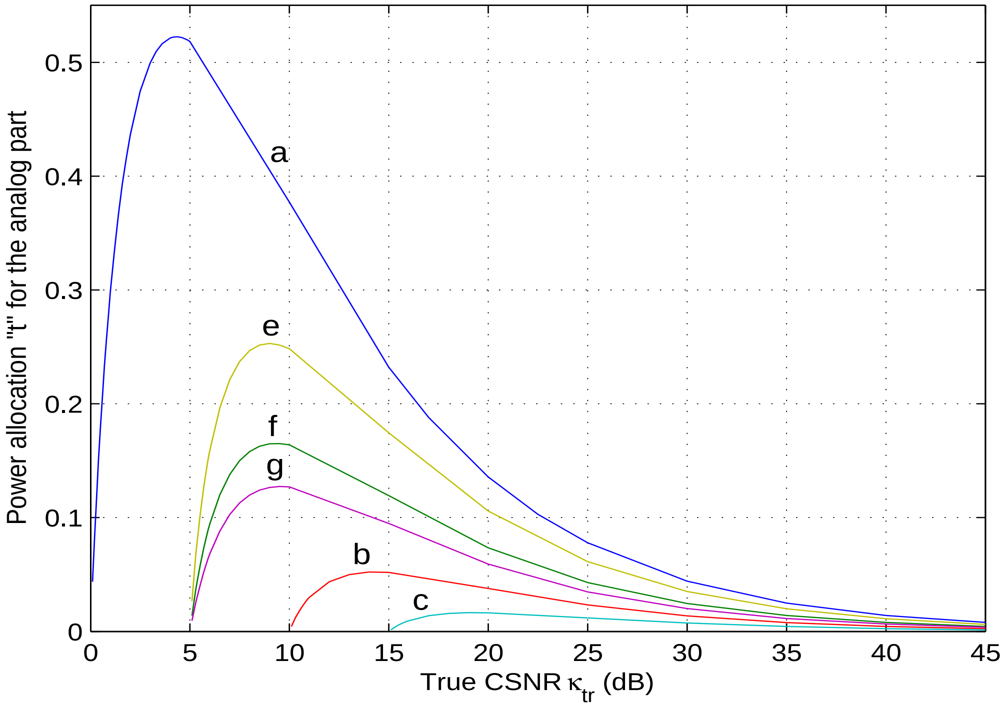


Figure 2.3: The best power allocation t (as a function of the true CSNR κ_{tr}) for different system parameters. For curves (a), (b) and (c), $r = 0.5$, $\kappa_{des} = 0$ dB, 10 dB and 15 dB, respectively. For curves (e), (f) and (g), $\kappa_{des} = 5$ dB, $r = 0.75$, 0.5 and 0.25 respectively.

CSNR is obtained by a "purely digital" design, or equivalently, by an optimal tandem coder which contains an optimal source code and an optimal channel code, as predicted by Shannon's theory [65].

- Curves (e), (f) and (g) show the best t for $\kappa_{des} = 5$ dB and coding rate of 0.75, 0.5 and 0.25, respectively. These curves demonstrate that t decreases as the coding rate r decreases. Indeed, as r decreases, less components of quantization error vectors are further coded via the analog part, which reduces the importance of the analog part relative to that of the tandem coding part.

In our system implementations, we fix a design CSNR κ_{des} and choose an adjusted value of t which is good over a large range of true CSNRs κ_{tr} ($> \kappa_{des}$); see Section 2.4 for details.

2.3 HDA System Design

We next consider a concrete implementation of the HDA scheme in Fig. 2.2. This system, which has low-complexity and low-delay as it avoids the use of channel coding in its digital part, is depicted in Fig. 2.4, and it employs VQ cascaded with BPSK modulation in the digital part, and uses linear coding in the analog part.

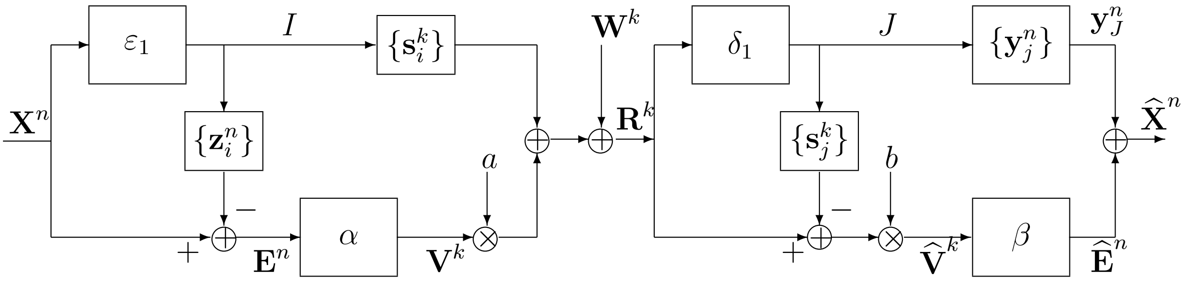


Figure 2.4: Proposed HDA system design with bandwidth compression.

2.3.1 System Description

The upper part, referred to as the digital part, is formed by a VQ cascaded with a binary symmetric channel (BSC) without the use of channel coding. An output index I of the k-bit n-dimensional VQ encoder ε_1 is assigned a k-dimensional channel symbol $\mathbf{s}_I^k$ from a set $\{\mathbf{s}_i^k\}$ of 2^k possible symbols. The index I also chooses a vector $\mathbf{z}_I^n$ from the *encoder codebook* $\{\mathbf{z}_i^n\}$, which is subtracted from $\mathbf{X}^n$ to form the error vector $\mathbf{E}^n$.

In the ideal case, for a memoryless source, the optimal source code (in the sense of asymptotically achieving the rate-distortion curve) splits source vectors into two asymptotically orthogonal components, the quantizer output and the quantization error (see, e.g., [53]). Furthermore, for memoryless Gaussian sources, the distribution of the quantization error is also approximately Gaussian as $n \to \infty$ (see the proof of Proposition 2.1). In the HDA system with linear analog coding, since the output of the linear analog encoder is just a scaled version of the quantization error, we model (as discussed in the proof of Proposition 2.1) the output of the linear encoder by a Gaussian random variable with variance tP which is independent of the source. Hence, for the digital part, a BSC is realized by using hard decision decoding on the BPSK-modulated AWGN channel with input power $(1-t)P$ and noise variance $tP + N_{des}$. Consequently, if the BPSK signals take values in $\{+\sqrt{(1-t)P}, -\sqrt{(1-t)P}\}$, the transition probabilities $\{P_{J|I}(j|i)\}$ of the BSC are $P_{J|I}(j|i) = q^{d_H(i,j)}(1-q)^{k-d_H(i,j)}$, where $d_H(i,j)$ denotes the Hamming distance between the binary representations of the integers i and j, and $q = Q(\sqrt{\kappa_{dig}})$ is the crossover probability, where $\kappa_{dig} \triangleq \frac{(1-t)\kappa_{des}}{t\kappa_{des}+1}$ is the effective CSNR of the digital part and $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} dt$. We remark that any memoryless modulation constellation can be used besides BPSK modulation. We choose BPSK modulation because it is simple and it performs comparatively well at low CSNRs.

Given an input error vector $\mathbf{E}^n$, the mapping α simply takes the first k components of $\mathbf{E}^n$ and forms a scaled vector $\mathbf{V}^k$ (to satisfy the average power constraint), which is added to $\mathbf{s}_I^k$ and sent over the AWGN channel. The received vector $\mathbf{R}^k$ is first fed to decoder δ_1 (which is a simple binary hard-decision demodulator), resulting in index J, and the corresponding reproduction $\mathbf{y}_J^n$ is chosen through a lookup table. The channel symbol $\mathbf{s}_J^k$ is then subtracted from $\mathbf{R}^k$ and scaled by a constant b, forming an estimate $\widehat{\mathbf{V}}^k$. The mapping β expands the message $\widehat{\mathbf{V}}^k$ back to n dimensions, by padding it with zeros in the corresponding locations. The resulting $\widehat{\mathbf{E}}^n$ is added back to $\mathbf{y}_J^n$ to form the reproduction $\widehat{\mathbf{X}}^n$.

2.3.2 System Design

For a total input power P, a fixed power allocation t and a design noise variance N_{des}, we derive an iterative training algorithm to optimize the source digital transmitter (both source encoder and source decoder) and both the digital decoder codebook and the analog decoder. Given an arbitrary encoder $\varepsilon_1, \{\mathbf{z}_i^n\}, \{\mathbf{s}_i^n\}, \{\mathbf{y}_j^n\}$, and a and b, the end-to-end average distortion can be expressed as

$$\begin{aligned}
&D_n(N_{des})\\
&= \frac{1}{n}\mathbb{E}\|\mathbf{X}^n - \widehat{\mathbf{X}}^n\|^2\\
&= \frac{1}{n}\mathbb{E}\left\|\begin{pmatrix} [\mathbf{X}^n]_1^k \\ [\mathbf{X}^n]_{k+1}^n \end{pmatrix} - \begin{pmatrix} [\mathbf{y}_J^n]_1^k \\ [\mathbf{y}_J^n]_{k+1}^n \end{pmatrix} - \begin{pmatrix} b\left(a([\mathbf{X}^n]_1^k - [\mathbf{z}_I^n]_1^k) + \mathbf{s}_I^k + \mathbf{W}^k - \mathbf{s}_J^k\right) \\ 0 \end{pmatrix}\right\|^2\\
&= \frac{1}{n}\underbrace{\mathbb{E}\left\|[\mathbf{X}^n]_1^k - [\mathbf{y}_J^n]_1^k - b\left(a([\mathbf{X}^n]_1^k - [\mathbf{z}_I^n]_1^k) + \mathbf{s}_I^k + \mathbf{W}^k - \mathbf{s}_J^k\right)\right\|^2}_{\triangleq D_n^1(N_{des})} + \frac{1}{n}\underbrace{\mathbb{E}\left\|[\mathbf{X}^n]_{k+1}^n - [\mathbf{y}_J^n]_{k+1}^n\right\|^2}_{\triangleq D_n^2(N_{des})}.
\end{aligned} \tag{2.35}$$

Assume that a is chosen such that the power constraint

$$a^2\mathbb{E}\left\|[\mathbf{X}^n]_1^k - [\mathbf{z}_I^n]_1^k\right\|^2 = ktP \tag{2.36}$$

is satisfied. Then

$$\begin{aligned}
D_n^1(N_{des}) &= \frac{1}{n}\mathbb{E}\left\|[\mathbf{X}^n]_1^k - [\mathbf{y}_J^n]_1^k - b\left(a([\mathbf{X}^n]_1^k - [\mathbf{z}_I^n]_1^k) + \mathbf{s}_I^k + \mathbf{W}^k - \mathbf{s}_J^k\right)\right\|^2\\
&= \frac{1}{n}\mathbb{E}\left\|[\mathbf{X}^n]_1^k - [\mathbf{y}_J^n]_1^k - b(\mathbf{s}_I^k - \mathbf{s}_J^k) - ba([\mathbf{X}^n]_1^k - [\mathbf{z}_I^n]_1^k)\right\|^2 + \frac{1}{n}b^2\mathbb{E}\|\mathbf{W}^k\|^2 \qquad (2.37)\\
&= \frac{1}{n}\mathbb{E}\left\|[\mathbf{X}^n]_1^k - [\mathbf{y}_J^n]_1^k - b(\mathbf{s}_I^k - \mathbf{s}_J^k)\right\|^2 + \frac{1}{n}b^2a^2\mathbb{E}\left\|[\mathbf{X}^n]_1^k - [\mathbf{z}_I^n]_1^k\right\|^2\\
&\quad -2ab\frac{1}{n}\mathbb{E}\left[\left([\mathbf{X}^n]_1^k - [\mathbf{y}_J^n]_1^k - b(\mathbf{s}_I^k - \mathbf{s}_J^k)\right)^T\left([\mathbf{X}^n]_1^k - [\mathbf{z}_I^n]_1^k\right)\right] + \frac{k}{n}b^2N_{des} \qquad (2.38)\\
&= \frac{1}{n}\mathbb{E}\left\|[\mathbf{X}^n]_1^k - [\mathbf{y}_J^n]_1^k - b(\mathbf{s}_I^k - \mathbf{s}_J^k)\right\|^2 + \frac{k}{n}b^2tP\\
&\quad -2ab\frac{1}{n}\mathbb{E}\left[\left([\mathbf{X}^n]_1^k - [\mathbf{y}_J^n]_1^k - b(\mathbf{s}_I^k - \mathbf{s}_J^k)\right)^T\left([\mathbf{X}^n]_1^k - [\mathbf{z}_I^n]_1^k\right)\right] + \frac{k}{n}b^2N_{des}. \qquad (2.39)
\end{aligned}$$

Lemma 2.1 Fix a set of encoder regions $\{Q_i\}$ of ϵ_1. For any digital decoder codebook $\{\mathbf{y}_j^n\}$ and b, the digital source decoder codebook $\{[\mathbf{z}_i^n]_1^k\}$ that minimizes the average distortion (2.35) is given by

$$[\mathbf{z}_i^n]_1^k = [\bar{\mathbf{y}}_i^n]_1^k + b(\mathbf{s}_i^k - \bar{\mathbf{s}}_i^k), \quad i = 0, \cdots, 2^k - 1. \tag{2.40}$$

For any $\{[\mathbf{z}_i^n]_1^k\}$, the average distortion (2.35) is minimized by choosing b and $\{\mathbf{y}_j^n\}$ as follows:

$$b = \frac{\mathbb{E}\left[\left([\mathbf{X}^n]_1^k - \mathbb{E}\left[[\mathbf{X}^n]_1^k \mid J\right]\right)^T \mathbf{U}^k\right]}{kN_{des} + \mathbb{E}||\mathbf{U}^k||^2}, \tag{2.41}$$

$$[\mathbf{y}_j^n]_1^k = \sum_{i=0}^{2^k-1} P_{I|J}(i|j)\left([\bar{\mathbf{x}}_i^n]_1^k - ba([\bar{\mathbf{x}}_i^n]_1^k - [\mathbf{z}_i^n]_1^k)\right) - b\left(\sum_{i=0}^{2^k-1} P_{I|J}(i|j)\mathbf{s}_i^k - \mathbf{s}_j^k\right), \tag{2.42}$$

$$[\mathbf{y}_j^n]_{k+1}^n = \sum_{i=0}^{2^k-1} P_{I|J}(i|j)[\bar{\mathbf{x}}_i^n]_{k+1}^n, \; j = 0, \cdots, 2^k - 1, \tag{2.43}$$

where

$$\mathbf{U}^k \triangleq a\left([\mathbf{X}^n]_1^k - \mathbb{E}\big[[\mathbf{X}^n]_1^k \mid J\big] - [\mathbf{z}_I^n]_1^k + \mathbb{E}\big[[\mathbf{z}_I^n]_1^k \mid J\big]\right) + \mathbf{s}_I^k - \mathbf{s}_J^k - \mathbb{E}\big[\mathbf{s}_I^k - \mathbf{s}_J^k \mid J\big], \tag{2.44}$$

$$\bar{\mathbf{x}}_i^n \triangleq \mathbb{E}\left[\mathbf{X}^n \mid I = i\right] = \int\limits_{\mathbf{x}^n \in Q_i} \mathbf{x}^n p(\mathbf{x}^n) d\mathbf{x}^n, \tag{2.45}$$

$$\bar{\mathbf{y}}_i^n \triangleq \mathbb{E}\left[\mathbf{y}_J^n \mid I = i\right] = \sum_{j=1}^{2^k-1} P_{J|I}(j|i)\mathbf{y}_j^n, \quad \bar{\mathbf{s}}_i^k \triangleq \mathbb{E}\left[\mathbf{s}_J^k \mid I = i\right] = \sum_{j=1}^{2^k-1} P_{J|I}(j|i)\mathbf{s}_j^k, \tag{2.46}$$

$$P_{I|J}(i|j) \triangleq \Pr(I = i|J = j) = P_{J|I}(j|i)P_I(i)/P_J(j), \tag{2.47}$$

$$P_I(i) \triangleq \Pr(I = i) = \Pr(\mathbf{X}^n \in Q_i), \quad P_J(j) \triangleq \Pr(J = j) = \sum_{i=1}^{2^k-1} P_I(i)P_{J|I}(j|i), \tag{2.48}$$

and $p(\mathbf{x}^n)$ is the pdf of $\mathbf{x}^n$.

Proof. We first focus on how the digital source decoder codebooks $\{[\mathbf{z}_i^n]_1^k\}$ should be chosen to minimize the distortion $D_n(N_{des})$ (note that the $\{[\mathbf{z}_i^n]_{k+1}^n\}$ are not needed since we only transmit the first k error components). We note that the only term in (2.39) that can be influenced by changing $\{[\mathbf{z}_i^n]_1^k\}$ is the third one. We have

$$\begin{aligned}
&\mathbb{E}\left[\left([\mathbf{X}^n]_1^k - [\mathbf{y}_J^n]_1^k - b(\mathbf{s}_I^k - \mathbf{s}_J^k)\right)^T \left([\mathbf{X}^n]_1^k - [\mathbf{z}_I^n]_1^k\right)\right] \\
&= \mathbb{E}\big\|[\mathbf{X}^n]_1^k\big\|^2 - \sum_{i=0}^{2^k-1} P_I(i) \left(\left([\bar{\mathbf{y}}_i^n]_1^k + b(\mathbf{s}_i^k - \bar{\mathbf{s}}_i^k)\right)^T \left([\bar{\mathbf{x}}_i^n]_1^k - [\mathbf{z}_i^n]_1^k\right) - [\bar{\mathbf{x}}_i^n]_1^{k^T} [\mathbf{z}_i^n]_1^k\right) \\
&= \mathbb{E}\left[\left([\mathbf{X}^n]_1^k - [\bar{\mathbf{y}}_I^n]_1^k - b(\mathbf{s}_I^k - \bar{\mathbf{s}}_I^k)\right)^T \left([\mathbf{X}^n]_1^k - [\mathbf{z}_I^n]_1^k\right)\right] \\
&\leq \sqrt{\mathbb{E}\big\|[\mathbf{X}^n]_1^k - [\bar{\mathbf{y}}_I^n]_1^k - b(\mathbf{s}_I^k - \bar{\mathbf{s}}_I^k)\big\|^2 \mathbb{E}\big\|[\mathbf{X}^n]_1^k - [\mathbf{z}_I^n]_1^k\big\|^2} \qquad (2.49)
\end{aligned}$$

where (2.49) holds by the Cauchy-Schwarz inequality. For arbitrary given $\{\mathbf{y}_j^n\}$ and b, equality holds when we choose $\{[\mathbf{z}_i^n]_1^k\}$ as in (2.40), thus minimizing the distortion $D_n^1(N_{des})$. Next, consider how the digital decoder codebook $\{\mathbf{y}_j^n\}$ should be chosen to minimize the average distortion $D_n(N_{des})$ in (2.35). Recall that

$$D_n^1(N_{des}) = \frac{1}{n}\mathbb{E}\big\|\left([\mathbf{X}^n]_1^k - ba([\mathbf{X}^n]_1^k - [\mathbf{z}_I^n]_1^k) - b(\mathbf{s}_I^k - \mathbf{s}_J^k + \mathbf{W}^k)\right) - [\mathbf{y}_J^n]_1^k\big\|^2, \qquad (2.50)$$

$$D_n^2(N_{des}) = \frac{1}{n}\mathbb{E}\big\|[\mathbf{X}^n]_{k+1}^n - [\mathbf{y}_J^n]_{k+1}^n\big\|^2. \qquad (2.51)$$

Thus, for arbitrary $\{[\mathbf{z}_i^n]_1^k\}$ and b, the $\{\mathbf{y}_j^n\}$ which minimize the average distortion (2.35) are obtained by letting $\{\mathbf{y}_j^n\}$ represent the MMSE estimator

$$\begin{aligned}
[\mathbf{y}_j^n]_1^k &= \mathbb{E}\left[[\mathbf{X}^n]_1^k - ba([\mathbf{X}^n]_1^k - [\mathbf{z}_I^n]_1^k) - b(\mathbf{s}_I^k - \mathbf{s}_J^k + \mathbf{W}^k) \mid J = j\right] \\
&= \sum_{i=0}^{2^k-1} P_{I|J}(i|j)\left([\bar{\mathbf{x}}_i^n]_1^k - ba([\bar{\mathbf{x}}_i^n]_1^k - [\mathbf{z}_i^n]_1^k)\right) - b\left(\sum_{i=0}^{2^k-1} P_{I|J}(i|j)\mathbf{s}_i^k - \mathbf{s}_j^k\right), \qquad (2.52) \\
[\mathbf{y}_j^n]_{k+1}^n &= \mathbb{E}\big[[\mathbf{X}^n]_{k+1}^n \mid J = j\big] = \sum_{i=0}^{2^k-1} P_{I|J}(i|j)[\bar{\mathbf{x}}_i^n]_{k+1}^n. \qquad (2.53)
\end{aligned}$$

Choosing $\{\mathbf{y}_j^n\}$ as above, and defining $\mathbf{U}^k$ as in (2.44), the distortion can be rewritten as

$$D_n(N_{des}) = \frac{1}{n}\mathbb{E}\big\|\mathbf{X}^n - \mathbb{E}[\mathbf{X}^n \mid J]\big\|^2 - \frac{1}{n}2b\mathbb{E}\left[\left([\mathbf{X}^n]_1^k - \frac{1}{n}\mathbb{E}\big[[\mathbf{X}^n]_1^k \mid J\big]\right)^T \mathbf{U}^k\right]$$

$$+\frac{1}{n}b^2\mathbb{E}\left\|\mathbf{U}^k\right\|^2+\frac{k}{n}b^2N_{des}.$$

Minimizing the above distortion by solving $\frac{\partial D_n(N_{des})}{\partial b}=0$ yields the expression of b given by (2.41). □

Lemma 2.2 For a fixed digital decoder codebook $\{\mathbf{y}_j^n\}$, a and b, fixed $\{[\mathbf{z}_i^n]_1^k\}$ as in (2.40), the optimal encoder regions $\{Q_i\}$ for ϵ_1 are given as follows:

$$\begin{aligned} Q_i=\Bigg\{\mathbf{x}^n\in\mathbb{R}^n: i=\arg\min_l\Big(&(ab-1)^2\left\|[\mathbf{x}^n]_1^k-[\mathbf{z}_l^n]_1^k\right\|^2+h_l \\ &+\sum_{j=0}^{2^k-1}\left\|[\mathbf{x}^n]_{k+1}^n-[\mathbf{y}_j^n]_{k+1}^n\right\|^2P_{J|I}(j|l)\Big)\Bigg\} \end{aligned} \tag{2.54}$$

where

$$h_l\triangleq\mathbb{E}\left[\left\|[\mathbf{y}_J^n]_1^k+b(\mathbf{s}_I^k-\mathbf{s}_J^k)\right\|^2\mid I=l\right]-\mathbb{E}\left[\left\|[\mathbf{z}_I^n]_1^k\right\|^2\right]. \tag{2.55}$$

Proof. The distortion $D_n^1(N_{des})$ in (2.35) can be written as

$$\begin{aligned} &D_n^1(N_{des}) \\ &= \frac{1}{n}\mathbb{E}\left\|[\mathbf{X}^n]_1^k-[\mathbf{z}_I^n]_1^k\right\|^2+\frac{1}{n}\mathbb{E}\left\|[\mathbf{y}_J^n]_1^k+b(\mathbf{s}_I^k-\mathbf{s}_J^k)\right\|^2-\frac{1}{n}\mathbb{E}\left\|[\mathbf{z}_I^n]_1^k\right\|^2+\frac{k}{n}b^2N_{des} \\ &\quad -\frac{1}{n}2ab\mathbb{E}\left[\left([\mathbf{X}^n]_1^k-[\bar{\mathbf{y}}_I^n]_1^k-b(\mathbf{s}_I^k-\bar{\mathbf{s}}_I^k)\right)^T\left([\mathbf{X}^n]_1^k-[\mathbf{z}_I^n]_1^k\right)\right]+\frac{1}{n}a^2b^2\mathbb{E}\left\|[\mathbf{X}^n]_1^k-[\mathbf{z}_I^n]_1^k\right\|^2 \\ &= \frac{1}{n}(ab-1)^2\mathbb{E}\left\|[\mathbf{X}^n]_1^k-[\mathbf{z}_I^n]_1^k\right\|^2+\frac{1}{n}\mathbb{E}\left\|[\mathbf{y}_J^n]_1^k+b(\mathbf{s}_I^k-\mathbf{s}_J^k)\right\|^2-\frac{1}{n}\mathbb{E}\left\|[\mathbf{z}_I^n]_1^k\right\|^2+\frac{k}{n}b^2N_{des} \\ &= \sum_{i=0}^{2^k-1}\frac{1}{n}\int_{Q_i}\left\{(ab-1)^2\left\|[\mathbf{x}^n]_1^k-[\mathbf{z}_i^n]_1^k\right\|^2+h_i\right\}p([\mathbf{x}^n]_1^k)\,d[\mathbf{x}^n]_1^k+\frac{k}{n}b^2N_{des} \end{aligned} \tag{2.56}$$

where h_i is defined as (2.55). Combining $D_n^1(N_{des})$ above with $D_n^2(N_{des})$ in (2.35) yields

$$D_n(N_{des}) \;=\; \sum_{i=0}^{2^k-1}\int_{Q_i}d[\mathbf{x}^n]_1^k\,p([\mathbf{x}^n]_1^k)\Big\{(ab-1)^2\left\|[\mathbf{x}^n]_1^k-[\mathbf{z}_i^n]_1^k\right\|^2+h_i$$

$$+\sum_{j=0}^{2^k-1}\left\|[\mathbf{x}^n]_{k+1}^n-[\mathbf{y}_j^n]_{k+1}^n\right\|^2 P_{J|I}(j|i)\Bigg\}+kb^2N_{des}.(2.57)$$

Therefore, the optimal encoder regions are given by (2.54). □

2.3.3 Some Special Cases

In Proposition 2.2, we derived the optimal power allocation coefficient t (with respect to $D_{hda}^{mis}(N_{tr}, N_{des})$) as a function of the design CSNR κ_{des}. Here we discuss the special cases of high and low κ_{des} regimes and examine how the power allocation coefficient t and the system distortion change with κ_{des} from the design point of view.

Assuming that the system is designed for a CSNR of $\kappa_{des} = P/N_{des}$ and a power allocation coefficient t, the digital channel has an effective CSNR of $\kappa_{dig} = \frac{(1-t)\kappa_{des}}{t\kappa_{des}+1}$, which means that the BSC transition probabilities $P_{J|I}(j|i)$ are calculated with the latter CSNR. Assume also that $\{Q_i\}$, $\{[\mathbf{z}_i^n]_1^k\}$, $\{\mathbf{y}_j^n\}$, and b are chosen according to the results of Section 2.3.2. We consider the following situations.

- *Low noise case, $\kappa_{des} \to \infty$.* In this case, $\kappa_{dig} \approx \frac{1-t}{t}$ and the $P_{J|I}(j|i)$'s no longer depend on κ_{des}. Since decoding the analog signal is dependent on the correct decoding of the digital signal, we can allocate more transmission power to the digital part (decrease t) to increase κ_{dig}, as long as $tP \gg N_{des}$. As a result, the distortion due to the digital transmission part decreases, which in turn makes the analog part more useful. This choice of t is consistent with the result of Proposition 2.2 (see Fig. 2.3). As more power is allocated to the digital part (e.g., as t decreases), $P_{J|I}(j|i) \to 0$ for $j \neq i$, hence, $\bar{\mathbf{s}}_I^k \to \mathbf{s}_I^k$, $[\mathbf{z}_I^n]_1^k \to [\bar{\mathbf{y}}_I^n]_1^k \to [\mathbf{y}_I^n]_1^k$, and $b \to \frac{1}{a}$. As a result, the encoder region $\{Q_i\}$ in (2.54) is simplified to $Q_I = \Big\{\mathbf{x}^n \in \mathbb{R}^n : I = \arg\min_l \Big(\left\|[\mathbf{x}^n]_{k+1}^n - [\mathbf{y}_l^n]_{k+1}^n\right\|^2\Big)\Big\}$ since $(ab-1)^2 \to 0$ and $h_l \to 0$. Thus the dominant distortion is the non-recoverable quantization error from the rest of the $n-k$ components of the source vectors. This observation is also justified by Proposition 2.1, where the first term of (2.5) goes to zero as $tP/N \to \infty$ (note that as $\kappa_{des} \to \infty$, we also have $\kappa_{tr} \to \infty$ since we assume that $\kappa_{tr} > \kappa_{des}$).

- *High noise case, $\kappa_{des} \to 0$.* In this case $b \to 0$, which means that we will not decode the analog signal because of its bad quality. Moreover, $[\mathbf{z}_I^n]_1^k \to [\bar{\mathbf{y}}_I^n]_1^k$ in (2.40) and $\mathbf{y}_j^n \to \sum\limits_i P_{I|J}(i|j)\bar{\mathbf{x}}_i^n$ in (2.42), (2.43). Since

$$\begin{aligned}\sum_i P_{I|J}(i|j)\bar{\mathbf{x}}_i^n &= \sum_i P_{I|J}(i|j)\mathbb{E}[\mathbf{X}^n|I=i] = \sum_i P_{I|J}(i|j)\mathbb{E}[\mathbf{X}^n|I=i, J=j]\\ &= \mathbb{E}[\mathbf{X}^n|J=j],\end{aligned}$$

we have $\mathbf{y}_J^n \to \mathbb{E}[\mathbf{X}^n \mid J]$, which means that the digital part approaches a COVQ [21]. In this case, it is best to allocate all the power to the digital part.

2.3.4 Training Algorithm

The results of Lemmas 2.1 and 2.2 can be used to formulate an iterative training algorithm as in [69] for codebooks design. The algorithm is summarized as follows: (1) Given the design noise variance N_{des}, total power P, power allocation coefficient t, and two thresholds γ_1, γ_2, calculate the corresponding transition probabilities $P_{J|I}(j|i)$ of the digital channel. Initialize the encoder regions[2]$\{Q_i\}$; (2) Determine the encoder centroids $\{\bar{\mathbf{x}}_i^n\}$ and the probabilities $\{P_I(i)\}$, initialize $[\mathbf{z}_I^n]_1^k = [\bar{\mathbf{x}}_I^n]_1^k$, initialize a to satisfy the power constraint; (3) Iteratively compute b, $\{\mathbf{y}_j^n\}$ and $\{[\mathbf{z}_i^n]_1^k\}$ using Lemma 2.1, update a after each iteration to satisfy power constraint, and stop when the changes of the codebooks $\{\mathbf{y}_j^n\}$ and $\{[\mathbf{z}_i^n]_1^k\}$ fall below the threshold γ_1; (4) Redefine the encoder regions

[2] Here we use the Voronoi regions of a VQ trained for a noiseless channel for the same source under consideration. An alternative way is to use the encoder of a COVQ [21] trained for the same digital channel $\{P_{J|I}(j|i)\}$).

$\{Q_i\}$ using Lemma 2.2, update a again, and estimate the average distortion; (5) Repeat steps (3) and (4) until the change of the average distortion falls below the threshold γ_2. In the simulations, $\gamma_1 = 10^{-5}$ and $\gamma_2 = 10^{-8}$ were used. We have the following remarks.

- Optimizing $\{[\mathbf{z}_i^n]_1^k\}$, $\{\mathbf{y}_j^n\}$ and b jointly is very complex. Instead, in the design we use Lemma 2.1 for an iterative approach similar to the one in [69]. First, we initialize $[\mathbf{z}_I^n]_1^k = [\bar{\mathbf{x}}_I^n]_1^k$. Then, we compute b using (2.41), and compute $\{\mathbf{y}_j^n\}$ using (2.42) and (2.43). We next update $\{[\mathbf{z}_i^n]_1^k\}$ using (2.40) with the new value of b and $\{\mathbf{y}_j^n\}$. The iterative algorithm is stopped when the changes of the codebooks $\{[\mathbf{z}_i^n]_1^k\}$ and $\{\mathbf{y}_j^n\}$ fall below a certain threshold.
- In our derivation, we assume that the power constraint (2.36) is satisfied with equality at all times. Strictly speaking, there is no guarantee for this to hold at all iterations. Therefore, convergence is not guaranteed. In our design, the coefficient a is updated after each computation of $\{[\mathbf{z}_i^n]_1^k\}$ to satisfy the power constraint. Our experimental studies suggest that the iterative algorithm does converge to a stable solution.
- In our design, all the codebooks are precomputed off line. During encoding, the digital encoder finds $\{Q_i\}$ using Lemma 2.2. It is easily seen from (2.54) that, $\{h_l\}$ can be precomputed. Given the input vector $\mathbf{x}^n$, most of the computation needed to find the encoder region involves the full COVQ-type search over the codebook $\{\mathbf{y}_j^n\}$ restricted to the last $n-k$ dimensions, i.e., we need to compute $\sum_{j=0}^{2^k-1} \left\| [\mathbf{x}^n]_{k+1}^n - [\mathbf{y}_j^n]_{k+1}^n \right\|^2 P_{J|I}(j|l)$. Thus, we can see that when a moderate block size n is used (e.g., $n = 24$ is used in the simulation of Section 2.4), the digital encoding part has low computational complexity and low delay. For the decoding part, since we use hard-decision demodulation, and the digital decoder codebook $\{\mathbf{y}_j^n\}$ is precomputed off line, we only need to perform table-lookup decoding.

Thus, the digital decoding complexity is low. As for the analog part, only k multiplications are needed for linear encoding/decoding.

2.4 Simulation Results

We evaluate the system's performance for the transmission of an i.i.d. Gaussian source over the AWGN channel. The source samples are grouped into vectors of dimension $n = 24$, and transmitted at an overall rate of 1/2 channel use per source sample. We implement our HDA optimized system using the training algorithm described in the previous section. Specifically, for a fixed input power $P = 1$ and design noise variance $N_{des} = 0.1$ (thus $\kappa_{des} = P/N_{des} = 10$), the training algorithm is implemented to generate the source digital transmitter and both the digital decoder codebook and the analog decoder. In light of Proposition 2.2 and curve (b) of Fig. 2.3, we choose $t = 0.05$ (this choice of t is expected to give good performance in the true CSNR range of 12 to 20 dB for the asymptotically achievable system). Apart from this choice of t, we carried out simulations with other choices of $t \in [0, 1]$ for the purpose of comparison. Motivated by a broadcast scenario, we assume (e.g., as in [69]) that the encoder is optimized for a given power allocation and fixed design CSNR κ_{des}, i.e., ε_1 and $\{\mathbf{z}_i^n\}$ are designed for a fixed κ_{des}, while the decoder knows the true CSNR κ_{tr} and adapts to it, i.e., $\{\mathbf{y}_j^n\}$ and b are adapted to κ_{tr}.

We present simulation results for the optimized HDA system with various power allocation coefficients t, as well as an unoptimized HDA system, a purely digital system, a purely analog system and the HDA-Turbo system of [70]. All systems have a transmission rate of 1/2 channel use per source sample.

- The optimized HDA system performance is shown in Figs. 2.5–2.8 for κ_{des} =10 dB and various values of t.

- The unoptimized HDA system uses the Linde-Buzo-Gray (LBG) algorithm [45] to design the digital encoder ε_1 and $\{\mathbf{z}_i^n\}$, and applies a linear encoder to the analog part. The digital decoder codebook $\{\mathbf{y}_j^n\}$ is adapted to the true CSNR κ_{tr}, and a linear MMSE decoder (also assuming knowledge of κ_{tr}) is applied to the analog part (its performance is shown in Fig. 2.7 for $t = 0.07$).

- The purely digital system, which solely employs the digital part of the HDA system, uses a COVQ source encoder [21] and a COVQ decoder codebook $\{\mathbf{y}_j^n\}$ adapted to the true CSNR κ_{tr} (its performance is shown in Fig. 2.7 for κ_{des}=10 dB).

- The purely analog system, which solely employs the analog part of the HDA system, transmits only half of each source vector using linear coding and employs a linear MMSE decoder with knowledge of the true CSNR (its performance is shown in Fig. 2.7).

- For the HDA-Turbo system of [70] (since the source is memoryless, the HDA-Turbo system does not employ Karhunen-Loéve processing), the digital part consists of a 24-dimensional 6-bit VQ designed using the LBG algorithm, and a high-delay ($k = 768, n = 1536$) rate 1/2 Turbo encoder with generator (37,21) (punctured to rate 1/2) and a random interleaver, and the analog part employs the same methods as the proposed HDA schemes. The digital decoder $\{\mathbf{y}_j\}$ and the analog decoder also has knowledge of κ_{tr} (its performance is shown in Fig. 2.8 for $t = 0.1$ and $t = 0.3$).

All systems are trained with 300,000 vectors, and tested with a different set of 100,000 vectors. For comparison purposes, we also present the following curves (shown in Figs. 2.7 and 2.8): the OPTA curve described by (2.4); the HDA bound described by (2.5) for fixed t; and the mismatched HDA bound described by (2.22) for fixed t and

κ_{des}. The performance results are presented in terms of the source signal-to-distortion ratio (SDR), which is defined by $\mathrm{SDR} = 10\log_{10}(\sigma_s^2/D)$ where D is the MSE distortion. We can observe the following:

- Figs. 2.5–2.6 indicate that the power allocation plays an important role in the performance of the optimized HDA system, especially for CSNRs above the design CSNR of 10 dB. Although we choose $t = 0.05$ based on Proposition 2.2, $t = 0.07$ turns out to be the best power allocation shown by the simulation results. In particular, the SDR increases as t increases from $t = 0$ (which is equivalent to the purely digital system) to about $t = 0.07$ (see Fig. 2.5) and then declines as t varies from $t = 0.07$ to $t = 1$ (which is equivalent to the purely analog system). While the optimal power allocation provided by Proposition 2.2 is derived for the ideal case (which assumes infinite block size), and the above numerical results are derived using a block size of 24, we note that the best choice (around $t = 0.07$) obtained by the numerical study is consistent with the value $t = 0.05$ suggested by Fig. 2.3. Another interesting observation is that when the true CSNR falls below 10 dB (κ_{des}), the SDR performance gets better as t increases. This is because the digital part degrades drastically when $\kappa_{tr} < \kappa_{des}$ (usually, the better the digital part performs at the design CSNR, the more drastic is its performance degradation for lower CSNRs).

- We observe from Fig. 2.7 that for $t = 0.07$, the optimized HDA system outperforms the unoptimized HDA system at all CSNRs. Moreover, it obtains a gain of 1 dB over the unoptimized HDA system, and is within 0.3 dB of the performance bound for the mismatched HDA system at high CSNRs (e.g., for CSNR $\geq$ 30 dB). The HDA systems present a smooth and robust performance for most CSNRs, and provide substantial improvements over the purely digital system from medium to high CSNRs. They also outperform the purely analog system for a wide range of CSNRs. We also note that the performance saturates at around 35 dB.

- In Fig. 2.8, we compare the optimized HDA system with the HDA-Turbo system of [70] for $t = 0.1$ and $t = 0.3$. We remark that for a proper choice of t, e.g., for $t = 0.1$, the optimized HDA system outperforms the HDA-Turbo system for CSNR $\geq$ 13 dB, and obtain a large gain for medium to high CSNRs. This behavior can be explained as follows. During the linear encoding process, we discard half of the components of each quantization error vector. For memoryless sources, all components of the error vectors have approximately the same variance. Since the optimized HDA system has higher quantization rate than that of the HDA-Turbo system (the HDA scheme does not employ channel coding while the HDA-Turbo system uses a rate 1/2 Turbo code), each component of the quantization error vector has a smaller variance than the corresponding quantization error component in the HDA-Turbo system. As a result, the distortion introduced in the optimized HDA system by this dropping-off process in the analog part is less severe than that for the HDA-Turbo system. On the other hand, the Turbo code plays an important role for CSNRs ranging from 5 to 10 dB. For CSNRs over 10 dB, channel coding becomes superfluous and most of the system distortion is due to quantization noise. Fig. 2.8 shows that in the CSNR range of 25 to 40 dB, the optimized HDA system has a gain around 1.5 dB over the HDA-Turbo system.

2.5 Some Remarks

Remark 1. Gauss-Markov sources. In [70], simulation results are presented for the transmission of Gauss-Markov sources with correlation coefficient 0.9 using the HDA-

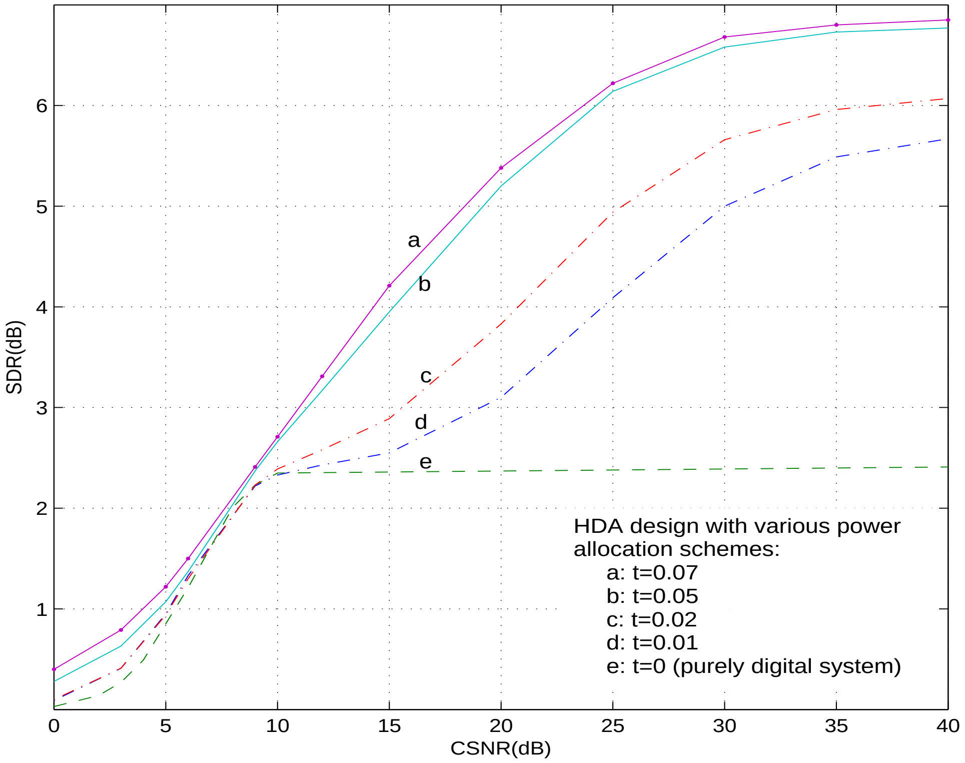


Figure 2.5: SDR performance (in dB) of optimized HDA systems with various power allocation coefficient t; i.i.d. Gaussian source over the AWGN channel, $\kappa_{des} = 10$ dB, $r = 1/2$ channel use/source sample.

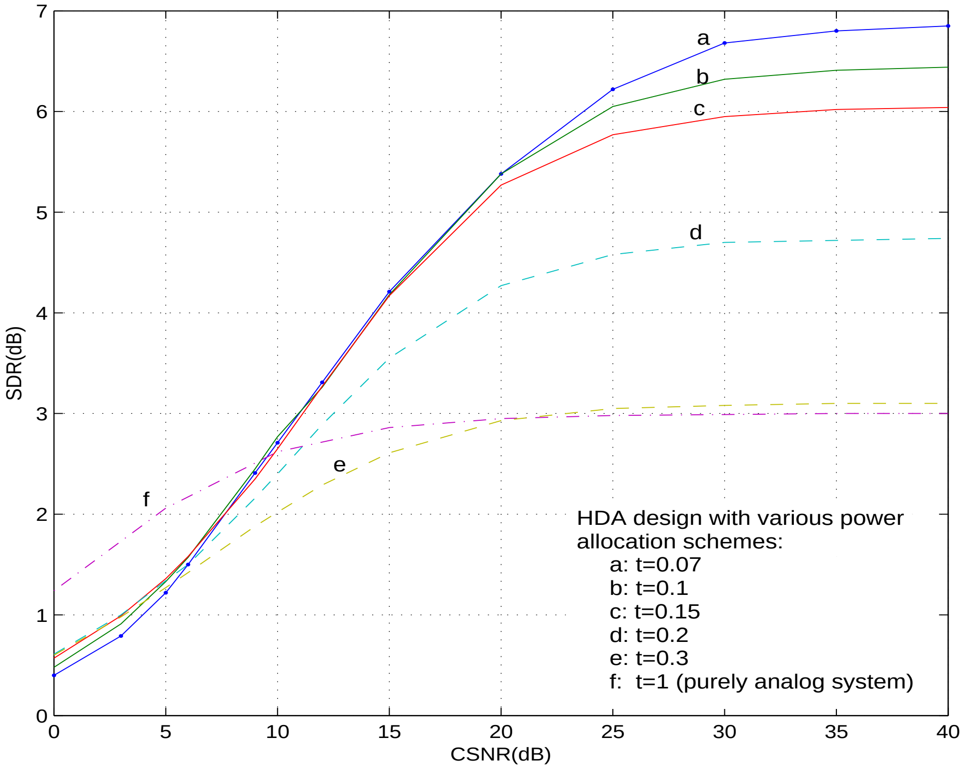


Figure 2.6: SDR performance (in dB) of optimized HDA systems with various power allocation coefficient t; i.i.d. Gaussian source over the AWGN channel, $\kappa_{des} = 10$ dB, $r = 1/2$ channel use/source sample.

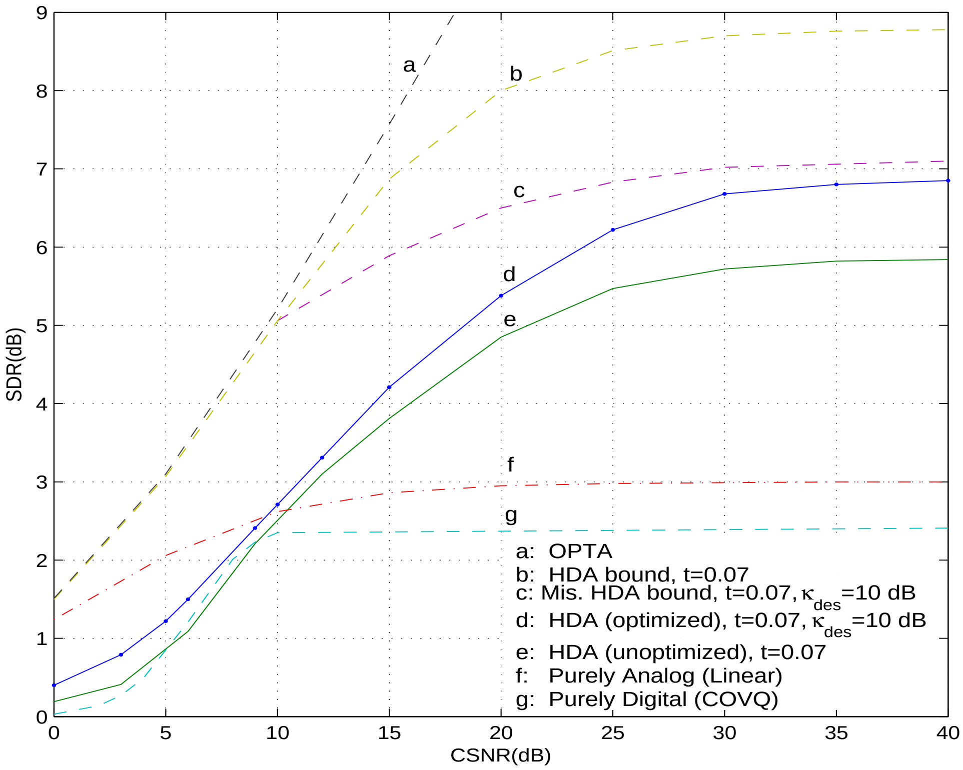


Figure 2.7: SDR performance (in dB) of the optimized HDA, the unoptimized HDA, the purely digital and the purely analog systems; i.i.d. Gaussian source over the AWGN channel, $r = 1/2$ channel use/source sample.

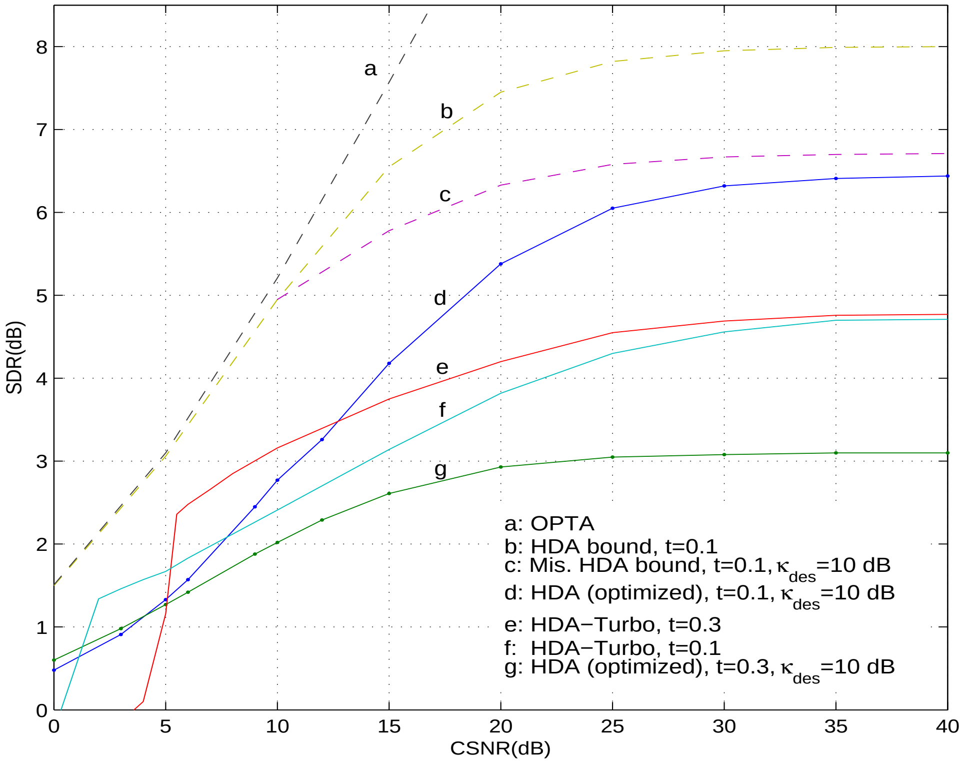


Figure 2.8: SDR performance (in dB) of various HDA systems; i.i.d. Gaussian source over the AWGN channel, $r = 1/2$ channel use/source sample.

Turbo system with bandwidth compression. We also applied our system to such sources. As in [70], we employed a KLT to decorrelate the source before HDA coding. Simulation results show that in this case, our HDA scheme performs marginally better than the HDA-Turbo system for high CSNRs, but it is inferior for low and medium CSNRs. This is due to the fact that the KLT concentrates most of the source power among the first few transform coefficients. This is in contrast to the memoryless source case where all source components have the same variance. Thus, even though the HDA-Turbo system has a lower quantization rate than our system, the distortion introduced by its analog part is significantly less than in the memoryless source case, since the most important components of the quantization error vector are not lost by the dropping-off process of the linear encoder. As a result the HDA-Turbo system performs better than our system at low and medium CSNRs due to the powerful error correcting capability of the Turbo code.

Remark 2: HDA system with nonlinear analog part. To improve the system performance at high CSNRs, we also implement an HDA system with nonlinear analog part which is similar to the ones studied in [70], [23]. Here the error vectors $\mathbf{E}^n$ are first quantized to some discrete values using a VQ and then mapped to a discrete set of signal points using pulse amplitude modulation (PAM). The nonlinear analog part is not strictly analog. When a high-level PAM is applied, it can be considered as "close to analog". The analog part is trained to minimize the end-to-end distortion

$$D_{analog} = \mathbb{E}\|\mathbf{E}^n - \hat{\mathbf{E}}^n\|^2 \tag{2.58}$$

with power $P_A = tP$ and noise variance N_{des}. The digital part employs using a COVQ (see e.g., [21]), which is trained with power $P_D = (1-t)P$ and noise variance $N_{des} + tP$.

In particular, we employ a system with rate 1/2 channel uses per source sample. We first decompose quantization error vector $\mathbf{E}^n$ into $n/2$ two-dimensional subvectors

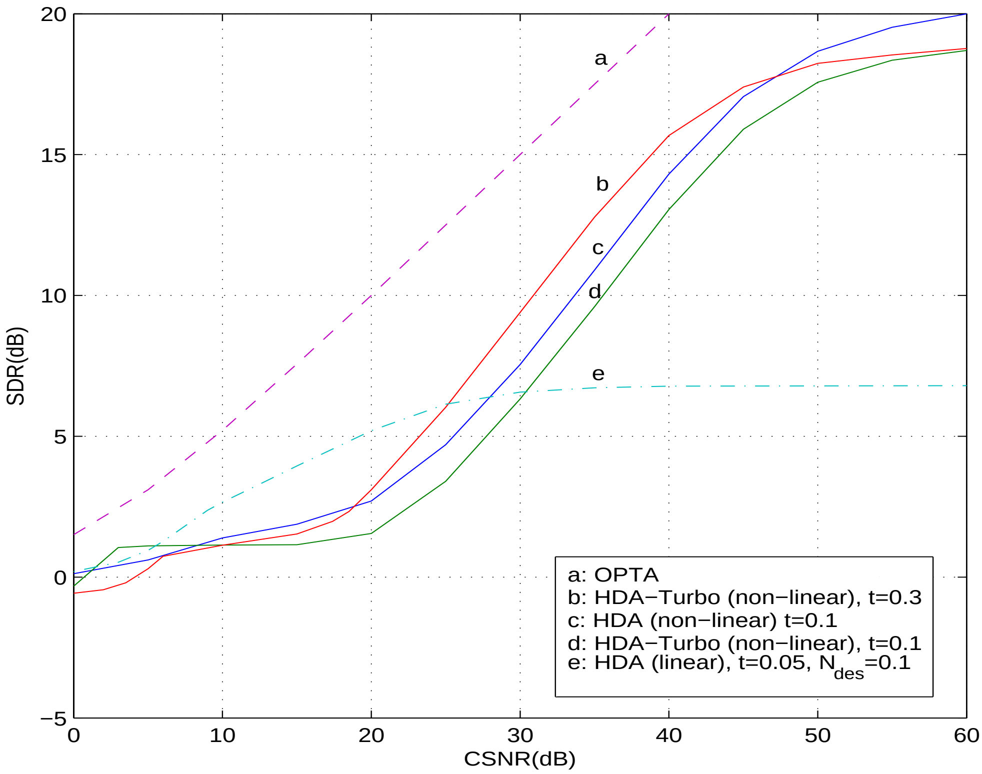


Figure 2.9: SDR performance (in dB) of various HDA systems with non-linear analog part; i.i.d. Gaussian source over the AWGN channel, $r = 1/2$ channel use/source sample.

(assuming that n is even), where each such subvector is quantized using a q-bit 2-dimensional vector quantizer. The output index is mapped to a channel symbol using a 2^q-level PAM. Therefore, all components of $\mathbf{E}^n$ are transmitted using this nonlinear analog coding method. The analog decoder implements hard-decision detection on the received signals, and performs a table lookup in a codebook, outputting $\hat{\mathbf{E}}^n$. The whole procedure can be modeled as a q-bit quantizer followed by a 2^q-input 2^q-output discrete memoryless channel (DMC). The channel transition probability matrix can hence be easily computed from the complementary error function given the CSNR. We note that the rate of the vector quantizer, as well as the size of the PAM constellation, can be increased to act more like analog coding. In the simulation, we choose $q = 8$.

Since the aim of this system is to focus on the performance at medium to high CSNRs, we assume $N_{des} = 0.0001$ and $t = 0.1$ (t is chosen via an experimental study). The digital part is designed using a COVQ algorithm for power $P_D = (1 - t)P = 0.9$ and noise variance 0.1001 (or around 9.5 dB). The non-linear code of the analog part is designed for power $P_A = tP = 0.1$ and noise variance 0.0001, and 256-level PAM signals are employed. Simulation results are given in Fig. 2.9. We observe that for the schemes with non-linear analog coding and $t = 0.1$, the HDA scheme outperforms the HDA-Turbo for CSNRs between 10 and 60 dB. The HDA scheme still outperforms the HDA-Turbo system with $t = 0.3$ at high CSNRs and saturate around 60 dB. We also observe from the simulation results that the CSNR for which the nonlinear mapping is designed is actually a trade-off between the quantization distortion and the distortion incurred by the channel noise. If we prefer to get good performance at high CSNRs, the design will focus on reducing the quantization distortion, as the channel noise is very small compared to the channel input power in this range. If we want to get more robust performance for medium CSNRs, we have to sacrifice some quantization accuracy in the quantizer design in order to combat the channel noise. This explanation has been justified by examining the distribution of the quantizer codebook: at lower CSNRs, the VQ codevectors are clustered more closely to each other while as the CSNR increases, the codevectors spread and yield a smaller overall quantization distortion.

2.6 Conclusions

An HDA joint source-channel system with bandwidth compression for the reliable communication of memoryless Gaussian sources over AWGN channels is studied. The system has a simple linear analog coding component. Information-theoretic distortion upper bounds (under both matched and mismatched channel conditions) and a power allocation scheme are established for the system. Then, a practical HDA scheme which employs a VQ cascaded with BPSK modulation in the digital part is designed and implemented. The system is similar to that considered in [70] but it is simpler as it does not use Turbo error-correcting coding. A training algorithm is presented to iteratively optimize the source digital transmitter (both source encoder and source decoder) and both the digital decoder codebook and the analog decoder. Numerical results show that the HDA scheme offers a robust and graceful performance improvement for a wide range of CSNRs (medium to high CSNRs), and substantially outperforms purely digital and purely analog systems for a large range of CSNRs. Furthermore, the performance of the HDA scheme approaches the theoretical distortion bound for high CSNRs. The advantages of the HDA scheme are as follows: (1) it has low complexity and low delay; (2) it guarantees a graceful performance improvement for high CSNRs; (3) the joint source-channel design of the codebooks enables smooth degradation for medium CSNRs.

Chapter 3

Design of VQ-Based Hybrid Digital-Analog Coder for Image Communication

This chapter is based on the paper presented at the *Data Compression Conference* (DCC'05), Snowbird, UT, USA, March 2005 [83].

3.1 Overview

In this chapter, we present an image communication application that illustrates the effectiveness of HDA coding. The image coding system combines the bandwidth compression system studied in Section 2.4 with the bandwidth expansion system of [69].

We consider the transmission of gray-scale still images over an AWGN channel. Fig. 3.1 shows the block diagram of the proposed image coding system. The images are first subjected to a wavelet decomposition. The wavelet coefficients are modeled as memoryless sources and are formed into three groups of vectors and transmitted using

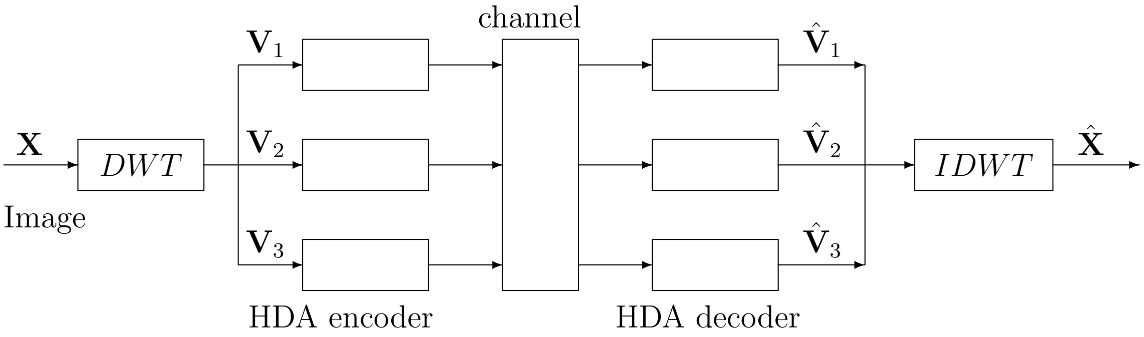


Figure 3.1: The structure of the HDA image coding system.

either a bandwidth compression system or a bandwidth expansion system, depending on the level of their importance. The channel outputs are decoded and placed back to form reconstruction images.

3.2 Structure of the Image Coder

Generally, we first subtract the average of all pixel values in the image from each pixel before wavelet decomposition to lower the average energy. For a gray scale image, we simply subtract the constant 128 from each entry instead of the actual average value in order to avoid additional side information. The results are decomposed using a two-dimensional separable discrete wavelet transform (DWT). Here we employ a lifting scheme with Antonni 9/7 biorthogonal wavelet filters (see [16], [22] for details on the lifting wavelet transform). The DWT is applied three times, each time on the lowest frequency subband of the previous resolution level, resulting in 10 subbands overall as shown in Fig. 3.2. The variance and mean of each subband are estimated by their empirical probabilities and all the wavelet coefficients are normalized to have zero mean and unit variance. The normalized wavelet coefficients are grouped into three classes of vectors as follows:

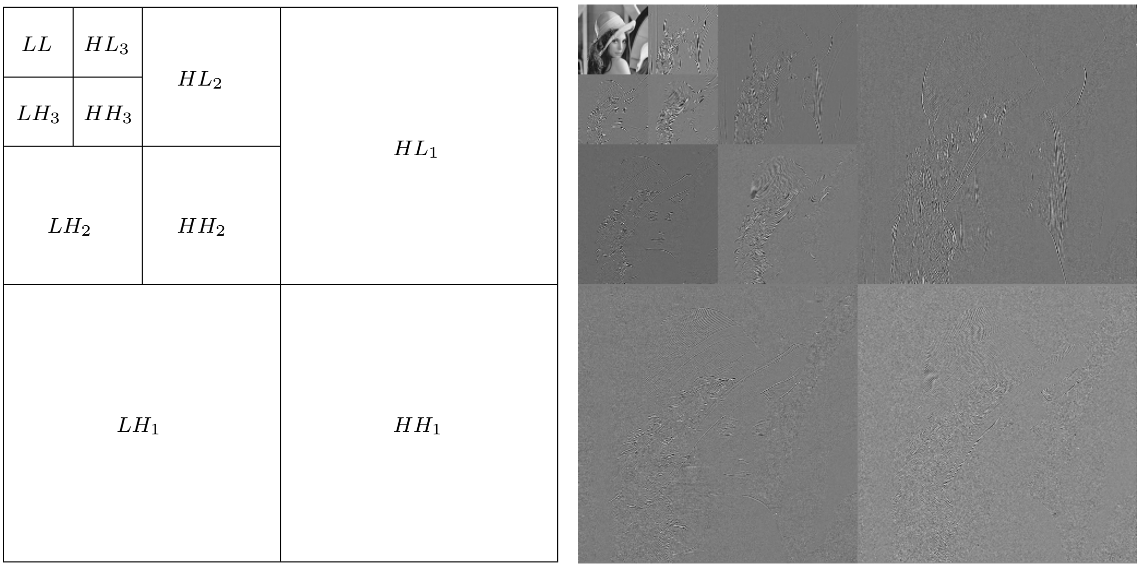


Figure 3.2: 3-level wavelet decomposition to Lena image.

- For the lowest frequency LL subband , each block of 2×2 coefficients form a vector of dimension 4, and is referred to as a class 1 vector.

- For the highest frequency levels (there are three such subbands in total), each block of 4×4 coefficients form a vector of dimension 16, and is referred to as a class 3 vector.

- For the remaining two frequency levels (six subbands in total with three subbands for each level), one coefficient from the coarser level and a block of 2×2 coefficients from the finer level (with the same frequency direction as the coarser one) form a vector of dimension 5, and is referred to as a class 2 vector.

Since the three classes of vectors have unequal roles in the reconstruction of the overall image, different coding strategies are employed in their processing and transmission. More precisely, we use a bandwidth expansion system to transmit the vectors of class 1 and class 2, since these classes of vectors involve the low and middle frequency components of the image, which are vital for the overall image quality. In total, 1/4 of the coefficients are coded using the bandwidth expansion system of [69]. The vectors of class 3 involve finer detail information of the image, which is also important when high quality image reconstruction is desired. Due to the large volume of this part (3/4 of the total coefficients), the proposed HDA system with bandwidth compression in Section 2.3 will be used for the class 3 vectors.

3.3 Probability Models

For wavelet image coding systems, there have been several assumptions concerning the distribution of the wavelet coefficients. A common assumption is that the distribution of the wavelet coefficients of each subband can be well approximated by the generalized Gaussian distribution (GGD) (e.g., [78], [27]) whose probability density function (pdf) is given by

$$f(x) = \frac{\alpha\eta(\alpha,\sigma_s)}{2\Gamma(1/\alpha)}\exp\{-[\eta(\alpha,\sigma_s)|x|]^\alpha\}$$

where $\eta(\alpha,\sigma_s) = \frac{1}{\sigma_s}\left(\frac{\Gamma(3/\alpha)}{\Gamma(1/\alpha)}\right)^{\frac{1}{2}}$, $\alpha > 0$ is a shaping parameter, σ_s is the standard deviation of the distribution, and $\Gamma(\cdot)$ is the Gamma function. The pdf of the GGD reduces to the Laplacian pdf when $\alpha = 1$ and yields the Gaussian pdf when $\alpha = 2$.

We have compared the empirical distribution of the wavelets coefficients to the GGD using the Kolmogorov-Smirnov (KS) test. For a given set of data $X = \{x_i\}_{i=1}^M$, the KS test compares the empirical distribution function $F_X(\cdot)$ to a given distribution function $F(\cdot)$. The empirical distribution function is defined as

$$F_X(t) = \begin{cases} 0, & t < x_{(1)} \\ \frac{i}{M}, & x_{(i)} \leq t < x_{(i+1)}, \quad i = 1, \cdots, M \\ 1, & t \geq x_{(M)} \end{cases} \tag{3.1}$$

where $x_{(i)}, i = 1, \cdots, M$ are the ascending-ordered version of the data X. The KS distance d_{KS} is then defined as

$$d_{KS} = \max_{i=1,\cdots,M} |F_X(x_i) - F(x_i)|. \tag{3.2}$$

When testing the data against several known distributions, the distribution which yields the smallest KS distance is the best fit for the data.

Here the KS test was carried out on a few gray-level test images. We have searched for the best value of the shape parameter α, in the sense of minimizing the KS distance in the range $0.1 \leq \alpha \leq 2$. The results are listed in Table 3.1. These show that $\alpha = 0.80$ is a good approximation of the shape parameter for the GGD for all the wavelets subbands except the LL subband. In the design, we use the Laplacian distribution for simplicity. For the LL subband, the Gaussian pdf tends to be the best fit for most test images we instigated. We then assume that the LL subband is modelled by the Gaussian distribution.

Since the HDA systems involve the transmission of quantization errors, the KS test is also carried out to approximate the distribution of the quantization errors. Results are shown in Table 3.2 which indicates that the Laplacian distribution is a good approximation.

3.4 Adaptive Decoding

As in [69], motivated by a broadcast scenario, we apply the training algorithms to a fixed-encoder adaptive-decoder optimized HDA system. For example, the optimized HDA bandwidth expansion system is designed for a fixed CSNR value (in decibels), yielding a fixed encoder ε_1, and fixed $\{\mathbf{z}_i\}$ and t, which are not modified as the true CSNR changes. On the other hand, the decoder has knowledge of the true CSNR and adapt to it by updating the values of $\{\mathbf{y}_j\}$ and b as the CSNR varies.

Test	Best values of α									
Images	Level 1			Level 2			Level 3			
	LH	HL	HH	LH	HL	HH	LH	HL	HH	LL
Baboon	0.75	0.85	0.95	0.75	0.90	0.85	0.85	1.00	0.90	2.00
Couple	0.70	1.05	1.15	0.55	0.70	0.85	0.60	0.60	0.70	1.55
Airplane	0.60	0.70	1.20	0.60	0.50	0.65	0.55	0.45	0.55	2.00
Girl	0.65	0.80	1.10	0.65	0.60	0.70	0.65	0.60	0.65	2.00
Goldhill	1.00	0.90	1.70	0.85	0.75	1.00	0.85	0.70	0.85	1.45
Lena	0.95	0.75	1.10	0.75	0.65	0.65	0.55	0.50	0.50	2.00
Pepper	0.80	0.85	1.55	0.60	0.60	0.80	0.60	0.55	0.65	2.00
Sailboat	0.85	0.95	1.40	0.60	0.65	0.75	0.55	0.65	0.65	2.00
Average	0.79	0.86	1.27	0.67	0.67	0.78	0.65	0.63	0.68	1.90
Average	0.80 (Average for all subbands except LL)									1.90

Table 3.1: The best α values chosen with the Kolmogorov-Smirnov test for all wavelets subbands.

Test Images	Best values of α									
	Level 1			Level 2			Level 3			
	LH	HL	HH	LH	HL	HH	LH	HL	HH	LL
Baboon	0.90	0.95	1.05	1.55	1.40	1.45	1.20	1.40	1.15	1.45
Couple	0.80	1.25	1.25	1.05	1.40	1.05	0.65	0.85	0.90	0.55
Airplane	0.80	0.80	1.35	1.35	1.45	1.40	0.90	0.65	0.75	1.05
Girl	0.85	0.90	1.20	1.00	1.15	1.20	0.75	0.75	0.85	0.85
Goldhill	1.05	0.95	1.80	1.30	1.10	1.25	1.20	0.85	1.05	1.60
Lena	1.00	0.85	1.15	1.15	1.35	1.20	0.70	0.85	0.70	0.80
Pepper	0.90	0.90	1.65	1.00	1.10	1.00	0.75	0.70	0.85	1.00
Sailboat	1.00	1.15	1.60	1.45	1.45	1.40	0.80	0.95	0.95	1.45
Average	0.91	0.97	1.38	1.23	1.30	1.24	0.87	0.88	0.90	1.10
Average	1.05 (Average for all subbands except LL)									1.10

Table 3.2: The best α values chosen with the Kolmogorov-Smirnov test for the quantization errors of all wavelets subbands.

3.5 Side Information

Certain side information must be reliably transmitted over the channel, including the mean and variance of each subband. By observing the statistical properties of the subband data for a variety of images, we found that the mean values for all subbands except for the LL subband are very small compared to the standard deviation. Thus, all these mean values (except the LL subband) are assumed to be zero in our design. For a 3-level 10-band octave decomposition, we use 12 bits to quantize the variance of each subband and 8 bits to quantize the mean value of the LL subband, resulting in a total of 128 bits. The image size also needs to be known at the receiver (it is encoded using the natural binary code). The side information is usually error protected before transmission. For an image of size 512×512 and a rate-1/2 error control code, the overhead consists of 292 bits in total, or equivalently around 0.001 bits per pixel. In the following discussion, we assume that the side information is transmitted error free, and we do not include it in the calculation of the overall system rate (as it is negligible compared to the rate).

3.6 Simulations

We next implement the proposed HDA image coding system for the transmission of gray-scale images over AWGN channels and test it for the images Lena and Goldhill, both of size 512×512. We denote this image coding system by VQHDA I. To improve the system performance at high CSNRs, we also propose a system, denoted by VQHDA II, where for the bandwidth compression part, an HDA design with nonlinear analog part is applied (c.f., Section 2.5 or [70] for details). For the sake of comparison, we also present a purely analog system based on linear coding, and two purely digital systems based on

COVQ and LBG-VQ designs respectively. All systems are trained with 300,000 training vectors.

A. HDA Systems

The VQHDA I system is carried out with the following parameters. The quantizers for the class 1, 2 and 3 vectors are 5 bits, 4 bits and 8 bits respectively (these bit allocations were determined experimentally). The overall transmission rate is 0.832 channel use/pixel. Since the power allocation result of Proposition 2.2 does not apply to Laplacian sources, we choose the power allocation coefficient $t = 0.1$ (based on a numerical study). The encoder is designed at a fixed CSNR of 10 dB, while (as in the previous section) the decoder is assumed to have knowledge of the true CSNR and adapts to it by updating the values of $\{\mathbf{y}_j\}$ and b as the CSNR varies. In particular, the quantizers of the class 1 and 2 vectors are trained using the algorithm proposed in [69], and the quantizer of the class 3 vectors is trained using the algorithm based on Lemmas 2.1 and 2.2 (see Section 2.3.2).

The VQHDA II system has the same structure as the VQHDA I, except that for the bandwidth compression part, an HDA system with nonlinear analog part is employed (see Section 2.5 for details). The nonlinear mappings are designed at two CSNRs: 25 dB and 50 dB. The other parts of the HDA system, i.e. the digital part of the compression system and the expansion systems are designed for a CSNR of 10 dB.

C. Purely Analog System

The purely analog system (denoted by Analog) is developed using a rate-one linear code. The system employs a similar rate allocation as that of the VQHDA I system. Class 1 and class 2 vectors are transmitted twice, and the receiver employs a linear minimum mean square error decoder. Class 3 vectors employ a similar method as in the bandwidth compression system, where half of the components of each vector is transmitted using linear mapping. The total rate of the system is around 0.875 channel use/pixel, which is comparable to our system.

D. Purely Digital Systems

Two purely digital systems are also investigated. The first digital system uses channel optimized vector quantization (COVQ). For this system, vectors are formed using the same vector schemes as in the VQHDA I system. A 4-dimensional 9-bit COVQ, a 5-dimensional 9-bit COVQ, and a 16-dimensional 8-bit COVQ are trained at a CSNR of 10 dB. The output indices of each VQ are then directly sent over the BPSK modulated channel. The channel input power per channel use is also set to unity. The receiver employs hard decision demodulation and adaptive COVQ decoding. The second digital system, denoted by LBG-VQ, uses the Linde, Buzo and Gray (LBG) vector quantization algorithm, where a 4-dimensional 9-bit LBG-VQ, a 5-dimensional 9-bit LBG-VQ, and a 16-dimensional 8-bit LBG-VQ are employed. The remaining system parts are identical to their counterparts in the COVQ system. Adaptive decoding is also employed. Both systems have an overall rate of 0.832 channel use/pixel.

E. Results and Discussion

In Figs. 3.3–3.6, we show simulation results for the images Lena and Goldhill in terms of the peak signal-to noise ratio (PSNR), which is defined (in dB) by

$$\mathrm{PSNR} = 10\log_{10}\frac{(255)^2}{D},$$

where D is the MSE distortion between the original and decoded images. Each image is tested 10 times and the average PSNR is reported.

We observe that the VQHDA I system outperforms the purely analog and LBG-VQ systems for most CSNRs, and provides substantial improvements over the purely digital

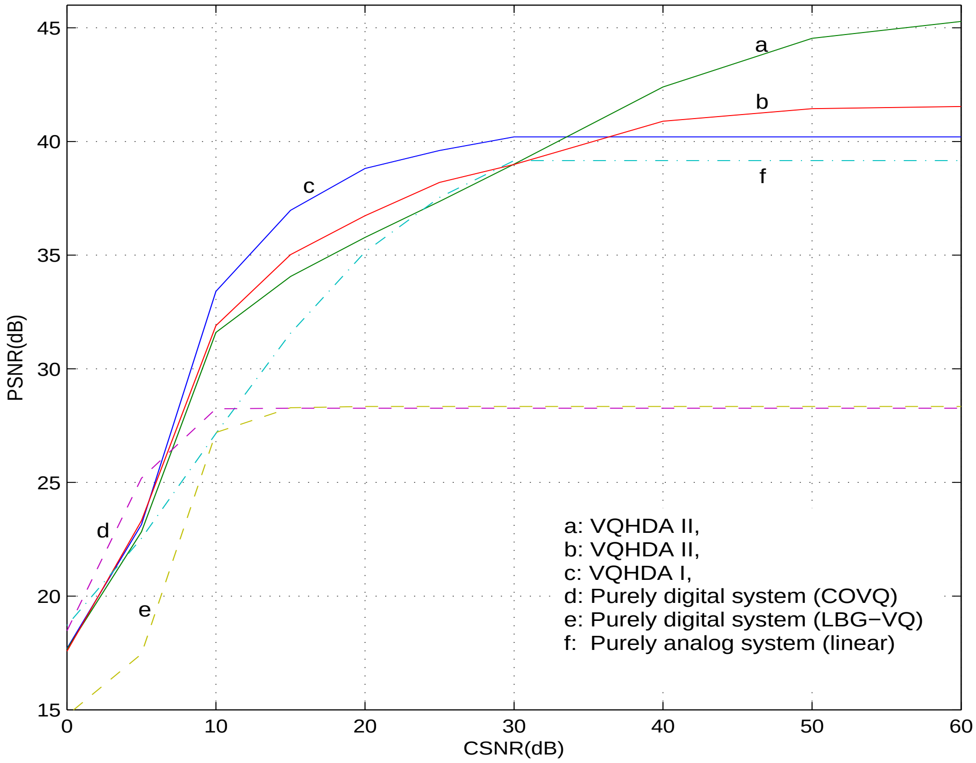


Figure 3.3: Image communication system performance for the Lena image. The VQHDA I and COVQ systems are designed at a CSNR of 10 dB. Curves (a) and (b) refer to VQHDA II systems, where for the compression part, the nonlinear mappings are designed at a CSNR of 50 dB and 25 dB respectively with t=0.1, while the other parts are designed at 10 dB. Adaptive decoding is used for all systems.

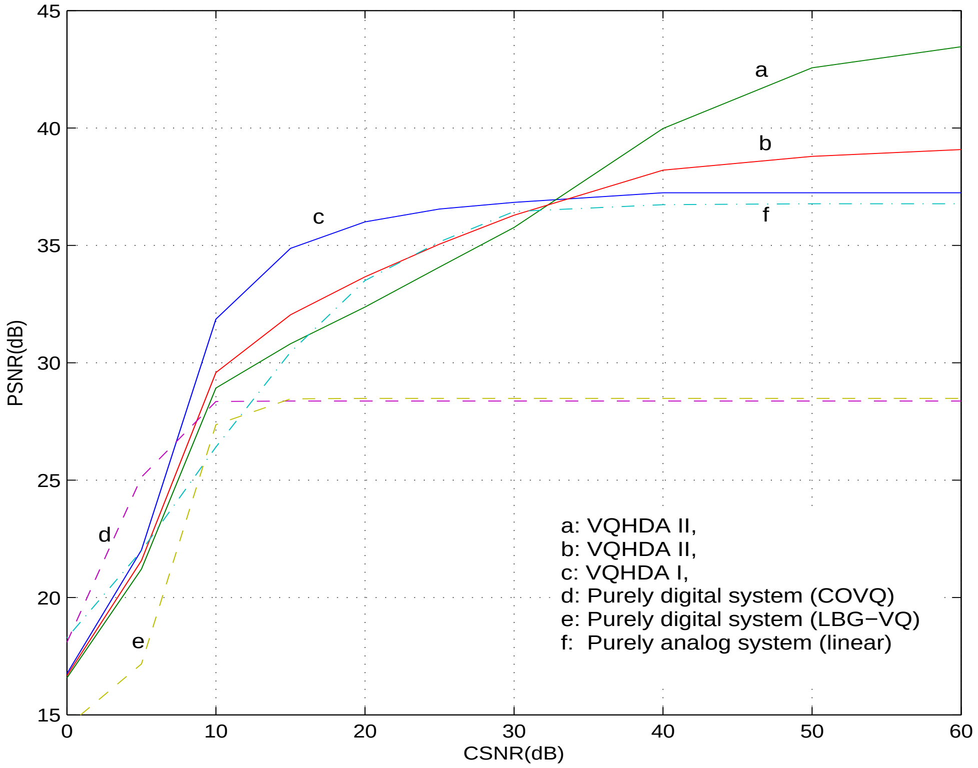
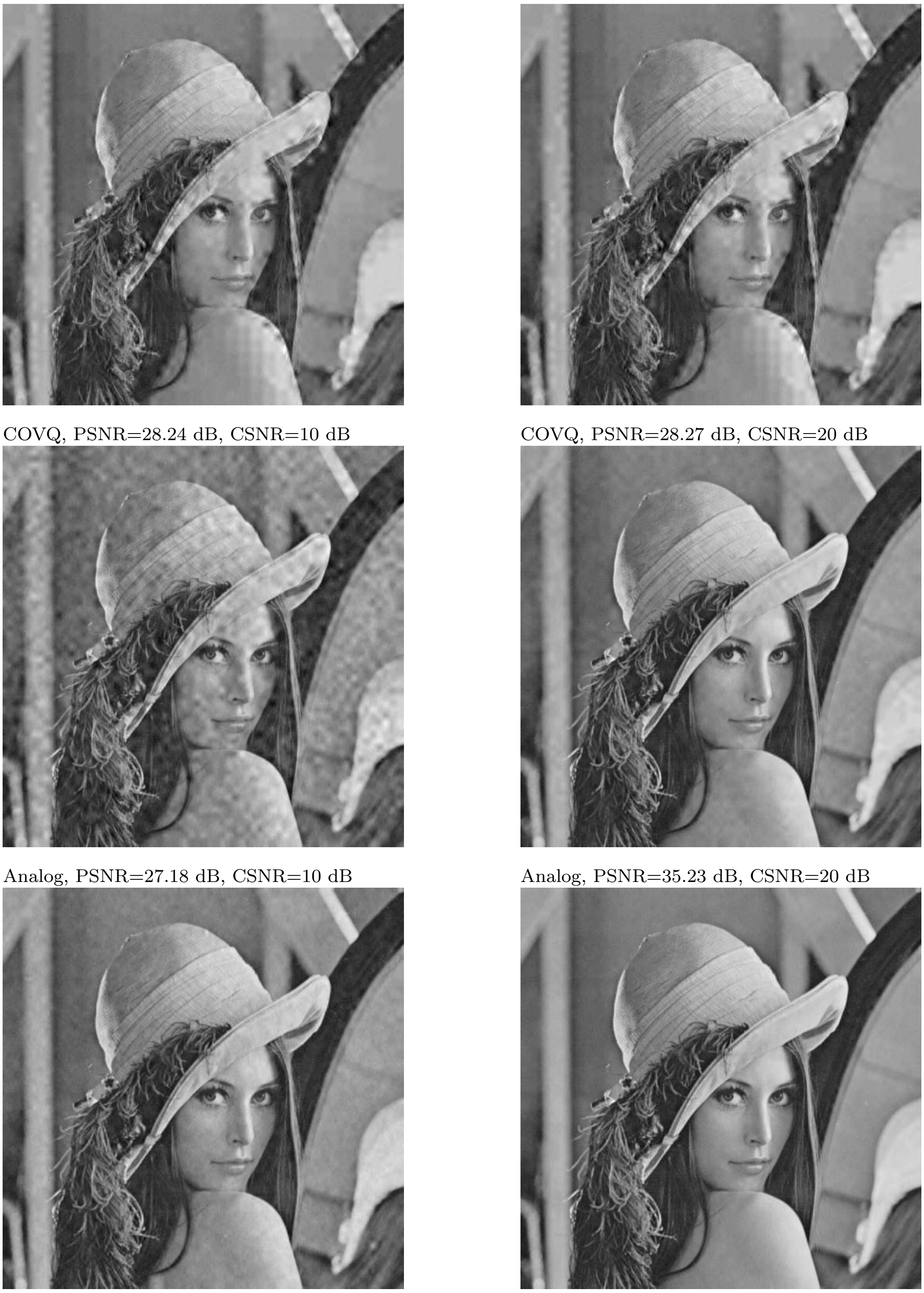


Figure 3.4: Image communication system performance for the Goldhill image. The VQHDA I and COVQ systems are designed at a CSNR of 10 dB. Curves (a) and (b) refer to VQHDA II systems, where for the compression part, the nonlinear mappings are designed at a CSNR of 50 dB and 25 dB respectively with t=0.1, while the other parts are designed at 10 dB. Adaptive decoding is used for all systems.



COVQ, PSNR=28.24 dB, CSNR=10 dB

COVQ, PSNR=28.27 dB, CSNR=20 dB

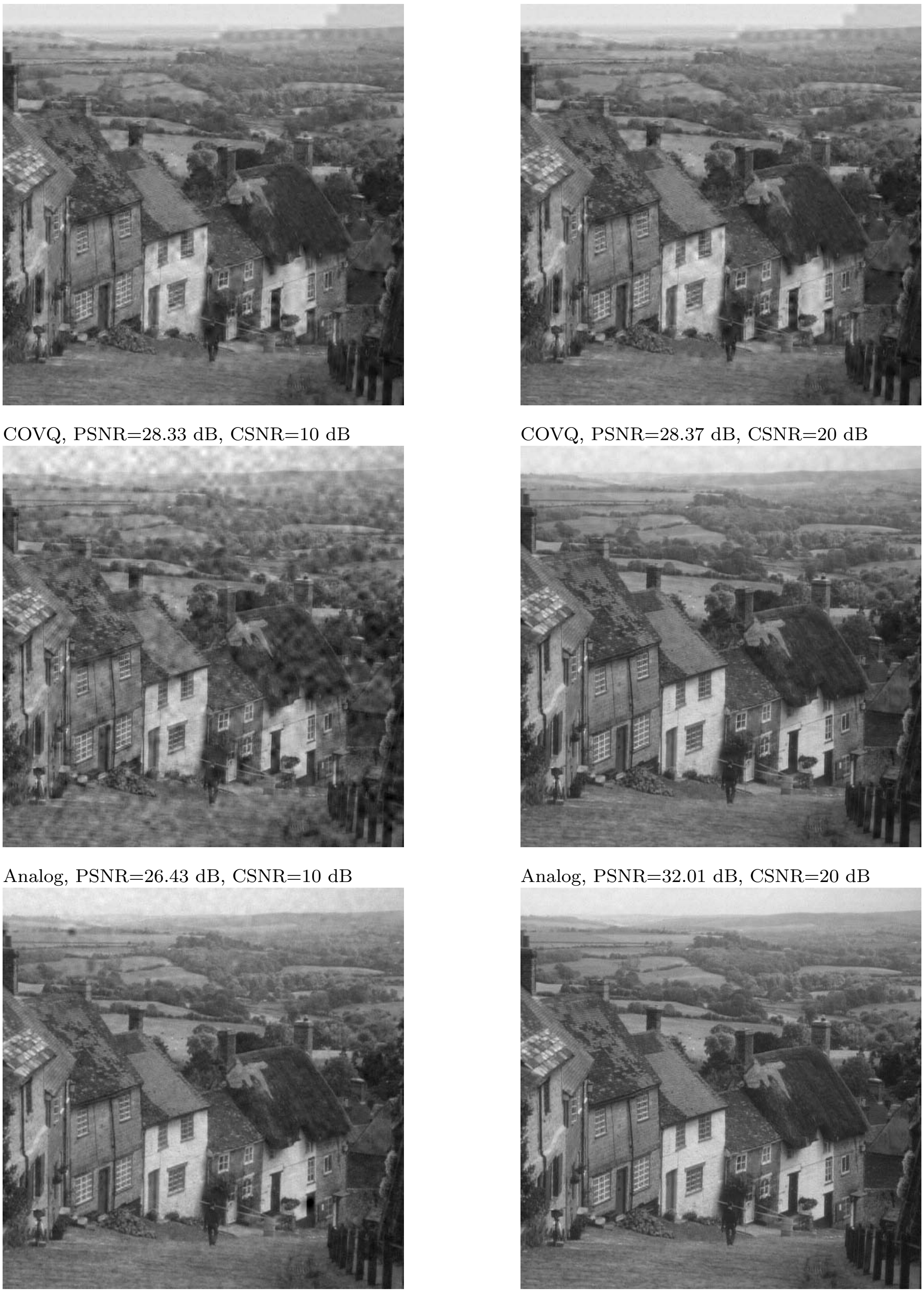
Analog, PSNR=27.18 dB, CSNR=10 dB

Analog, PSNR=35.23 dB, CSNR=20 dB

VQHDA I, PSNR=33.69 dB, CSNR=10 dB

VQHDA I, PSNR=39.33 dB, CSNR=20 dB

Figure 3.5: Comparison between different systems, where the VQHDA and COVQ systems are designed at CSNR of 10 dB. Adaptive decoding is used in all systems.



COVQ, PSNR=28.33 dB, CSNR=10 dB

COVQ, PSNR=28.37 dB, CSNR=20 dB

Analog, PSNR=26.43 dB, CSNR=10 dB

Analog, PSNR=32.01 dB, CSNR=20 dB

VQHDA I, PSNR=33.69 dB, CSNR=10 dB

VQHDA I, PSNR=37.03 dB, CSNR=20 dB

Figure 3.6: Comparison between different systems, where the VQHDA I and COVQ systems are designed at CSNR of 10 dB. Adaptive decoding is used in all systems.

systems from medium to high CSNRs. However its performance saturates at around CSNR of 30 dB; this is due to the non-reversible analog linear map in the bandwidth compression system. We also note that for the VQHDA II system, replacing the linear analog map with a nonlinear map enables the system to saturate at higher CSNRs. In fact, the performance of the VQHDA II can be made to saturate at an arbitrary high CSNR by increasing the resolution of the nonlinear mappings. However, the VQHDA II is inferior to VQHDA I for low to medium CSNRs due to the breakdown of the nonlinear maps in this range. We also remark that the COVQ system performs better than the proposed VQHDA systems at low CSNRs; this can be remedied by using soft-decision COVQ (e.g., [57]) in the digital part of the HDA systems.

3.7 Conclusions

An image communication system using VQ-based HDA JSC coding for AWGN channels is proposed. This system is robust and enjoys graceful improvement characteristics for a large range of channel conditions. Both bandwidth expansion and compression HDA systems are used for the coding and transmission of the image wavelet coefficients: bandwidth expansion is applied on the low and medium frequency subbands, while bandwidth compression is applied on the high frequency subbands. Numerical results show that the proposed system is superior to purely analog and purely digital systems for a wide range of CSNRs. Future work may focus on optimizing the rate allocation among the different subbands and optimizing the power allocation between the digital and analog parts.

Part II

Information Hiding:
Information-Theoretic Perspectives

Chapter 4

Error Exponent Analysis of Single-User Joint Compression and Private Watermarking with Gaussian Attacks

This chapter is based on the paper presented at the *10th Canadian Workshop on Information Theory* (CWIT'07), Edmonton, AB, Canada, June 6-8, 2007 [86].

4.1 Introduction

In a joint compression and embedding information-hiding model, the watermarker encodes a watermark and a covertext jointly to output a (compressed) stegotext. Denoting the *compression rate* by R_c and the *watermarking rate* by R_w, the main goal is to determine the achievable rate pairs (R_c, R_w) under transparency and robustness constraints on the system.

Karakos and Papamarcou [31], [32], [33] study the tradeoff between R_c and R_w for Gaussian host data: the case where the stegotext is not subjected to attacks is studied in [31, 32]; the case where the stegotext is subjected to additive memoryless Gaussian attacks is examined in [33]. The main result of [33] is a coding theorem which establishes the achievable region for rate pairs (R_c, R_w) under transparency and robustness constraints for a private scenario. Maor and Merhav [46], [47] study a similar tradeoff problem for discrete memoryless sources in a public scenario. The work in [46] focuses on the attack-free problem of joint watermarking and lossy compression, where the host data, the watermark and the stegotext are drawn from finite alphabets, while [47] extends the model in [46] to include a stationary memoryless discrete attack channel operating on the stegotext. In both works, coding theorems are established in which a single-letter expressions involving the maximum achievable watermarking rate, the compression rate and the distortion threshold are obtained. Yang and Sun [89] study a similar joint compression/watermarking problem with abstract alphabets in a private scenario. Other related works include [5], [39], [50], [59].

In this chapter, we focus on the problem introduced and investigated in [33], i.e., the joint compression and watermarking of memoryless Gaussian sources under additive white Gaussian noise (AWGN) attacks in a private scenario. We refine the analysis of the probability of error in decoding the watermarks for any achievable rate pairs (R_c, R_w). Using a random coding technique that incorporates Gallager's method [24], we obtain a computable exponential upper bound on the error probability of watermark decoding. In a sense, our problem can be described as a joint source-channel coding problem with side information at both the encoder and the decoder, and we study this problem from an error exponent viewpoint.

It is worth pointing out that Merhav [48] and Somekh-Baruch and Merhav [73] studied the error exponent performance for systems with finite alphabets in a private

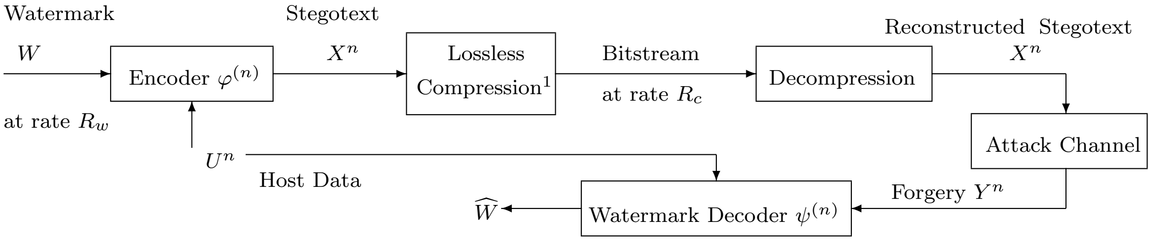


Figure 4.1: A model for joint compression and watermarking in a private scenario.

scenario. In [48], a single-letter expression of the Gallager random coding lower bound to the error exponent is obtained, while in [73], an asymptotic expression for the exact error exponent is derived. Note that the results of [48], [73] do not apply to the Gaussian Karakos-Papamarcou setup studied here, as they depend on the finiteness of the covertext and attack channel alphabets. Furthermore in [48], [73], different distortion constraints are imposed at the encoder.

The rest of this chapter is organized as follows. In Section 4.2, we give a formal description of the joint compression and watermarking problem and some preliminary results are presented. The main result is presented in Section 4.3. Section 4.4 provides some numerical examples. All proofs are left to Sections 4.5 –4.7. Finally, some concluding remarks are given in Section 4.8.

4.2 Problem Description

A general model for joint compression and watermarking in a private scenario is given in Fig. 4.1. Let $\{U_i\}_{i=1}^{\infty}$ be an independent and identically distributed (i.i.d.) sequence of zero mean Gaussian random variables with variance σ_u^2. Let $\mathcal{U} = \mathcal{X} = \mathbb{R}$, and

[1] Here the term "lossless compression" means a one-to-one binary representation of the stegotext X^n.

$d : \mathcal{U} \times \mathcal{X} \to [0, \infty)$ be a single-letter distortion measure. For $\mathbf{u} \in \mathcal{U}^n$ and $\mathbf{x} \in \mathcal{X}^n$, define

$$d(\mathbf{u}, \mathbf{x}) = \sum_{i=1}^{n} d(u_i, x_i). \tag{4.1}$$

In this paper, we consider the squared distortion measure, i.e.,

$$d(\mathbf{u}, \mathbf{x}) = \|\mathbf{u} - \mathbf{x}\|^2 = \sum_{i=1}^{n} (u_i - x_i)^2. \tag{4.2}$$

Let $A_{Y|X}$ be an additive white Gaussian noise (AWGN) channel with input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$ ($\mathcal{Y} = \mathcal{X} = \mathbb{R}$) so that $Y = X + Z$, where Z is Gaussian with mean zero and variance $D_a > 0$ which is independent of X.

Definition 4.1 An (R_c, R_w, n) joint compression and watermarking code consists of an encoder-decoder pair $(\varphi^{(n)}, \psi^{(n)})$:

$$\varphi^{(n)} \ : \ \mathcal{W} \times \mathcal{U}^n \to \mathcal{X}^n, \tag{4.3}$$

$$\psi^{(n)} \ : \ \mathcal{Y}^n \times \mathcal{U}^n \to \mathcal{W}, \tag{4.4}$$

where $\mathcal{W} = \{1, 2, \ldots, M_w\}$ is the watermark set and $M_w \triangleq \lceil e^{nR_w} \rceil$[2]. Given $w \in \mathcal{W}$ and $\mathbf{u} \in \mathcal{U}^n$, the stegotext $\mathbf{x}$ takes values from a set $\mathbf{c}$ of $M_c \triangleq \lceil e^{nR_c} \rceil$ codevectors, i.e., $\mathbf{c} \triangleq \{\mathbf{x}(1), \mathbf{x}(2), \ldots, \mathbf{x}(M_c)\}$.

Definition 4.2 Given an (R_c, R_w, n) code, the conditional probability of error in decoding a watermark index w is given by

$$P_{e,w}^{(n)} = \Pr \left\{ \widehat{w} \neq w | \, w \text{ is embedded} \right\}, \tag{4.5}$$

where $\widehat{w}$ is the decoded watermark message. Furthermore, if we assume that all watermark indices are equiprobable, the average probability of decoding error is given by

$$P_e^{(n)} = \frac{1}{M_w} \sum_{w=1}^{M_w} P_{e,w}^{(n)}. \tag{4.6}$$

[2] Throughout this chapter, all logarithms and exponentials are in the natural base.

Definition 4.3 Given an (R_c, R_w, n) joint compression and watermarking code, the average distortion between the host data and the stegotext is given by

$$D^{(n)} \triangleq \mathbb{E}\left[\frac{1}{n}d\Big(U^n, \varphi^{(n)}(W, U^n)\Big)\right]. \tag{4.7}$$

Definition 4.4 The transparency and robustness conditions for a sequence of (R_c, R_w, n) joint compression and watermarking codes require that for $D_c > 0$ and any $\epsilon, \delta > 0$,

$$D^{(n)} \leq D_c + \delta, \tag{4.8}$$

$$P_e^{(n)} \leq \epsilon, \tag{4.9}$$

for n sufficiently large.

Definition 4.5 A quadruple $(R_c, R_w; D_c, D_a)$ is said to be achievable if for every $\epsilon, \delta > 0$, there exists a sequence of (R_c, R_w, n) joint compression and watermarking codes such that $P_e^{(n)} \leq \epsilon$ and $D^{(n)} \leq D_c + \delta$ for n sufficiently large. Given (D_c, D_a), denote by $\mathcal{R}_{D_c,D_a}$ the set of all rate pairs (R_c, R_w) such that $(R_c, R_w; D_c, D_a)$ is achievable.

The achievable rate region has been derived for memoryless Gaussian sources and memoryless Gaussian attacks. The main result is summarized in the following theorem.

Theorem 4.1 [33]. The achievable rate region is given by

$$\begin{aligned}\mathcal{R}_{D_c,D_a} = \Bigg\{ &(R_c, R_w) : \\ &R_c \geq \left[\frac{1}{2}\log\Big(\frac{\sigma_u^2}{D_c}\Big)\right]^+, \\ &R_w \leq \max_{\gamma \in \left[\max\left\{1, \frac{\sigma_u^2}{D_c}\right\}, e^{2R_c}\right]} \min\Big[R_c - \frac{1}{2}\log\gamma, \frac{1}{2}\log\Big(1 + \frac{P_w(\gamma)}{D_a}\Big)\Big]\Bigg\},\end{aligned} \tag{4.10}$$

where $[a]^+ \triangleq \max\{a, 0\}$ and

$$P_w(\gamma) \triangleq \frac{\gamma(\sigma_u^2 + D_c) - 2\sigma_u^2 + 2\sqrt{\sigma_u^2(\gamma D_c - \sigma_u^2)(\gamma - 1)}}{\gamma^2}. \tag{4.11}$$

4.3 Main Results

Given an i.i.d. Gaussian covertext $\{U_i\}_{i=1}^{\infty}$, a distortion threshold D_c, and a Gaussian attack variance D_a, consider a rate pair $(R_c, R_w) \in \mathcal{R}_{D_c,D_a}$ (we assume that $D_c < \sigma_u^2$, which is a reasonable assumption in most practical applications). We refine the analysis of probability of error in detecting watermarks when $(R_c, R_w) \in \mathcal{R}_{D_c,D_a}$. Using a random coding argument and maximum-likelihood decoding technique, we obtain an exponential upper bound on the probability of error in decoding the watermark. The main result is the following theorem.

Theorem 4.2 Given $\delta > 0$, $s \geq 0$, $\rho \in [0, 1]$, $\beta^2 \in (D_c, \sigma_u^2)$, and $\gamma \in (\sigma_u^2/D_c, e^{2(R_c-R_w)})$, there exists a sequence of (R_c, R_w, n) joint compression and watermarking codes such that

$$D^{(n)} \leq D_c + \delta, \tag{4.12}$$

$$P_e^{(n)} \leq 4\exp\Big\{-n\big[\Lambda(\gamma, \beta, \rho, s) - o(1)\big]\Big\}, \tag{4.13}$$

for n sufficiently large, where $\Lambda(\gamma, \beta, \rho, s) \triangleq \min\big\{\Lambda_1(\gamma, \beta, \rho, s), \Lambda_2(\gamma, \beta)\big\}$, where

$$\begin{aligned}\Lambda_1(\gamma, \beta, \rho, s) \quad = \quad & \frac{1}{2}\log\left(\frac{1 + 2s\beta^2(\gamma-1)D_c + 2s(1+\rho)\sigma_u^2\theta}{\gamma}\right) \\ & + \frac{\rho}{2}\log\left(\frac{1 + 2s\beta^2(\gamma-1)D_c + \frac{(\gamma-1)D_c}{(1+\rho)D_a}}{\gamma}\right) - \rho R_w,\end{aligned} \tag{4.14}$$

with $\theta = \beta^2 - D_c - 2s\beta^2(\gamma-1)D_c^2$, and

$$\Lambda_2(\gamma, \beta) \quad = \quad \min\left\{\frac{1}{2}\Big(\frac{\beta^2}{\sigma_u^2} - 1 - \log\frac{\beta^2}{\sigma_u^2}\Big), \frac{1}{2}\Big(\frac{\gamma D_c}{\sigma_u^2} - 1 - \log\frac{\gamma D_c}{\sigma_u^2}\Big)\right\}, \tag{4.15}$$

and $o(1) \to 0$ as $n \to \infty$.

Proof. See Section 4.5.

Remark:

- From (4.15), it is clear that $\Lambda_2(\gamma, \beta)$ is always positive by the choice of $\gamma > \sigma_u^2/D_c$ and $\beta^2 < \sigma_u^2$. The condition $\gamma < e^{2(R_c - R_w)}$ is equivalent to $R_w < R_c - \frac{1}{2}\log\gamma$.

- The term $\Lambda_1(\gamma, \beta, \rho, s)$ is similar to the random coding lower bound derived in [24, pp. 337–343] for AWGN channels. However, here we deal with a distortion constraint at the channel input instead of a power constraint. The term $\Lambda_2(\gamma, \beta)$ is somewhat similar to the reliability function for Gaussian sources with respect to the rate-distortion pair $(R_c - R_w, D_c)$ [30].

$\Lambda(\gamma, \beta, \rho, s)$ can be tightened by optimizing it with respect to γ, β, ρ and s. In particular, we have the following result.

Corollary 4.1 . $\Lambda(\gamma, \beta, \rho, s)$ is maximized over $s \geq 0$ by[3]

$$s^* = \frac{1 - 2abc + \sqrt{(1-2abc)^2 + 4ac(\rho a + b)\frac{2+\rho}{1+\rho}}}{4ac(2+\rho)} \tag{4.16}$$

where

$$a \triangleq \frac{1}{\beta^2(\gamma-1)D_c + (1+\rho)\sigma_u^2(\beta^2 - D_c)}, \tag{4.17}$$

$$b \triangleq \frac{1}{\beta^2(\gamma-1)D_c} + \frac{1}{(1+\rho)\beta^2 D_a}, \tag{4.18}$$

$$c \triangleq \sigma_u^2\beta^2(\gamma-1)D_c^2. \tag{4.19}$$

Proof. See Section 4.6.

Corollary 4.2 . Let

$$E_R(R_c, R_w; D_c, D_a) \triangleq \sup_{\gamma \in (\frac{\sigma_u^2}{D_c}, e^{2(R_c - R_w)}),\, \beta^2 \in (D_c, \sigma_u^2),\, \rho \in [0,1],\, s \geq 0} \Lambda(\gamma, \beta, \rho, s).$$

[3] An analytical derivation of the other three optimizing parameters seems difficult. However the optimization can be carried numerically (e.g., see Section 4.4).

$E_R(R_c, R_w; D_c, D_a)$ is positive for all R_w satisfying

$$R_w < \min\left\{ R_c - \frac{1}{2}\log\frac{\sigma_u^2}{D_c}, \frac{1}{2}\log\left(1 + \frac{D_c - D_c^2/\sigma_u^2}{D_a}\right)\right\}. \tag{4.20}$$

Proof. See Section 4.7.

4.4 Examples

We next present some numerical examples to illustrate the results of the previous section. Figs. 4.2 and 4.3 show the random coding error exponent versus the watermarking rate R_w for various compression rates R_c and channel noise levels D_a. Fig. 4.4 shows a typical region of rate pairs where the random coding error exponent is positive, in addition to the overall achievable region $\mathcal{R}_{D_c,D_a}$ of Theorem 4.1 [33]. We note that $E_R(R_c, R_w; D_c, D_a) > 0$ nearly everywhere in $\mathcal{R}_{D_c,D_a}$.

Here, point A is given by $R_c = \frac{1}{2}\log(\frac{\sigma_u^2}{D_c}), R_w = 0$; B is given by $R_c = \frac{1}{2}\log(\frac{\sigma_u^2}{D_c} + \frac{\sigma_u^2 - D_c}{D_a}), R_w = \frac{1}{2}\log(1 + \frac{D_c - D_c^2/\sigma_u^2}{D_a})$; and C is given by $R_c = \frac{1}{2}\log(1 + \frac{\sigma_u^2}{D_c} + \frac{\sigma_u^2 + D_c}{D_a}), R_w = \frac{1}{2}\log(1 + \frac{D_c}{D_a})$ [33]. The figure shows that we can achieve all rates under the line segments AB and BB_∞. In fact, for segment AB, i.e., for $R_c < \frac{1}{2}\log(\frac{\sigma_u^2}{D_c} + \frac{\sigma_u^2 - D_c}{D_a})$, given any $(R_c, R_w) \in \mathcal{R}_{D_c,D_w}$, we have that $\sup_{\rho,s,\beta}\Lambda_1(\gamma,\beta,\rho,s) > 0$ for any given γ. Since $\Lambda_2(\gamma,\beta) > 0$ implies that $R_w < R_c - \frac{1}{2}\log\gamma$, we can approach segment AB by letting $\gamma \to \big(\frac{\sigma_u^2}{D_c}\big)^{+}$[4]. For segment BB_∞, if $R_w \geq \frac{1}{2}\log(1 + \frac{D_c - D_c^2/\sigma_u^2}{D_a})$, we will have $\Lambda_1(\gamma,\beta,\rho,s) \leq 0$ for any γ, β, ρ and s, which means $\Lambda(\gamma,\beta,\rho,s) \leq 0$. On the other hand, for $R_w < \frac{1}{2}\log(1 + \frac{D_c - D_c^2/\sigma_u^2}{D_a})$, we have $\sup_{\gamma,\beta,\rho,s}\Lambda_1(\gamma,\beta,\rho,s) > 0$. Since $\Lambda_2(\gamma,\beta)$ is always positive, we get a positive random coding error exponent. In this case, when

[4] Here $a \to (b)^+$ means $a \to b$ in such a way that $a > b$; similarly, $a \to (b)^-$ means $a \to b$ such that $a < b$.

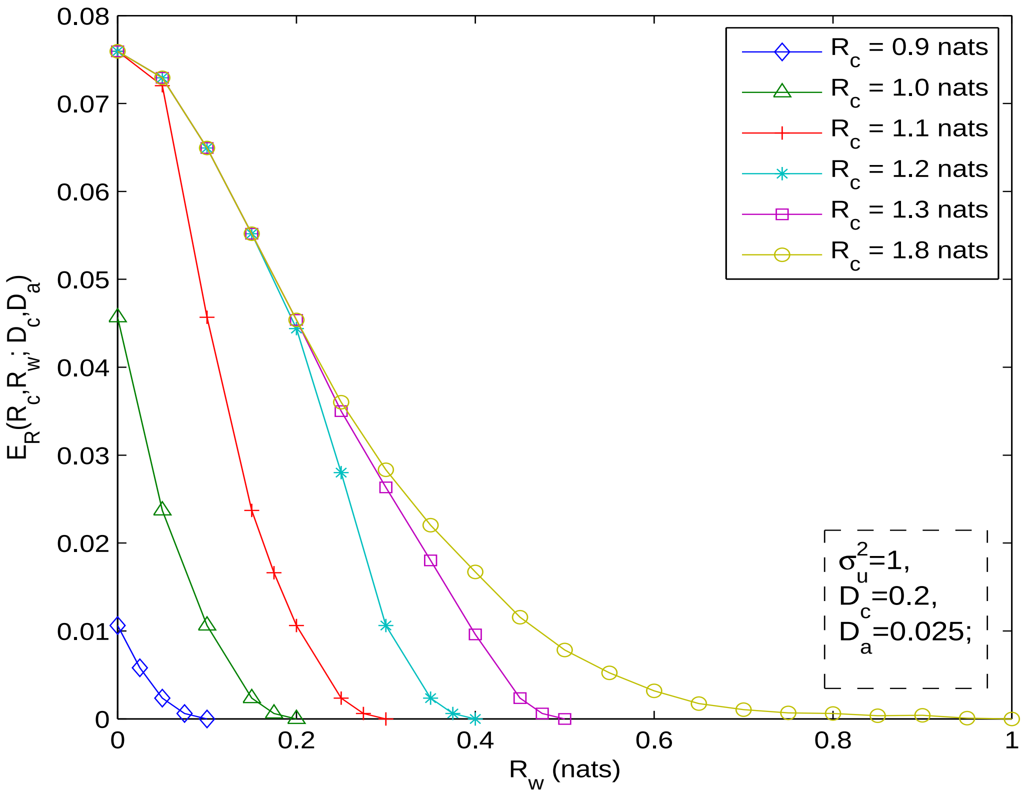


Figure 4.2: $E_R(R_c, R_w; D_c, D_a)$ v.s. R_w for various values of R_c.

we choose s as in (4.16), and letting $\gamma \to \left(\frac{\sigma_u^2}{D_c}\right)^+$, $\beta^2 \to \left(\sigma_u^2\right)^-$, and $\rho \to 0$, we can approach segment BB_∞ with $R_w \to \left(\frac{1}{2}\log(1 + \frac{D_c - D_c^2/\sigma_u^2}{D_a})\right)^-$.

Remark: As $\frac{D_c}{\sigma_u^2} \to 0$, point B approaches point C. In this case, the exponent is positive in the whole rate region. On the other hand, as $\frac{D_c}{\sigma_u^2} \to 1$, point B approaches point A. In this case, the positive exponent region becomes empty.

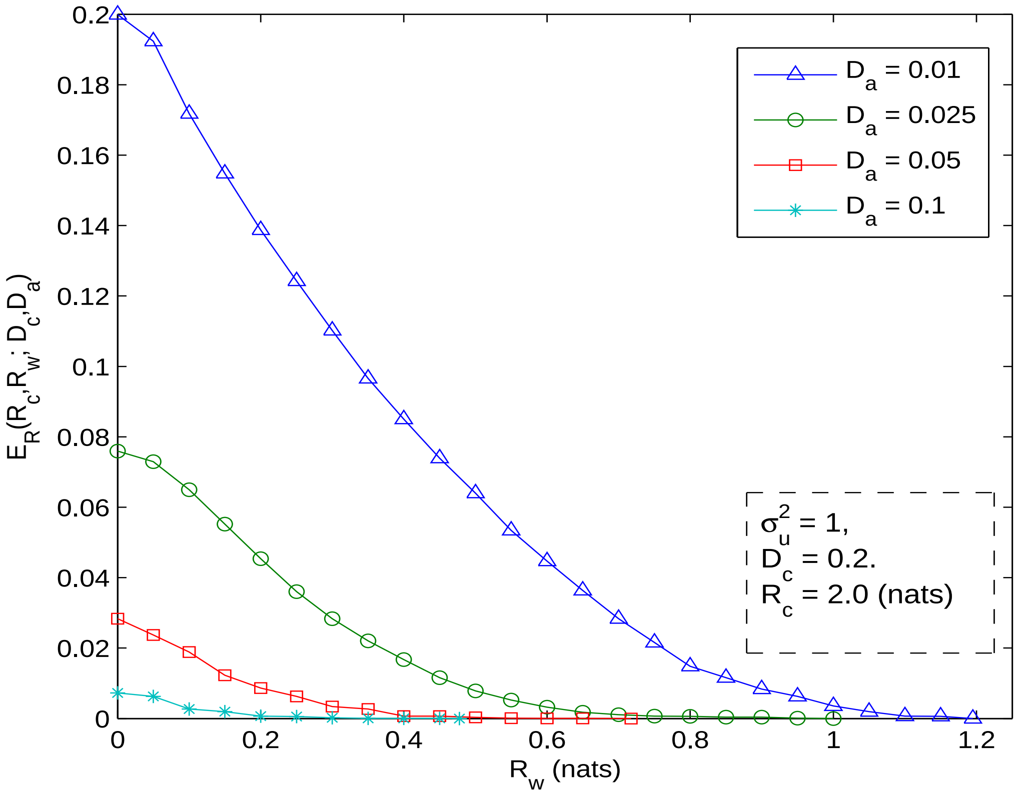


Figure 4.3: $E_R(R_c, R_w; D_c, D_a)$ v.s. R_w for various values of D_a.

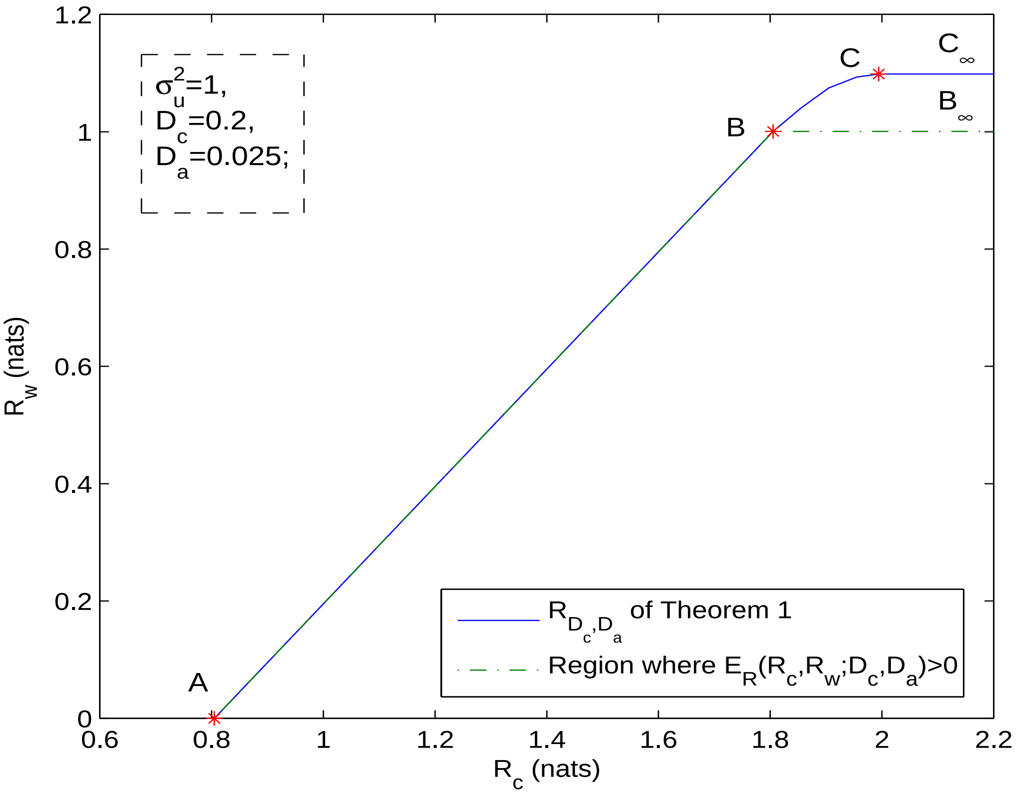


Figure 4.4: $\mathcal{R}_{D_c,D_a}$ of Theorem 4.1 and the region where the exponent $E_R(R_c, R_w; D_c, D_a)$ is positive.

4.5 Proof of Theorem 4.2

4.5.1 Outline of the Proof

We construct a class of random codes, and show that the average distortion and the average probability of error are satisfied with (4.12) and (4.13), respectively. We then show that there exists at least one such code satisfying (4.12) and (4.13) simultaneously. The analysis for the average distortion (Section 4.5.3) applies some properties of stationary memoryless Gaussian sources (see Lemma 4.1) and some techniques used to derive the Gaussian source reliability function [30]. The analysis for the average probability of error (Section 4.5.4) incorporates Gallager's random coding technique [24] and some bounds obtained in Section 4.5.3.

4.5.2 Code Construction

Given an i.i.d. Gaussian covertext $\{U_i\}_{i=1}^{\infty}$ with mean zero and variance σ_u^2, a distortion threshold D_c, and a Gaussian attack variance D_a, assume that $(R_c, R_w) \in \mathcal{R}_{D_c,D_a}$. Let $M \triangleq \lceil e^{n(R_c-R_w)} \rceil$, choose $\gamma \in (\frac{\sigma_u^2}{D_c}, e^{2(R_c-R_w)})$ and $\beta^2 \in (D_c, \sigma_u^2)$. Now consider a code $\mathbf{c}$ described as follows.

Random Code Generation. The code $\mathbf{c}$ contains $M_w = \lceil e^{nR_w} \rceil$ "subcodes" $\mathbf{c} \triangleq \{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_{M_w}\}$ which are assigned a product density function $q(\mathbf{c}) = \prod_{i=1}^{M_w} q(\mathbf{c}_i)$. Each $\mathbf{c}_i$ contains M codewords, i.e., $\mathbf{c}_i = \{\mathbf{x}(i,1), \ldots, \mathbf{x}(i,M)\}$, where each codeword $\mathbf{x}(i,j)$ is drawn i.i.d. according to $q(\mathbf{x}) = \prod_{k=1}^{n} q(x_k)$. Here $q(x_k)$ is the Gaussian density with mean zero and variance $\sigma_x^2 = (\gamma-1)D_c$. Thus for each $\mathbf{c}_i$, we have $q(\mathbf{c}_i) = \prod_{j=1}^{M} q(\mathbf{x}(i,j))$. Given the watermark index w, the subcode $\mathbf{c}_w$ will be used for quantizing the covertext $\mathbf{u} \in \mathcal{U}^n$.

Encoding. Given a watermarking index w and a covertext $\mathbf{u}$, the encoder chooses the first codeword $\mathbf{x}(w,t)$ in $\mathbf{c}_w$ such that $\|\mathbf{u}-\mathbf{x}(w,t)\|^2 \leq nD_c$, i.e.,

$$\begin{aligned}
&\|\mathbf{u}-\mathbf{x}(w,i)\|^2 > nD_c,\ i=1,\ldots,t-1,\\
&\|\mathbf{u}-\mathbf{x}(w,t)\|^2 \leq nD_c,\ t \leq M.
\end{aligned} \tag{4.21}$$

Denote the chosen codevector by $\mathbf{x}(\mathbf{c}_w,\mathbf{u})$. If no such $\mathbf{x}(w,t)$ exists, an error is declared and $\mathbf{x}(\mathbf{c}_w,\mathbf{u}) = \mathbf{0} \triangleq \mathbf{x}(0)$ will be sent.

Decoding. The decoder has full knowledge of $\mathbf{u}$, and thus can generate all possible watermarked versions $\{\mathbf{x}(\mathbf{c}_i,\mathbf{u})\}_{i=1}^{M_w}$. Upon receiving the "forgery" $\mathbf{y} = \mathbf{x}(\mathbf{c}_w,\mathbf{u}) + \mathbf{v}$, the decoder compares it with all $\{\mathbf{x}(\mathbf{c}_i,\mathbf{u})\}_{i\in\mathcal{I}}$, where $\mathcal{I} \triangleq \{i \in \mathcal{W} : \|\mathbf{u}-\mathbf{x}(\mathbf{c}_i,\mathbf{u})\|^2 \leq nD_c\}$, and chooses an output $\hat{w}$ using the maximum-likelihood decoding criterion:

$$\hat{w} = \arg\max_{i\in\mathcal{I}} f\big(\mathbf{y}|\mathbf{x}(\mathbf{c}_i,\mathbf{u})\big), \tag{4.22}$$

where $f(\cdot|\cdot)$ is the pdf for the additive Gaussian noise channel, i.e.,

$$f(\mathbf{y}|\mathbf{x}) = \prod_{j=1}^{n} f(y_j|x_j) = \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi D_a}} \exp\Big(-\frac{(y_j-x_j)^2}{2D_a}\Big). \tag{4.23}$$

4.5.3 Analysis for the Average Distortion

Define the following events

$$\begin{aligned}
\mathcal{E}_0(\mathbf{c}_w) &\triangleq \{\mathbf{u}\in\mathcal{U}^n :\ \text{embedding the watermark index } w \text{ into } \mathbf{u} \text{ with } \mathbf{c}_w \text{ is unsuccessful}\}\\
&= \{\mathbf{u}\in\mathcal{U}^n : \|\mathbf{u}-\mathbf{x}(w,i)\|^2 > nD_c,\ \mathbf{x}(w,i)\in\mathbf{c}_w, \forall\, i=1,2,\ldots,M\},\\
\mathcal{E}_1(\mathbf{u}) &\triangleq \{\mathbf{c}_w :\ \text{embedding the watermark index } w \text{ into } \mathbf{u} \text{ with } \mathbf{c}_w \text{ is unsuccessful}\}\\
&= \{\mathbf{c}_w : \|\mathbf{u}-\mathbf{x}(w,i)\|^2 > nD_c,\ \mathbf{x}(w,i)\in\mathbf{c}_w, \forall\, i=1,2,\ldots,M\}.
\end{aligned}$$

Given any $\mathbf{c} = \{\mathbf{c}_1,\ldots,\mathbf{c}_{M_w}\}$, the average distortion can be written as

$$D^{(n)}(\mathbf{c}) = \frac{1}{n}\mathbb{E}\left[\left\|U^n - \varphi^{(n)}(W,U^n)\right\|^2 \middle| \mathbf{c}\right]$$

$$
\begin{aligned}
&= \frac{1}{n}\sum_{w=1}^{M}\frac{1}{M}\int\limits_{\mathcal{U}^n} p(\mathbf{u})\|\mathbf{u}-\mathbf{x}(\mathbf{c}_w,\mathbf{u})\|^2\,d\mathbf{u}\\
&= \frac{1}{n}\sum_{w=1}^{M}\frac{1}{M}\left[\int\limits_{(\mathcal{E}_0(\mathbf{c}_w))^c} p(\mathbf{u})\|\mathbf{u}-\mathbf{x}(\mathbf{c}_w,\mathbf{u})\|^2\,d\mathbf{u}+\int\limits_{\mathcal{E}_0(\mathbf{c}_w)} p(\mathbf{u})\|\mathbf{u}-\mathbf{x}(\mathbf{c}_w,\mathbf{u})\|^2\,d\mathbf{u}\right]\\
&\leq \frac{1}{n}\sum_{w=1}^{M}\frac{1}{M}\left[\int\limits_{(\mathcal{E}_0(\mathbf{c}_w))^c} p(\mathbf{u})nD_c\,d\mathbf{u}+\int\limits_{\mathcal{E}_0(\mathbf{c}_w)} p(\mathbf{u})\|\mathbf{u}-\mathbf{0}\|^2\,d\mathbf{u}\right]\\
&\leq D_c+\frac{1}{n}\sum_{w=1}^{M}\frac{1}{M}\int\limits_{\mathcal{E}_0(\mathbf{c}_w)} p(\mathbf{u})\|\mathbf{u}\|^2\,d\mathbf{u}. \qquad (4.24)
\end{aligned}
$$

Then, the distortion averaged over the random choice of $\mathbf{c}$ is upper bounded by

$$
\begin{aligned}
\overline{D}^{(n)} &= \frac{1}{n}\mathbb{E}\left[\left\|U^n-\varphi^{(n)}(W,U^n)\right\|^2\right]\\
&= \int\limits_{\mathbf{c}} q(\mathbf{c})\frac{1}{n}\mathbb{E}\left[\left\|U^n-\varphi^{(n)}(W,U^n)\right\|^2\middle|\,\mathbf{c}\right]d\mathbf{c}\\
&\leq D_c+\frac{1}{n}\sum_{w=1}^{M}\frac{1}{M}\int\limits_{\mathbf{c}_w} q(\mathbf{c}_w)\int\limits_{\mathcal{E}_0(\mathbf{c}_w)} p(\mathbf{u})\|\mathbf{u}\|^2\,d\mathbf{u}\,d\mathbf{c}_w \qquad (4.25)
\end{aligned}
$$

where $d\mathbf{c}_w = d\mathbf{x}(w,1)\cdots d\mathbf{x}(w,M)$. Note that the second term in (4.25) is the average distortion over those $\mathbf{u}$'s for which the embedding of watermark index w into $\mathbf{u}$ is unsuccessful, averaged over all randomly chosen $\mathbf{c}_w$. By changing the order of the integration, we can also interpret this as the average distortion over those randomly chosen $\mathbf{c}_w$ such that the embedding of w into $\mathbf{u}$ is unsuccessful, averaged with respected to $p(\mathbf{u})$. Thus, we have

$$
\int\limits_{\mathbf{c}_w} q(\mathbf{c}_w)\int\limits_{\mathcal{E}_0(\mathbf{c}_w)} p(\mathbf{u})\|\mathbf{u}\|^2\,d\mathbf{u}\,d\mathbf{c}_w=\int\limits_{\mathcal{U}^n} p(\mathbf{u})\|\mathbf{u}\|^2\int\limits_{\mathcal{E}_1(\mathbf{u})} q(\mathbf{c}_w)\,d\mathbf{c}_w\,d\mathbf{u}. \qquad (4.26)
$$

We have

$$
\int\limits_{\mathcal{E}_1(\mathbf{u})} q(\mathbf{c}_w)\,d\mathbf{c}_w = \prod_{j=1}^{M}\left[\int\limits_{\mathcal{X}^n} q(\mathbf{x}(w,j))\big[1-\Phi_{D_c}(\mathbf{x}(w,j);\mathbf{u})\big]\,d\mathbf{x}(w,j)\right]
$$

$$= \left[\int\limits_{\mathcal{X}^n} q(\mathbf{x}) \big[1 - \Phi_{D_c}(\mathbf{x};\mathbf{u})\big] \, d\mathbf{x} \right]^M \tag{4.27}$$

where we have defined

$$\Phi_{D_c}(\mathbf{x};\mathbf{u}) = \begin{cases} 1, & d(\mathbf{x},\mathbf{u}) \le nD_c, \\ 0, & d(\mathbf{x},\mathbf{u}) > nD_c, \end{cases}$$

and (4.27) holds since the codewords are drawn independently according the same distribution $q(\mathbf{x})$.

We need the following lemma.

Lemma 4.1 [30] Let $\{X_i\}$ be an i.i.d. Gaussian source with distribution $X \sim \mathcal{N}(0,\sigma^2)$. For any $\Delta > 0$,

(a) if $a^2 = \sigma^2 + \Delta$, we have

$$\lim_{n\to\infty} \frac{1}{n} \log \Pr\left(\frac{1}{n} \sum_{i=1}^{n} |X_i - a|^2 \le \Delta \right) = -\frac{1}{2} \log \frac{a^2}{\Delta}; \tag{4.28}$$

(b) if $0 < \beta < \sigma$, then

$$\lim_{n\to\infty} \frac{1}{n} \log \Pr\left(\frac{1}{n} \|X^n\|^2 < \beta^2 \right) = -\frac{1}{2} \left(\frac{\beta^2}{\sigma^2} - 1 - \log \frac{\beta^2}{\sigma^2} \right); \tag{4.29}$$

(c) if $\alpha > \sigma$, then

$$\lim_{n\to\infty} \frac{1}{n} \log \Pr\left(\frac{1}{n} \|X^n\|^2 > \alpha^2 \right) = -\frac{1}{2} \left(\frac{\alpha^2}{\sigma^2} - 1 - \log \frac{\alpha^2}{\sigma^2} \right). \tag{4.30}$$

Now define

$$P_{ex}(\mathbf{u}, X^n) \triangleq \int\limits_{\mathcal{X}^n} q(\mathbf{x}) \big[1 - \Phi_{D_c}(\mathbf{x};\mathbf{u})\big] \, d\mathbf{x} = 1 - \Pr\left(\frac{1}{n} \|X^n - \mathbf{u}\|^2 \le D_c \right). \tag{4.31}$$

Let $\alpha^2 \triangleq \sigma_x^2 + D_c = \gamma D_c$ (recall that $\gamma > \sigma_u^2/D_c$ implies $\alpha^2 > \sigma_u^2$), $\beta_1^2 \in (D_c, \sigma_u^2)$, $\delta_0 \triangleq \sigma_u^2 - \beta_1^2$, and define $\mathcal{B}_n(\alpha, \beta_1) = \{\mathbf{u} \in \mathcal{U}^n : n\beta_1^2 \le \|\mathbf{u}\|^2 \le n\alpha^2\}$. Since the pdf of the n-tuple X^n is strictly decreasing in $\|\mathbf{x}\|^2$, we know that, for $\mathbf{u} \in \mathcal{B}_n(\alpha, \beta_1)$,

$$\Pr\left(\frac{1}{n}\|X^n - \mathbf{u}\|^2 \leq D_c\right) \geq \Pr\left(\frac{1}{n}\sum_{i=1}^{n} \mid X_i - \alpha \mid^2 \leq D_c\right). \tag{4.32}$$

Applying Lemma 4.1, we have

$$\lim_{n\to\infty} \frac{1}{n} \log \Pr\left(\frac{1}{n}\sum_{i=1}^{n} \mid X_i - \alpha \mid^2 \leq D_c\right) = -\frac{1}{2}\log\frac{\alpha^2}{D_c} = -\frac{1}{2}\log\gamma. \tag{4.33}$$

Thus, given any $0 < \epsilon_1^{(n)} < \epsilon_0^{(n)}$ such that $R_c - R_w > \frac{1}{2}\log\gamma + \epsilon_0^{(n)}$ (here $\gamma < e^{2(R_c - R_w)}$ guarantees the existence of such $\epsilon_0^{(n)}$), there exists a positive integer N_1 such that for $n \geq N_1$,

$$\Pr\left(\frac{1}{n}\sum_{i=1}^{n} |X_i - \alpha|^2 \leq D_c\right) \geq \exp\Big\{-n\Big(\frac{1}{2}\log\gamma + \epsilon_1^{(n)}\Big)\Big\}. \tag{4.34}$$

Therefore,

$$\Pr\left(\frac{1}{n}\|X^n - \mathbf{u}\|^2 \leq D_c\right) \geq \exp\Big\{-n\Big(\frac{1}{2}\log\gamma + \epsilon_1^{(n)}\Big)\Big\}. \tag{4.35}$$

Hence, for $n \geq N_1$ and $\mathbf{u} \in \mathcal{B}_n(\alpha, \beta_1)$, we have

$$\begin{aligned}
&\big[P_{ex}(\mathbf{u}, X^n)\big]^M \\
&= \Big[1 - \Pr\Big(\frac{1}{n}\|X^n - \mathbf{u}\|^2 \leq D_c\Big)\Big]^{e^{n(R_c - R_w)}} \\
&\leq \exp\left\{-\Pr\Big\{\frac{1}{n}\|X^n - \mathbf{u}\|^2 \leq D_c\Big\} e^{n(R_c - R_w)}\right\} && (4.36) \\
&\leq \exp\left\{-\exp\Big\{-n\Big(\frac{1}{2}\log\gamma + \epsilon_1^{(n)}\Big)\Big\}\exp\Big\{n\Big(\frac{1}{2}\log\gamma + \epsilon_0^{(n)}\Big)\Big\}\right\} && (4.37) \\
&= \exp\Big\{-\exp\big(n(\epsilon_0^{(n)} - \epsilon_1^{(n)})\big)\Big\} \triangleq \delta_1^{(n)}, && (4.38)
\end{aligned}$$

where we have used the inequality $(1-t)^k \leq e^{-tk}$ for $0 \leq t \leq 1$, $k > 0$ in (4.36), and $\delta_1^{(n)} \to 0$ as $n \to \infty$.

Now combining (4.25)–(4.31) and (4.38), we obtain

$$\overline{D}^{(n)} \leq D_c + \frac{1}{n}\int_{\mathcal{U}^n} p(\mathbf{u})\|\mathbf{u}\|^2 \big[P_{ex}(\mathbf{u}, X^n)\big]^M d\mathbf{u}$$

$$
\begin{aligned}
&\le D_c + \frac{1}{n}\int\limits_{\mathcal{B}_n(\alpha,\beta_1)} p(\mathbf{u})\|\mathbf{u}\|^2\big[P_{ex}(\mathbf{u},X^n)\big]^M\,d\mathbf{u}\\
&\quad + \frac{1}{n}\int\limits_{(\mathcal{B}_n(\alpha,\beta_1))^c} p(\mathbf{u})\|\mathbf{u}\|^2\big[P_{ex}(\mathbf{u},X^n)\big]^M\,d\mathbf{u}\\
&\le D_c + \delta_1^{(n)}\frac{1}{n}\int\limits_{\mathcal{B}_n(\alpha,\beta_1)} p(\mathbf{u})\|\mathbf{u}\|^2\,d\mathbf{u} + \frac{1}{n}\int\limits_{(\mathcal{B}_n(\alpha,\beta_1))^c} p(\mathbf{u})\|\mathbf{u}\|^2\,d\mathbf{u}\\
&\le D_c + \delta_1^{(n)}\sigma_u^2 + \frac{1}{n}\int\limits_{(\mathcal{B}_n(\alpha,\beta_1))^c} p(\mathbf{u})\|\mathbf{u}\|^2\,d\mathbf{u}. \qquad (4.39)
\end{aligned}
$$

Observe that

$$
\begin{aligned}
&\frac{1}{n}\int\limits_{(\mathcal{B}_n(\alpha,\beta_1))^c} p(\mathbf{u})\|\mathbf{u}\|^2\,d\mathbf{u}\\
&= \sigma_u^2 - \frac{1}{n}\int\limits_{\mathcal{B}_n(\alpha,\beta_1)} p(\mathbf{u})\|\mathbf{u}\|^2\,d\mathbf{u}\\
&\le \sigma_u^2 - \frac{1}{n}\int\limits_{\mathcal{B}_n(\alpha,\beta_1)} p(\mathbf{u})n\beta_1^2\,d\mathbf{u}\\
&= \sigma_u^2 - \beta_1^2\Big(1-\Pr\big\{\|U^n\|^2 < n\beta_1^2\big\} - \Pr\big\{\|U^n\|^2 > n\alpha^2\big\}\Big)\\
&\le \delta_0 + \sigma_u^2\Big(\Pr\big\{\|U^n\|^2 < n\beta_1^2\big\} + \Pr\big\{\|U^n\|^2 > n\alpha^2\big\}\Big). \qquad (4.40)
\end{aligned}
$$

By Lemma 4.1, there exists $\epsilon_2^{(n)} > 0$ and a positive integer N_2 such that, for $\forall\ n \ge N_2$, we have

$$
\Pr\big\{\|U^n\|^2 < n\beta_1^2\big\} \le \exp\left\{-n\Big[\frac{1}{2}\Big(\frac{\beta_1^2}{\sigma_u^2} - 1 - \log\frac{\beta_1^2}{\sigma_u^2}\Big) - \epsilon_2^{(n)}\Big]\right\} \triangleq \delta_2^{(n)} \qquad (4.41)
$$

and

$$
\Pr\big\{\|U^n\|^2 > n\alpha^2\big\} \le \exp\left\{-n\Big[\frac{1}{2}\Big(\frac{\alpha^2}{\sigma_u^2} - 1 - \log\frac{\alpha^2}{\sigma_u^2}\Big) - \epsilon_2^{(n)}\Big]\right\} \triangleq \delta_3^{(n)}. \qquad (4.42)
$$

Plugging the above bounds back into (4.39), and choosing $n \ge \max\{N_1, N_2\}$, we obtain

$$
\overline{D}^{(n)} \;\le\; D_c + \delta_0 + \sigma_u^2\left(\delta_1^{(n)} + \delta_2^{(n)} + \delta_3^{(n)}\right) \triangleq D_c + \bar{\delta}^{(n)}, \qquad (4.43)
$$

where $\bar{\delta}^{(n)}$ can be made arbitrarily small by choosing $\sigma_u^2 - \beta_1^2$ sufficiently small and n sufficiently large.

4.5.4 Analysis for the Average Probability of Error

Recall that $\mathcal{E}_1(\mathbf{u}) = \{\mathbf{c}_w : \text{embedding } w \text{ into } \mathbf{u} \text{ with } \mathbf{c}_w \text{ is unsuccessful}\}$. Given a randomly chosen codebook $\mathbf{c} = \{\mathbf{c}_1, \ldots, \mathbf{c}_{M_w}\}$, denote by $\Pr(\text{error}|w, \mathbf{x}(\mathbf{c}_w, \mathbf{u}), \mathbf{y})$ the probability of decoding error conditioned, first, on w and $\mathbf{u}$ entering the encoder, second, on the selection of a codeword $\mathbf{x}(w, i) \in \mathbf{c}_w$, denoted as $\mathbf{x}(\mathbf{c}_w, \mathbf{u})$, and on the channel output $\mathbf{y}$. Let $\beta^2 \in (D_c, \sigma_u^2)$ and define $\mathcal{B}_n(\alpha, \beta) = \{\mathbf{u} \in \mathcal{U}^n : n\beta^2 \leq \|\mathbf{u}\|^2 \leq n\alpha^2\}$. Then the probability of decoding error given that watermark index w was embedded, averaged over the random choice of $\mathbf{c}$, satisfies

$$\begin{aligned}
\overline{P}_{e,w}^{(n)} &= \int\limits_{\mathcal{U}^n} p(\mathbf{u}) \int\limits_{\mathbf{c}_w} q(\mathbf{c}_w) \int\limits_{\mathcal{Y}^n} f(\mathbf{y}|\mathbf{x}(\mathbf{c}_w, \mathbf{u})) \Pr(\text{error}|w, \mathbf{x}(\mathbf{c}_w, \mathbf{u}), \mathbf{y})\, d\mathbf{y}\, d\mathbf{c}_w\, d\mathbf{u} \\
&\leq \int\limits_{\mathcal{B}_n(\alpha,\beta)} p(\mathbf{u}) \int\limits_{(\mathcal{E}_1(\mathbf{u}))^c} q(\mathbf{c}_w) \int\limits_{\mathcal{Y}^n} f(\mathbf{y}|\mathbf{x}(\mathbf{c}_w, \mathbf{u})) \Pr(\text{error}|w, \mathbf{x}(\mathbf{c}_w, \mathbf{u}), \mathbf{y})\, d\mathbf{y}\, d\mathbf{c}_w\, d\mathbf{u} \\
&\quad + \int\limits_{\mathcal{B}_n(\alpha,\beta)} p(\mathbf{u}) \int\limits_{\mathcal{E}_1(\mathbf{u})} q(\mathbf{c}_w)\, d\mathbf{c}_w\, d\mathbf{u} + \int\limits_{(\mathcal{B}_n(\alpha,\beta))^c} p(\mathbf{u})\, d\mathbf{u} \\
&\triangleq P_0 + P_1 + P_2
\end{aligned} \tag{4.44}$$

Following Gallager's technique for deriving the random coding lower bound for the channel error exponent [24], we can upper bound P_0. Given w, $\mathbf{x}(\mathbf{c}_w, \mathbf{u})$ and $\mathbf{y}$, define $\mathcal{E}_{w'}$ as the event that

$$f\big(\mathbf{y}|\mathbf{x}(\mathbf{c}_{w'}, \mathbf{u})\big) \geq f\big(\mathbf{y}|\mathbf{x}(\mathbf{c}_w, \mathbf{u})\big) \tag{4.45}$$

where $\mathbf{x}(\mathbf{c}_{w'}, \mathbf{u})$ is the codeword of embedding w' into $\mathbf{u}$ via the codebook $\mathbf{c}_{w'}$. Then we have for any $\rho \in [0, 1]$ and $r > 0$,

$$\Pr(\text{error}|w, \mathbf{x}(\mathbf{c}_w, \mathbf{u}), \mathbf{y})$$

$$
\begin{aligned}
&\leq \Pr\left(\bigcup_{w' \neq w} \mathcal{E}_{w'}\right) \\
&\leq \left[\sum_{w' \neq w} \Pr(\mathcal{E}_{w'})\right]^{\rho} \\
&\leq \left[\sum_{w' \neq w} \int\limits_{(\mathcal{E}_1(\mathbf{u}))^c} q(\mathbf{c}_{w'}) \left(\frac{f(\mathbf{y}|\mathbf{x}(\mathbf{c}_{w'}, \mathbf{u}))}{f(\mathbf{y}|\mathbf{x}(\mathbf{c}_{w}, \mathbf{u}))}\right)^{r} d\mathbf{c}_{w'}\right]^{\rho} \\
&= \left[(e^{nR_w} - 1) \int\limits_{(\mathcal{E}_1(\mathbf{u}))^c} q(\mathbf{c}_{w'}) \left(\frac{f(\mathbf{y}|\mathbf{x}(\mathbf{c}_{w'}, \mathbf{u}))}{f(\mathbf{y}|\mathbf{x}(\mathbf{c}_{w}, \mathbf{u}))}\right)^{r} d\mathbf{c}_{w'}\right]^{\rho}. \qquad (4.46)
\end{aligned}
$$

Plugging (4.46) into (4.44), we get

$$
\begin{aligned}
P_0 \;\leq\; & e^{n\rho R_w} \int\limits_{\mathcal{B}_n(\alpha,\beta)} p(\mathbf{u}) \int\limits_{\mathcal{Y}^n} \left[\int\limits_{(\mathcal{E}_1(\mathbf{u}))^c} q(\mathbf{c}_w) f(\mathbf{y}|\mathbf{x}(\mathbf{c}_w, \mathbf{u}))^{1-r\rho}\, d\mathbf{c}_w\right] \\
& \left[\int\limits_{(\mathcal{E}_1(\mathbf{u}))^c} q(\mathbf{c}_{w'}) f(\mathbf{y}|\mathbf{x}(\mathbf{c}_{w'}, \mathbf{u}))^{r} d\mathbf{c}_{w'}\right]^{\rho} d\mathbf{y}\, d\mathbf{u}. \qquad (4.47)
\end{aligned}
$$

Substituting $r = 1/(1+\rho)$, we get

$$
P_0 \leq e^{n\rho R_w} \int\limits_{\mathcal{B}_n(\alpha,\beta)} p(\mathbf{u}) \int\limits_{\mathcal{Y}^n} \left[\int\limits_{(\mathcal{E}_1(\mathbf{u}))^c} q(\mathbf{c}_w) f(\mathbf{y}|\mathbf{x}(\mathbf{c}_w, \mathbf{u}))^{\frac{1}{1+\rho}}\, d\mathbf{c}_w\right]^{1+\rho} d\mathbf{y}\, d\mathbf{u}. \qquad (4.48)
$$

Let

$$
\eta(\mathbf{u}, \mathbf{y}) \triangleq \int\limits_{(\mathcal{E}_1(\mathbf{u}))^c} q(\mathbf{c}_w) f(\mathbf{y}|\mathbf{x}(\mathbf{c}_w, \mathbf{u}))^{\frac{1}{1+\rho}}\, d\mathbf{c}_w. \qquad (4.49)
$$

We have

$$
\begin{aligned}
&\eta(\mathbf{u}, \mathbf{y}) \\
&= \int\limits_{\mathcal{X}^n} \cdots \int\limits_{\mathcal{X}^n} \left(1 - \prod_{i=1}^{M}\Big(1 - \Phi_{D_c}\big(\mathbf{x}(w,i), \mathbf{u}\big)\Big)\right) \prod_{i=1}^{M} q\big(\mathbf{x}(w,j)\big)
\end{aligned}
$$

$$f\big(\mathbf{y}|\mathbf{x}(\mathbf{c}_w,\mathbf{u})\big)^{\frac{1}{1+\rho}}\,d\mathbf{x}(w,1)\ldots d\mathbf{x}(w,M)$$

$$= \sum_{i=1}^{M}\prod_{j=1}^{i-1}\int_{\mathcal{X}^n} q\big(\mathbf{x}(w,j)\big)\big[1-\Phi_{D_c}(\mathbf{x}(w,j);\mathbf{u})\big]\,d\mathbf{x}(w,j)$$

$$\int_{\mathcal{X}^n} q\big(\mathbf{x}(w,i)\big)\Phi_{D_c}\big(\mathbf{x}(w,i);\mathbf{u}\big)f\big(\mathbf{y}|\mathbf{x}(w,i)\big)^{\frac{1}{1+\rho}}\,d\mathbf{x}(w,i)$$

$$= \sum_{i=1}^{M}\left[\int_{\mathcal{X}^n} q(\mathbf{x})\big[1-\Phi_{D_c}\big(\mathbf{x};\mathbf{u}\big)\big]\,d\mathbf{x}\right]^{i-1}\int_{\mathcal{X}^n} q(\mathbf{x})\Phi_{D_c}(\mathbf{x};\mathbf{u})f(\mathbf{y}|\mathbf{x})^{\frac{1}{1+\rho}}\,d\mathbf{x}$$

$$= \sum_{i=1}^{M}\big[P_{ex}(\mathbf{u},X^n)\big]^{i-1}\int_{\mathcal{X}^n} q(\mathbf{x})\Phi_{D_c}(\mathbf{x};\mathbf{u})f(\mathbf{y}|\mathbf{x})^{\frac{1}{1+\rho}}\,d\mathbf{x} \tag{4.50}$$

$$= \frac{1-\big[P_{ex}(\mathbf{u},X^n)\big]^M}{1-P_{ex}(\mathbf{u},X^n)}\int_{\mathcal{X}^n} q(\mathbf{x})\Phi_{D_c}(\mathbf{x};\mathbf{u})f(\mathbf{y}|\mathbf{x})^{\frac{1}{1+\rho}}\,d\mathbf{x}$$

$$\leq \frac{1}{1-P_{ex}(\mathbf{u},X^n)}\int_{\mathcal{X}^n} q(\mathbf{x})\Phi_{D_c}(\mathbf{x};\mathbf{u})f(\mathbf{y}|\mathbf{x})^{\frac{1}{1+\rho}}\,d\mathbf{x}, \tag{4.51}$$

where in the second equality we have used the fact that

$$1-\prod_{i=1}^{M}\Big(1-\Phi_{D_c}\big(\mathbf{x}(w,i),\mathbf{u}\big)\Big) = \sum_{i=1}^{M}\prod_{j=1}^{i-1}\Big(1-\Phi_{D_c}\big(\mathbf{x}(w,j),\mathbf{u}\big)\Big)\Phi_{D_c}\big(\mathbf{x}(w,i),\mathbf{u}\big).$$

Applying the inequality $\Phi_{D_c}(\mathbf{x},\mathbf{u}) \leq \exp\Big\{s\big[nD_c-d(\mathbf{x},\mathbf{u})\big]\frac{\|\mathbf{u}\|^2}{n}\Big\}$ $(s\geq 0)$, we have

$$\int_{\mathcal{X}^n} q(\mathbf{x})\Phi_{D_c}(\mathbf{x};\mathbf{u})f(\mathbf{y}|\mathbf{x})^{\frac{1}{1+\rho}}\,d\mathbf{x}$$

$$\leq \int_{\mathcal{X}^n} q(\mathbf{x})\exp\Big\{s\big[nD_c-d(\mathbf{x},\mathbf{u})\big]\frac{\|\mathbf{u}\|^2}{n}\Big\}f(\mathbf{y}|\mathbf{x})^{\frac{1}{1+\rho}}\,d\mathbf{x}$$

$$\leq \int_{\mathcal{X}^n} q(\mathbf{x})\exp\Big\{s\big[D_c\|\mathbf{u}\|^2-\beta^2 d(\mathbf{x},\mathbf{u})\big]\Big\}f(\mathbf{y}|\mathbf{x})^{\frac{1}{1+\rho}}\,d\mathbf{x} \tag{4.52}$$

$$= \prod_{i=1}^{n}\int_{\mathcal{X}} q(x_i)\exp\Big\{s\big[D_c u_i^2-\beta^2(x_i-u_i)^2\big]\Big\}f(y_i|x_i)^{\frac{1}{1+\rho}}\,dx_i$$

$$= \prod_{i=1}^{n}\int_{\mathcal{X}} \frac{1}{\sqrt{2\pi\sigma_x^2}}\left(\frac{1}{\sqrt{2\pi D_a}}\right)^{\frac{1}{1+\rho}}$$

$$\exp\left[-\frac{x_i^2}{2\sigma_x^2}+s\big(D_c u_i^2-\beta^2(x_i-u_i)^2\big)-\frac{(y_i-x_i)^2}{2(1+\rho)D_a}\right]\,dx_i$$

$$= \prod_{i=1}^{n}\tau\left(\frac{1}{\sqrt{2\pi D_a}}\right)^{\frac{1}{1+\rho}}$$

$$\exp\Big\{\big(2\sigma_x^2\tau^2s^2\beta^4-s\beta^2+sD_c\big)u_i^2+\big(2\sigma_x^2\kappa^2\tau^2-\kappa\big)y_i^2+4\kappa s\tau^2\beta^2\sigma_x^2u_iy_i\Big\} \quad (4.53)$$

where

$$\kappa\triangleq\frac{1}{2(1+\rho)D_a},\quad \tau\triangleq\frac{1}{\sqrt{1+2(\kappa+s\beta^2)\sigma_x^2}}, \quad (4.54)$$

and in (4.52) we have used the fact that $\|\mathbf{u}\|^2\geq n\beta^2$ for $\mathbf{u}\in\mathcal{B}_n(\alpha,\beta)$. Using this bound when integrating (4.51) over $\mathbf{y}$, we get

$$\int\limits_{\mathcal{Y}^n}\left[\int\limits_{\mathcal{X}^n}q(\mathbf{x})\Phi_{D_c}(\mathbf{x};\mathbf{u})f(\mathbf{y}|\mathbf{x})^{\frac{1}{1+\rho}}d\mathbf{x}\right]^{1+\rho}d\mathbf{y}$$

$$\leq \prod_{i=1}^{n}\tau^{1+\rho}\exp\Big\{(1+\rho)\big[(2\sigma_x^2\tau^2s^2\beta^4-s\beta^2+sD_c)u_i^2\big]\Big\}$$

$$\times\int\limits_{\mathcal{Y}}\frac{1}{\sqrt{2\pi D_a}}\exp\left\{\frac{(2\sigma_x^2\kappa\tau^2-1)y_i^2+4\sigma_x^2\tau^2s\beta^2u_iy_i}{2D_a}\right\}dy_i. \quad (4.55)$$

Since

$$\int\limits_{\mathcal{Y}}\frac{1}{\sqrt{2\pi D_a}}\exp\left\{\frac{(2\sigma_x^2\kappa\tau^2-1)y_i^2+4\sigma_x^2\tau^2s\beta^2u_iy_i}{2D_a}\right\}dy_i$$

$$=\exp\left\{\frac{4\sigma_x^4\tau^4s^2\beta^4u_i^2}{2D_a(1-2\sigma_x^2\kappa\tau^2)}\right\}\int\limits_{\mathcal{Y}}\frac{1}{\sqrt{2\pi D_a}}\exp\left\{-\frac{[y_i+(2\sigma_x^2\tau^2s\beta^2u_i)/(2\sigma_x^2\kappa\tau^2-1)]^2}{2D_a/(1-2\sigma_x^2\kappa\tau^2)}\right\}dy_i$$

$$=\exp\left\{\frac{4\sigma_x^4\tau^4s^2\beta^4u_i^2}{2D_a(1-2\sigma_x^2\kappa\tau^2)}\right\}\frac{1}{\sqrt{1-2\sigma_x^2\kappa\tau^2}}.$$

Plugging the above expression back into (4.55), we get

$$\int\limits_{\mathcal{Y}^n}\left[\int\limits_{\mathcal{X}^n}q(\mathbf{x})\Phi_{D_c}(\mathbf{x};\mathbf{u})f(\mathbf{y}|\mathbf{x})^{\frac{1}{1+\rho}}\,d\mathbf{x}\right]^{1+\rho}d\mathbf{y}$$

$$\leq \prod_{i=1}^{n} \frac{\tau^{\rho}}{\sqrt{1+2s\beta^2\sigma_x^2}} \exp\left\{-\left(\frac{(1+\rho)s\beta^2}{1+2s\beta^2\sigma_x^2} - s(1+\rho)D_c\right)u_i^2\right\}. \tag{4.56}$$

Substituting this into (4.48), we obtain

$$\begin{aligned}
P_0 &\leq e^{n\rho R_w} \int\limits_{\mathcal{B}_n(\alpha,\beta)} p(\mathbf{u}) \int\limits_{\mathcal{Y}^n} \left[\int\limits_{(\mathcal{E}_1(\mathbf{u}))^c} q(\mathbf{c}_w) f(\mathbf{y}|\mathbf{x}(\mathbf{c}_w,\mathbf{u}))^{\frac{1}{1+\rho}}\, d\mathbf{c}_w\right]^{1+\rho} d\mathbf{y}\, d\mathbf{u} \\
&\leq e^{n\rho R_w}\left[\frac{\tau^{\rho}}{\sqrt{1+2s\beta^2\sigma_x^2}}\right]^n \int\limits_{\mathcal{B}_n(\alpha,\beta)} p(\mathbf{u}) \frac{1}{\big(1-P_{ex}(\mathbf{u},X^n)\big)^{1+\rho}} \\
&\qquad \prod_{i=1}^{n} \exp\left\{-\left(\frac{(1+\rho)s\beta^2}{1+2s\beta^2\sigma_x^2} - s(1+\rho)D_c\right)u_i^2\right\} d\mathbf{u} \\
&\leq \left[\frac{\tau^{\rho}\exp\Big\{(1+\rho)(\frac{1}{2}\log\gamma + \epsilon_1^{(n)}) + \rho R_w\Big\}}{\sqrt{1+2s\beta^2\sigma_x^2}}\right]^n \\
&\qquad \int\limits_{\mathcal{U}^n} p(\mathbf{u}) \prod_{i=1}^{n} \exp\left\{-\left(\frac{(1+\rho)s\beta^2}{1+2s\beta^2\sigma_x^2} - s(1+\rho)D_c\right)u_i^2\right\} d\mathbf{u} \\
&= \left[\frac{\tau^{\rho}\exp\Big\{(1+\rho)(\frac{1}{2}\log\gamma + \epsilon_1^{(n)}) + \rho R_w\Big\}}{\sqrt{1+2s\beta^2\sigma_x^2 + 2s\beta^2(1+\rho)\sigma_u^2 - 2s(1+\rho)(1+2s\beta^2\sigma_x^2)\sigma_u^2 D_c}}\right]^n
\end{aligned} \tag{4.57}$$

where in the second inequality we used (4.35).

To bound P_2, recall that for $n \geq N_2$, we have

$$\begin{aligned}
P_2 &= \int\limits_{(\mathcal{B}_n(\alpha,\beta))^c} p(\mathbf{u})\, d\mathbf{u} \\
&= \Pr\big\{\|U^n\|^2 < n\beta^2\big\} + \Pr\big\{\|U^n\|^2 > n\alpha^2\big\} \\
&\leq \exp\left\{-n\Big[\frac{1}{2}\Big(\frac{\beta^2}{\sigma_u^2} - 1 - \log\frac{\beta^2}{\sigma_u^2}\Big) - \epsilon_2^{(n)}\Big]\right\} \\
&\quad + \exp\left\{-n\Big[\frac{1}{2}\Big(\frac{\alpha^2}{\sigma_u^2} - 1 - \log\frac{\alpha^2}{\sigma_u^2}\Big) - \epsilon_2^{(n)}\Big]\right\}.
\end{aligned} \tag{4.58}$$

To bound P_1, we use (4.27) and (4.38) to obtain

$$\begin{aligned} P_1 &= \int\limits_{\mathcal{B}_n(\alpha,\beta)} p(\mathbf{u}) \int\limits_{\mathcal{E}_1(\mathbf{u})} q(\mathbf{c}_w)\, d\mathbf{c}_w\, d\mathbf{u} \\ &\leq \exp\Big\{-\exp\big(n(\epsilon_0^{(n)} - \epsilon_1^{(n)})\big)\Big\}. \end{aligned} \tag{4.59}$$

Since the right hand side of (4.59) vanishes at a double exponential speed by choosing $0 < \epsilon_1^{(n)} < \epsilon_0^{(n)}$, the exponential converges to zero. Thus the exponential of $\overline{P}_{e,w}^{(n)}$ is dominated by (4.57) and (4.58).

Noting that $\epsilon_1^{(n)}, \epsilon_2^{(n)} > 0$ can be arbitrarily small thus be absorbed into γ, α and β, define

$$\begin{aligned} &\Lambda_1(\gamma, \beta, \rho, s) \\ &\triangleq \frac{1}{2}\log\Big(1 + 2s\beta^2\sigma_x^2 + 2s(1+\rho)\sigma_u^2\Big(\beta^2 - D_c - 2sD_c\beta^2\sigma_x^2\Big)\Big) \\ &\quad -\rho\log\tau - \frac{1+\rho}{2}\log\gamma - \rho R_w \\ &= \frac{1}{2}\log\left(\frac{1 + 2s\beta^2(\gamma-1)D_c + 2s(1+\rho)\sigma_u^2\big(\beta^2 - D_c - 2s\beta^2(\gamma-1)D_c^2\big)}{\gamma}\right) \\ &\qquad + \frac{\rho}{2}\log\left(\frac{1 + 2s\beta^2(\gamma-1)D_c + \frac{(\gamma-1)D_c}{(1+\rho)D_a}}{\gamma}\right) - \rho R_w, \end{aligned} \tag{4.60}$$

and

$$\Lambda_2(\gamma, \beta) = \min\left\{\frac{1}{2}\Big(\frac{\beta^2}{\sigma_u^2} - 1 - \log\frac{\beta^2}{\sigma_u^2}\Big), \frac{1}{2}\Big(\frac{\gamma D_c}{\sigma_u^2} - 1 - \log\frac{\gamma D_c}{\sigma_u^2}\Big)\right\}. \tag{4.61}$$

We obtain

$$\begin{aligned} \overline{P}_{e,w}^{(n)} &\leq P_0 + P_1 + P_2 \\ &\leq \exp\big\{-n\Lambda_1(\gamma,\beta,\rho,s)\big\} + \exp\big\{-\exp\big(n(\epsilon_0^{(n)} - \epsilon_1^{(n)})\big)\big\} + 2\exp\big\{-n\Lambda_2(\gamma,\beta)\big\} \\ &\leq 4\exp\Big\{-n\Big(\min\big[\Lambda_1(\gamma,\beta,\rho,s), \Lambda_2(\gamma,\beta)\big]\Big)\Big\} \triangleq \bar{\epsilon}^{(n)} \end{aligned} \tag{4.62}$$

for n sufficiently large. Since the above bound is independent of watermark w, we then obtain a random coding upper bound for $\overline{P}_e^{(n)}$.

4.5.5 The Existence of a Sequence of (R_c, R_w, n) Codes

In Section 4.5.3 and 4.5.4, we show that for any $\bar{\epsilon}^{(n)}, \bar{\delta}^{(n)} > 0$, the average probability of error and the average distortion with respect to the random codebook $\mathbf{c}$ are bounded by $\overline{P}_e^{(n)} \leq \bar{\epsilon}^{(n)}$ and $\overline{D}^{(n)} \leq D_c + \bar{\delta}^{(n)}$, respectively, for n sufficiently large. Now we show that for given $\delta > 0$, there exists at least one code $\widetilde{\mathbf{c}}$ such that $P_e^{(n)}(\widetilde{\mathbf{c}}) < (\bar{\epsilon}^{(n)})^{1-\frac{1}{\sqrt{n}}}$ and simultaneously $D^{(n)}(\widetilde{\mathbf{c}}) \leq D_c + \delta$ for n sufficiently large. Let $\mathcal{A}$ be the set of all the codes with $P_e^{(n)}(\mathbf{c}) \leq (\bar{\epsilon}^{(n)})^{1-\frac{1}{\sqrt{n}}}$, i.e.,

$$\mathcal{A} \triangleq \big\{\mathbf{c} : P_e^{(n)}(\mathbf{c}) \leq (\bar{\epsilon}^{(n)})^{1-\frac{1}{\sqrt{n}}}\big\}.$$

Since $P_e^{(n)}(\mathbf{c})$ is a random variable (a function of the random code $\mathbf{c}$), it follows from Markov's inequality that $\Pr\big\{P_e^{(n)}(\mathbf{c}) > (\bar{\epsilon}^{(n)})^{1-\frac{1}{\sqrt{n}}}\big\} \leq (\bar{\epsilon}^{(n)})^{\frac{1}{\sqrt{n}}}$ or $\Pr(\mathcal{A}) \geq 1 - (\bar{\epsilon}^{(n)})^{\frac{1}{\sqrt{n}}}$ for n sufficiently large. Therefore, we have

$$\begin{aligned} \int_{\mathcal{A}} \frac{q(\mathbf{c})}{\Pr(\mathcal{A})} D^{(n)}(\mathbf{c})\, d\mathbf{c} &\leq \frac{1}{\Pr(\mathcal{A})} \int q(\mathbf{c}) D^{(n)}(\mathbf{c})\, d\mathbf{c} \\ &\leq \frac{D_c + \bar{\delta}^{(n)}}{1 - (\bar{\epsilon}^{(n)})^{\frac{1}{\sqrt{n}}}}. \end{aligned} \tag{4.63}$$

Since $(\bar{\epsilon}^{(n)})^{\frac{1}{\sqrt{n}}} \leq 4^{\frac{1}{\sqrt{n}}} \exp\big\{-\sqrt{n}\big(\min\big[\Lambda_1(\gamma, \beta, \rho, s), \Lambda_2(\gamma, \beta)\big]\big)\big\}$, which goes to 0 as $n \to \infty$, we can make

$$\frac{D_c + \bar{\delta}^{(n)}}{1 - (\bar{\epsilon}^{(n)})^{\frac{1}{\sqrt{n}}}} \leq D_c + \delta \tag{4.64}$$

for n sufficiently large. This demonstrates that, there exists at least one sequence of codes $\{\widetilde{\mathbf{c}}\}$ satisfying $P_e^{(n)}(\widetilde{\mathbf{c}}) < (\bar{\epsilon}^{(n)})^{1-\frac{1}{\sqrt{n}}}$ and $D^{(n)}(\widetilde{\mathbf{c}}) \leq D_c + \delta$ simultaneously for n sufficiently large.

□

4.6 Proof of Corollary 4.1

We first get a stationary point with respect to s by letting

$$\begin{aligned}\frac{\partial\Lambda_1(\gamma,\beta,\rho,s)}{\partial s} &= \frac{\beta^2(\gamma-1)D_c+(1+\rho)\sigma_u^2(\beta^2-D_c)-4s(1+\rho)\sigma_u^2\beta^2(\gamma-1)D_c^2}{1+2s\beta^2(\gamma-1)D_c+2s(1+\rho)\sigma_u^2\big(\beta^2-D_c-2s\beta^2(\gamma-1)D_c^2\big)} \\ &\quad+\frac{\rho\beta^2(\gamma-1)D_c}{1+2s\beta^2(\gamma-1)D_c+\frac{\gamma-1)D_c}{(1+\rho)D_a}} \qquad (4.65)\\ &= 0. \qquad (4.66)\end{aligned}$$

It is convenient to introduce the following notation:

$$a \triangleq \frac{1}{\beta^2(\gamma-1)D_c+(1+\rho)\sigma_u^2(\beta^2-D_c)}, \tag{4.67}$$

$$b \triangleq \frac{1}{\beta^2(\gamma-1)D_c}+\frac{1}{(1+\rho)\beta^2D_a}, \tag{4.68}$$

$$c \triangleq \sigma_u^2\beta^2(\gamma-1)D_c^2. \tag{4.69}$$

Then we obtain

$$\frac{1-4s(1+\rho)ac}{a+2s-4s^2(1+\rho)ac}+\frac{\rho}{b+2s}=0. \tag{4.70}$$

or equivalently

$$2ac(2+\rho)s^2+(2abc-1)s-\frac{\rho a+b}{2(1+\rho)}=0. \tag{4.71}$$

Solving the above equation, we obtain

$$s=\frac{1-2abc\pm\sqrt{(1-2abc)^2+4ac(\rho a+b)\frac{2+\rho}{1+\rho}}}{4ac(2+\rho)}. \tag{4.72}$$

It can be easily checked that only

$$s^*=\frac{1-2abc+\sqrt{(1-2abc)^2+4ac(\rho a+b)\frac{2+\rho}{1+\rho}}}{4ac(2+\rho)}. \tag{4.73}$$

satisfies the constraint $s \geq 0$. In fact, the above s^* maximizes $\Lambda_1(\gamma, \beta, \rho, s)$ over $s \geq 0$ since

$$\begin{aligned}
\frac{\partial^2 \Lambda_1(\gamma, \beta, \rho, s)}{\partial s^2} &= \frac{-4(1+\rho)ac\big(a + 2s - 4s^2(1+\rho)ac\big) - 2\big(1 - 4s(1+\rho)ac\big)^2}{\big(a + 2s - 4s^2(1+\rho)ac\big)^2} \\
&\quad - \frac{2\rho}{(b+2s)^2} \\
&= -\frac{1 + 4(1+\rho)a^2c + \big(1 - 4s(1+\rho)ac\big)^2}{\big(a + 2s - 4s^2(1+\rho)ac\big)^2} - \frac{2\rho}{(b+2s)^2} \\
&< 0. \qquad (4.74)
\end{aligned}$$

4.7 Proof of Corollary 4.2

We first maximize $\Lambda_1(\gamma, \beta, \rho, s)$ over $\rho \in [0, 1]$. Denote $\theta = \beta^2 - D_c - 2s\beta^2(\gamma - 1)D_c^2$, we have

$$\begin{aligned}
&\frac{\partial \Lambda_1(\gamma, \beta, \rho, s)}{\partial \rho} \\
&= \frac{s\sigma_u^2\theta}{1 + 2s\beta^2(\gamma-1)D_c + 2s(1+\rho)\sigma_u^2\theta} + \frac{1}{2}\log\left(\frac{1 + 2s\beta^2(\gamma-1)D_c + \frac{(\gamma-1)D_c}{(1+\rho)D_a}}{\gamma}\right) \\
&\quad - \frac{\frac{\rho(\gamma-1)D_c}{2(1+\rho)^2D_a}}{1 + 2s\beta^2(\gamma-1)D_c + \frac{(\gamma-1)D_c}{(1+\rho)D_a}} - R_w \qquad (4.75)
\end{aligned}$$

and

$$\begin{aligned}
&\frac{\partial^2 \Lambda_1(\gamma, \beta, \rho, s)}{\partial \rho^2} \\
&= -\frac{2\big(s\sigma_u^2\theta\big)^2}{\big(1 + 2s\beta^2(\gamma-1)D_c + 2s(1+\rho)\sigma_u^2\theta\big)^2} - \frac{1}{2}\frac{\frac{\gamma(\gamma-1)D_c}{(1+\rho)^2D_a}}{1 + 2s\beta^2(\gamma-1)D_c + \frac{(\gamma-1)D_c}{(1+\rho)D_a}} \\
&\quad - \frac{\frac{(\gamma-1)D_c(1+2\rho-\rho^2)}{2D_a(1+\rho)^4}\big(1 + 2s\beta^2(\gamma-1)D_c + \frac{(\gamma-1)D_c}{(1+\rho)D_a}\big) + \frac{\rho}{2}\left(\frac{(\gamma-1)D_c}{(1+\rho)^2D_a}\right)^2}{\big(1 + 2s\beta^2(\gamma-1)D_c + \frac{(\gamma-1)D_c}{(1+\rho)D_a}\big)^2} \\
&\leq 0. \qquad (4.76)
\end{aligned}$$

Thus by solving the equation

$$\frac{\partial \Lambda_1(\gamma,\beta,\rho,s)}{\partial \rho} = 0,$$

we can obtain a ρ which maximizes $\Lambda_1(\gamma,\beta,\rho,s)$, or equivalently

$$\begin{aligned} R_w &= \frac{1}{2}\log\left(\frac{1+2s\beta^2(\gamma-1)D_c+\frac{(\gamma-1)D_c}{(1+\rho)D_a}}{\gamma}\right) \\ &\quad -\frac{\rho}{2s+b}\frac{1}{2(1+\rho)^2D_a\beta^2}+\frac{sa\sigma_u^2\theta}{a+2s-4s^2(1+\rho)ac}. \end{aligned} \tag{4.77}$$

For the s^* satisfying (4.66), this reduces to

$$\begin{aligned} R_w &= \frac{1}{2}\log\left(\frac{1+2s^*\beta^2(\gamma-1)D_c+\frac{(\gamma-1)D_c}{(1+\rho)D_a}}{\gamma}\right) \\ &\quad -\frac{\rho}{2s^*+b}\left(\frac{1}{2(1+\rho)^2D_a\beta^2}+\frac{s^*a\sigma_u^2\theta}{1-4s^*(1+\rho)ac}\right). \end{aligned} \tag{4.78}$$

It can be easily shown that R_w is a decreasing function of ρ. That is, the maximum value of R_w, denoted by R_w^*, can be obtained by letting $\rho = 0$ (in other words, when the maximization of $\Lambda_1(\gamma,\beta,\rho,s)$ over ρ is achieved by $\rho = 0$, and we know that the R_w which satisfies (4.78) with $\rho = 0$ is the maximum value we can get, since R_w is a decreasing function of ρ), which results in

$$R_w \;\leq\; R_w^* \triangleq \frac{1}{2}\log\left(\frac{1+2s^*\beta^2(\gamma-1)D_c+\frac{(\gamma-1)D_c}{D_a}}{\gamma}\right). \tag{4.79}$$

Noting that by setting $\rho = 0$, s^* reduces to

$$s^* = \frac{1}{4ac}. \tag{4.80}$$

Plugging the above back into (4.79) and replacing a, c with (4.67) and (4.69) respectively, we obtain

$$R_w^* = \frac{1}{2}\log\left(\frac{1+\frac{\frac{\beta^2}{\sigma_u^2}(\gamma-1)+(\frac{\beta^2}{D_c}-1)}{2}+\frac{(\gamma-1)D_c}{D_a}}{\gamma}\right). \tag{4.81}$$

Furthermore, noting that R_w^* is a decreasing function of γ and an increasing function of β^2, and γ and β^2 are independent of each other, if we choose $\gamma \to \left(\frac{\sigma_u^2}{D_c}\right)^+$ and $\beta^2 \to \left(\sigma_u^2\right)^-$, we have

$$R_w \;\; \leq \;\; R_w^* \to \left(\frac{1}{2}\log\left(1 + \frac{D_c - D_c^2/\sigma_u^2}{D_a}\right)\right)^-. \tag{4.82}$$

It is easily seen that in this case we have $\sup_{\rho,s,\beta,\gamma} \Lambda_1(\gamma,\beta,\rho,s) > 0$ for any given R_w satisfying (4.82).

On the other hand, the condition $\gamma < e^{2(R_c - R_w)}$ is equivalent to

$$R_w < R_c - \frac{1}{2}\log\gamma. \tag{4.83}$$

Noting that $R_c - \frac{1}{2}\log\gamma$ is also a decreasing function of γ, as $\gamma \to \left(\frac{\sigma_u^2}{D_c}\right)^+$, we have

$$R_w \to \left(R_c - \frac{1}{2}\log\frac{\sigma_u^2}{D_c}\right)^-. \tag{4.84}$$

Combining (4.82) with (4.84), we obtain that the error exponent is positive for all R_w satisfying

$$R_w < \min\left\{R_c - \frac{1}{2}\log\frac{\sigma_u^2}{D_c}, \frac{1}{2}\log\left(1 + \frac{D_c - D_c^2/\sigma_u^2}{D_a}\right)\right\}. \tag{4.85}$$

By the continuity of $\Lambda_1(\gamma,\beta,\rho,s)$ with respect to all the parameters, it is easy to see that we can always find γ,β,ρ,s such that $\Lambda_1(\gamma,\beta,\rho,s)$ is positive as long as R_w satisfies (4.85). The proof is finished by noting that $\Lambda_2(\gamma,\beta)$ is always positive, thus $E_R(R_c,R_w;D_c,D_a)$ is positive as long as R_w satisfies (4.85).

□

4.8 Conclusions

In this chapter, a computable error exponent for a joint compression and private watermarking system for memoryless Gaussian sources under AWGN attacks is obtained.

The error exponent is derived by applying random coding technique that uses Gallager's method, and by incorporating techniques for the derivation of Gaussian source reliability functions. Numerical results show that the random coding exponent is positive within almost the entire achievable region [33]. In future work, we plan to refine our analysis to obtain a random coding exponent which is positive in the entire achievable region.

Chapter 5

Achievable Rate Region for Multi-User Joint Compression and Private Watermarking Under Multiple Access Attacks

5.1 Introduction

In this chapter, we extend the joint compression and information hiding problem from a single-user (e.g., [33], [46], [47], [76], [89]) to a multi-user scenario. Our model is depicted in Fig. 5.1. Assume that two users separately embed their secret messages W_1 and W_2 (at rates R_w^1 and R_w^2 respectively) into two correlated DMS's (U_1^n, U_2^n). Each user can only access one of the two host sources. Due to bandwidth and/or storage constraints, two compressed stegotexts X_1^n and X_2^n are obtained at rates R_c^1 and R_c^2 respectively. The stegotexts are corrupted by a discrete memoryless multiple access

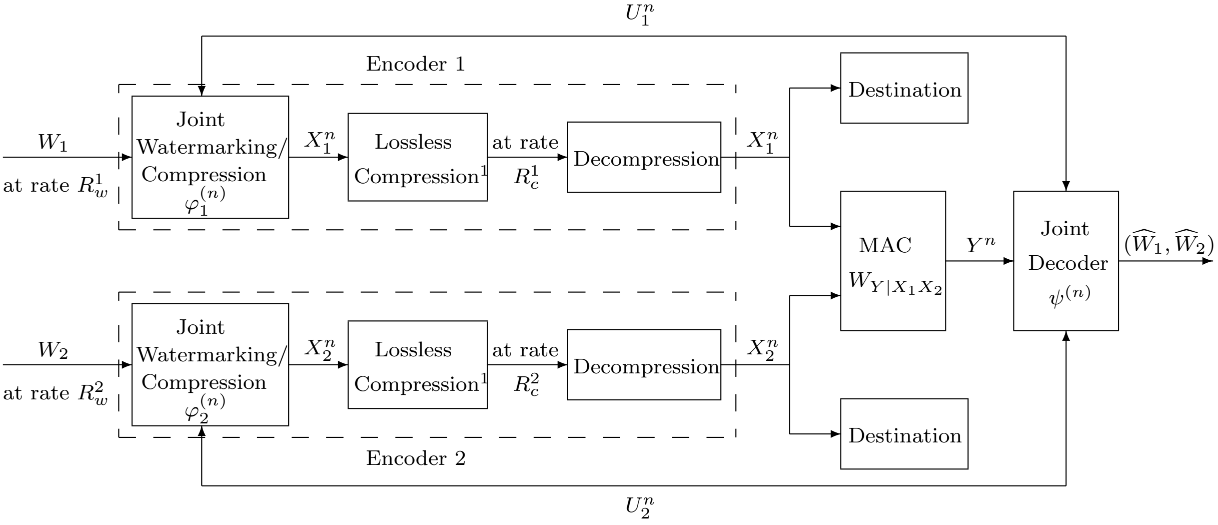


Figure 5.1: A general model of joint watermarking and compression for multi-user information hiding.

channel (MAC) $W_{Y|X_1X_2}$. As a result, a forgery Y^n is produced at the output of the channel. The authorized receiver retrieves the hidden information from the received forgery. Throughout the paper, we focus on the private scenario, i.e., we assume the decoder has perfect knowledge of the original host sources (U_1^n, U_2^n).

For this two-user private information hiding system, we are interested in determining the rate region of all achievable rate quadruples $(R_w^1, R_w^2, R_c^1, R_c^2)$ for given distortion levels (D_1, D_2). We find that the multiple embedding problem is strongly related to the lossy multi-terminal source coding problem for correlated sources, where separate encoders are designed in order to guarantee the joint typicality of the codewords with respect to the correlated source sequences. An inner bound for the achievable rate region is obtained (see Theorem 5.1) based on a ϵ-strong typicality coding/decoding argument.

[1] Here the term “lossless compression” means an invertible binary representation of the stegotext X_i^n, $i = 1, 2$.

More specifically, we employ a generalized rate-distortion encoding scheme to ensure that $(\mathbf{u}_1, \mathbf{u}_2, \mathbf{x}_1, \mathbf{x}_2)$ are jointly strongly typical with high probability. The generalized rate-distortion encoding scheme, introduced in [54] for Gaussian multi-terminal source coding (see also [79], [26]), can be briefly described as follows. One of the encoders, say Encoder 1, chooses a codeword $\mathbf{x}_1$ such that conditioned on $(\mathbf{u}_1, \mathbf{x}_1)$, $(\mathbf{u}_1, \mathbf{x}_1, U_2^n, X_2^n)$ is ϵ-strongly typical with high probability. The other encoder, Encoder 2, which is assumed to know the codebook of Encoder 1 $(\varphi_1^{(n)}(W_1, U_1^n))$, generates a codeword $\mathbf{x}_2$ such that $(U_1^n, \varphi_1^{(n)}(W_1, U_1^n), \mathbf{u}_2, \mathbf{x}_2)$ is ϵ-strongly typical with high probability. To this end, an extended Markov lemma (see Lemma 5.5) ensures that the codewords $\mathbf{x}_1$ and $\mathbf{x}_2$, although generated from separate encoders, are ϵ-strongly typical with the source sequences $(\mathbf{u}_1, \mathbf{u}_2)$ with high probability.

We also derive an outer bound for the achievable rate region with single-letter characterization (see Theorem 5.2), which follows from Fano's inequality and standard information-theoretical bounding arguments.

The rest of this chapter is organized as follows. In Section 5.2, we review the definition of ϵ-strong typicality and its properties, which are widely used in the rest of this thesis. In Section 5.3, we formulate our problem and establish an inner bound and an outer bound for the achievable rate region. All the proofs are given in Section 5.4. Finally, we draw conclusions in Section 5.5.

5.2 Preliminary: Jointly Typical Sequences

Let $V \triangleq (X_1, X_2, ..., X_m)$ be a superletter (a collection of RV's) taking values in a finite set $\mathcal{V} \triangleq \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_m$ with joint distribution $P_V(x_1, ..., x_m)$, which for simplicity we also denote by $P_V(v)$.

Definition 5.1 [10] For any $0 < \epsilon < 1$, a vector of n-length sequences $\mathbf{v} \triangleq (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_m) \in \mathcal{V}^n$ is called ϵ-strongly typical with respect to P_V if

1. For all $v \in \mathcal{V}$ with $P_V(v) > 0$, we have
$$\left| \frac{N(v|\mathbf{v})}{n} - P_V(v) \right| \leq \frac{\epsilon}{|\mathcal{V}|},$$
where $N(v|\mathbf{v})$ is the number of occurrences of v in $\mathbf{v}$;

2. For all $v \in \mathcal{V}$ with $P_V(v) = 0$, $N(v|\mathbf{v}) = 0$.

Denote by $T_\epsilon^{(n)}(V)$ or $T_\epsilon^{(n)}$ the set of all ϵ-strongly typical sequences $(\mathbf{x}_1, \ldots, \mathbf{x}_m)$ with respect to the joint distribution $P_V(v)$. Let $I_V \triangleq \{1, 2, ..., m\}$, and $I_G \subseteq I_V$. We then let $G = (X_{g_1}, X_{g_2}, ..., X_{g_{|I_G|}}) \in \mathcal{G}$ be a "sub-superletter" corresponding to I_G such that $g_i \in I_G$. Let G, K, and L be sub-superletters of V such that I_G, I_K, I_L are disjoint, and let P_G, P_K and $P_{G|K}$ be the marginal and conditional distributions induced by P_V, respectively. Denote by $T_\epsilon^{(n)}(G)$ the projection of $T_\epsilon^{(n)}(V)$ to the coordinates of G. Given any $\mathbf{k} \in \mathcal{K}^n$, denote $T_\epsilon^{(n)}(G|\mathbf{k}) \triangleq \left\{ \mathbf{g} \in \mathcal{G}^n : (\mathbf{g}, \mathbf{k}) \in T_\epsilon^{(n)}(G, K) \right\}$. Clearly $T_\epsilon^{(n)}(G|\mathbf{k}) = \emptyset$ if $\mathbf{k} \notin T_\epsilon^{(n)}(K)$. The following lemma (see, e.g., [10, pp. 342–343][2]) restates the well known exponential bounds for the cardinality of ϵ-strongly typical sets.

Lemma 5.1 ([10]) For any disjoint subsets $G, K \subseteq V$, let G^n, K^n and V^n be i.i.d. drawn according to $P_G^{(n)}$, $P_K^{(n)}$ and $P_V^{(n)}$. The following properties hold for sufficiently large n.

1. $P_V^{(n)} \left\{ V^n \in \mathcal{T}_\epsilon^{(n)} \right\} \geq 1 - \eta$. Moreover $P_K^{(n)} \left\{ K^n \in \mathcal{T}_\epsilon^{(n)}(K) \right\} \geq 1 - \eta$,

2. For any $\mathbf{k} \in \mathcal{T}_\epsilon^{(n)}(K)$, $\left| \frac{1}{n} \log P_K^{(n)}(\mathbf{k}) + H(K) \right| \leq \eta$,

[2] Note that the sets $T_\epsilon^{(n)}(G|\mathbf{k})$ are only implicitly defined in [10].

3. $2^{n(H(K)-\eta)} \leq \left|\mathcal{T}_\epsilon^{(n)}(K)\right| \leq 2^{n(H(K)+\eta)}$,

4. For any $\mathbf{k} \in \mathcal{T}_\epsilon^{(n)}(K)$, $P_G^{(n)}\left\{G^n \in \mathcal{T}_\epsilon^{(n)}(G|\mathbf{k})\right\} \geq 1-\eta$,

5. For any $\mathbf{k} \in \mathcal{T}_\epsilon^{(n)}(K)$, $2^{n(H(G|K)-\eta)} \leq \left|\mathcal{T}_\epsilon^{(n)}(G|\mathbf{k})\right| \leq 2^{n(H(G|K)+\eta)}$,

6. For $(\mathbf{g},\mathbf{k}) \in \mathcal{T}_\epsilon^{(n)}(G,K)$, $\left|\frac{1}{n}\log P_{G|K}^{(n)}(\mathbf{g}|\mathbf{k}) + H(G|K)\right| \leq \eta$,

where $\eta \triangleq \eta(\epsilon, n)$ is a generic positive term such that $\lim_{\epsilon\to 0}\lim_{n\to\infty}\eta(\epsilon,n) = 0$.

Finally, we recall the Markov lemma for ϵ-strong typicality.

Lemma 5.2 (Markov Lemma [10, p. 579]) Let $G \to K \to L$ form a Markov chain in this order. For any $0 < \epsilon_0 < 1$ and $(\mathbf{g},\mathbf{k}) \in T_\epsilon^{(n)}(G,K)$,

$$P_{L|K}^{(n)}\left(\mathbf{l} : (\mathbf{g},\mathbf{k},\mathbf{l}) \in T_\epsilon^{(n)}(G,K,L)\middle|\mathbf{k}\right) > 1-\epsilon_0$$

for n sufficiently large, independently of $(\mathbf{g},\mathbf{k})$.

5.3 Problem Formulation and Main Results

Let the pair of finite-alphabets discrete memoryless correlated host sources $\{(U_{1j},U_{2j})\}_{j=1}^{\infty}$ have marginal distribution $Q_{U_1U_2}$. The secret messages w_1 and w_2 are independently and uniformly chosen from the sets $\mathcal{W}_1 \triangleq \{1,2,...,M_w^1\}$ and $\mathcal{W}_2 \triangleq \{1,2,...,M_w^2\}$, respectively. The attack channel is modeled as a two-sender one-receiver discrete memoryless MAC $W_{Y|X_1X_2}$ having finite input alphabets $\mathcal{X}_1$ and $\mathcal{X}_2$, finite output alphabet $\mathcal{Y}$, and a transition probability distribution $W_{Y|X_1X_2}(y|x_1,x_2)$. The probability of receiving $\mathbf{y} \in \mathcal{Y}^n$ conditioned on sending $\mathbf{x}_1 \in \mathcal{X}_1^n$ and $\mathbf{x}_2 \in \mathcal{X}_2^n$ is hence given by $W_{Y|X_1X_2}^{(n)}(\mathbf{y}|\mathbf{x}_1,\mathbf{x}_2)$.

Let $d_i : \mathcal{U}_i \times \mathcal{X}_i \to [0,\infty)$ be single-letter distortion measures and define $d_i^{max} \triangleq \max_{u_i,x_i} d_i(u_i,x_i)$ for $i=1,2$. For $\mathbf{u}_i \in \mathcal{U}_i^n$ and $\mathbf{x}_i \in \mathcal{X}_i^n$, let $d_i(\mathbf{u}_i,\mathbf{x}_i) = \sum_{j=1}^n d_i(u_{ij},x_{ij})$.

Definition 5.2 Given $P_{U_1U_2}$ and $W_{Y|X_1X_2}$, a $(R_w^1, R_w^2, R_c^1, R_c^2; n)$ joint compression and private watermarking (JCPW) code with block length n (see Fig. 5.1) consists of two encoders

$$\varphi_1^{(n)} : \mathcal{W}_1 \times \mathcal{U}_1^n \longrightarrow \mathcal{X}_1^n,$$

$$\varphi_2^{(n)} : \mathcal{W}_2 \times \mathcal{U}_2^n \longrightarrow \mathcal{X}_2^n,$$

and a private (in the sense that the host sources are available at the decoder) decoder

$$\psi^{(n)} : \mathcal{Y}^n \times \mathcal{U}_1^n \times \mathcal{U}_2^n \longrightarrow \mathcal{W}_1 \times \mathcal{W}_2.$$

Let $i \in \{1, 2\}$. The watermarking rate for encoder i is defined as $R_w^i = \frac{\log_2 M_w^i}{n}$. The stegotext $\mathbf{x}_i = \varphi_i^{(n)}(w_i, \mathbf{u}_i)$ takes values from a set $\mathcal{C}^{(i)} \subseteq \mathcal{X}_i^n$ of M_c^i codevectors. The compression rate for encoder i is defined as $R_c^i = \frac{\log_2 M_c^i}{n}$.

Definition 5.3 The probability of error in reproducing the secret sources is given by

$$\begin{aligned} & P_e^{(n)} \\ \triangleq \ & \Pr\left(\psi^{(n)}(Y^n, U_1^n, U_2^n) \neq (W_1, W_2)\right) \\ = \ & \frac{1}{2^{n(R_w^1+R_w^2)}} \sum_{w_1=1}^{M_w^1} \sum_{w_2=1}^{M_w^2} \sum_{\mathcal{U}_1^n \times \mathcal{U}_2^n} Q_{U_1U_2}^{(n)}(\mathbf{u}_1, \mathbf{u}_2) W_{Y|X_1X_2}^{(n)}\left(\mathbf{y} : \psi^{(n)}(\mathbf{y}, \mathbf{u}_1, \mathbf{u}_2) \neq (w_1, w_2) \middle| \mathbf{x}_1, \mathbf{x}_2\right) \end{aligned}$$

where $\mathbf{x}_i \triangleq \varphi_i^{(n)}(w_i, \mathbf{u}_i)$ $(i = 1, 2)$.

Definition 5.4 Given $Q_{U_1U_2}$ and $W_{Y|X_1X_2}$, a quadruple $(R_w^1, R_w^2, R_c^1, R_c^2)$ is said to be achievable with respect to distortion levels (D_1, D_2) if there exists a sequence of $(R_w^1, R_w^2, R_c^1, R_c^2; n)$ JCPW codes such that

$$\lim_{n\to\infty} P_e^{(n)} = 0$$

and

$$\limsup_{n\to\infty} \frac{1}{n}\mathbb{E}\Big[d_i\big(U_i^n, \varphi_i^{(n)}(W_i, U_i^n)\big)\Big] \le D_i, \quad i = 1, 2.$$

Definition 5.5 The achievable rate region $\mathcal{R}(D_1, D_2)$ is the closure of the set of achievable rate quadruples $\mathbf{R} \triangleq (R_w^1, R_w^2, R_c^1, R_c^2)$.

Remark: It can be shown by using a time-sharing argument [10] that $\mathcal{R}(D_1, D_2)$ is convex.

Definition 5.6 Given $Q_{U_1U_2}$, $W_{Y|X_1X_2}$, and a pair of distortion levels (D_1, D_2), let $\mathcal{P}_{D_1,D_2}$ be the set of random variable tuples $(U_1, U_2, X_1, X_2, Y) \in \mathcal{U}_1 \times \mathcal{U}_2 \times \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}$ such that the joint distribution $P_{U_1U_2X_1X_2Y}$ satisfies: (1) it can be factorized as

$$P_{U_1U_2X_1X_2Y} = Q_{U_1U_2}P_{X_1|U_1}P_{X_2|U_2}W_{Y|X_1X_2},$$

and (2) $\mathbb{E}[d_i(U_i, X_i)] \leq D_i$, for $i = 1, 2$.

Theorem 5.1 Let $\mathcal{R}_{in}(D_1, D_2)$ be the closure of the convex hull of all $(R_w^1, R_w^2, R_c^1, R_c^2)$ satisfying

$$R_w^1 \quad < \quad \min\Big\{R_c^1 - I(U_1; X_1),\ I(X_1; Y|X_2, U_1, U_2)\Big\}, \tag{5.1}$$

$$R_w^2 \quad < \quad \min\Big\{R_c^2 - I(U_2; X_2),\ I(X_2; Y|X_1, U_1, U_2)\Big\}, \tag{5.2}$$

$$R_w^1 + R_w^2 \quad < \quad I(X_1, X_2; Y|U_1, U_2). \tag{5.3}$$

for some $(U_1, U_2, X_1, X_2, Y) \in \mathcal{P}_{D_1,D_2}$. Then $\mathcal{R}_{in}(D_1, D_2) \subseteq \mathcal{R}(D_1, D_2)$.

By introducing an auxiliary random variable V, we establish the following outer bound.

Definition 5.7 Given P_V, $Q_{U_1U_2}$, $W_{Y|X_1X_2}$, and a pair of distortion levels (D_1, D_2), let $\widetilde{\mathcal{P}}_{D_1,D_2}$ be the set of random variable tuples $(V, U_1, U_2, X_1, X_2, Y) \in \mathcal{V} \times \mathcal{U}_1 \times \mathcal{U}_2 \times \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}$ such that the joint distribution $P_{VU_1U_2X_1X_2Y}$ satisfies: (1) it can be factorized as

$$P_{VU_1U_2X_1X_2Y} = P_V Q_{U_1U_2}P_{X_1|U_1V}P_{X_2|U_2V}W_{Y|X_1X_2},$$

and (2) $\mathbb{E}[d_i(U_i, X_i)] \leq D_i$, for $i = 1, 2$.

Theorem 5.2 Let $\mathcal{R}_{out}(D_1, D_2)$ be the closure of all $(R_w^1, R_w^2, R_c^1, R_c^2)$ satisfying

$$\begin{aligned} R_w^1 &< \min\Big\{R_c^1 - I(U_1; X_1|V),\, I(X_1; Y|X_2, U_1, U_2, V)\Big\}, && (5.4)\\ R_w^2 &< \min\Big\{R_c^2 - I(U_2; X_2|V),\, I(X_2; Y|X_1, U_1, U_2, V)\Big\}, && (5.5)\\ R_w^1 + R_w^2 &< I(X_1, X_2; Y|U_1, U_2, V). && (5.6)\end{aligned}$$

for some $(V, U_1, U_2, X_1, X_2, Y) \in \widetilde{\mathcal{P}}_{D_1,D_2}$. Then $\mathcal{R}(D_1, D_2) \subseteq \mathcal{R}_{out}(D_1, D_2)$. Furthermore, the cardinality of the alphabet of the auxiliary RV V can be bounded as $|\mathcal{V}| \leq |\mathcal{U}_1||\mathcal{U}_2||\mathcal{X}_1||\mathcal{X}_2| + 4$.

5.3.1 Special Cases

1. *Single User Case.* There is only one secret message W_1, i.e., Encoder 2 is turned off and the MAC reduces to a (single-user) discrete memoryless channel. Define $U_1 = U, X_1 = X, D_1 = D, R_c^1 = R_c$ and $R_w^1 = R_w$, (5.1)–(5.3) reduce to

$$R_w < \min\Big\{R_c - I(U; X);\, I(X; Y|U)\Big\}. \tag{5.7}$$

Therefore, given a compression rate R_c, the maximum watermarking rate R_w^* is obtained by

$$R_w^* = \max_{P_{X|U}: \mathbb{E}[d(U,X)] \leq D} \min\big\{R_c - I(U; X),\, I(X; Y|U)\big\}. \tag{5.8}$$

Equivalently, the minimum rate compression rate R_c^* to achieve a target watermarking rate R_w is obtained by

$$R_c^* = \min_{P_{X|U}: \mathbb{E}[d(U,X)] \leq D,\, R_w < I(X;Y|U)} I(U; X) + R_w. \tag{5.9}$$

It is easy to see that these results reduces to the single-user private joint compression and watermarking scenarios (see, e.g., [33], [89]). Furthermore, in this single-user case, it can be shown that $\mathcal{R}_{in}(D) = \mathcal{R}_{out}(D) = \mathcal{R}(D)$ (see e.g., [34, Theorem 2.1], for the case of discrete alphabets).

2. *Attack-Free Channel.* Let $l : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathcal{Y}$ be a bijection and let $Y = l(X_1, X_2)$. (5.1)–(5.3) reduce to

$$\begin{aligned} R_w^1 &< \min\Big\{R_c^1 - I(U_1; X_1),\, H(X_1|U_1)\Big\}, && (5.10)\\ R_w^2 &< \min\Big\{R_c^2 - I(U_2; X_2),\, H(X_2|U_2)\Big\}, && (5.11)\\ R_w^1 + R_w^2 &< H(X_1|U_1) + H(X_2|U_2). && (5.12) \end{aligned}$$

Note that the last inequality is a redundant condition and thus can be removed. As a result, this reduces to two parallel joint compression and watermarking problems with no attacks. For this case, we can easily show that $\mathcal{R}_{in}(D_1, D_2) = \mathcal{R}_{out}(D_1, D_2) = \mathcal{R}(D_1, D_2)$. Furthermore, $\mathcal{R}(D_1, D_2)$ is the Cartesian product of $\mathcal{R}(D_1)$ and $\mathcal{R}(D_2)$, where $\mathcal{R}(D_i) \triangleq \max_{P_{X_i|U_i}:\, \mathbb{E}[d_i(U_i,X_i)] \le D_i} \min\big\{R_c^i - I(U_i; X_i),\, H(X_i|U_i)\big\}$, $i = 1, 2$.

5.4 Proofs

5.4.1 Proof of Theorem 5.1

We first give an outline of the proof. We need to show that for given $Q_{U_1U_2}$, $W_{Y|X_1X_2}$, and any $(R_w^1, R_w^2, R_c^1, R_c^2) \in \mathcal{R}_{in}(D_1, D_2)$, there exists a sequence of JCPW codes $(\varphi_1^{(n)}, \varphi_2^{(n)}, \psi^{(n)})$ such that $P_e^{(n)} \to 0$ as $n \to \infty$ and for any $\delta > 0$, $\frac{1}{n}\mathbb{E}[d_i(U_i^n, \varphi_i^{(n)}(W_i, U_i^n))] \le D_i + \delta$, $i = 1, 2$, for n sufficiently large.

Fix $(P_{X_1|U_1}, P_{X_2|U_2})$ such that the following are satisfied for some $\epsilon' > 0$ (ϵ' will be specified later),

$$\begin{aligned} &R_w^1 < \min\Big\{R_c^1 - I(U_1; X_1) - \epsilon',\, I(X_1; Y|X_2, U_1, U_2) - \epsilon'\Big\}, && (5.13)\\ &R_w^2 < \min\Big\{R_c^2 - I(U_2; X_2) - \epsilon',\, I(X_2; Y|X_1, U_1, U_2) - \epsilon'\Big\}, && (5.14)\\ &R_w^1 + R_w^2 < I(X_1, X_2; Y|U_1, U_2) - \epsilon', && (5.15) \end{aligned}$$

$$\mathbb{E}[d_i(U_i, X_i)] \leq D_i,\ i = 1, 2. \tag{5.16}$$

Define

$$P_i^{(n)} \triangleq \Pr\Big(\frac{1}{n} d_i\big(U_i^n, \varphi_i^{(n)}(W_i, U_i^n)\big) > D_i + \epsilon d_i^{max}\Big),\ i = 1, 2.$$

We will prove that for any $\epsilon_1 > 0$, the following probabilities, which are averaged probabilities over a family of random codes $(\varphi_1^{(n)}, \varphi_2^{(n)})$, satisfy

$$\mathbb{E}[P_e^{(n)}] \leq \epsilon_1, \quad \mathbb{E}[P_1^{(n)}] \leq \epsilon_1, \quad \mathbb{E}[P_2^{(n)}] \leq \epsilon_1$$

for n sufficiently large. Then $\mathbb{E}\{P_e^{(n)} + P_1^{(n)} + P_2^{(n)}\} \leq 3\epsilon_1$, which guarantees that there exists at least one pair of codes $(\varphi_1^{(n)}, \varphi_2^{(n)})$ such that $P_e^{(n)} + P_1^{(n)} + P_2^{(n)} \leq 3\epsilon_1$ and hence $P_e^{(n)} \leq 3\epsilon_1$, $P_1^{(n)} \leq 3\epsilon_1$, $P_2^{(n)} \leq 3\epsilon_1$ are simultaneously satisfied for n sufficiently large. Finally, it can be easily shown that $P_i^{(n)} \leq 3\epsilon_1$ implies for n sufficiently large that

$$\frac{1}{n}\mathbb{E}\left[d_i(U_i^n, \varphi_i^{(n)}(W_i, U_i^n))\right] \leq D_i + \epsilon d_i^{max} + P_i^{(n)} d_i^{max} \leq D_i + \delta.$$

Random Code Design

In what follows, the ϵ-strongly typical set $\mathcal{T}_\epsilon^{(n)}$ is defined under the joint distribution

$$P_{U_1U_2X_1X_2Y} = Q_{U_1U_2} P_{X_1|U_1} P_{X_2|U_2} W_{Y|X_1X_2}$$

and all the marginal and conditional distributions, e.g., $P_{U_2X_2}$, $P_{U_1|U_2X_2}$, etc, are induced by $P_{U_1U_2X_1X_2Y}$ defined in the above. The parameter ϵ, which is chosen sufficiently small, will be specified later in the proof.

Random code generation. Let $i = 1, 2$. Denote $L_i \triangleq 2^{n(R_c^i - R_w^i)}$. For every $w_i \in \mathcal{W}_i$, randomly generate a codebook

$$\mathcal{C}_{w_i} \triangleq \{\mathbf{x}_i(w_i, t_i);\ t_i = 1, \cdots, L_i\}$$

where each codeword $\mathbf{x}_i(w_i, t_i)$ is independently and uniformly drawn from the typical set $\mathcal{T}_\epsilon^{(n)}(X_i)$. Denote the whole codebook for Encoder i by $\mathcal{C}^{(i)} = \{\mathcal{C}_{w_i}\}_{w_i=1}^{M_w^i}$, where we recall that $M_w^i = 2^{nR_w^i}$. Reveal the codebooks to both encoders and the decoder.

Encoder $\varphi_1^{(n)}$. To define the encoder $\varphi_1^{(n)}$, we need some new notation. Following [54], we introduce a conditional probability

$$A^{(n)}(\mathbf{u}_1, \mathbf{x}_1) \triangleq P_{U_2X_2|U_1X_1}^{(n)} \left\{(\mathbf{u}_2, \mathbf{x}_2) : (\mathbf{u}_2, \mathbf{x}_2) \in \mathcal{T}_\epsilon^{(n)}(U_2, X_2|\mathbf{u}_1, \mathbf{x}_1)\big|\mathbf{u}_1, \mathbf{x}_1\right\}. \tag{5.17}$$

For $\mu \in (0, 1)$, let

$$\mathcal{F}_{\mu,\epsilon}^{(n)}(U_1, X_1) \triangleq \left\{(\mathbf{u}_1, \mathbf{x}_1) : A^{(n)}(\mathbf{u}_1, \mathbf{x}_1) \geq 1 - \mu\right\}, \tag{5.18}$$

$$\mathcal{F}_{\mu,\epsilon}^{(n)}(U_1|\mathbf{x}_1) \triangleq \left\{\mathbf{u}_1 : (\mathbf{u}_1, \mathbf{x}_1) \in \mathcal{F}_{\mu,\epsilon}^{(n)}(U_1, X_1)\right\}. \tag{5.19}$$

By definition, we have $\mathcal{F}_{\mu,\epsilon}^{(n)}(U_1, X_1) \subseteq \mathcal{T}_\epsilon^{(n)}(U_1, X_1)$. Moreover, we introduce the following sets based on $\mathcal{F}_{\mu,\epsilon}^{(n)}(U_1, X_1)$ for later in the proof,

$$\begin{aligned}
\mathcal{F}_{\mu,\epsilon}^{(n)}(X_1|\mathbf{u}_1) &\triangleq \left\{\mathbf{x}_1 : (\mathbf{u}_1, \mathbf{x}_1) \in \mathcal{F}_{\mu,\epsilon}^{(n)}(U_1, X_1)\right\}, \\
\mathcal{F}_{\mu,\epsilon}^{(n)}(U_1) &\triangleq \left\{\mathbf{u}_1 : (\mathbf{u}_1, \mathbf{x}_1) \in \mathcal{F}_{\mu,\epsilon}^{(n)}(U_1, X_1) \text{ for some } \mathbf{x}_1\right\}, \\
\widetilde{\mathcal{F}}_{\mu,\epsilon}^{(n)}(U_1) &\triangleq \left\{\mathbf{u}_1 : P_{X_1|U_1}^{(n)}\left\{\mathbf{x}_1 : \mathbf{x}_1 \in \mathcal{F}_{\mu,\epsilon}^{(n)}(X_1|\mathbf{u}_1)\big|\mathbf{u}_1\right\} \geq 1 - \mu\right\}.
\end{aligned}$$

Given $w_1 \in \{1, \ldots, M_w^1\}$ and $\mathbf{u}_1$, $\varphi_1^{(n)}$ seeks the first codeword $\mathbf{x}_1(w_1, t_1)$ with $t_1 \leq L_1 - 1$ in $\mathcal{C}_{w_1}$ such that $\mathbf{u}_1 \in \mathcal{F}_{\mu,\epsilon}^{(n)}(U_1|\mathbf{x}_1(w_1, t_1))$. If no such codeword is found, choose $\mathbf{x}_1(w_1, L_1)$. The resulting $\mathbf{x}_1(w_1, t_1)$ (or $\mathbf{x}_1(w_1, L_1)$), known as the stegotext $\varphi_1^{(n)}(w_1, \mathbf{u}_1)$ or denoted by $\mathbf{x}_1(w_1, \mathbf{u}_1)$, is then sent to the (attack) channel.

Encoder $\varphi_2^{(n)}$. To define the encoder $\varphi_2^{(n)}$, we introduce the following notation. For the above given $\varphi_1^{(n)}$, let

$$B_{\varphi_1}^{(n)}(\mathbf{u}_2, \mathbf{x}_2) \triangleq \frac{1}{2^{nR_w^1}} \sum_{w_1=1}^{M_w^1} P_{U_1|U_2X_2}^{(n)} \left\{\mathbf{u}_1 : \left(\mathbf{u}_1, \varphi_1^{(n)}(w_1, \mathbf{u}_1)\right) \in \mathcal{T}_\epsilon^{(n)}(U_1, X_1|\mathbf{u}_2, \mathbf{x}_2)\big|\mathbf{u}_2, \mathbf{x}_2\right\}.$$

For $\nu \in (0,1)$, let

$$\mathcal{F}^{(n)}_{\varphi_1,\nu,\epsilon}(U_2, X_2) \triangleq \left\{(\mathbf{u}_2, \mathbf{x}_2) : B^{(n)}_{\varphi_1}(\mathbf{u}_2, \mathbf{x}_2) \geq 1-\nu\right\}, \tag{5.20}$$

$$\mathcal{F}^{(n)}_{\varphi_1,\nu,\epsilon}(U_2|\mathbf{x}_2) \triangleq \left\{\mathbf{u}_2 : (\mathbf{u}_2, \mathbf{x}_2) \in \mathcal{F}^{(n)}_{\varphi_1,\nu,\epsilon}(U_2, X_2)\right\}. \tag{5.21}$$

By definition, it is seen that $\mathcal{F}^{(n)}_{\varphi_1,\nu,\epsilon}(U_2, X_2) \subseteq \mathcal{T}^{(n)}_{\epsilon}(U_2, X_2)$. Moreover, we introduce the following sets based on $\mathcal{F}^{(n)}_{\varphi_1,\nu,\epsilon}(U_2, X_2)$ for later in the proof,

$$\begin{aligned}
&\mathcal{F}^{(n)}_{\varphi_1,\nu,\epsilon}(X_2|\mathbf{u}_2) \triangleq \left\{\mathbf{x}_2 : (\mathbf{u}_2, \mathbf{x}_2) \in \mathcal{F}^{(n)}_{\varphi_1,\nu,\epsilon}(U_2, X_2)\right\}, \\
&\mathcal{F}^{(n)}_{\varphi_1,\nu,\epsilon}(U_2) \triangleq \left\{\mathbf{u}_2 : (\mathbf{u}_2, \mathbf{x}_2) \in \mathcal{F}^{(n)}_{\varphi_1,\nu,\epsilon}(U_2, X_2) \text{ for some } \mathbf{x}_2\right\}, \\
&\widetilde{\mathcal{F}}^{(n)}_{\varphi_1,\nu,\epsilon}(U_2) \triangleq \left\{\mathbf{u}_2 : P^{(n)}_{X_2|U_2}\left\{\mathbf{x}_2 : \mathbf{x}_2 \in \mathcal{F}^{(n)}_{\varphi_1,\nu,\epsilon}(X_2|\mathbf{u}_2) \big| \mathbf{u}_2\right\} \geq 1-\nu\right\}.
\end{aligned}$$

Given $w_2 \in \{1, 2, ..., M^2_w\}$ and $\mathbf{u}_2$, $\varphi^{(n)}_2$ seeks the first codeword $\mathbf{x}_2(w_2, t_2)$ with $t_2 \leq L_2$ in $\mathcal{C}_{w_2}$ such that $\mathbf{u}_2 \in \mathcal{F}^{(n)}_{\varphi_1,\nu,\epsilon}(U_2|\mathbf{x}_2(w_2, t_2))$. If no such codeword is found, choose $\mathbf{x}_2(w_2, L_2)$. The resulting $\mathbf{x}_2(w_2, t_2)$ (or $\mathbf{x}_2(w_2, L_2)$), known as the stegotext $\varphi^{(n)}_2(w_2, \mathbf{u}_2)$ or simply denoted by $\mathbf{x}_2(w_2, \mathbf{u}_2)$, is then sent to the (attack) channel.

Decoder $\psi^{(n)}$. The decoder has full knowledge of $(\mathbf{u}_1, \mathbf{u}_2)$, and thus can generate all possible stegotexts $\{\varphi^{(n)}_i(w_i, \mathbf{u}_i)\}^{M^i_w}_{w_i=1}$, $i = 1, 2$. Upon receiving the sequence $\mathbf{y}$, the decoder finds the unique pair $(\widehat{w}_1, \widehat{w}_2)$ such that, $(\widehat{\mathbf{x}}_1, \widehat{\mathbf{x}}_2, \mathbf{y}) \in \mathcal{T}^{(n)}_{\epsilon}(X_1, X_2, Y|\mathbf{u}_1, \mathbf{u}_2)$, where $\widehat{\mathbf{x}}_1 = \varphi^{(n)}_1(\widehat{w}_1, \mathbf{u}_1)$ and $\widehat{\mathbf{x}}_2 = \varphi^{(n)}_2(\widehat{w}_2, \mathbf{u}_2)$. A decoding error occurs if at least one of the following event occurs:

1. $E_0 : (\mathbf{u}_1, \mathbf{u}_2, \mathbf{x}_1(w_1, \mathbf{u}_1), \mathbf{x}_2(w_2, \mathbf{u}_2), \mathbf{y}) \notin \mathcal{T}^{(n)}_{\epsilon}(U_1, U_2, X_1, X_2, Y)$;

2. E_1: there exist $w'_1 \neq w_1$ such that

$$(\mathbf{u}_1, \mathbf{u}_2, \mathbf{x}_1(w'_1, \mathbf{u}_1), \mathbf{x}_2(w_2, \mathbf{u}_2), \mathbf{y}) \in \mathcal{T}^{(n)}_{\epsilon}(U_1, U_2, X_1, X_2, Y);$$

3. E_2: there exist $w'_2 \neq w_2$ such that

$$(\mathbf{u}_1, \mathbf{u}_2, \mathbf{x}_1(w_1, \mathbf{u}_1), \mathbf{x}_2(w'_2, \mathbf{u}_2), \mathbf{y}) \in \mathcal{T}^{(n)}_{\epsilon}(U_1, U_2, X_1, X_2, Y);$$

4. E_3: there exist $w_1' \neq w_1$ and $w_2' \neq w_2$ such that

$$(\mathbf{u}_1, \mathbf{u}_2, \mathbf{x}_1(w_1', \mathbf{u}_1), \mathbf{x}_2(w_2', \mathbf{u}_2), \mathbf{y}) \in \mathcal{T}_\epsilon^{(n)}(U_1, U_2, X_1, X_2, Y).$$

Bounding $\mathbb{E}_{\mathcal{C}^{(1)},\mathcal{C}^{(2)}}[P_i^{(n)}]$

We first give two useful lemmas.

Lemma 5.3 Given $\mu \in (0,1)$ and $\mathbf{u}_1 \in \mathcal{F}_{\mu,\epsilon}^{(n)}(U_1)$,

$$|\mathcal{F}_{\mu,\epsilon}^{(n)}(X_1|\mathbf{u}_1)| \geq \left(1 - \frac{\eta}{\mu}\right) 2^{n(H(X_1|U_1)-\eta)} \tag{5.22}$$

for n sufficiently large. For any $\mathbf{u}_2 \in \widetilde{\mathcal{F}}_{\varphi_1,\nu,\epsilon}^{(n)}(U_2)$,

$$|\mathcal{F}_{\varphi_1,\nu,\epsilon}^{(n)}(X_2|\mathbf{u}_2)| \geq (1-\nu)\, 2^{n(H(X_2|U_2)-\eta)} \tag{5.23}$$

for n sufficiently large.

Proof: The proof is given in Section 5.4.3.

Lemma 5.4 Given $\mu, \nu \in (0,1)$, we have

$$\sum_{(\widetilde{\mathcal{F}}_{\mu,\epsilon}^{(n)}(U_1))^c} Q_{U_1}^{(n)}(\mathbf{u}_1) \leq \frac{\eta}{\mu^2} \tag{5.24}$$

and

$$\mathbb{E}_{\mathcal{C}^{(1)}}\left[\sum_{(\widetilde{\mathcal{F}}_{\varphi_1,\nu,\epsilon}^{(n)}(U_2))^c} Q_{U_2}^{(n)}(\mathbf{u}_2)\right] \leq \frac{\mu + (1-\mu)\left(\frac{\eta}{\mu} + \lambda + \frac{\eta}{\mu^2}\right)}{\nu^2} \tag{5.25}$$

for n sufficiently large, where $\lambda \triangleq e^{-2^{n(\epsilon'-2\eta)}} e^{-2^{-n(I(X_1;U_1)+2\eta)}}$.

Proof: The proof is given in Section 5.4.4.

Note that the two encoders are designed asymmetrically. Encoder 1 works with codebook $\mathcal{C}^{(1)}$, while Encoder 2 works with both $\mathcal{C}^{(1)}$ and $\mathcal{C}^{(2)}$. Thus, we need to bound $P_1^{(n)}$ and $P_2^{(n)}$ separately.

We first bound $\mathbb{E}_{\mathcal{C}^{(1)},\mathcal{C}^{(2)}}[P_1^{(n)}]$. Since Encoder 1 only depends on $\mathcal{C}^{(1)}$, we may also write it as

$$\mathbb{E}_{\mathcal{C}^{(1)}}[P_1^{(n)}] = \mathbb{E}_{\mathcal{C}^{(1)}}\Pr\Big(\frac{1}{n}d_1\big(U_1^n, \varphi_1^{(n)}(W_1, U_1^n)\big) > D_1 + \epsilon d_1^{max}\Big).$$

For a randomly chosen subcodebook $\mathcal{C}_{w_1} = \{\mathbf{x}_1(w_1, t_1), t_1 = 1, 2, ..., L_1\}$, define

$$\theta_1 \triangleq \left\{t_1 : t_1 \le L_1 - 1, \mathbf{u}_1 \in \mathcal{F}_{\mu,\epsilon}^{(n)}(U_1|\mathbf{x}_1(w_1, t_1))\right\}.$$

Given w_1 and $\mathbf{u}_1$, define the waiting time for finding a codeword in $\mathcal{C}_{w_1}$ by

$$T(\mathbf{u}_1, \mathcal{C}_{w_1}) \triangleq \begin{cases} \min\{t_1 : t_1 \in \theta_1\}, & \text{if } \theta_1 \neq \emptyset, \\ L_1, & \text{otherwise .} \end{cases} \tag{5.26}$$

Given $\mathbf{u}_1$ and $\mathcal{C}_{w_1}$, if $T(\mathbf{u}_1, \mathcal{C}_{w_1}) \le L_1 - 1$, then we have $\big(\mathbf{u}_1, \varphi_1^{(n)}(w_1, \mathbf{u}_1)\big) \in \mathcal{F}_{\mu,\epsilon}^{(n)}(U_1, X_1) \subseteq \mathcal{T}_\epsilon^{(n)}(U_1, X_1)$. For every$\big(\mathbf{u}_1, \varphi_1^{(n)}(w_1, \mathbf{u}_1)\big) \in \mathcal{T}_\epsilon^{(n)}(U_1, X_1)$, we have

$$\frac{1}{n}d_1\big(\mathbf{u}_1, \varphi_1^{(n)}(w_1, \mathbf{u}_1)\big) \le \mathbb{E}[d_1(U_1, X_1)] + \epsilon d_1^{max} \le D_1 + \epsilon d_1^{max}$$

for n large enough, where the first inequality holds by the definition of ϵ-strong typicality, and the second inequality follows by (5.16). Thus

$$\begin{aligned} \mathbb{E}_{\mathcal{C}^{(1)}}[P_1^{(n)}] &= \mathbb{E}_{\mathcal{C}^{(1)}}\left[\frac{1}{2^{nR_w^1}}\sum_{w_1=1}^{M_w^1}\sum_{\mathcal{U}_1^n} Q_{U_1}^{(n)}(\mathbf{u}_1)\Phi\left\{T(\mathbf{u}_1, \mathcal{C}_{w_1}) \ge L_1\right\}\right] \\ &\le \sum_{\mathcal{F}_{\mu,\epsilon}^{(n)}(U_1)} Q_{U_1}^{(n)}(\mathbf{u}_1)\sum_{\mathcal{C}_{w_1}}\Pr(\mathcal{C}_{w_1})\mathbb{1}\left\{T(\mathbf{u}_1, \mathcal{C}_{w_1}) \ge L_1\right\} \\ &\quad + \sum_{(\mathcal{F}_{\mu,\epsilon}^{(n)}(U_1))^c} Q_{U_1}^{(n)}(\mathbf{u}_1). \end{aligned} \tag{5.27}$$

By Lemma 5.4, we have

$$\sum_{(\mathcal{F}_{\mu,\epsilon}^{(n)}(U_1))^c} Q_{U_1}^{(n)}(\mathbf{u}_1) \le \frac{\eta}{\mu^2} \tag{5.28}$$

for sufficiently large n. Let P_0 be the uniform distribution on $\mathcal{T}_\epsilon^{(n)}(X_1)$. For given $\mathbf{u}_1 \in \widetilde{\mathcal{F}}_{\mu,\epsilon}^{(n)}(U_1)$, we have

$$\begin{aligned}
&\sum_{\mathcal{C}_{w_1}} \Pr(\mathcal{C}_{w_1}) \mathbb{1}\left\{T(\mathbf{u}_1, \mathcal{C}_{w_1}) \geq L_1\right\} \\
&= \prod_{t_1=1}^{L_1-1} \left[1 - P_0\left\{\mathbf{x}_1(w_1,t_1) : \mathbf{u}_1 \in \mathcal{F}_{\mu,\epsilon}^{(n)}(U_1|\mathbf{x}_1(w_1,t_1))\right\}\right] && (5.29)\\
&= \left[1 - P_0\left\{\mathbf{x}_1(w_1,1) : \mathbf{u}_1 \in \mathcal{F}_{\mu,\epsilon}^{(n)}(U_1|\mathbf{x}_1(w_1,1))\right\}\right]^{L_1-1} \\
&= \left[1 - P_0\left\{\mathbf{x}_1(w_1,1) \in \mathcal{F}_{\mu,\epsilon}^{(n)}(X_1|\mathbf{u}_1)\right\}\right]^{L_1-1} \\
&= \left[1 - \frac{|\mathcal{F}_{\mu,\epsilon}^{(n)}(X_1|\mathbf{u}_1)|}{|\mathcal{T}_\epsilon^{(n)}(X_1)|}\right]^{L_1-1} \\
&\leq \left[1 - \frac{(1-\frac{\eta}{\mu})2^{n(H(X_1|U_1)-\eta)}}{2^{n(H(X_1)+\eta)}}\right]^{L_1-1} && (5.30)\\
&= \left[1 - (1-\frac{\eta}{\mu})2^{-n(I(X_1;U_1)+2\eta)}\right]^{2^{n(R_c^1-R_w^1)}-1} \\
&\leq 1 - (1-\frac{\eta}{\mu}) + \exp\{-2^{n(R_c^1-R_w^1-I(X_1;U_1)-2\eta)} - 2^{-n(I(X_1;U_1)+2\eta)}\} && (5.31)\\
&\leq \frac{\eta}{\mu} + e^{-2^{n(\epsilon'-2\eta)}} e^{-2^{-n(I(X_1;U_1)+2\eta)}} && (5.32)\\
&\triangleq \frac{\eta}{\mu} + \lambda, && (5.33)
\end{aligned}$$

where (5.30) follows from Lemmas 5.1 and 5.3, (5.31) follows from the inequality $(1-xy)^k \leq 1 - x + e^{-ky}$ for $0 \leq x, y \leq 1$, $k > 0$, (5.32) holds from (5.13), and we define $\lambda = e^{-2^{n(\epsilon'-2\eta)}} e^{-2^{-n(I(X_1;U_1)+2\eta)}}$ in (5.33), which goes to 0 as $n \to \infty$. Consequently, plugging (5.33) and (5.28) back into the average distortion expression (5.27) yields

$$\mathbb{E}_{\mathcal{C}^{(1)}}[P_1^{(n)}] \leq \frac{\eta}{\mu} + \lambda + \frac{\eta}{\mu^2} \leq \epsilon_1 \qquad (5.34)$$

for n sufficiently large (recalling that $\eta \to 0$ as $\epsilon \to 0$ and $n \to \infty$, we can make ϵ_1 arbitrarily small for n sufficiently large and ϵ sufficiently small).

Next we bound

$$\mathbb{E}_{\mathcal{C}^{(1)},\mathcal{C}^{(2)}}[P_1^{(n)}] = \mathbb{E}_{\mathcal{C}^{(1)},\mathcal{C}^{(2)}}\Pr\Big(\frac{1}{n}d_1\big(U_2^n, \varphi_1^{(n)}(W_2, U_2^n)\big) > D_2 + \epsilon d_2^{max}\Big).$$

By introducing a waiting time variable $T(\mathbf{u}_2, \mathcal{C}_{w_2})$ defined similarly as $T(\mathbf{u}_1, \mathcal{C}_{w_1})$, we know that for any $\mathbf{u}_2$ and $\mathcal{C}_{w_2}$ such that $T(\mathbf{u}_2, \mathcal{C}_{w_2}) \leq L_2 - 1$, we have $\big(\mathbf{u}_2, \varphi_2^{(n)}(w_2, \mathbf{u}_2)\big) \in \mathcal{F}^{(n)}_{\varphi_1,\nu,\epsilon}(U_2, X_2) \subseteq \mathcal{T}_\epsilon^{(n)}(U_2, X_2)$, and for any $\big(\mathbf{u}_2, \varphi_2^{(n)}(w_2, \mathbf{u}_2)\big) \in \mathcal{T}_\epsilon^{(n)}(U_2, X_2)$, and thus

$$\frac{1}{n}d_2\big(\mathbf{u}_2, \mathbf{x}_2(\mathbf{s}_2, \mathbf{u}_2)\big) \leq \mathbb{E}\big[d_2(U_2, X_2)\big] + \epsilon d_2^{max} < D_2 + \epsilon d_2^{max}$$

for n large enough. Then we have

$$\begin{aligned}
\mathbb{E}_{\mathcal{C}^{(1)},\mathcal{C}^{(2)}}[P_2^{(n)}] &= \mathbb{E}_{\mathcal{C}^{(1)},\mathcal{C}^{(2)}}\left[\frac{1}{2^{nR_w^2}}\sum_{w_2=1}^{M_w^2}\sum_{\mathcal{U}_2^n} Q_{U_2}^{(n)}(\mathbf{u}_2)\mathbb{1}\left\{T(\mathbf{u}_2, \mathcal{C}_{w_2}) \geq L_2\right\}\right] \\
&\leq \mathbb{E}_{\mathcal{C}^{(1)}}\left[\frac{1}{2^{nR_w^2}}\sum_{w_2=1}^{M_w^2}\sum_{\widetilde{\mathcal{F}}^{(n)}_{\varphi_1,\nu,\epsilon}(U_2)} Q_{U_2}^{(n)}(\mathbf{u}_2)\sum_{\mathcal{C}_{w_2}}\Pr(\mathcal{C}_{w_2})\mathbb{1}\left\{T(\mathbf{u}_2, \mathcal{C}_{w_2}) \geq L_2\right\}\right] \\
&\quad + \mathbb{E}_{\mathcal{C}^{(1)}}\left[\sum_{\left(\widetilde{\mathcal{F}}^{(n)}_{\varphi_1,\nu,\epsilon}(U_2)\right)^c} Q_{U_2}^{(n)}(\mathbf{u}_2)\right]. \qquad (5.35)
\end{aligned}$$

The first term in the right side of the above inequality can be upper bounded in a similar manner as in (5.29)–(5.33) to obtain

$$\sum_{\mathcal{C}_{w_2}}\Pr(\mathcal{C}_{w_2})\mathbb{1}\big\{T(\mathbf{u}_2, \mathcal{C}_{w_2}) > L_2\big\} \leq \nu + \lambda \qquad (5.36)$$

for any w_2 and $\mathbf{u}_2 \in \widetilde{\mathcal{F}}^{(n)}_{\varphi_1,\nu,\epsilon}(U_2)$. The second term can be upper bounded using Lemma 5.4.

Therefore, we obtain

$$\begin{aligned}
\mathbb{E}_{\mathcal{C}^{(1)},\mathcal{C}^{(2)}}[P_2^{(n)}] &\leq \nu + \lambda + \frac{\mu + (1-\mu)\Big(\frac{\eta}{\mu} + \lambda + \frac{\eta}{\mu^2}\Big)}{\nu^2} \\
&\leq \epsilon_1 \qquad (5.37)
\end{aligned}$$

for μ, ν sufficiently small and n sufficiently large.

Bounding $\mathbb{E}_{\mathcal{C}^{(1)},\mathcal{C}^{(2)}}[P_e^{(n)}]$

To analyze the average probability of error, we need the following lemma.

Lemma 5.5 For any $\epsilon_0, \epsilon \in (0,1)$, one can choose $\mu, \nu \in (0,1)$ sufficiently small such that

$$\mathbb{E}_{\mathcal{C}^{(1)},\mathcal{C}^{(2)}}\left[\frac{1}{2^{n(R_w^1+R_w^2)}}\sum_{w_1=1}^{M_w^1}\sum_{w_2=1}^{M_w^2} P_{U_1U_2}^{(n)}\left(\mathbf{v}\in\mathcal{T}_\epsilon^{(n)}(T_1,U_1,U_2,T_2)\right)\right] \geq 1-\epsilon_0$$

for n sufficiently large, where $\mathbf{v} \triangleq (\varphi_1^{(n)}(w_1,\mathbf{u}_1),\mathbf{u}_1,\mathbf{u}_2,\varphi_2^{(n)}(w_2,\mathbf{u}_2))$, and the expectation is taken with respect to the random codes $\mathcal{C}^{(1)}$ and $\mathcal{C}^{(2)}$.

Proof: The proof is very similar to the proof of the extended Markov Lemma in [54, Lemma 3] for correlated Gaussian sources. A self-contained proof is provided in Section 5.4.5 for the sake of completeness. □

We now bound the average probability of error

$$\begin{aligned} P_e^{(n)} &= \Pr\{\psi^{(n)}(Y^n,U_1^n,U_2^n)\neq(W_1,W_2)\} \\ &\leq \Pr(A_0)+\Pr\left(\{\psi^{(n)}(Y^n,U_1^n,U_2^n)\neq(W_1,W_2)\}\big|A_0^c\right), \end{aligned} \tag{5.38}$$

where A_0 is the event

$$A_0 : (\mathbf{x}_1(w_1,\mathbf{u}_1),\mathbf{u}_1,\mathbf{u}_2,\mathbf{x}_2(w_2,\mathbf{u}_2)) \notin \mathcal{T}_\epsilon^{(n)}(X_1,U_1,U_2,X_2).$$

Consequently, taking expectation in (5.38) and using the union bound, we have

$$\mathbb{E}_{\mathcal{C}^{(1)},\mathcal{C}^{(2)}}[P_e^{(n)}] \leq \mathbb{E}_{\mathcal{C}^{(1)},\mathcal{C}^{(2)}}\Pr(A_0)+\mathbb{E}_{\mathcal{C}^{(1)},\mathcal{C}^{(2)}}\Pr(E_0|A_0^c)+\sum_{k=1}^{3}\mathbb{E}_{\mathcal{C}^{(1)},\mathcal{C}^{(2)}}\Pr(E_k|A_0^c). \tag{5.39}$$

First of all, it immediately follows from Lemma 5.5 (with $\mu=\mu(\epsilon_0)$ and $\nu=\nu(\epsilon_0)$) that

$$\mathbb{E}_{\mathcal{C}^{(1)},\mathcal{C}^{(2)}}\Pr(A_0)\leq\epsilon_0 \tag{5.40}$$

for n sufficiently large. In the following, we set $\epsilon_0 = \epsilon_1/5$ throughout the proof. When A_0^c holds, since $\mathbf{y}$ is drawn from the conditional distribution $W^{(n)}_{Y|X_1X_2}(\cdot|\mathbf{x}_1, \mathbf{x}_2)$, it follows from Lemma 5.2 that

$$\mathbb{E}_{\mathcal{C}^{(1)},\mathcal{C}^{(2)}}\mathrm{Pr}\left(E_0\middle|A_0^c\right) \leq \epsilon_0 \tag{5.41}$$

for n sufficiently large. It remains to bound $\mathbb{E}_{\mathcal{C}^{(1)},\mathcal{C}^{(2)}}\mathrm{Pr}\left(E_k\middle|A_0^c\right)$ for $k = 1, 2, 3$. Using the union bound we write

$$\begin{aligned}
&\mathbb{E}_{\mathcal{C}^{(1)},\mathcal{C}^{(2)}}\mathrm{Pr}\left(E_1\middle|A_0^c\right)\\
&\quad\leq \sum_{w_1'\neq w_1}\mathrm{Pr}\Big(\big\{\big(U_1^n, U_2^n, X_1^n(w_1', U_1^n), X_2^n(w_2, U_2^n), Y^n\big) \in \mathcal{T}_\epsilon^{(n)}(U_1, U_2, X_1, X_2, Y)\big\}\big|A_0^c\Big).
\end{aligned}$$

Since $X_1^n(w_1', U_1^n) \to U_1^n \to U_2^n \to X_2^n(w_2, U_2^n)$, and by construction, $X_1^n(w_1', U_1^n)$ is independent of Y^n given U_1^n if $w_1' \neq w_1$, we have

$$\begin{aligned}
&\mathrm{Pr}\Big(\big\{\big(U_1^n, U_2^n, X_1^n(w_1', U_1^n), X_2^n(w_2, U_2^n), Y^n\big) \in \mathcal{T}_\epsilon^{(n)}(U_1, U_2, X_1, X_2, Y)\big\}\big|A_0^c\Big)\\
&\quad= \sum_{(\mathbf{u}_1,\mathbf{u}_2,\mathbf{x}_2,\mathbf{y})\in\mathcal{T}_\epsilon^{(n)}(U_1,U_2,X_2,Y)}\ \sum_{\mathbf{x}_1\in\mathcal{T}_\epsilon^{(n)}(X_1|\mathbf{u}_1,\mathbf{u}_2,\mathbf{x}_2,\mathbf{y})}\\
&\qquad\qquad \mathrm{Pr}\left(U_1^n = \mathbf{u}_1, U_2^n = \mathbf{u}_2, X_2^n(w_2, U_2^n) = \mathbf{x}_2, Y^n = \mathbf{y}\middle|A_0^c\right)\\
&\qquad\qquad \times \mathrm{Pr}\left(X_1^n(w_1', U_1^n) = \mathbf{x}_1\middle|U_1^n = \mathbf{u}_1, U_2^n = \mathbf{u}_2, X_2^n(w_2, U_2^n) = \mathbf{x}_2, Y^n = \mathbf{y}, A_0^c\right)\\
&\quad= \sum_{(\mathbf{u}_1,\mathbf{u}_2,\mathbf{x}_2,\mathbf{y})\in\mathcal{T}_\epsilon^{(n)}(U_1,U_2,X_2,Y)} \mathrm{Pr}\left(U_1^n = \mathbf{u}_1, U_2^n = \mathbf{u}_2, X_2^n(w_2, U_2^n) = \mathbf{x}_2, Y^n = \mathbf{y}\middle|A_0^c\right)\\
&\qquad\qquad \sum_{\mathbf{x}_1\in\mathcal{T}_\epsilon^{(n)}(X_1|\mathbf{u}_1,\mathbf{u}_2,\mathbf{x}_2,\mathbf{y})} \mathrm{Pr}\left(X_1^n(w_1', U_1^n) = \mathbf{x}_1\middle|U_1^n = \mathbf{u}_1\right).
\end{aligned} \tag{5.42}$$

Recalling the definition of $T(\mathbf{u}_1, \mathcal{C}_{w_1})$ in (5.26), we have

$$\begin{aligned}
&\mathrm{Pr}\big\{X_1^n(w_1', U_1^n) = \mathbf{x}_1\big|U_1^n = \mathbf{u}_1\big\}\\
&\quad= \mathrm{Pr}\big\{T(\mathbf{u}_1, \mathcal{C}_{w_1'}) < L_1\big\}\mathrm{Pr}\big\{X_1^n(w_1', U_1^n) = \mathbf{x}_1\big|U_1^n = \mathbf{u}_1, T(\mathbf{u}_1, \mathcal{C}_{w_1'}) < L_1\big\}\\
&\qquad+ \mathrm{Pr}\big\{T(\mathbf{u}_1, \mathcal{C}_{w_1'}) \geq L_1\big\}\mathrm{Pr}\big\{X_1^n(w_1', U_1^n) = \mathbf{x}_1\big|U_1^n = \mathbf{u}_1, T(\mathbf{u}_1, \mathcal{C}_{w_1'}) \geq L_1\big\}
\end{aligned}$$

$$
\begin{aligned}
&= \Pr\big\{T(\mathbf{u}_1, \mathcal{C}_{w_1'}) < L_1\big\}\frac{1}{|\mathcal{F}^{(n)}_{\mu,\epsilon}(X_1|\mathbf{u}_1)|} + \Pr\big\{T(\mathbf{u}_1, \mathcal{C}_{w_1'}) \geq L_1\big\}\frac{1}{|\mathcal{T}^{(n)}_{\epsilon}(X_1)|} && (5.43)\\
&\leq \frac{1}{|\mathcal{F}^{(n)}_{\mu,\epsilon}(X_1|\mathbf{u}_1)|} + \Big(\frac{\eta}{\mu} + \lambda\Big)\frac{1}{|\mathcal{T}^{(n)}_{\epsilon}(X_1)|} && (5.44)\\
&\leq \frac{1}{\big(1 - \frac{\eta}{\mu}\big)2^{n[H(X_1|U_1)-\eta]}} + \Big(\frac{\eta}{\mu} + \lambda\Big)\frac{1}{2^{n[H(X_1)-\eta]}} && (5.45)\\
&\leq \frac{1}{\big(1 - \frac{\eta}{\mu}\big)2^{n[H(X_1|U_1)-\eta]}} + \Big(\frac{\eta}{\mu} + \lambda\Big)\frac{1}{2^{n[H(X_1|U_1)-\eta]}} && (5.46)\\
&= \Big[\frac{1}{1 - \frac{\eta}{\mu}} + \frac{\eta}{\mu} + \lambda\Big]\frac{1}{2^{n[H(X_1|U_1)-\eta]}} && (5.47)
\end{aligned}
$$

for n sufficiently large, where (5.44) holds by applying (5.33), (5.45) follows by applying Lemma 5.3 and Lemma 5.1, and (5.46) holds since conditioning reduces entropy. Letting $\tau_1(\eta, \mu, \lambda) \triangleq \frac{1}{1-\frac{\eta}{\mu}} + \frac{\eta}{\mu} + \lambda$, and plugging (5.47) back into (5.42), we obtain

$$
\begin{aligned}
&\Pr\Big(\big\{\big(U_1^n, U_2^n, X_1^n(w_1', U_1^n), X_2^n(w_2, U_2^n), Y^n\big) \in \mathcal{T}^{(n)}_{\epsilon}(U_1, U_2, X_1, X_2, Y)\big\}\big|A_0^c\Big)\\
&\leq \sum_{(\mathbf{u}_1,\mathbf{u}_2,\mathbf{x}_2,\mathbf{y})\in\mathcal{T}^{(n)}_{\epsilon}(U_1,U_2,X_2,Y)} \Pr\big(U_1^n = \mathbf{u}_1, U_2^n = \mathbf{u}_2, X_2^n(w_2, U_2^n) = \mathbf{x}_2, Y^n = \mathbf{y}\big|A_0^c\big)\\
&\qquad \sum_{\mathbf{x}_1\in\mathcal{T}^{(n)}_{\epsilon}(X_1|\mathbf{u}_1,\mathbf{u}_2,\mathbf{x}_2,\mathbf{y})} \frac{\tau_1(\eta,\mu,\lambda)}{2^{n[H(X_1|U_1)-\eta]}}\\
&\leq \big|\mathcal{T}^{(n)}_{\epsilon}(X_1|\mathbf{u}_1, \mathbf{u}_2, \mathbf{x}_2, \mathbf{y})\big|\frac{\tau_1(\eta,\mu,\lambda)}{2^{n[H(X_1|U_1)-\eta]}} && (5.48)\\
&\leq 2^{n[H(X_1|U_1,U_2,X_2,Y)+\eta]}\frac{\tau_1(\eta,\mu,\lambda)}{2^{n[H(X_1|U_1)-\eta]}} && (5.49)\\
&\leq \tau_1(\eta,\mu,\lambda)2^{-n[I(X_1;U_2,X_2,Y|U_1)-2\eta]}\\
&= \tau_1(\eta,\mu,\lambda)2^{-n[I(X_1;U_2,X_2|U_1)+I(X_1;Y|U_1,U_2,X_2)-2\eta]}\\
&= \tau_1(\eta,\mu,\lambda)2^{-n[I(X_1;Y|U_1,U_2,X_2)-2\eta]} && (5.50)
\end{aligned}
$$

where the last inequality holds since $X_1 \to U_1 \to U_2 \to X_2$. Thus

$$
\begin{aligned}
\mathbb{E}_{\mathcal{C}^{(1)},\mathcal{C}^{(2)}}\Pr\big(E_1\big|A_0^c\big) &\leq \tau_1(\eta,\mu,\lambda)2^{nR_w^1}2^{-n(I(X_1;Y|U_1,U_2,X_2)-2\eta]}\\
&\leq \tau_1(\eta,\mu,\lambda)2^{n[I(X_1;Y|U_1,U_2,X_2)-\epsilon'-I(X_1;Y|U_1,U_2,X_2)+2\eta]} && (5.51)\\
&= \tau_1(\eta,\mu,\lambda)2^{-n(\epsilon'-2\eta)}
\end{aligned}
$$

$$\leq \epsilon_0 \tag{5.52}$$

for ϵ, μ small enough and n sufficiently large, where (5.51) follows from (5.13). Similarly we can obtain

$$\Pr\big\{X_2^n(w_2', U_2^n) = \mathbf{x}_2 \big| U_2^n = \mathbf{u}_2\big\} \;\leq\; \left[\frac{1}{1-\nu} + \nu + \lambda\right]\frac{1}{2^{n[H(X_2|U_2)-\eta]}}. \tag{5.53}$$

Defining $\tau_2(\nu,\lambda) \triangleq \frac{1}{1-\nu} + \nu + \lambda$, we have

$$\mathbb{E}_{\mathcal{C}^{(1)},\mathcal{C}^{(2)}}\Pr\left(E_2\big|A_0^c\right) \leq \tau_2(\nu,\lambda)2^{-n(\epsilon'-2\eta)} \leq \epsilon_0 \tag{5.54}$$

for ϵ small enough and n sufficiently large.

It remains to bound $\mathbb{E}_{\mathcal{C}^{(1)},\mathcal{C}^{(2)}}\Pr\left(E_3\big|A_0^c\right)$. Using the union bound we write

$$\begin{aligned}&\mathbb{E}_{\mathcal{C}^{(1)},\mathcal{C}^{(2)}}\Pr\left(E_3\big|A_0^c\right)\\ &\leq \sum_{w_1'\neq w_1}\sum_{w_2'\neq w_2}\Pr\Big(\big\{\big(U_1^n, U_2^n, X_1^n(w_1', U_1^n), X_2^n(w_2', U_2^n), Y^n\big) \in \mathcal{T}_\epsilon^{(n)}(U_1, U_2, X_1, X_2, Y)\big\}\big|A_0^c\Big).\end{aligned}$$

Since $X_1^n(w_1', U_1^n) \to U_1^n \to U_2^n \to X_2^n(w_2', U_2^n)$, and by construction, $X_1^n(w_1', U_1^n)$ and $X_2^n(w_2', U_2^n)$ are independent of Y^n given (U_1^n, U_2^n) by noting that $w_1' \neq w_1$ and $w_2' \neq w_2$, we have

$$\begin{aligned}&\Pr\Big(\big\{\big(U_1^n, U_2^n, X_1^n(w_1', U_1^n), X_2^n(w_2', U_2^n), Y^n\big) \in \mathcal{T}_\epsilon^{(n)}(U_1, U_2, X_1, X_2, Y)\big\}\big|A_0^c\Big)\\
&= \sum_{(\mathbf{u}_1,\mathbf{u}_2,\mathbf{y})\in\mathcal{T}_\epsilon^{(n)}(U_1,U_2,Y)}\;\sum_{(\mathbf{x}_1,\mathbf{x}_2)\in\mathcal{T}_\epsilon^{(n)}(X_1,X_2|\mathbf{u}_1,\mathbf{u}_2,\mathbf{y})}\Pr\big\{U_1^n = \mathbf{u}_1, U_2^n = \mathbf{u}_2, Y^n = \mathbf{y}\big|A_0^c\big\}\\
&\qquad\times\Pr\big\{X_1^n(w_1', U_1^n) = \mathbf{x}_1, X_2^n(w_2', U_2^n) = \mathbf{x}_2\big|U_1^n = \mathbf{u}_1, U_2^n = \mathbf{u}_2, Y^n = \mathbf{y}, A_0^c\big\}\\
&= \sum_{(\mathbf{u}_1,\mathbf{u}_2,\mathbf{y})\in\mathcal{T}_\epsilon^{(n)}(U_1,U_2,Y)}\;\sum_{(\mathbf{x}_1,\mathbf{x}_2)\in\mathcal{T}_\epsilon^{(n)}(X_1,X_2|\mathbf{u}_1,\mathbf{u}_2,\mathbf{y})}\Pr\big\{U_1^n = \mathbf{u}_1, U_2^n = \mathbf{u}_2, Y^n = \mathbf{y}\big|A_0^c\big\}\\
&\qquad\times\Pr\big\{X_1^n(w_1', U_1^n) = \mathbf{x}_1\big|U_1^n = \mathbf{u}_1\big\}\Pr\big\{X_2^n(w_2', U_2^n) = \mathbf{x}_2\big|U_2^n = \mathbf{u}_2\big\}\\
&\leq \tau_1(\eta,\mu,\lambda)\tau_2(\nu,\lambda)\frac{2^{n[H(X_1,X_2|U_1,U_2,Y)+\eta]}}{2^{n[H(X_1|U_1)-\eta]}2^{n(H(X_2|U_2)-\eta)}}\\
&= \tau_1(\eta,\mu,\lambda)\tau_2(\nu,\lambda)\frac{2^{n[H(X_1,X_2|U_1,U_2,Y)+\eta]}}{2^{n[H(X_1,X_2|U_1,U_2)-2\eta]}}\end{aligned} \tag{5.55}$$

$$= \tau_1(\eta,\mu,\lambda)\tau_2(\nu,\lambda)2^{-n[I(X_1,X_2;Y|U_1,U_2)-3\eta]} \tag{5.56}$$

where (5.55) holds since $X_1 \to U_1 \to U_2 \to X_2$. Thus

$$\begin{aligned}
&\mathbb{E}_{\mathcal{C}^{(1)},\mathcal{C}^{(2)}}\Pr\left(E_3\middle|A_0^c\right)\\
&\leq \tau_1(\eta,\mu,\lambda)\tau_2(\nu,\lambda)2^{n(R_w^1+R_w^2)}2^{-n(I(X_1;Y|U_1,U_2,X_2)-3\eta]}\\
&\leq \tau_1(\eta,\mu,\lambda)\tau_2(\nu,\lambda)2^{n[I(X_1,X_2;Y|U_1,U_2)-\epsilon'-I(X_1,X_2;Y|U_1,U_2)+3\eta]} && (5.57)\\
&= \tau_1(\eta,\mu,\lambda)\tau_2(\nu,\lambda)2^{-n(\epsilon'-3\eta)}\\
&\leq \epsilon_0 && (5.58)
\end{aligned}$$

for ϵ small enough and n sufficiently large, where (5.57) follows from (5.15).

Finally, substituting (5.40), (5.41), (5.52), (5.54), and (5.58) into (5.39) yields

$$\mathbb{E}_{\mathcal{C}^{(1)},\mathcal{C}^{(2)}}[P_e^{(n)}] \leq 5\epsilon_0 = \epsilon_1 \tag{5.59}$$

for ϵ sufficiently small and n substantially large.

Completing the Proof

By (5.34), (5.37) and (5.59), We obtain

$$\mathbb{E}_{\mathcal{C}^{(1)},\mathcal{C}^{(2)}}\{P_e^{(n)}+P_1^{(n)}+P_2^{(n)}\} = \mathbb{E}_{\mathcal{C}^{(1)},\mathcal{C}^{(2)}}[P_e^{(n)}]+\mathbb{E}_{\mathcal{C}^{(1)},\mathcal{C}^{(2)}}[P_1^{(n)}]+\mathbb{E}_{\mathcal{C}^{(1)},\mathcal{C}^{(2)}}[P_2^{(n)}] \leq 3\epsilon_1,$$

which implies that there exists a pair of codes $(\mathcal{C}^{(1)},\mathcal{C}^{(2)})$ such that $P_e^{(n)} \leq 3\epsilon_1$, $P_1^{(n)} \leq 3\epsilon_1$, $P_2^{(n)} \leq 3\epsilon_1$ are simultaneously satisfied for n sufficiently large.

Furthermore, if $P_i^{(n)} \leq 3\epsilon_1$, we have

$$\begin{aligned}
\frac{1}{n}\mathbb{E}[d_i(U_i^n,X_i^n)] &\leq D_i+\epsilon d_i^{max}+\mathbb{E}_{\mathcal{C}^{(1)},\mathcal{C}^{(2)}}[P_i^{(n)}]d_i^{max}\\
&\leq D_i+\delta
\end{aligned}$$

as $n \to \infty$. □

5.4.2 Proof of Theorem 5.2

We need to show that any sequence of achievable JCPW codes $(\varphi_1^{(n)}, \varphi_2^{(n)}, \psi^{(n)})$ with rate quadruple $(R_w^1, R_w^2, R_c^1, R_c^2)$ must satisfy (5.4)–(5.6) for some auxiliary RV's V with joint distribution $P_{VU_1U_2X_1X_2Y} \in \widetilde{\mathcal{P}}_{D_1,D_2}$. It follows from Fano's inequality that

$$H(W_1, W_2|Y^n, U_1^n, U_2^n) \leq n(R_w^1 + R_w^2)P_e^{(n)} + H(P_e^{(n)}) \triangleq n\epsilon_n. \tag{5.60}$$

It is clear that $\epsilon_n \to 0$ if $P_e^{(n)} \to 0$. Since

$$\begin{aligned} H(W_1|U_1^n, U_2^n, X_2^n, Y^n) &= H(W_1|U_1^n, U_2^n, Y^n) - I(W_1; X_2^n|U_1^n, U_2^n, Y^n) \\ &\leq H(W_1|U_1^n, U_2^n, Y^n) \\ &\leq H(W_1, W_2|U_1^n, U_2^n, Y^n), \end{aligned} \tag{5.61}$$

we have

$$H(W_1|U_1^n, U_2^n, X_2^n, Y^n) \leq H(W_1, W_2|Y^n, U_1^n, U_2^n) \leq n\epsilon_n. \tag{5.62}$$

Similarly,

$$H(W_2|U_1^n, U_2^n, X_1^n, Y^n) \leq n\epsilon_n. \tag{5.63}$$

Because W_1 is uniformly drawn from the message set $\{1, 2, ..., 2^{nR_w^1}\}$ and is independent of (U_1^n, U_2^n, X_2^n), we have

$$\begin{aligned} nR_w^1 &= H(W_1) \\ &= H(W_1|U_1^n, U_2^n, X_2^n) \\ &= I(W_1; Y^n|U_1^n, U_2^n, X_2^n) + H(W_1|U_1^n, U_2^n, X_2^n, Y^n) \\ &\leq I(X_1^n; Y^n|U_1^n, U_2^n, X_2^n) + n\epsilon_n \end{aligned} \tag{5.64}$$

where the last inequality follows from the data processing inequality and (5.62). Note that

$$I(X_1^n; Y^n|U_1^n, U_2^n, X_2^n) = H(Y^n|U_1^n, U_2^n, X_2^n) - H(Y^n|X_1^n, U_1^n, U_2^n, X_2^n)$$

$$\begin{aligned}
&= \sum_{j=1}^{n} H(Y_j|Y^{j-1}, U_1^n, U_2^n, X_2^n) - H(Y^n|X_1^n, X_2^n) \\
&\leq \sum_{j=1}^{n} H(Y_j|U_{1j}, U_{2j}, X_{2j}) - \sum_{j=1}^{n} H(Y_j|Y^{j-1}, X_1^n, X_2^n) \quad (5.65)\\
&= \sum_{j=1}^{n} H(Y_j|U_{1j}, U_{2j}, X_{2j}) - \sum_{j=1}^{n} H(Y_j|X_{1j}, X_{2j}) \quad (5.66)\\
&= \sum_{j=1}^{n} H(Y_j|U_{1j}, U_{2j}, X_{2j}) - \sum_{j=1}^{n} H(Y_j|X_{1j}, X_{2j}, U_{1j}, U_{2j}) \\
&= \sum_{j=1}^{n} I(X_{1j}; Y_j|U_{1j}, U_{2j}, X_{2j}),
\end{aligned}$$

where (5.65) holds since conditioning reduces entropy, and (5.66) follows by the memoryless property of the channel.

On the other hand, we have

$$\begin{aligned}
nR_w^1 &= H(W_1) \\
&= H(W_1|U_1^n) \quad (5.67)\\
&= I(W_1; X_1^n|U_1^n) + H(W_1|X_1^n, U_1^n) \\
&= I(W_1; X_1^n|U_1^n) + H(W_1|X_1^n, U_1^n, U_2^n, X_2^n) \quad (5.68)\\
&\leq I(W_1; X_1^n|U_1^n) + H(W_1|Y^n, U_1^n, U_2^n, X_2^n) \quad (5.69)\\
&\leq I(W_1; X_1^n|U_1^n) + n\epsilon_n \quad (5.70)\\
&= H(X_1^n|U_1^n) - H(X_1^n|U_1^n, W_1) + n\epsilon_n \\
&= H(X_1^n|U_1^n) + n\epsilon_n \quad (5.71)\\
&= H(X_1^n) - \big(H(X_1^n) - H(X_1^n|U_1^n)\big) + n\epsilon_n \\
&\leq nR_c^1 - I(X_1^n; U_1^n) + n\epsilon_n \quad (5.72)\\
&= nR_c^1 - H(U_1^n) + H(U_1^n|X_1^n) + n\epsilon_n \\
&\leq nR_c^1 - \sum_{j=1}^{n} H(U_{1j}) + \sum_{j=1}^{n} H(U_{1j}|X_{1j}) + n\epsilon_n \quad (5.73)
\end{aligned}$$

$$= \quad nR_c^1 - \sum_{j=1}^{n} I(U_{1j}; X_{1j}) + n\epsilon_n \tag{5.74}$$

where (5.67) holds since W_1 is independent of U_1^n; (5.68) holds from the Markov chain relationship $W_1 \to (X_1^n, U_1^n) \to (X_2^n, U_2^n)$; (5.69) holds from the data processing inequality and the Markov chain relationship $W_1 \to (U_1^n, U_2^n, X_1^n, X_2^n) \to (U_1^n, U_2^n, X_2^n, Y^n)$; (5.70) holds from (5.62); (5.71) follows from the fact that X_1^n is a deterministic function of U_1^n and W_1; (5.72) holds since $nR_c^1 \geq H(X_1^n)$; and (5.73) holds since conditioning reduces entropy.

Hence we obtain

$$R_w^1 \leq \min\Big\{R_c^1 - \frac{1}{n}\sum_{j=1}^{n} I(U_{1j}; X_{1j}),\ \frac{1}{n}\sum_{j=1}^{n} I(X_{1j}; Y_j | U_{1j}, U_{2j}, X_{2j})\Big\} + \epsilon_n. \tag{5.75}$$

Similarly, we have

$$R_w^2 \leq \min\Big\{R_c^2 - \frac{1}{n}\sum_{j=1}^{n} I(U_{2j}; X_{2j}),\ \frac{1}{n}\sum_{j=1}^{n} I(X_{2j}; Y_j | U_{1j}, U_{2j}, X_{1j})\Big\} + \epsilon_n. \tag{5.76}$$

To bound the sum of the rates, we have

$$\begin{aligned} n(R_w^1 + R_w^2) &= H(W_1) + H(W_2) \\ &= H(W_1, W_2 | U_1^n, U_2^n) \\ &= I(W_1, W_2; Y^n | U_1^n, U_2^n) + H(W_1, W_2 | U_1^n, U_2^n, Y^n) \\ &\leq I(X_1^n, X_2^n; Y^n | U_1^n, U_2^n) + n\epsilon_n \qquad (5.77) \end{aligned}$$

where (5.77) follows the data processing inequality and (5.60). Now

$$\begin{aligned} & I(X_1^n, X_2^n; Y^n | U_1^n, U_2^n) \\ &= H(Y^n | U_1^n, U_2^n) - H(Y^n | U_1^n, U_2^n, X_1^n, X_2^n) \\ &= \sum_{j=1}^{n} H(Y_j | Y^{j-1}, U_1^n, U_2^n) - H(Y^n | X_1^n, X_2^n) \end{aligned}$$

$$\begin{aligned}
&\leq \sum_{j=1}^{n} H(Y_j|U_{1j},U_{2j}) - \sum_{j=1}^{n} H(Y_j|Y^{j-1},X_1^n,X_2^n) && (5.78)\\
&= \sum_{j=1}^{n} H(Y_j|U_{1j},U_{2j}) - \sum_{j=1}^{n} H(Y_j|X_{1j},X_{2j}) && (5.79)\\
&= \sum_{j=1}^{n} H(Y_j|U_{1j},U_{2j}) - \sum_{j=1}^{n} H(Y_j|X_{1j},X_{2j},U_{1j},U_{2j})\\
&= \sum_{j=1}^{n} I(X_{1j},X_{2j};Y_j|U_{1j},U_{2j})
\end{aligned}$$

where (5.78) holds since conditioning reduces entropy, and (5.79) follows from the memoryless property of the channel. Therefore, we have

$$R_w^1 + R_w^2 \leq \frac{1}{n}\sum_{j=1}^{n} I(X_{1j},X_{2j};Y_j|U_{1j},U_{2j}) + \epsilon_n. \qquad (5.80)$$

We next introduce an auxiliary RV to simplify the bounds (5.75), (5.76), and (5.80) with single-letter characterization. Define a RV V with alphabet $\{1,2,...,n\}$ and distribution $P_V(v) = 1/n$. We next introduce RV's U_1 and U_2 which are independent of V such that

$$\Pr(U_1 = u_1, U_2 = u_2) = \Pr(U_{1j} = u_1, U_{2j} = u_2) = Q_{U_1U_2}(u_1,u_2)$$

for all $(u_1,u_2) \in \mathcal{U}_1 \times \mathcal{U}_2$, $j = 1,2,\ldots,n$. Furthermore, we define new RV's X_1, X_2, and Y by

$$\Pr(X_1 = x_1, X_2 = x_2, Y = y|V = j) = \Pr(X_{1j} = x_1, X_{2j} = x_2, Y_j = y)$$

for all $(x_1,x_2,y) \in \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}$. It follows that

$$\frac{1}{n}\sum_{j=1}^{n} I(X_{1j};Y_j|U_{1j},U_{2j},X_{2j}) = I(X_1;Y|U_1,U_2,X_2,V).$$

This shows that

$$R_w^1 \leq I(X_1;Y|U_1,U_2,X_2,V) + \epsilon_n. \qquad (5.81)$$

By a similar argument, we can show that

$$R_w^2 \leq I(X_2; Y|U_1, U_2, X_1, V) + \epsilon_n, \tag{5.82}$$

$$R_w^1 \leq R_c^1 - I(X_1; U_1|V) + \epsilon_n, \tag{5.83}$$

$$R_w^2 \leq R_c^2 - I(X_2; U_2|V) + \epsilon_n, \tag{5.84}$$

$$R_w^1 + R_w^2 \leq I(X_1, X_2; Y|U_1, U_2, V) + \epsilon_n. \tag{5.85}$$

For such RV $(V, U_1, U_2, X_1, X_2, Y)$, we have the Markov chain relationship $(V, U_1, U_2) \to (X_1, X_2) \to Y$. In fact,

$$\begin{aligned}
&\Pr(Y = y|V = j, U_1 = u_1, U_2 = u_2, X_1 = x_1, X_2 = x_2) \\
&= \Pr(Y_j = y|U_{1j} = u_1, U_{2j} = u_2, X_{1j} = x_1, X_{2j} = x_2) \\
&= \Pr(Y_j = y|X_{1j} = x_1, X_{2j} = x_2) \\
&= W_{Y|X_1X_2}(y|x_1, x_2).
\end{aligned}$$

Similarly, we can prove that the Markov chain relationship $X_1 \to (V, U_1) \to (V, U_2) \to X_2$ holds. Therefore, the joint distribution $P_{VU_1U_2X_1X_2Y}$ can be factorized as

$$P_{VU_1U_2X_1X_2Y} = P_V Q_{U_1U_2} P_{X_1|U_1V} P_{X_2|U_2V} W_{Y|X_1X_2}.$$

Next we bound the distortions $\mathbb{E}[d_i(U_i, X_i)]$. Since $(R_w^1, R_w^2, R_c^1, R_c^2)$ is achievable under the sequence of codes $(\varphi_1^{(n)}, \varphi_2^{(n)}, \psi^{(n)})$, this implies that for any $\delta > 0$ and n large enough, we have

$$\begin{aligned}
D_i + \delta &\geq \frac{1}{n}\mathbb{E}\big[d_i(U_i^n, \varphi_i^{(n)}(W_i, U_i^n))\big] \\
&= \frac{1}{n}\frac{1}{2^{nR_w^i}} \sum_{w_i=1}^{M_w^i} \sum_{\mathcal{U}_i^n} Q_{U_i}^{(n)}(\mathbf{u}_i) d_i\big(\mathbf{u}_i, \varphi_i^{(n)}(w_i, \mathbf{u}_i)\big) \\
&= \frac{1}{n} \sum_{\mathcal{U}_i^n \times \mathcal{X}_i^n} \Pr(U_i^n = \mathbf{u}_i, X_i^n = \mathbf{x}_i) d_i(\mathbf{u}_i, \mathbf{x}_i)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{n}\sum_{j=1}^{n}\sum_{\mathcal{U}_i^n\times\mathcal{X}_i^n}\Pr(U_i^n=\mathbf{u}_i,X_i^n=\mathbf{x}_i)d_i(u_{ij},x_{ij})\\
&= \sum_{j=1}^{n}P_V(V=j)\sum_{\mathcal{U}_i\times\mathcal{X}_i}\Pr(U_{ij}=u_{ij},X_{ij}=x_{ij})d_i(u_{ij},x_{ij})\\
&= \sum_{j=1}^{n}P_V(V=j)\sum_{\mathcal{U}_i\times\mathcal{X}_i}\Pr(U_i=u_i,X_i=x_i|V=j)d_i(u_i,x_i)\\
&= \sum_{j=1}^{n}\sum_{\mathcal{U}_i\times\mathcal{X}_i}\Pr(U_i=u_i,X_i=x_i,V=j)d_i(u_i,x_i)\\
&= \sum_{\mathcal{U}_i\times\mathcal{X}_i}P_{U_iX_i}(u_i,x_i)d_i(u_i,x_i).
\end{aligned}
$$

Thus we obtained that $\mathbb{E}[d_i(U_i,X_i)]\leq D_i+\delta$ for $i=1,2$. Combined with (5.81)–(5.85) and recalling that $\lim_{n\to\infty}\epsilon_n=0$ and that $\mathcal{R}(D_1,D_2)$ is closed, we conclude that $\mathcal{R}(D_1,D_2)\subset\mathcal{R}_{out}(D_1+\delta,D_2+\delta)$ for any $\delta>0$. This, and the fact that in the definition of $\mathcal{R}_{out}$, the random variable V can be taken to have a fixed finite alphabet (as we show next) implies that $\bigcap_{\delta>0}\mathcal{R}_{out}(D_1+\delta,D_2+\delta)=\mathcal{R}_{out}(D_1,D_2)$. Thus $\mathcal{R}(D_1,D_2)\subset\mathcal{R}_{out}(D_1,D_2)$ as claimed.

It remains to show that the alphabet of the random variable V can be limited by $|\mathcal{V}|\leq|\mathcal{U}_1||\mathcal{U}_2||\mathcal{X}_1||\mathcal{X}_2|+4$. To this end, we will need the following support lemma, which is based on Carathéodory's theorem.

Lemma 5.6 ([15, Support lemma, p. 311]) Let $\mathcal{P}(\mathcal{X})$ be the set of distributions defined on a finite set $\mathcal{X}$ (represented as the probability simplex in $\mathbb{R}^{|\mathcal{X}|}$) and let f_j, $j=1,2,...,k$ be real-valued continuous functions on $\mathcal{P}(\mathcal{X})$. For any probability measure μ on the Borel σ-algebra of $\mathcal{P}(\mathcal{X})$, there exist k elements $P_1,P_2,...,P_k$ of $\mathcal{P}(\mathcal{X})$ and k non-negative reals $\alpha_1,\alpha_2,...\alpha_k$ with $\sum_{i=1}^{k}\alpha_i=1$ such that for every $j=1,2,...,k$

$$\int_{\mathcal{P}(\mathcal{X})}f_j(P)\mu(dP)=\sum_{i=1}^{k}\alpha_if_j(P_i).$$

Before we actually apply the support lemma, we first rewrite all relevant mutual informations of (5.81)–(5.85)

$$\begin{aligned}
I(X_1;Y|U_1,U_2,X_2,V) &= H(Y|U_1,U_2,X_2,V) - H(Y|U_1,U_2,X_1,X_2,V) \\
&= H(Y|U_1,U_2,X_2,V) - H(Y|X_1,X_2); \qquad (5.86) \\
I(X_2;Y|U_1,U_2,X_1,V) &= H(Y|U_1,U_2,X_1,V) - H(Y|U_1,U_2,X_1,X_2,V) \\
&= H(Y|U_1,U_2,X_1,V) - H(Y|X_1,X_2); \qquad (5.87) \\
R_c^1 - I(X_1;U_1|V) &= R_c^1 - H(U_1|V) + H(U_1|X_1,V) \\
&= R_c^1 - H(U_1) + H(U_1|X_1,V); \qquad (5.88) \\
R_c^2 - I(X_2;U_2|V) &= R_c^2 - H(U_2|V) + H(U_2|X_2,V) \\
&= R_c^2 - H(U_2) + H(U_2|X_2,V); \qquad (5.89) \\
I(X_1,X_2;Y|U_1,U_2,V) &= H(Y|U_1,U_2,V) - H(Y|U_1,U_2,X_1,X_2,V) \\
&= H(Y|U_1,U_2,V) - H(Y|X_1,X_2). \qquad (5.90)
\end{aligned}$$

Note that $H(Y|X_1,X_2)$, $H(U_1)$, and $H(U_2)$ are unaffected by V since $V \to (X_1,X_2) \to Y$ forms a Markov chain relationship, and U_1 and U_2 are independent of V. Thus, it is sufficient to preserve the values $H(Y|U_1,U_2,X_1,V)$, $H(Y|U_1,U_2,X_2,V)$, $H(U_1|X_1,V)$, $H(U_2|X_2,V)$ and $H(Y|U_1,U_2,V)$.

Now define the following real-valued continuous functions of a generic distribution P over $\mathcal{U}_1 \times \mathcal{U}_2 \times \mathcal{X}_1 \times \mathcal{X}_2$ for fixed $v \in \mathcal{V}$, where $\mathcal{U}_1 \times \mathcal{U}_2 \times \mathcal{X}_1 \times \mathcal{X}_2$ is assumed to be $\{1,2,\ldots,m\}$, $m \triangleq |\mathcal{U}_1||\mathcal{U}_2||\mathcal{X}_1||\mathcal{X}_2|$ without loss of generality:

$$\begin{aligned}
f_i(P(\cdot|v)) &\triangleq P_{U_1U_2X_1X_2|V}(u_1,u_2,x_1,x_2|v),\ i \triangleq (u_1,u_2,x_1,x_2) = 1,\ldots,m-1; \\
f_m(P(\cdot|v)) &\triangleq H(Y|U_1,U_2,X_1,V=v); \\
f_{m+1}(P(\cdot|v)) &\triangleq H(Y|U_1,U_2,X_2,V=v); \\
f_{m+2}(P(\cdot|v)) &\triangleq H(U_1|X_1,V=v);
\end{aligned}$$

$$
\begin{aligned}
f_{m+3}(P(\cdot|v)) &\triangleq H(U_2|X_2, V=v);\\
f_{m+4}(P(\cdot|v)) &\triangleq H(Y|U_1, U_2, V=v).
\end{aligned}
$$

It is easy to see that the f_i, $i = 1, \ldots, m+4$, are continuous in $P(\cdot|v)$. Applying the support lemma, there must exist a new RV $\widehat{V}$ (jointly distributed with (U_1, U_2, X_1, X_2)), whose alphabet size is $|\widehat{\mathcal{V}}| = m + 4 = |\mathcal{U}_1||\mathcal{U}_2||\mathcal{X}_1||\mathcal{X}_2| + 4$, and it satisfies:

$$
\begin{aligned}
P_{U_1U_2X_1X_2}(u_1, u_2, x_1, x_2) &= \sum_{\widehat{v}\in\widehat{\mathcal{V}}} P_V(v) f_i(P(\cdot|v));\\
H(Y|U_1, U_2, X_1, \widehat{V}) &= \sum_{\widehat{v}\in\widehat{\mathcal{V}}} P_V(v) f_m(P(\cdot|v));\\
H(Y|U_1, U_2, X_2, \widehat{V}) &= \sum_{\widehat{v}\in\widehat{\mathcal{V}}} P_V(v) f_{m+1}(P(\cdot|v));\\
H(U_1|X_1, \widehat{V}) &= \sum_{\widehat{v}\in\widehat{\mathcal{V}}} P_V(v) f_{m+2}(P(\cdot|v));\\
H(U_2|X_2, \widehat{V}) &= \sum_{\widehat{v}\in\widehat{\mathcal{V}}} P_V(v) f_{m+3}(P(\cdot|v));\\
H(Y|U_1, U_2, \widehat{V}) &= \sum_{\widehat{v}\in\widehat{\mathcal{V}}} P_V(v) f_{m+4}(P(\cdot|v)).
\end{aligned}
$$

Furthermore, this RV $\widehat{V}$ maintains the distortion level $\mathbb{E}[d_i(U_i, X_i)] \leq D_i + \delta$, since the joint distribution $P_{U_1U_2X_1X_2}(u_1, u_2, x_1, x_2)$ is preserved. This completes the proof. □

5.4.3 Proof of Lemma 5.3

Using Lemma 5.1, we have

$$
\begin{aligned}
&P^{(n)}_{X_1|U_1}\left\{\mathbf{x}_1 : \mathbf{x}_1 \in \mathcal{F}^{(n)}_{\mu,\epsilon}(X_1|\mathbf{u}_1)\big|\mathbf{u}_1\right\}\\
&= \sum_{\mathbf{x}_1\in\mathcal{F}^{(n)}_{\mu,\epsilon}(X_1|\mathbf{u}_1)} P^{(n)}_{X_1|U_1}(\mathbf{x}_1|\mathbf{u}_1)\\
&\leq |\mathcal{F}^{(n)}_{\mu,\epsilon}(X_1|\mathbf{u}_1)| \cdot 2^{-n(H(X_1|U_1)-\eta)}. \qquad (5.91)
\end{aligned}
$$

On the other hand, by the Markov inequality, we have

$$\begin{aligned}
&P^{(n)}_{X_1|U_1}\left\{\mathbf{x}_1 : \mathbf{x}_1 \in \mathcal{F}^{(n)}_{\mu,\epsilon}(X_1|\mathbf{u}_1)\middle|\mathbf{u}_1\right\} \\
&= P^{(n)}_{X_1|U_1}\left\{\mathbf{x}_1 : A^{(n)}(\mathbf{u}_1,\mathbf{x}_1) \geq 1-\mu\middle|\mathbf{u}_1\right\} \\
&= 1 - P^{(n)}_{X_1|U_1}\left\{\mathbf{x}_1 : 1 - A^{(n)}(\mathbf{u}_1,\mathbf{x}_1) > \mu\middle|\mathbf{u}_1\right\} \\
&\geq 1 - \frac{1 - \mathbb{E}_{X_1^n|U_1^n}\left[A^{(n)}(\mathbf{u}_1, X_1^n)\middle|\mathbf{u}_1\right]}{\mu}, \qquad (5.92)
\end{aligned}$$

where

$$\begin{aligned}
&\mathbb{E}_{X_1^n|U_1^n}\left[A^{(n)}(\mathbf{u}_1, X_1^n)\middle|\mathbf{u}_1\right] \\
&= \Pr\left\{(\mathbf{u}_1, X_1^n, U_2^n, X_2^n) \in \mathcal{T}^{(n)}_{\epsilon}(U_1, X_1, U_2, X_2)\middle|\mathbf{u}_1\right\} \\
&\geq 1 - \eta. \qquad (5.93)
\end{aligned}$$

Combining (5.91)–(5.93), we get (5.22). Similarly, by the definition of $\mathcal{F}^{(n)}_{\varphi_1,\nu,\epsilon}(X_2|\mathbf{u}_2)$ and Lemma 5.1, we have

$$\begin{aligned}
&P^{(n)}_{X_2|U_2}\left\{\mathbf{x}_2 : \mathbf{x}_2 \in \mathcal{F}^{(n)}_{\varphi_1,\nu,\epsilon}(X_2|\mathbf{u}_2)|\mathbf{u}_2\right\} \\
&= \sum_{\mathbf{x}_2 \in \mathcal{F}^{(n)}_{\varphi_1,\nu,\epsilon}(X_2|\mathbf{u}_2)} P^{(n)}_{X_2|U_2}(\mathbf{x}_2|\mathbf{u}_2) \\
&\leq |\mathcal{F}^{(n)}_{\varphi_1,\nu,\epsilon}(X_2|\mathbf{u}_2)| \cdot 2^{-n(H(X_2|U_2)-\eta)}. \qquad (5.94)
\end{aligned}$$

for n sufficiently large. Given any $\mathbf{u}_2 \in \widetilde{\mathcal{F}}^{(n)}_{\varphi_1,\nu,\epsilon}(U_2)$, recall the definition of $\widetilde{\mathcal{F}}^{(n)}_{\varphi_1,\nu,\epsilon}(U_2)$, we have

$$P^{(n)}_{X_2|U_2}\left\{\mathbf{x}_2 : \mathbf{x}_2 \in \mathcal{F}^{(n)}_{\varphi_1,\nu,\epsilon}(X_2|\mathbf{u}_2)\middle|\mathbf{u}_2\right\} \geq 1 - \nu. \qquad (5.95)$$

(5.23) immediately holds by combining (5.94) and (5.95). □

5.4.4 Proof of Lemma 5.4

By definition

$$\Pr\left\{(U_1^n, X_1^n) \in \widetilde{\mathcal{F}}^{(n)}_{\mu,\epsilon}(U_1, X_1)\right\}$$

$$\begin{aligned}
&= \Pr\left\{(\mathbf{u}_1, \mathbf{x}_1) : A^{(n)}(\mathbf{u}_1, \mathbf{x}_1) \geq 1-\mu\right\} \\
&= 1-\Pr\left\{(\mathbf{u}_1, \mathbf{x}_1) : A^{(n)}(\mathbf{u}_1, \mathbf{x}_1) < 1-\mu\right\} \\
&= 1-\Pr\left\{(\mathbf{u}_1, \mathbf{x}_1) : 1-A^{(n)}(\mathbf{u}_1, \mathbf{x}_1) > \mu\right\} \\
&\geq 1-\frac{\mathbb{E}_{U_1^n X_1^n}[1-A^{(n)}(U_1^n, X_1^n)]}{\mu}, \qquad (5.96)
\end{aligned}$$

where the last inequality is from Markov's inequality. Since

$$\begin{aligned}
&\mathbb{E}_{U_1^n X_1^n}[1-A^{(n)}(U_1^n, X_1^n)] \\
&\quad = 1-\mathbb{E}_{U_1^n X_1^n}[A^{(n)}(U_1^n, X_1^n)] \\
&\quad = 1-\sum_{\mathcal{U}_1^n \times \mathcal{X}_1^n} P_{U_1 X_1}^{(n)}(\mathbf{u}_1, \mathbf{x}_1) A^{(n)}(\mathbf{u}_1, \mathbf{x}_1) \\
&\quad = 1-\Pr\Big\{(U_1^n, X_1^n, U_2^n, X_2^n) \in T_\epsilon^{(n)}(U_1, X_1, U_2, X_2)\Big\} \\
&\quad \leq \eta, \qquad (5.97)
\end{aligned}$$

we have

$$\Pr\Big\{(U_1^n, X_1^n) \in \widetilde{\mathcal{F}}_{\mu,\epsilon}^{(n)}(U_1, X_1)\Big\} \geq 1-\frac{\eta}{\mu}. \qquad (5.98)$$

Similarly,

$$\begin{aligned}
&\Pr\Big\{U_1^n \in \widetilde{\mathcal{F}}_{\mu,\epsilon}^{(n)}(U_1)\Big\} \\
&\quad = 1-\Pr\left\{(\mathbf{u}_1) : 1-\Pr\left\{X_1^n \in \mathcal{F}_{\mu,\epsilon}^{(n)}(X_1|\mathbf{u}_1)\big|\mathbf{u}_1\right\} > \mu\right\} \\
&\quad \geq 1-\frac{E_{P_{U_1^n}^{(n)}}\Big[1-\Pr\Big\{X_1^n \in \mathcal{F}_{\mu,\epsilon}^{(n)}(X_1|U_1^n)\big|U_1^n\Big\}\Big]}{\mu} \\
&\quad = 1-\frac{1-\Pr\Big\{(U_1^n, X_1^n) \in \mathcal{F}_{\mu,\epsilon}^{(n)}(U_1, X_1)\Big\}}{\mu} \\
&\quad \geq 1-\frac{1-(1-\frac{\eta}{\mu})}{\mu} \\
&\quad = 1-\frac{\eta}{\mu^2}. \qquad (5.99)
\end{aligned}$$

or equivalently,

$$\sum_{(\widetilde{\mathcal{F}}^{(n)}_{\mu,\epsilon}(U_1))^c} Q^{(n)}_{U_1}(\mathbf{u}_1) \leq \frac{\eta}{\mu^2}. \tag{5.100}$$

Next we prove (5.25). It follows from the definition that

$$\begin{aligned}
&\Pr\Big\{(U_2^n, X_2^n) \in \mathcal{F}^{(n)}_{\varphi_1,\nu,\epsilon}(U_2, X_2)\Big\} \\
&= \Pr\left\{(\mathbf{u}_2, \mathbf{x}_2) : B^{(n)}_{\varphi_1}(\mathbf{u}_2, \mathbf{x}_2) \geq 1-\nu\right\} \\
&= 1 - \Pr\left\{(\mathbf{u}_2, \mathbf{x}_2) : 1 - B^{(n)}_{\varphi_1}(\mathbf{u}_2, \mathbf{x}_2) > \nu\right\} \\
&\geq 1 - \frac{1 - \mathbb{E}_{U_2^n X_2^n}[B^{(n)}_{\varphi_1}(U_2^n, X_2^n)]}{\nu},
\end{aligned} \tag{5.101}$$

where the last inequality is from Markov's inequality. Since

$$\begin{aligned}
&\mathbb{E}_{U_2^n X_2^n}[B^{(n)}_{\varphi_1}(U_2^n, X_2^n)] \\
&= \sum_{\mathcal{U}_2^n \times \mathcal{X}_2^n} P^{(n)}_{U_2X_2}(\mathbf{u}_2, \mathbf{x}_2) B^{(n)}_{\varphi_1}(\mathbf{u}_2, \mathbf{x}_2) \\
&= \frac{1}{2^{nR_w^1}} \sum_{w_1=1}^{M_w^1} \Pr\Big\{(U_1^n, \varphi_1^{(n)}(w_1, U_1^n), U_2^n, X_2^n) \in \mathcal{T}^{(n)}_\epsilon(U_1, X_1, U_2, X_2)\Big\} \\
&= \frac{1}{2^{nR_w^1}} \sum_{w_1=1}^{M_w^1} \sum_{\mathcal{U}_1^n} Q^{(n)}_{U_1}(\mathbf{u}_1) \Pr\Big\{(U_2^n, X_2^n) \in \mathcal{T}^{(n)}_\epsilon(U_2, X_2 | \mathbf{u}_1, \varphi_1^{(n)}(w_1, \mathbf{u}_1)) \big| \mathbf{u}_1\Big\} \\
&= \frac{1}{2^{nR_w^1}} \sum_{w_1=1}^{M_w^1} \mathbb{E}_{U_1^n}\Big[A^{(n)}(U_1^n, \varphi_1^{(n)}(w_1, U_1^n))\Big] \\
&\geq (1-\mu)\Pr\Big\{\mathbf{u}_1 : A^{(n)}(\mathbf{u}_1, \varphi_1^{(n)}(w_1, \mathbf{u}_1)) \geq 1-\mu\Big\},
\end{aligned} \tag{5.102}$$

where (5.102) follows from Markov's inequality. Note that the probability in (5.102) is exactly the probability of successfully encoding for Encoder 1, i.e., $\Pr\big\{\mathbf{u}_1 : T(\mathbf{u}_1, \mathcal{C}_{w_1}) \leq L_1\big\}$, where $T(\mathbf{u}_1, \mathcal{C}_{w_1})$ is defined by (5.26). Taking the average over $\mathcal{C}^{(1)}$, we obtain

$$\mathbb{E}_{\mathcal{C}^{(1)}}\Big[\Pr\Big\{\mathbf{u}_1 : A^{(n)}(\mathbf{u}_1, \varphi_1^{(n)}(w_1, \mathbf{u}_1)) \geq 1-\mu\Big\}\Big]$$

$$
\begin{aligned}
&= \mathbb{E}_{\mathcal{C}_{w_1}}\Big[\Pr\Big\{\mathbf{u}_1 : T(\mathbf{u}_1, \mathcal{C}_{w_1}) \leq L_1\Big\}\Big] \\
&= 1 - \mathbb{E}_{\mathcal{C}_{w_1}}\Big[\Pr\Big\{\mathbf{u}_1 : T(\mathbf{u}_1, \mathcal{C}_{w_1}) > L_1\Big\}\Big] \\
&\geq 1 - \left(\frac{\eta}{\mu} + \lambda + \frac{\eta}{\mu^2}\right) \qquad (5.103)
\end{aligned}
$$

for n sufficiently large where the last inequality holds from (5.27)–(5.34). Hence,

$$
\begin{aligned}
&\mathbb{E}_{\mathcal{C}^{(1)}}\left[\mathbb{E}_{U_2^n X_2^n}[B_{\varphi_1}^{(n)}(U_2^n, X_2^n)]\right] \\
&\quad\geq (1-\mu)\left(1 - \Big(\frac{\eta}{\mu} + \lambda + \frac{\eta}{\mu^2}\Big)\right) \\
&\quad= 1 - \left(\mu + (1-\mu)\Big(\frac{\eta}{\mu} + \lambda + \frac{\eta}{\mu^2}\Big)\right) \qquad (5.104)
\end{aligned}
$$

for n sufficiently large. Plugging (5.104) back into (5.101), we have

$$
\begin{aligned}
&\mathbb{E}_{\mathcal{C}^{(1)}}\Big[\Pr\Big\{(U_2^n, X_2^n) \in \mathcal{F}_{\varphi_1,\nu,\epsilon}^{(n)}(U_2, X_2)\Big\}\Big] \\
&\quad\geq 1 - \frac{\mu + (1-\mu)\Big(\frac{\eta}{\mu} + \lambda + \frac{\eta}{\mu^2}\Big)}{\nu} \qquad (5.105)
\end{aligned}
$$

for n sufficiently large. Similarly, we have

$$
\begin{aligned}
&\Pr\Big\{U_2^n \in \widetilde{\mathcal{F}}_{\varphi_1,\nu,\epsilon}^{(n)}(U_2)\Big\} \\
&= 1 - \Pr\Big\{\mathbf{u}_2 : 1 - P_{X_2|U_2}^{(n)}\left\{X_2^n \in \mathcal{F}_{\varphi_1,\nu,\epsilon}^{(n)}(X_2|\mathbf{u}_2)\big|\mathbf{u}_2\right\} > \nu\Big\} \\
&\geq 1 - \frac{\mathbb{E}_{U_2^n}\Big[1 - P_{X_2|U_2}^{(n)}\Big\{X_2^n \in \mathcal{F}_{\varphi_1,\nu,\epsilon}^{(n)}(X_2|\mathbf{u}_2)\big|\mathbf{u}_2\Big\}\Big]}{\nu} \\
&= 1 - \frac{1 - \Pr\Big\{(U_2^n, X_2^n) \in \mathcal{F}_{\varphi_1,\nu,\epsilon}^{(n)}(U_2, X_2)\Big\}}{\nu}. \qquad (5.106)
\end{aligned}
$$

Thus

$$
\begin{aligned}
&\mathbb{E}_{\mathcal{C}^{(1)}}\Big[\Pr\Big\{U_2^n \in \widetilde{\mathcal{F}}_{\varphi_1,\nu,\epsilon}^{(n)}(U_2)\Big\}\Big] \\
&\geq 1 - \frac{1 - \mathbb{E}_{\mathcal{C}^{(1)}}\Big[\Pr\Big\{(U_2^n, X_2^n) \in \mathcal{F}_{\varphi_1,\nu,\epsilon}^{(n)}(U_2, X_2)\Big\}\Big]}{\nu}
\end{aligned}
$$

$$\geq \;\; 1-\frac{\mu+(1-\mu)\Big(\frac{\eta}{\mu}+\lambda+\frac{\eta}{\mu^2}\Big)}{\nu^2} \tag{5.107}$$

for n sufficiently large. □

5.4.5 Proof of Lemma 5.5

Step 1: We first state how we are going to prove this lemma. Define the following quantities

$$\begin{aligned}
1-\theta_0^{(n)} &\triangleq P^{(n)}_{X_1U_1U_2X_2}\left((\mathbf{x}_1,\mathbf{u}_1,\mathbf{u}_2,\mathbf{x}_2)\in\mathcal{T}_\epsilon^{(n)}(X_1,U_1,U_2,X_2)\right),\\
1-\theta_1^{(n)} &\triangleq \frac{1}{2^{nR_w^1}}\sum_{w_1=1}^{M_w^1} P^{(n)}_{U_1U_2X_2}\Big((\varphi_1^{(n)}(w_1,\mathbf{u}_1),\mathbf{u}_1,\mathbf{u}_2,\mathbf{x}_2)\in\mathcal{T}_\epsilon^{(n)}(X_1,U_1,U_2,X_2)\Big),\\
1-\theta_2^{(n)} &\triangleq \frac{1}{2^{n(R_w^1+R_w^2)}}\sum_{w_1=1}^{M_w^1}\sum_{w_2=1}^{M_w^2} P^{(n)}_{U_1U_2}\Big((\varphi_1^{(n)}(w_1,\mathbf{u}_1),\mathbf{u}_1,\mathbf{u}_2,\varphi_2^{(n)}(w_2,\mathbf{u}_2))\in\mathcal{T}_\epsilon^{(n)}(X_1,U_1,U_2,X_2)\Big).
\end{aligned}$$

Clearly, by the property of typicality, $\theta_0^{(n)}$ vanishes as n goes to infinity. In the following steps we upper bound $\mathbb{E}_{\mathcal{C}^{(1)}}[\theta_1^{(n)}]$ and $\mathbb{E}_{\mathcal{C}^{(1)},\mathcal{C}^{(2)}}[\theta_2^{(n)}]$ recursively.

Step 2: Upper bounding $\mathbb{E}_{\mathcal{C}^{(1)}}[\theta_1^{(n)}]$. Note that

$$1-\theta_0^{(n)}=\sum_{\mathcal{X}_1^n\times\mathcal{U}_1^n}P^{(n)}_{X_1U_1}(\mathbf{x}_1,\mathbf{u}_1)A^{(n)}(\mathbf{u}_1,\mathbf{x}_1).$$

It then follows from Markov's inequality that

$$\begin{aligned}
1-P^{(n)}_{X_1U_1}\left(\mathcal{F}^{(n)}_{\mu,\epsilon}(U_1,X_1)\right) &= P^{(n)}_{X_1U_1}\left((\mathbf{x}_1,\mathbf{u}_1): A^{(n)}(\mathbf{u}_1,\mathbf{x}_1)<1-\mu\right)\\
&= P^{(n)}_{X_1U_1}\left((\mathbf{x}_1,\mathbf{u}_1): 1-A^{(n)}(\mathbf{u}_1,\mathbf{x}_1)>\mu\right)\\
&\leq \frac{\theta_0^{(n)}}{\mu}. \qquad (5.108)
\end{aligned}$$

Thus

$$P^{(n)}_{X_1U_1}\left(\mathcal{F}^{(n)}_{\mu,\epsilon}(U_1,X_1)\right)\geq 1-\frac{\theta_0^{(n)}}{\mu}. \tag{5.109}$$

Note also that

$$P^{(n)}_{X_1U_1}\left(\mathcal{F}^{(n)}_{\mu,\epsilon}(U_1,X_1)\right)=\sum_{\mathcal{U}_1^n}P^{(n)}_{U_1}(\mathbf{u}_1)P^{(n)}_{X_1|U_1}\left(\mathcal{F}^{(n)}_{\mu,\epsilon}(X_1|\mathbf{u}_1)|\mathbf{u}_1\right).$$

Recall the definition of $\widetilde{\mathcal{F}}^{(n)}_{\mu,\epsilon}(U_1)$ that

$$\widetilde{\mathcal{F}}^{(n)}_{\mu,\epsilon}(U_1)\triangleq\left\{\mathbf{u}_1:P^{(n)}_{X_1|U_1}\left(\mathcal{F}^{(n)}_{\mu,\epsilon}(X_1|\mathbf{u}_1)|\mathbf{u}_1\right)\geq 1-\mu\right\}.$$

It then follows from Markov's inequality again and the use of (5.109) that

$$\begin{aligned}1-P^{(n)}_{U_1}(\widetilde{\mathcal{F}}^{(n)}_{\mu,\epsilon}(U_1)) &= P^{(n)}_{U_1}\left(\mathbf{u}_1:P^{(n)}_{X_1|U_1}\left(\mathcal{F}^{(n)}_{\mu,\epsilon}(X_1|\mathbf{u}_1)|\mathbf{u}_1\right)<1-\mu\right)\\ &= P^{(n)}_{U_1}\left(\mathbf{u}_1:1-P^{(n)}_{X_1|U_1}\left(\mathcal{F}^{(n)}_{\mu,\epsilon}(X_1|\mathbf{u}_1)|\mathbf{u}_1\right)>\mu\right)\\ &\leq \frac{1-\mathbb{E}_{P^{(n)}_{U_1}}\left[P^{(n)}_{X_1|U_1}\left(\mathcal{F}^{(n)}_{\mu,\epsilon}(X_1|\mathbf{u}_1)|\mathbf{u}_1\right)\right]}{\mu}\\ &\leq \frac{1-(1-\frac{\theta^{(n)}_0}{\mu})}{\mu}=\frac{\theta^{(n)}_0}{\mu^2},\end{aligned}$$

which yields

$$P^{(n)}_{U_1}(\widetilde{\mathcal{F}}^{(n)}_{\mu,\epsilon}(U_1))\geq 1-\frac{\theta^{(n)}_0}{\mu^2}. \tag{5.110}$$

On the other hand, define

$$\mathcal{A}(w_1,l_1)\triangleq\left\{\mathbf{u}_1\in\widetilde{\mathcal{F}}^{(n)}_{\mu,\epsilon}(U_1):\varphi^{(n)}_1(w_1,\mathbf{u}_1)=\mathbf{x}_1(w_1,l_1)\right\}.$$

Then we can write

$$\begin{aligned}&1-\theta^{(n)}_1\\ &=\frac{1}{2^{nR^1_w}}\sum_{w_1=1}^{M^1_w}\sum_{\mathbf{u}_1\in\mathcal{U}_1^n}P^{(n)}_{U_1}(\mathbf{u}_1)P^{(n)}_{U_2X_2|U_1}\left((\mathbf{u}_2,\mathbf{x}_2)\in\mathcal{T}^{(n)}_\epsilon\big(U_2,X_2|\mathbf{u}_1,\varphi^{(n)}_1(w_1,\mathbf{u}_1)\big)\big|\mathbf{u}_1\right)\\ &\geq\frac{1}{2^{nR^1_w}}\sum_{w_1=1}^{M^1_w}\sum_{l_1=1}^{L_1-1}\sum_{\mathbf{u}_1\in\mathcal{A}(w_1,l_1)}P^{(n)}_{U_1}(\mathbf{u}_1)P^{(n)}_{U_2X_2|U_1}\left((\mathbf{u}_2,\mathbf{x}_2)\in\mathcal{T}^{(n)}_\epsilon\big(U_2,X_2|\mathbf{u}_1,\varphi^{(n)}_1(w_1,\mathbf{u}_1)\big)\big|\mathbf{u}_1\right)\end{aligned}$$

$$= \frac{1}{2^{nR_w^1}} \sum_{w_1=1}^{M_w^1} \sum_{l_1=1}^{L_1-1} \sum_{\mathbf{u}_1 \in \mathcal{A}(w_1,l_1)} P_{U_1}^{(n)}(\mathbf{u}_1) A^{(n)}\big(\mathbf{u}_1, \mathbf{x}_1(w_1,l_1)\big) \tag{5.111}$$

where the last equality holds since $X_1 \to U_1 \to U_2 \to X_2$. By definition, $\widetilde{\mathcal{F}}_{\mu,\epsilon}^{(n)}(U_1) \subseteq \mathcal{F}_{\mu,\epsilon}^{(n)}(U_1) \subseteq \mathcal{T}_\epsilon^{(n)}(U_1)$ and $\mathcal{A}(w_1,l_1) \subseteq \widetilde{\mathcal{F}}_{\mu,\epsilon}^{(n)}(U_1)$. Given $l_1 \in \{1,\ldots,L_1-1\}$ and $\mathbf{u}_1 \in \mathcal{A}(w_1,l_1)$, noting that $\varphi_1^{(n)}(w_1,\mathbf{u}_1) = \mathbf{x}_1(w_1,l_1)$, we have

$$A^{(n)}(\mathbf{u}_1, \mathbf{x}_1(w_1,l_1)) = A^{(n)}(\mathbf{u}_1, \varphi_1^{(n)}(w_1,\mathbf{u}_1)) \geq 1-\mu$$

where the inequality follows from the definition of the encoder $\varphi_1^{(n)}$. Plugging this into (5.111), we have

$$\begin{aligned}
1-\theta_1^{(n)} &\geq (1-\mu)\frac{1}{2^{nR_w^1}} \sum_{w_1=1}^{M_w^1} \sum_{l_1=1}^{L_1-1} \sum_{\mathbf{u}_1 \in \mathcal{A}(w_1,l_1)} P_{U_1}^{(n)}(\mathbf{u}_1) \\
&= (1-\mu)\frac{1}{2^{nR_w^1}} \sum_{w_1=1}^{M_w^1} \sum_{l_1=1}^{L_1-1} P_{U_1}^{(n)}\big(\mathcal{A}(w_1,l_1)\big) \\
&\overset{(a)}{=} (1-\mu)\frac{1}{2^{nR_w^1}} \sum_{w_1=1}^{M_w^1} P_{U_1}^{(n)}\left(\bigcup_{l_1=1}^{L_1-1} \mathcal{A}(w_1,l_1)\right) \\
&= (1-\mu)\frac{1}{2^{nR_w^1}} \sum_{w_1=1}^{M_w^1} P_{U_1}^{(n)}\left(\widetilde{\mathcal{F}}_{\mu,\epsilon}^{(n)}(U_1) \setminus \mathcal{A}(w_1,L_1)\right) \\
&= (1-\mu)\frac{1}{2^{nR_w^1}} \sum_{w_1=1}^{M_w^1} \left[P_{U_1}^{(n)}(\widetilde{\mathcal{F}}_{\mu,\epsilon}^{(n)}(U_1)) - P_{U_1}^{(n)}\big(\mathcal{A}(w_1,L_1)\big)\right] \\
&\overset{(b)}{\geq} 1-\mu-\frac{\theta_0^{(n)}}{\mu^2} - \frac{1}{2^{nR_w^1}} \sum_{w_1=1}^{M_w^1} P_{U_1}^{(n)}\big(\mathcal{A}(w_1,L_1)\big),
\end{aligned}$$

where (a) holds since $\mathcal{A}(w_1,l_1)$ and $\mathcal{A}(w_1,\tilde{l}_1)$ are disjoint for any $l_1 \neq \tilde{l}_1$, $l_1,\tilde{l}_1 \in \{2,\ldots,L_1\}$, and (b) holds from (5.110). Equivalently,

$$\theta_1^{(n)} \leq \mu + \frac{\theta_0^{(n)}}{\mu^2} + \frac{1}{2^{nR_w^1}} \sum_{w_1=1}^{M_w^1} P_{U_1}^{(n)}\big(\mathcal{A}(w_1,L_1)\big). \tag{5.112}$$

Consequently, to upper bound $\mathbb{E}_{\mathcal{C}^{(1)}}[\theta_1^{(n)}]$, it suffices to upper bound

$$\begin{aligned}
&\mathbb{E}_{\mathcal{C}^{(1)}}\left[\frac{1}{2^{nR_w^1}}\sum_{w_1=1}^{M_w^1} P_{U_1}^{(n)}\big(\mathcal{A}(w_1,L_1)\big)\right]\\
&= \sum_{\mathbf{u}_1\in\widetilde{\mathcal{F}}_{\mu,\epsilon}^{(n)}(U_1)} P_{U_1}^{(n)}(\mathbf{u}_1)\mathbb{E}_{\mathcal{C}^{(1)}}\left[\frac{1}{2^{nR_w^1}}\sum_{w_1=1}^{M_w^1}\mathbb{1}\left\{\mathbf{u}_1\notin\bigcup_{l_1=1}^{L_1-1}\mathcal{F}_{\mu,\epsilon}^{(n)}\big(U_1|\mathbf{x}_1(w_1,l_1)\big)\right\}\right].\qquad(5.113)
\end{aligned}$$

Note that for $\mathbf{u}_1\in\widetilde{\mathcal{F}}_{\mu,\epsilon}^{(n)}(U_1)$, we have

$$\begin{aligned}
&\mathbb{E}_{\mathcal{C}^{(1)}}\left[\frac{1}{2^{nR_w^1}}\sum_{w_1=1}^{M_w^1}\mathbb{1}\left\{\mathbf{u}_1\notin\bigcup_{l_1=1}^{L_1-1}\mathcal{F}_{\mu,\epsilon}^{(n)}\big(U_1|\mathbf{x}_1(w_1,l_1)\big)\right\}\right]\\
&= \sum_{\mathcal{C}_{w_1}}\Pr(\mathcal{C}_{w_1})\mathbb{1}\left\{T(\mathbf{u}_1,\mathcal{C}_{w_1})\geq L_1\right\}\\
&\leq \frac{\eta}{\mu}+\lambda, \qquad (5.114)
\end{aligned}$$

where $T(\mathbf{u}_1,\mathcal{C}_{w_1})$ is defined by (5.26), and the inequality follows from (5.33). Now let n be sufficiently large so that $\theta_0^{(n)}\leq\mu^3$, it then follows from (5.112) that $\mathbb{E}_{\mathcal{C}^{(1)}}[\theta_1^{(n)}]\leq 2\mu+\frac{\eta}{\mu}+\lambda$ for n large enough.

Step 3: Upper bounding $\mathbb{E}_{\mathcal{C}^{(1)},\mathcal{C}^{(2)}}[\theta_2^{(n)}]$. We first write

$$\begin{aligned}
1-\theta_1^{(n)} &= \frac{1}{2^{nR_w^1}}\sum_{w_1=1}^{M_w^1} P_{U_2X_2}^{(n)}(\mathbf{u}_2,\mathbf{x}_2)P_{U_1|U_2X_2}^{(n)}\Big\{(\mathbf{u}_1,\varphi_1^{(n)}(w_1,\mathbf{u}_1))\in\mathcal{T}_\epsilon^{(n)}(U_1,X_1|\mathbf{u}_2,\mathbf{x}_2)\Big|\mathbf{u}_2,\mathbf{x}_2\Big\}\\
&= P_{U_2X_2}^{(n)}(\mathbf{u}_2,\mathbf{x}_2)B_{\varphi_1}^{(n)}(\mathbf{u}_2,\mathbf{x}_2).
\end{aligned}$$

It then follows from Markov's inequality

$$\begin{aligned}
1-P_{U_2X_2}^{(n)}\left(\mathcal{F}_{\varphi_1,\nu,\epsilon}^{(n)}(U_2,X_2)\right) &= P_{U_2X_2}^{(n)}\left((\mathbf{u}_2,\mathbf{x}_2):1-B_{\varphi_1}^{(n)}(\mathbf{u}_2,\mathbf{x}_2)>\nu\right)\\
&\leq \frac{\theta_1^{(n)}}{\nu}, \qquad (5.115)
\end{aligned}$$

hence

$$P_{U_2X_2}^{(n)}\left(\mathcal{F}_{\varphi_1,\nu,\epsilon}^{(n)}(U_2,X_2)\right)\geq 1-\frac{\theta_1^{(n)}}{\nu}.$$

Observing that

$$P^{(n)}_{U_2X_2}\left(\mathcal{F}^{(n)}_{\varphi_1,\nu,\epsilon}(U_2,X_2)\right)=\sum_{\mathcal{U}_2^n}P^{(n)}_{U_2}(\mathbf{u}_2)P^{(n)}_{X_2|U_2}\left(\mathcal{F}^{(n)}_{\varphi_1,\nu,\epsilon}(X_2|\mathbf{u}_2)|\mathbf{u}_2\right),$$

and recall the definition of $\widetilde{\mathcal{F}}^{(n)}_{\varphi_1,\nu,\epsilon}(U_2)$ that

$$\widetilde{\mathcal{F}}^{(n)}_{\varphi_1,\nu,\epsilon}(U_2)=\left\{\mathbf{u}_2:P^{(n)}_{X_2|U_2}\left(\mathcal{F}^{(n)}_{\varphi_1,\nu,\epsilon}(X_2|\mathbf{u}_2)|\mathbf{u}_2\right)\geq 1-\nu\right\},$$

we can obtain by using Markov's inequality again and (5.115) that

$$P^{(n)}_{U_2}(\widetilde{\mathcal{F}}^{(n)}_{\varphi_1,\nu,\epsilon}(U_2))\geq 1-\frac{\theta_1^{(n)}}{\nu^2}. \tag{5.116}$$

Similarly, define

$$\mathcal{B}(w_2,l_2)\triangleq\left\{\mathbf{u}_2\in\widetilde{\mathcal{F}}^{(n)}_{\varphi_1,\nu,\epsilon}(U_2):\varphi_2^{(n)}(w_2,\mathbf{u}_2)=\mathbf{x}_2(w_2,l_2)\right\},$$

we can lower bound

$$\begin{aligned}
&1-\theta_2^{(n)}\\
&=\frac{1}{2^{n(R_w^1+R_w^2)}}\sum_{w_1=1}^{M_w^1}\sum_{w_2=1}^{M_w^2}\sum_{\mathcal{U}_2^n}P^{(n)}_{U_2}(\mathbf{u}_2)\\
&\quad P^{(n)}_{U_1|U_2}\left((\varphi_1^{(n)}(w_1,\mathbf{u}_1),\mathbf{u}_1,\mathbf{u}_2,\varphi_2^{(n)}(w_2,\mathbf{u}_2))\in\mathcal{T}_\epsilon^{(n)}(X_1,U_1,U_2,X_2)\big|\mathbf{u}_2\right)\\
&\geq\frac{1}{2^{n(R_w^1+R_w^2)}}\sum_{w_1=1}^{M_w^1}\sum_{w_2=1}^{M_w^2}\sum_{l_2=1}^{L_2-1}\sum_{\mathbf{u}_2\in\mathcal{B}(w_2,l_2)}P^{(n)}_{U_2}(\mathbf{u}_2)\\
&\quad P^{(n)}_{U_1|U_2}\left(\left(\varphi_1^{(n)}(w_1,\mathbf{u}_1),\mathbf{u}_1,\mathbf{u}_2,\mathbf{x}_2(w_2,l_2)\right)\in\mathcal{T}_\epsilon^{(n)}(X_1,U_1,U_2,X_2)\big|\mathbf{u}_2\right)\\
&=\frac{1}{2^{nR_w^2}}\sum_{w_2=1}^{M_w^2}\sum_{l_2=1}^{L_2-1}\sum_{\mathbf{u}_2\in\mathcal{B}(w_2,l_2)}P^{(n)}_{U_2}(\mathbf{u}_2)B^{(n)}_{\varphi_1}(\mathbf{u}_2,\mathbf{x}_2(w_2,l_2))
\end{aligned}\tag{5.117}$$

where the last equality holds since $X_2\to U_2\to U_1\to X_1$. By definition, $\widetilde{\mathcal{F}}^{(n)}_{\varphi_1,\nu,\epsilon}(U_2)\subseteq\mathcal{F}^{(n)}_{\varphi_1,\nu,\epsilon}(U_2)$ and $\mathcal{B}(w_2,l_2)\subseteq\widetilde{\mathcal{F}}^{(n)}_{\varphi_1,\nu,\epsilon}(U_2)$. Given $l_2\in\{1,\ldots,L_2-1\}$ and $\mathbf{u}_2\in\mathcal{B}(w_2,l_2)$, noting that $\varphi_2^{(n)}(w_2,\mathbf{u}_2)=\mathbf{x}_2(w_2,l_2)$, we have

$$B^{(n)}_{\varphi_1}(\mathbf{u}_2,\mathbf{x}_2(w_2,l_2))=B^{(n)}_{\varphi_1}(\mathbf{u}_2,\varphi_2^{(n)}(w_2,\mathbf{u}_2))\geq 1-\nu$$

Plugging it into (5.117), we have

$$\begin{aligned}
1-\theta_2^{(n)} & \\
\geq\ & (1-\nu)\frac{1}{2^{nR_w^2}}\sum_{w_2=1}^{M_w^2}\sum_{l_2=1}^{L_2-1}\sum_{\mathbf{u}_2\in\mathcal{B}(w_2,l_2)}P_{U_2}^{(n)}(\mathbf{u}_2) \\
=\ & (1-\nu)\frac{1}{2^{nR_w^2}}\sum_{w_2=1}^{M_w^2}\sum_{l_2=1}^{L_2-1}P_{U_2}^{(n)}\big(\mathcal{B}(w_2,l_2)\big) \\
=\ & (1-\nu)\frac{1}{2^{nR_w^2}}\sum_{w_2=1}^{M_w^2}P_{U_2}^{(n)}\left(\bigcup_{l_2=1}^{L_2-1}\mathcal{B}(w_2,l_2)\right) \\
=\ & (1-\nu)\frac{1}{2^{nR_w^2}}\sum_{w_2=1}^{M_w^2}P_{U_2}^{(n)}\left(\widetilde{\mathcal{F}}_{\varphi_1,\nu,\epsilon}^{(n)}(U_2)\setminus\mathcal{B}(w_2,L_2)\right) \\
=\ & (1-\nu)\frac{1}{2^{nR_w^2}}\sum_{w_2=1}^{M_w^2}\left[P_{U_2}^{(n)}(\widetilde{\mathcal{F}}_{\varphi_1,\nu,\epsilon}^{(n)}(U_2))-P_{U_2}^{(n)}\big(\mathcal{B}(w_2,L_2)\big)\right] \\
\geq\ & 1-\nu-\frac{\theta_1^{(n)}}{\nu^2}-\frac{1}{2^{nR_w^2}}\sum_{w_2=1}^{M_w^2}P_{U_2}^{(n)}\big(\mathcal{B}(w_2,L_2)\big), \qquad (5.118)
\end{aligned}$$

where the last inequality follows from (5.116). Therefore, we have

$$\theta_2^{(n)}\leq\nu+\frac{\theta_1^{(n)}}{\nu^2}+\frac{1}{2^{nR_w^2}}\sum_{w_2=1}^{M_w^2}P_{U_2}^{(n)}\big(\mathcal{B}(w_2,L_2)\big).$$

As in (5.113)–(5.114), by using the result (5.36), we obtain

$$\begin{aligned}
\mathbb{E}_{\mathcal{C}^{(1)},\mathcal{C}^{(2)}}[\theta_2^{(n)}] &\leq \nu+\frac{\mathbb{E}_{\mathcal{C}^{(1)}}[\theta_1^{(n)}]}{\nu^2}+\mathbb{E}_{\mathcal{C}^{(1)},\mathcal{C}^{(1)}}\left[\frac{1}{2^{nR_w^2}}\sum_{w_2=1}^{M_w^2}P_{U_2}^{(n)}\big(\mathcal{B}(w_2,L_2)\big)\right] \\
&\leq \nu+\frac{2\mu+\frac{\eta}{\mu}+\lambda}{\nu^2}+\nu+\lambda \qquad (5.119)
\end{aligned}$$

for n large enough. Finally, choosing μ and ν small enough and n large enough, we obtain $\mathbb{E}_{\mathcal{C}^{(1)},\mathcal{C}^{(2)}}[\theta_2^{(n)}]\geq\epsilon_0$. □

5.5 Conclusions

In this chapter, we studied the joint compression and private watermarking problem with a multi-user setting, where two users separately embed their secret messages W_1 and W_2 (at rates R_w^1 and R_w^2 respectively) into two correlated DMS's (U_1^n, U_2^n), and transmit the compressed stegotexts X_1^n and X_2^n (at rates R_c^1 and R_c^2 respectively) over a memoryless MAC. We established an inner bound and an outer bound with single-letter characterization for the rate region of all achievable rate quadruples $(R_w^1, R_w^2, R_c^1, R_c^2)$ with respect to the distortion levels (D_1, D_2). We do not provide conditions for which the inner bound and the outer bound are tight. However, for the special single-user case and no-attack case, we observe that the bounds are actually tight.

Chapter 6

Private Information Hiding of Correlated Sources Under Multiple Access Attacks

This chapter is based on the paper presented at the *IEEE International Symposium on Information Theory* (ISIT'07), Nice, France, June 24-29, 2007 [87].

6.1 Introduction

In practical situations (e.g., instant (online) data-hiding), in order to reduce the complexity of coding, we may need to directly hide an information source (or correlated sources) with a nonuniform distribution. In this chapter, instead of transmitting two independent watermark messages, we consider the private information hiding of two correlated secret sources. Our model is depicted in Fig. 6.1. Instead of embedding uniformly distributed indices, two encoders independently embed two (arbitrarily distributed) discrete memoryless correlated sources (S_1, S_2) into a common memoryless host source U, and transmit the resulting sequences to a common destination in the presence of discrete memoryless multiple access channel (MAC) attacks. One possible application of this scenario is that two agents separately embed noisy observations of the same source, and transmit the hidden information over a MAC attack channel.

Given the secret sources (S_1, S_2), a MAC $W_{Y|X_1X_2}$, the host source U, and a distortion level pair (D_1, D_2), one may ask whether there exists a coding scheme, such that (S_1, S_2) can be embedded in U within distortion levels (D_1, D_2), and transmitted over $W_{Y|X_1X_2}$ with an arbitrarily small probability of error. To begin, we note that, especially in a multi-user system, jointly source coding and embedding the sources (S_1, S_2) into U might perform better than the traditional separate coding (i.e., concatenating lossless data compression and embedding). In this section, we investigate whether (S_1, S_2) can be successfully transmitted under the MAC attacks by joint source coding and embedding codes. In particular, we establish a sufficient condition for successfully embedding (S_1, S_2) into U under the MAC $W_{Y|X_1X_2}$; see Theorem 6.1. Note that our problem can be viewed as a generalization of the problem of transmitting correlated sources over ordinary MAC channels [1], [9], [71], [72].

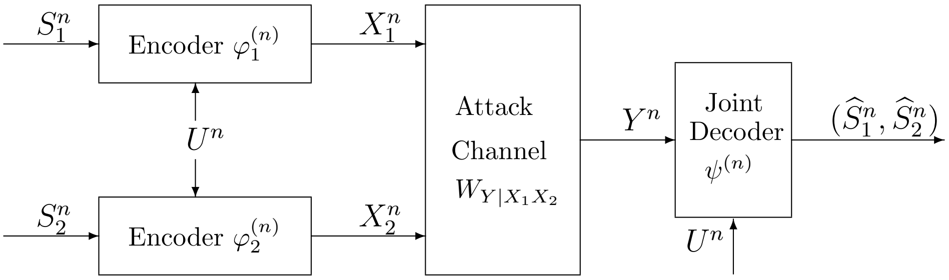


Figure 6.1: A joint source coding and embedding model for multi-user information hiding.

Related works on multi-user information hiding include [36], [76]. They are different from this work in the following aspects: first, they study a public information-embedding scenario; and second, the secret sources (watermarks) are independent and uniformly distributed. To the best of our knowledge, the private multi-user information hiding problem with correlated secret sources has not been addressed before.

6.2 Problem Formulation and Main Results

Let the pair of memoryless correlated secret sources $\{(S_{1j}, S_{2j})\}_{j=1}^{\infty}$ have marginal distribution $P_{S_1S_2}$ and denote the marginal distribution of the host source $\{U_j\}_{j=1}^{\infty}$ by P_U. Assume (S_1, S_2) and U are independent. The attack channel is modeled as a two-sender one-receiver discrete memoryless MAC $W_{Y|X_1X_2}$ having input alphabets $\mathcal{X}_1$ and $\mathcal{X}_2$, output alphabet $\mathcal{Y}$, and a transition probability distribution $W_{Y|X_1X_2}(y|x_1, x_2)$. The probability of receiving $\mathbf{y} \in \mathcal{Y}^n$ conditioned on sending $\mathbf{x}_1 \in \mathcal{X}_1^n$ and $\mathbf{x}_2 \in \mathcal{X}_2^n$ is hence given by $W^{(n)}_{Y|X_1X_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2) = \prod_{j=1}^n W_{Y|X_1X_2}(y_j|x_{1j}, x_{2j})$. Let $d_i : \mathcal{U} \times \mathcal{X}_i \to [0, \infty)$ be single-letter distortion measures and define $d_i^{max} \triangleq \max\limits_{u,x_i} d_i(u, x_i)$ for $i = 1, 2$. For $\mathbf{u} \in \mathcal{U}^n$ and $\mathbf{x}_i \in \mathcal{X}_i^n$, let $d_i(\mathbf{u}, \mathbf{x}_i) = \sum_{j=1}^n d_i(u_j, x_{ij})$. All alphabets are finite.

A joint source coding and embedding (JSCE) code $(\varphi_1^{(n)}, \varphi_2^{(n)}, \psi^{(n)})$ with block length n consists of two encoders $\varphi_1^{(n)} : \mathcal{S}_1^n \times \mathcal{U}^n \to \mathcal{X}_1^n$ and $\varphi_2^{(n)} : \mathcal{S}_2^n \times \mathcal{U}^n \to \mathcal{X}_2^n$ and a decoder $\psi^{(n)} : \mathcal{Y}^n \times \mathcal{U}^n \to \mathcal{S}_1^n \times \mathcal{S}_2^n$; see Fig. 6.1. The probability of error in reproducing the secret sources is given by

$$\begin{aligned} P_e^{(n)} &= \Pr\big\{\psi^{(n)}(Y^n, U^n) \neq (S_1^n, S_2^n)\big\} \\ &= \sum_{\mathcal{S}_1^n \times \mathcal{S}_2^n \times \mathcal{U}^n} P^{(n)}_{S_1S_2}(\mathbf{s}_1, \mathbf{s}_2) P^{(n)}_U(\mathbf{u}) \sum_{\mathbf{y}:\psi^{(n)}(\mathbf{y},\mathbf{u}) \neq (\mathbf{s}_1,\mathbf{s}_2)} W^{(n)}_{Y|X_1X_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2) \end{aligned}$$

where $\mathbf{x}_i \triangleq \varphi_i^{(n)}(\mathbf{s}_i, \mathbf{u})$ $(i = 1, 2)$.

Definition 6.1 Given P_U and distortion levels $D_1 > 0$ and $D_2 > 0$, we say that the secret sources $\{(S_{1j}, S_{2j})\}$ are (D_1, D_2)-*admissible* with respect to the MAC $W_{Y|X_1X_2}$, if there exists a sequence of codes $(\varphi_1^{(n)}, \varphi_2^{(n)}, \psi^{(n)})$ such that $P_e^{(n)} \to 0$ as $n \to \infty$ and $\limsup_{n\to\infty} \frac{1}{n}\mathbb{E}\big[d_i(U^n, \varphi_i^{(n)}(S_i^n, U^n))\big] \le D_i$ for $i = 1, 2$.

Theorem 6.1 $\{(S_{1j}, S_{2j})\}$ are (D_1, D_2)-*admissible* with respect to the MAC $W_{Y|X_1X_2}$ if there exist some RV Q and a pair of conditional distributions $(P_{X_1|S_1UQ}, P_{X_2|S_2UQ})$ such that

$$H(S_1|S_2) < I(X_1; Y|X_2, S_2, U, Q), \tag{6.1}$$

$$H(S_2|S_1) < I(X_2; Y|X_1, S_1, U, Q), \tag{6.2}$$

$$H(S_1, S_2) < I(X_1, X_2; Y|U, Q), \tag{6.3}$$

$$\mathbb{E}[d_i(U, X_i)] \le D_i,\ i = 1, 2, \tag{6.4}$$

where the above entropies, mutual informations, and expectations are taken with respect to the joint distribution

$$P_Q P_{S_1S_2} P_U P_{X_1|S_1UQ} P_{X_2|S_2UQ} W_{Y|X_1X_2}. \tag{6.5}$$

We remark that the RV Q serves as a time-sharing RV and the cardinality of its alphabet can be bounded by $|\mathcal{Q}| \le 5$ ([10]).

The proof of the theorem, which employs a joint strong typicality coding argument [9] under additional distortion constraints, is deferred to Section 6.3.1. Note that if U is removed in (6.1)–(6.3), then the inequalities reduce to the sufficient condition under which the sources $\{(S_{1j}, S_{2j})\}$ can be reliably transmitted over the MAC $W_{Y|X_1X_2}$ obtained in [9], [1].

Although we are unable to obtain a converse to Theorem 6.1 in single-letter form, we can still obtain an "n-dimensional" embedding theorem.

Theorem 6.2 $\{S_{1j}, S_{2j}\}_{j=1}^{\infty}$ can be sent with (asymptotically) arbitrarily small probability of error over the MAC $W_{Y|X_1X_2}$ with block codes $\big\{(\varphi_1^{(n)}(S_1^n, U^n), \varphi_2^{(n)}(S_2^n, U^n))\big\}$ satisfying $\frac{1}{n}\mathbb{E}[d_i(U^n, \varphi_i^{(n)}(S_i^n, U^n))] \le D_i, i = 1, 2$, if and only if

$$\big(H(S_1|S_2), H(S_2|S_1), H(S_1, S_2)\big) \in cl\bigg(\bigcup_{n=1}^{\infty} \mathcal{R}_n\bigg), \tag{6.6}$$

where $cl(B)$ denotes the closure of a set $B \subset \mathbb{R}^3$ and

$$\begin{aligned}\mathcal{R}_n = \bigcup_{\substack{P^{(n)}_{X_i|S_iU}:\\ \mathbb{E}[d_i(U^n,X_i^n)]\le nD_i,\, i=1,2}} \bigg\{(R_1, R_2, R_3) : R_1 &< \frac{1}{n}I(X_1^n; Y^n|S_2^n, U^n, X_2^n),\\ R_2 &< \frac{1}{n}I(X_2^n; Y^n|S_1^n, U^n, X_1^n),\\ R_3 &< \frac{1}{n}I(X_1^n, X_2^n; Y^n|U^n) \quad \bigg\}.\end{aligned} \tag{6.7}$$

for some joint distribution

$$\prod_{j=1}^{n} P_{S_1S_2}(s_{1j}, s_{2j})P_U(u_j)\Pr(\mathbf{x}_1|\mathbf{s}_1, \mathbf{u}_1)\Pr(\mathbf{x}_2|\mathbf{s}_2, \mathbf{u}_2)\prod_{j=1}^{n} P_{Y|X_1X_2}(y_j|x_{1j}, x_{2j}).$$

6.2.1 Special Cases

Uniform and Independent Sources

Suppose that the sources are independent and uniform, i.e., $P_{S_1}(s_1) = 1/|\mathcal{S}_1|$, $P_{S_2}(s_2) = 1/|\mathcal{S}_2|$ and $P_{S_1S_2}(s_1, s_2) = P_{S_1}(s_1)P_{S_2}(s_2)$ for any $(s_1, s_2) \in \mathcal{S}_1 \times \mathcal{S}_2$. Define $\widetilde{R}_1 = H(S_1) = \log|\mathcal{S}_1|$ and $\widetilde{R}_2 = H(S_2) = \log|\mathcal{S}_2|$ to be the rates of the sources. By Theorem 6.1, $\{(S_{1j}, S_{2j})\}$ are (D_1, D_2)-admissible with respect to the MAC $W_{Y|X_1X_2}$ if there exists some RV Q with $|\mathcal{Q}| \le 5$, and a pair of distributions $(P_{X_1|UQ}, P_{X_2|UQ})$ such that

$$\widetilde{R}_1 < I(X_1; Y|X_2, U, Q), \tag{6.8}$$

$$\widetilde{R}_2 < I(X_2; Y|X_1, U, Q), \tag{6.9}$$

$$\widetilde{R}_1 + \widetilde{R}_2 < I(X_1, X_2; Y|U, Q), \tag{6.10}$$

$$\mathbb{E}[d_i(U, X_i)] \leq D_i,\ i = 1, 2, \tag{6.11}$$

where the above mutual informations and expectations are taken with respect to the joint distribution $P_Q P_U P_{X_1|UQ} P_{X_2|UQ} W_{Y|X_1X_2}$. If we further set $D_1 \geq d_1^{max}$ and $D_2 \geq d_2^{max}$ and let U be deterministic, inequalities (6.8)–(6.11) give the capacity region of the MAC [10].

Parallel Attack Channels

Assume that the attack MAC is composed of two independent discrete memoryless channels $W_{Y|X_1X_2}(y|x_1, x_2) = W_{Y_1|X_1}(y_1|x_1) \times W_{Y_2|X_2}(y_2|x_2)$ where $W_{Y_i|X_i}$ has input alphabet $\mathcal{X}_i$ and output alphabet $\mathcal{Y}_i$ such that $\mathcal{Y}_1 \times \mathcal{Y}_2 = \mathcal{Y}$, $i = 1, 2$. This can be interpreted as two attackers separately attacking the stegotexts. In this case, the condition given by Theorem 6.1 for successful embedding is equivalent to the following (see Section 6.3.3 for the proof): $\{(S_{1j}, S_{2j})\}$ are (D_1, D_2)-admissible with respect to the MAC $W_{Y|X_1X_2}$ if

$$H(S_1|S_2) < C(W^{(1)}, D_1), \tag{6.12}$$

$$H(S_2|S_1) < C(W^{(2)}, D_2), \tag{6.13}$$

$$H(S_1, S_2) < C(W^{(1)}, D_1) + C(W^{(2)}, D_2), \tag{6.14}$$

where $C(W^{(i)}, D_i) = \max_{P_{X_i|U}:\, \mathbb{E}[d_i(U,X_i)] \leq D_i} I(X_i; Y_i|U)$, $i = 1, 2$, is the information-hiding capacity of the attack channel $W_{Y_i|X_i}$ with distortion threshold D_i [51].

Attack-Free Channel

Let $l : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathcal{Y}$ be a bijection and let $Y = l(X_1, X_2)$. In this case, Theorem 6.1 implies that $\{(S_{1j}, S_{2j})\}$ are (D_1, D_2)-admissible with respect to the MAC $W_{Y|X_1X_2}$ if

$$H(S_1|S_2) < H(X_1|X_2, S_2, U, Q), \tag{6.15}$$

$$H(S_2|S_1) < H(X_2|X_1, S_1, U, Q), \tag{6.16}$$

$$H(S_1, S_2) < H(X_1, X_2|U, Q), \tag{6.17}$$

$$\mathbb{E}[d_i(U, X_i)] \leq D_i, \ i = 1, 2, \tag{6.18}$$

where the entropies are taken under the joint distribution $P_{S_1 S_2} P_U P_{X_1|S_1 U} P_{X_2|S_2 U}$. Note also that conditions (6.15)–(6.18) give the Slepian-Wolf lossless data compression region [10], [71] if we set $D_1 \geq d_1^{max}$, $D_2 \geq d_2^{max}$, and let U be deterministic.

6.3 Proofs

6.3.1 Proof of Theorem 6.1

We first give an outline of the proof. We need to show that for given $P_{S_1 S_2}$, P_U, and $W_{Y|X_1 X_2}$, there exists a sequence of JSCE codes $(\varphi_1^{(n)}, \varphi_2^{(n)}, \psi^{(n)})$ such that $P_e^{(n)} \to 0$ as $n \to \infty$ and for any $\delta > 0$, $\frac{1}{n}\mathbb{E}\big[d_i(U^n, \varphi_i^{(n)}(S_i^n, U^n))\big] \leq D_i + \delta$, $i = 1, 2$, for n sufficiently large. Fix $(P_Q, P_{X_1|S_1 UQ}, P_{X_2|S_2 UQ})$ such that the following are satisfied for some $\epsilon > 0$,

$$H(S_1|S_2) < I(X_1; Y|X_2, S_2, U, Q) - 7\epsilon, \tag{6.19}$$

$$H(S_2|S_1) < I(X_2; Y|X_1, S_1, U, Q) - 7\epsilon, \tag{6.20}$$

$$H(S_1, S_2) < I(X_1, X_2; Y|U, Q) - 7\epsilon, \tag{6.21}$$

$$\mathbb{E}[d_i(U, X_i)] \leq D_i, \ i = 1, 2. \tag{6.22}$$

Define $P_i^{(n)} \triangleq \Pr\{\frac{1}{n} d_i(U^n, \varphi_i^{(n)}(S_i^n, U^n)) > D_i + \epsilon d_i^{max}\}$, $i = 1, 2$. We will prove that for any $\epsilon_1 > 0$, the following probabilities, which are averaged over a family of random codes $(\mathcal{C}_1, \mathcal{C}_2)$, $i = 1, 2$, satisfy

$$\mathbb{E}_{\mathcal{C}_1, \mathcal{C}_2}[P_e^{(n)}] \leq \epsilon_1, \quad \mathbb{E}_{\mathcal{C}_1, \mathcal{C}_2}[P_1^{(n)}] \leq \epsilon_1, \quad \mathbb{E}_{\mathcal{C}_1, \mathcal{C}_2}[P_2^{(n)}] \leq \epsilon_1$$

for n sufficiently large. Then $\mathbb{E}_{\mathcal{C}_1,\mathcal{C}_2}\{P_e^{(n)} + P_1^{(n)} + P_2^{(n)}\} \leq 3\epsilon_1$, which guarantees that there exists at least one pair $(\mathcal{C}_1, \mathcal{C}_2)$ such that $P_e^{(n)} + P_1^{(n)} + P_2^{(n)} \leq 3\epsilon_1$ and hence $P_e^{(n)} \leq 3\epsilon_1$, $P_1^{(n)} \leq 3\epsilon_1$, $P_2^{(n)} \leq 3\epsilon_1$ are simultaneously satisfied for n sufficiently large. Finally, it can be easily shown that $P_i^{(n)} \leq 3\epsilon_1$ implies for n sufficiently large that

$$\frac{1}{n}\mathbb{E}\big[d_i(U^n, \varphi_i^{(n)}(S_i^n, U^n))\big] \leq D_i + \epsilon d_i^{max} + P_i^{(n)} d_i^{max} \leq D_i + \delta.$$

Random Code Design

Random Code Generation. Let $i \in \{1, 2\}$. Choose a typical sequence $\mathbf{q} = (q_1, q_2, ..., q_n)$ arbitrarily in $\mathcal{T}_\epsilon^{(n)}(Q)$. The sequence serves as a time sharing sequence and it is known at both the encoders and the decoder. For any sequences $\mathbf{s}_i, \mathbf{u}$ and the fixed $\mathbf{q}$, generate one $\mathbf{x}_i(\mathbf{s}_i, \mathbf{u}, \mathbf{q})$ sequence according to $\prod_{j=1}^{n} P_{X_i|S_iUQ}(x_{ij}|s_{ij}, u_j, q_j)$. Define codebook $\mathcal{C}_i$ as $\mathcal{C}_i \triangleq \{\mathbf{x}_i(\mathbf{s}_i, \mathbf{u}, \mathbf{q}) : (\mathbf{s}_i, \mathbf{u}) \in \mathcal{S}_i^n \times \mathcal{U}^n\}$. Reveal the codebooks to both the encoders and the decoder.

Encoding. Given $(\mathbf{s}_i, \mathbf{u}) \in \mathcal{S}_i^n \times \mathcal{U}^n$, Encoder i sends $\mathbf{x}_i(\mathbf{s}_i, \mathbf{u}, \mathbf{q})$.

Decoding. The decoder has full knowledge of $\mathbf{u}$ (and also the time sharing sequence $\mathbf{q}$). Upon receiving sequence $\mathbf{y}$, the decoder finds the only pair $(\widehat{\mathbf{s}}_1, \widehat{\mathbf{s}}_2) \in \mathcal{T}_\epsilon^{(n)}(S_1, S_2)$, such that $\mathbf{y} \in \mathcal{T}_\epsilon^{(n)}(Y|\widehat{\mathbf{s}}_1, \widehat{\mathbf{s}}_2, \mathbf{u}, \mathbf{q}, \widehat{\mathbf{x}}_1, \widehat{\mathbf{x}}_2)$, where $\widehat{\mathbf{x}}_1 = \mathbf{x}_1(\widehat{\mathbf{s}}_1, \mathbf{u}, \mathbf{q})$ and $\widehat{\mathbf{x}}_2 = \mathbf{x}_2(\widehat{\mathbf{s}}_2, \mathbf{u}, \mathbf{q})$. If there is no or more than one such pair of sequences $(\widehat{\mathbf{s}}_1, \widehat{\mathbf{s}}_2)$, an error is declared.

For the sake of convenience, define the events

$$\begin{aligned} A_0 &: (\mathbf{s}_1, \mathbf{s}_2, \mathbf{u}) \in \mathcal{T}_\epsilon^{(n)}(S_1, S_2, U|\mathbf{q}) \\ A_1 &: (\mathbf{s}_1, \mathbf{s}_2, \mathbf{u}, X_1^n(\mathbf{s}_1, \mathbf{u}, \mathbf{q}), X_2^n(\mathbf{s}_2, \mathbf{u}, \mathbf{q})) \in \mathcal{T}_\epsilon^{(n)}(\cdot|\mathbf{q}). \end{aligned}$$

The following result is a consequence of the Markov lemma (Lemma 5.2).

Lemma 6.1 For any $\epsilon, \epsilon_2 \in (0, 1)$, $\mathbb{E}_{\mathcal{C}_1,\mathcal{C}_2}\left[\Pr(A_1^c|A_0)\right] \leq \epsilon_2$ for n sufficiently large, where the expectation is taken with respect to the random codes $\mathcal{C}_1$ and $\mathcal{C}_2$.

Bounding $\mathbb{E}_{\mathcal{C}_1,\mathcal{C}_2}[P_e^{(n)}]$

$$\begin{aligned}
&\mathbb{E}_{\mathcal{C}_1,\mathcal{C}_2}[P_e^{(n)}]\\
&\quad\leq \sum_{(\mathcal{T}_\epsilon^{(n)}(S_1,S_2,U|\mathbf{q}))^c} P_{S_1S_2}^{(n)}(\mathbf{s}_1,\mathbf{s}_2)P_U^{(n)}(\mathbf{u})\\
&\qquad+\sum_{\mathcal{T}_\epsilon^{(n)}(S_1,S_2,U|\mathbf{q})} P_{S_1S_2}^{(n)}(\mathbf{s}_1,\mathbf{s}_2)P_U^{(n)}(\mathbf{u})\mathbb{E}_{\mathcal{C}_1,\mathcal{C}_2}\left[\sum_{\mathbf{y}:\psi^{(n)}(\mathbf{y},\mathbf{u})\neq(\mathbf{s}_1,\mathbf{s}_2)} W_{Y|X_1X_2}^{(n)}(\mathbf{y}|\mathbf{x}_1,\mathbf{x}_2)\right].
\end{aligned}$$

The first term vanishes for n sufficiently large by Lemma 5.1. It suffices to bound the expectation in the second term. Given $(\mathbf{s}_1,\mathbf{s}_2,\mathbf{u})\in\mathcal{T}_\epsilon^{(n)}(S_1,S_2,U|\mathbf{q})$, we have the following four error events:

$$\begin{aligned}
E_0 &: (\mathbf{s}_1,\mathbf{s}_2,\mathbf{u},X_1^n(\mathbf{s}_1,\mathbf{u},\mathbf{q}),X_2^n(\mathbf{s}_2,\mathbf{u},\mathbf{q}),Y^n)\notin\mathcal{T}_\epsilon^{(n)}(\cdot|\mathbf{q}),\\
E_1 &: \exists\ \ \widehat{\mathbf{s}}_1\neq\mathbf{s}_1 \text{ such that}\\
&\qquad(\widehat{\mathbf{s}}_1,\mathbf{s}_2,\mathbf{u},X_1^n(\widehat{\mathbf{s}}_1,\mathbf{u},\mathbf{q}),X_2^n(\mathbf{s}_2,\mathbf{u},\mathbf{q}),Y^n)\in\mathcal{T}_\epsilon^{(n)}(\cdot|\mathbf{q}),\\
E_2 &: \exists\ \widehat{\mathbf{s}}_2\neq\mathbf{s}_2 \text{ such that}\\
&\qquad(\mathbf{s}_1,\widehat{\mathbf{s}}_2,\mathbf{u},X_1^n(\mathbf{s}_1,\mathbf{u},\mathbf{q}),X_2^n(\widehat{\mathbf{s}}_2,\mathbf{u},\mathbf{q}),Y^n)\in\mathcal{T}_\epsilon^{(n)}(\cdot|\mathbf{q}),\\
E_3 &: \exists\ \widetilde{\mathbf{s}}_1\neq\mathbf{s}_1,\widetilde{\mathbf{s}}_2\neq\mathbf{s}_2 \text{ such that}\\
&\qquad(\widetilde{\mathbf{s}}_1,\widetilde{\mathbf{s}}_2,\mathbf{u},X_1^n(\widetilde{\mathbf{s}}_1,\mathbf{u},\mathbf{q}),\mathbf{u}),X_2^n(\widetilde{\mathbf{s}}_2,\mathbf{u},\mathbf{q}),Y^n)\in\mathcal{T}_\epsilon^{(n)}(\cdot|\mathbf{q}).
\end{aligned}$$

It then immediately follows from the union bound that

$$\mathbb{E}_{\mathcal{C}_1,\mathcal{C}_2}\left[\sum_{\mathbf{y}:\psi^{(n)}(\mathbf{y},\mathbf{u})\neq(\mathbf{s}_1,\mathbf{s}_2)} W_{Y|X_1X_2}^{(n)}(\mathbf{y}|\mathbf{x}_1,\mathbf{x}_2)\right] \leq \sum_{j=0}^{3}\mathbb{E}_{\mathcal{C}_1,\mathcal{C}_2}\Big[\Pr\{E_j|A_0\}\Big]. \tag{6.23}$$

To bound $\mathbb{E}_{\mathcal{C}_1,\mathcal{C}_2}\Big[\Pr\{E_0|A_0\}\Big]$, it follows from Lemma 5.2 that

$$\begin{aligned}
\mathbb{E}_{\mathcal{C}_1,\mathcal{C}_2}\Big[\Pr\{E_0|A_0\}\Big] &\leq \mathbb{E}_{\mathcal{C}_1,\mathcal{C}_2}\Big[\Pr(A_1^c|A_0)\Big]+\mathbb{E}_{\mathcal{C}_1,\mathcal{C}_2}\Big[\Pr\{E_0|A_0,A_1\}\Big]\\
&\leq \frac{\epsilon_0}{2}+\frac{\epsilon_0}{2}=\epsilon_0
\end{aligned} \tag{6.24}$$

if n sufficiently large, where ϵ_0 will be specified later.

To bound $\mathbb{E}_{\mathcal{C}_1,\mathcal{C}_2}\Big[\Pr\left\{ E_1 \middle| A_0\right\}\Big]$, write

$$\mathbb{E}_{\mathcal{C}_1,\mathcal{C}_2}\Big[\Pr\{E_1|A_0\}\Big] \leq \sum_{\widehat{\mathbf{s}}_1\neq\mathbf{s}_1:\widehat{\mathbf{s}}_1\in\mathcal{T}_\epsilon^{(n)}(S_1|\mathbf{s}_2)} \mathbb{E}_{\mathcal{C}_1,\mathcal{C}_2}\left[\Pr\left\{ \mathbf{v}_1\in\mathcal{T}_\epsilon^{(n)} \middle| A_0\right\}\right] \tag{6.25}$$

where $\mathbf{v}_1 = (\widehat{\mathbf{s}}_1, \mathbf{s}_2, \mathbf{u}, \mathbf{q}, X_1^n(\widehat{\mathbf{s}}_1, \mathbf{u}, \mathbf{q}), X_2^n(\mathbf{s}_2, \mathbf{u}, \mathbf{q}), Y^n)$ and the expectation can be upper bounded by

$$\begin{aligned}
&\mathbb{E}_{\mathcal{C}_1,\mathcal{C}_2}\left[\Pr\left\{ \mathbf{v}_1\in\mathcal{T}_\epsilon^{(n)} \middle| A_0\right\}\right] \\
\leq\ & \sum_{\mathcal{X}_1^n\times\mathcal{X}_2^n} P^{(n)}_{X_1|S_1UQ}(\widehat{\mathbf{x}}_1|\widehat{\mathbf{s}}_1,\mathbf{u},\mathbf{q}) P^{(n)}_{X_2|S_2UQ}(\mathbf{x}_2|\mathbf{s}_2,\mathbf{u},\mathbf{q}) \\
& \sum_{\mathbf{y}\in\mathcal{T}_\epsilon^{(n)}(Y|\widehat{\mathbf{s}}_1,\mathbf{s}_2,\mathbf{u},\mathbf{q},\widehat{\mathbf{x}}_1,\mathbf{x}_2)} P^{(n)}_{Y|S_2UQX_2}(\mathbf{y}|\mathbf{s}_2,\mathbf{u},\mathbf{q},\mathbf{x}_2) \\
\leq\ & \left|\mathcal{T}_\epsilon^{(n)}(Y|\widehat{\mathbf{s}}_1,\mathbf{s}_2,\mathbf{u},\mathbf{q},\widehat{\mathbf{x}}_1,\mathbf{x}_2)\right| 2^{-n(H(Y|S_2,U,Q,X_2)-2\epsilon)} && (6.26) \\
\leq\ & 2^{n(H(Y|X_1,X_2)+2\epsilon)} 2^{-n(H(Y|S_2,U,Q,X_2)-2\epsilon)} && (6.27) \\
=\ & 2^{n(H(Y|X_1,X_2,S_2,U,Q)+2\epsilon)} 2^{-n(H(Y|S_2,U,Q,X_2)-2\epsilon)} \\
=\ & 2^{-n(I(X_1;Y|X_2,S_2,U,Q)-4\epsilon)}, && (6.28)
\end{aligned}$$

where $\widehat{\mathbf{x}}_1 = \mathbf{x}_1(\widehat{\mathbf{s}}_1,\mathbf{u},\mathbf{q})$, $\mathbf{x}_2 = \mathbf{x}_2(\mathbf{s}_2,\mathbf{u},\mathbf{q})$, and (6.26) and (6.27) follow from Lemma 5.1. It then follows from (6.25), Lemma 5.1 and (6.19) that

$$\begin{aligned}
\mathbb{E}_{\mathcal{C}_1,\mathcal{C}_2}\Big[\Pr\left\{ E_1 \middle| A_0\right\}\Big] &\leq \left|\mathcal{T}_\epsilon^{(n)}(S_1|\mathbf{s}_2)\right| 2^{-n(I(X_1;Y|X_2,S_2,U,Q)-4\epsilon)} \\
&\leq 2^{n(H(S_1|S_2)+2\epsilon)} 2^{-n(I(X_1;Y|X_2,S_2,U,Q)-4\epsilon)} \\
&= 2^{-n(I(X_1;Y|X_2,S_2,U,Q)-H(S_1|S_2)-6\epsilon)} \\
&\leq 2^{-n\epsilon} \leq \epsilon_0, && (6.29)
\end{aligned}$$

for n sufficiently large. Similarly, we can bound using (6.20)

$$\mathbb{E}_{\mathcal{C}_1,\mathcal{C}_2}\Big[\Pr\left\{ E_2 \middle| A_0\right\}\Big] \leq \epsilon_0, \tag{6.30}$$

for n sufficiently large.

It remains to bound $\mathbb{E}_{\mathcal{C}_1,\mathcal{C}_2}\Big[\Pr\left\{E_3 | A_0\right\}\Big]$. Write

$$\mathbb{E}_{\mathcal{C}_1,\mathcal{C}_2}\Big[\Pr\{E_3|A_0\}\Big] \leq \sum_{\widetilde{\mathbf{s}}_1\neq\mathbf{s}_1,\widetilde{\mathbf{s}}_2\neq\mathbf{s}_2:(\widetilde{\mathbf{s}}_1,\widetilde{\mathbf{s}}_2)\in\mathcal{T}_\epsilon^{(n)}(S_1,S_2)} \mathbb{E}_{\mathcal{C}_1,\mathcal{C}_2}\left[\Pr\left\{\mathbf{v}_2\in\mathcal{T}_\epsilon^{(n)}\middle| A_0\right\}\right], \quad (6.31)$$

where $\mathbf{v}_2 = (\widetilde{\mathbf{s}}_1, \widetilde{\mathbf{s}}_2, \mathbf{u}, \mathbf{q}, X_1^n(\widetilde{\mathbf{s}}_1, \mathbf{u}, \mathbf{q}), X_2^n(\widetilde{\mathbf{s}}_2, \mathbf{u}, \mathbf{q}), Y^n)$ and

$$\begin{aligned}
&\mathbb{E}_{\mathcal{C}_1,\mathcal{C}_2}\left[\Pr\left\{\mathbf{v}_2\in\mathcal{T}_\epsilon^{(n)}\middle| A_0\right\}\right]\\
\leq\ & \sum_{\mathcal{X}_1^n\times\mathcal{X}_2^n} P^{(n)}_{X_1|S_1UQ}(\widetilde{\mathbf{x}}_1|\widetilde{\mathbf{s}}_1,\mathbf{u},\mathbf{q})P^{(n)}_{X_2|S_2UQ}(\widetilde{\mathbf{x}}_2|\widetilde{\mathbf{s}}_2,\mathbf{u},\mathbf{q})\\
&\qquad \sum_{\mathbf{y}\in\mathcal{T}_\epsilon^{(n)}(Y|\widetilde{\mathbf{s}}_1,\widetilde{\mathbf{s}}_2,\mathbf{u},\mathbf{q},\widetilde{\mathbf{x}}_1,\widetilde{\mathbf{x}}_2)} P^{(n)}_{Y|UQ}(\mathbf{y}|\mathbf{u},\mathbf{q})\\
\leq\ & \left|\mathcal{T}_\epsilon^{(n)}(Y|\widetilde{\mathbf{s}}_1,\widetilde{\mathbf{s}}_2,\mathbf{u},\mathbf{q},\widetilde{\mathbf{x}}_1,\widetilde{\mathbf{x}}_2)\right| 2^{-n(H(Y|U,Q)-2\epsilon)} &(6.32)\\
\leq\ & 2^{n(H(Y|U,Q,X_1,X_2)+2\epsilon)}2^{-n(H(Y|U,Q)-2\epsilon)} &(6.33)\\
=\ & 2^{-n(I(X_1,X_2;Y|U,Q)-4\epsilon)}
\end{aligned}$$

where $\widetilde{\mathbf{x}}_1 = \mathbf{x}_1(\widetilde{\mathbf{s}}_1, \mathbf{u}, \mathbf{q})$ and $\widetilde{\mathbf{x}}_2 = \mathbf{x}_2(\widetilde{\mathbf{s}}_2, \mathbf{u}, \mathbf{q})$, and (6.32) and (6.33) follow Lemma 5.1. It then follows that,

$$\begin{aligned}
\mathbb{E}_{\mathcal{C}_1,\mathcal{C}_2}\Big[\Pr\{E_3|A_0\}\Big] &\leq \sum_{\widetilde{\mathbf{s}}_1\neq\mathbf{s}_1,\widetilde{\mathbf{s}}_2\neq\mathbf{s}_2:(\widetilde{\mathbf{s}}_1,\widetilde{\mathbf{s}}_2)\in\mathcal{T}_\epsilon^{(n)}(S_1,S_2)} 2^{-n(I(X_1,X_2;Y|U)-4\epsilon)}\\
&\leq \left|\mathcal{T}_\epsilon^{(n)}(S_1,S_2)\right| 2^{-n(I(X_1,X_2;Y|U)-4\epsilon)}\\
&\leq 2^{n(H(S_1,S_2)+2\epsilon)}2^{-n(I(X_1;Y|X_2,S_2,U)-4\epsilon)}\\
&= 2^{-n(I(X_1;Y|X_2,S_2,U)-H(S_1,S_2)-6\epsilon)}\\
&\leq 2^{-n\epsilon} \leq \epsilon_0 &(6.34)
\end{aligned}$$

for n sufficiently large. Now plugging (6.24), (6.29), (6.30), and (6.34) back into (6.23), and setting $\epsilon_0 = \frac{\epsilon_1}{4}$, we see that $\mathbb{E}_{\mathcal{C}_1,\mathcal{C}_2}[P_e^{(n)}] \leq \epsilon_1$ for n sufficiently large.

Bounding $\mathbb{E}_{\mathcal{C}_1,\mathcal{C}_2}[P_i^{(n)}]$

Since the encoding is separately performed, Encoder 1 is independent of $\mathcal{C}_2$. Thus it suffices to show that $\mathbb{E}_{\mathcal{C}_1}[P_1^{(n)}] \le \epsilon_1$ for n sufficiently large.

Clearly, if $(\mathbf{s}_1, \mathbf{u}, \mathbf{x}_1) \in \mathcal{T}_\epsilon^{(n)}(S_1, U, X_1|\mathbf{q})$, then

$$\frac{1}{n} d_1(\mathbf{u}, \mathbf{x}_1(\mathbf{s}_1, \mathbf{u}, \mathbf{q})) \le \mathbb{E}[d_1(U, X_1)] + \epsilon d_1^{max} \le D_1 + \epsilon d_1^{max}$$

for n sufficiently large, where the first inequality follows from the definition of strong typicality and the second inequality follows from (6.22). According to Lemma 5.1,

$$\begin{aligned} \mathbb{E}_{\mathcal{C}_1}[P_1^{(n)}] \ \le\ & \sum_{(\mathcal{T}_\epsilon^{(n)}(S_1,U|\mathbf{q}))^c} P_{S_1U}^{(n)}(\mathbf{s}_1, \mathbf{u}) \\ & + \sum_{\mathcal{T}_\epsilon^{(n)}(S_1,U|\mathbf{q})} P_{S_1U}^{(n)}(\mathbf{s}_1, \mathbf{u}) \mathbb{E}_{\mathcal{C}_1}\big[\mathbb{1}\big\{\mathbf{v}_3 \notin \mathcal{T}_\epsilon^{(n)}(S_1, U, Q, X_1)\big\}\big] \\ \le\ & \frac{\epsilon_1}{2} + \frac{\epsilon_1}{2} = \epsilon_1 \end{aligned} \tag{6.35}$$

for n sufficiently large, where $\mathbf{v}_3 = (\mathbf{s}_1, \mathbf{u}, \mathbf{q}, X_1^n(\mathbf{s}_1, \mathbf{u}, \mathbf{q}))$.

Completing the Proof

As we mentioned in the beginning of the section,

$$\mathbb{E}_{\mathcal{C}_1,\mathcal{C}_2}\{P_e^{(n)} + P_1^{(n)} + P_2^{(n)}\} \le 3\epsilon_1,$$

implies that there exists a pair of codes $(\mathcal{C}_1, \mathcal{C}_2)$ such that $P_e^{(n)} \le 3\epsilon_1$, $P_1^{(n)} \le 3\epsilon_1$, $P_2^{(n)} \le 3\epsilon_1$ are simultaneously satisfied for n sufficiently large. Furthermore, if $P_i^{(n)} \le 3\epsilon_1$, we have

$$\frac{1}{n}\mathbb{E}\big[d_i(U^n, \varphi_i^{(n)}(S_i^n, U^n))\big] \le D_i + \epsilon d_i^{max} + P_i^{(n)} d_i^{max} \le D_i + \delta_i,$$

as $n \to \infty$, by setting $\delta_i = \epsilon + 3\epsilon_1 d_i^{max}$. Thus the distortion constraint is satisfied. This completes the proof of Theorem 6.1.

6.3.2 Proof of Theorem 6.2

Suppose $(H(S_1|S_2), H(S_2|S_1), H(S_1, S_2)) \in \mathcal{R}_n$. Then we can replace the channel by its nth extension. Thus, the achievability follows directly from Theorem 1. We now prove the converse part. We wish to show that for any sequence of encoder-decoder triplets $(\varphi_1^{(n)}, \varphi_2^{(n)}, \psi^{(n)})$ that achieves $P_e^{(n)} < \epsilon$ and $\frac{1}{n}\mathbb{E}[d(U^n, \varphi_i^{(n)}(S_i^n, U^n))] \leq D + \delta$ for $i = 1, 2$ for n sufficiently large, (6.6) holds. By Fano's inequality, we have

$$H(S_1^n, S_2^n|U^n, Y^n) \leq P_e^{(n)} \log |(S_1^n, S_2^n)| + H(P_e^{(n)}) \triangleq n\epsilon_n. \tag{6.36}$$

Since

$$H(S_1^n|S_2^n, U^n, Y^n, X_2^n) \leq H(S_1^n|S_2^n, U^n, Y^n) \leq H(S_1^n, S_2^n|U^n, Y^n) \tag{6.37}$$

by data processing inequality, we also have

$$H(S_1^n|S_2^n, U^n, Y^n, X_2^n) \leq n\epsilon_n, \tag{6.38}$$

$$H(S_2^n|S_1^n, U^n, Y^n, X_1^n) \leq n\epsilon_n. \tag{6.39}$$

We now bound $H(S_1|S_2)$ as

$$\begin{aligned}
& nH(S_1|S_2) \\
&= H(S_1^n|S_2^n) \\
&= H(S_1^n|S_2^n, U^n) && (6.40) \\
&= H(S_1^n|S_2^n, U^n, X_2^n) + I(S_1^n; X_2^n|S_2^n, U^n) && (6.41) \\
&= I(S_1^n; Y^n|S_2^n, U^n, X_2^n) + H(S_1^n|S_2^n, U^n, X_2^n, Y^n) + I(S_1^n; X_2^n|S_2^n, U^n) && (6.42) \\
&\leq I(X_1^n; Y^n|S_2^n, U^n, X_2^n) + H(S_1^n|S_2^n, U^n, X_2^n, Y^n) && (6.43) \\
&\leq I(X_1^n; Y^n|S_2^n, U^n, X_2^n) + n\epsilon_n && (6.44)
\end{aligned}$$

where (6.43) follows from the data processing inequality and the fact that S_1^n and X_2^n are conditional independent given U^n and S_2^n, and (6.44) follows from (6.38). Therefore, we obtain that

$$H(S_1|S_2) \leq \frac{1}{n} I(X_1^n; Y^n|S_2^n, U^n, X_2^n) + \epsilon_n. \tag{6.45}$$

Similarly, we have

$$H(S_2|S_1) \leq \frac{1}{n} I(X_2^n; Y^n|S_1^n, U^n, X_1^n) + \epsilon_n. \tag{6.46}$$

To bound $H(S_1, S_2)$, we have

$$\begin{aligned} nH(S_1, S_2) &= H(S_1^n, S_2^n) \\ &= H(S_1^n, S_2^n|U^n) && (6.47) \\ &= I(S_1^n, S_2^n; Y^n|U^n) + H(S_1^n, S_2^n|U^n, Y^n) && (6.48) \\ &\leq I(X_1^n, X_2^n; Y^n|U^n) + n\epsilon_n && (6.49) \end{aligned}$$

where the last inequality follows from (6.36). Hence, we have

$$H(S_1, S_2) \leq \frac{1}{n} I(X_1^n, X_2^n; Y^n|U^n) + \epsilon_n. \tag{6.50}$$

Now if $P_e^{(n)} \to 0$, we have $\epsilon_n \to 0$ and $\delta \to 0$ as n$\to \infty$. It follows from (6.45), (6.46), and (6.50) that

$$(H(S_1|S_2), H(S_2|S_1), H(S_1, S_2)) \in \bigcup_{n=1}^{\infty} \mathcal{R}_n. \tag{6.51}$$

which proves the converse part.

6.3.3 Proof of the Case of Parallel Attack Channels

When $W_{Y|X_1X_2} = W_{Y_1|X_1} \times W_{Y_2|X_2}$, we see that (6.12)–(6.14) imply (6.1)–(6.4). In fact, if the maximums in (6.12)–(6.14) are achieved by $P^*_{X_1|U}(x_1|u)$ and $P^*_{X_2|U}(x_2|u)$, then simply letting $|\mathcal{Q}| = 1$, $P_{X_1|S_1U}(x_1|s_1,u) = P^*_{X_1|U}(x_1|u)$ and $P_{X_2|S_2U}(x_2|s_2,u) = P^*_{X_2|U}(x_2|u)$, we see that with this choice,

$$\begin{aligned} I(X_1;Y|X_2,S_2,U,Q) &= I(X_1;Y_1|S_2,U,Q) \\ &= I(X_1;Y_1|U) \\ &= \max_{P_{X_1|U}:\mathbb{E}[d_1(U,X_1)]\le D_1} I(X_1;Y_1|U). \end{aligned}$$

Similarly,

$$I(X_2;Y|X_1,S_1,U,Q) = \max_{P_{X_2|U}:\mathbb{E}[d_2(U,X_2)]\le D_2} I(X_2;Y_2|U),$$

and

$$I(X_1,X_2;Y_1,Y_2|U,Q) \;=\; \max_{P_{X_1|U}:\mathbb{E}[d_1(U,X_1)]\le D_1} I(X_1;Y_1|U) + \max_{P_{X_2|U}:\mathbb{E}[d_2(U,X_2)]\le D_2} I(X_2;Y_2|U).$$

We next show that (6.1)–(6.4) imply (6.12)–(6.14). We only need to show that for any $P_{X_1|S_1UQ}$ satisfying $\mathbb{E}[d_1(U,X_1)] < D_1$, the right hand side of (6.1) is upper bounded by (6.12). Since $(Q,S_1,U) \to X_1 \to Y_1$ form a Markov chain in this order,

$$\begin{aligned} I(X_1;Y_1|U) &= H(Y_1|U) - H(Y_1|X_1,U) \\ &\ge H(Y_1|S_2,U,Q) - H(Y_1|X_1,S_2,U,Q) \\ &= I(X_1;Y_1|S_2,U,Q). \end{aligned}$$

For any $P_{X_1|S_1UQ}$ satisfying $\mathbb{E}[d_1(U,X_1)] < D_1$, set

$$\widehat{P}_{X_1|U}(x_1|u) = \sum_{\mathcal{S}_1\times\mathcal{Q}} P_{S_1}(s_1)P_Q(q)P_{X_1|S_1UQ}(x_1|s_1,u,q).$$

Under the corresponding $\widehat{P}_{X_1|U}(x_1|u)$, we have

$$\begin{aligned} I(X_1;Y_1|U,S_2,Q) &\le I(X_1;Y_1|U) \\ &\le \max_{P_{X_1|U}:\mathbb{E}[d_1(U,X_1)]\le D_1} I(X_1;Y_1|U). \end{aligned}$$

We can similarly show that (6.2)–(6.3) imply (6.13)–(6.14). □

6.4 Conclusions

We presented a multi-user information hiding model for the transmission of two correlated secret sources over memoryless multiple attack access channel with common host data. The achievable rate region is studied for the case where lossless compression of stegotexts is jointly performed with information hiding. Based on our definition for reliable transmission (admissibility), we derived a sufficient condition with single-letter characterizations for hiding correlated sources against MAC attacks. An uncomputable (and somewhat trivial) outer bound (converse condition) is formulated by applying Fano's inequality in terms of a sequence of n-dimensional joint distributions. In the future, we are interested to study the embedding of correlated sources with joint embedding/compression rate constraints. Our next step is to answer the question: when (S_1, S_2) are (D_1, D_2)-admissible with respect to $W_{Y|X_1X_2}$, what is the compression limit for the sources (S_1, S_2) and U?

Chapter 7

Capacity Region for Multi-User Public Information Hiding Under Multiple Access Attacks

7.1 Introduction

In [88], an information embedding model was considered for hiding secret information in a discrete memoryless source (DMS) which is then transmitted through a discrete memoryless channel (DMC); see Fig. 7.1. The secret message W, independently and uniformly drawn from a message set of 2^{nR} elements, is embedded into n-length sequences U^n (referred to as host messages) generated by the DMS $\{U\}$. Since the secret messages should not interfere perceptually with the host messages, a distortion constraint is placed between the encoder output (referred to as stegotext) X^n and the original host message U^n. This secret information is to be recovered from a noisy version of the sequence in which the information is embedded. The noise is used to model the effects of standard data-processing or a malicious attack. The embedding model is motivated by practical information hiding problems such as watermarking or copyright protection, where a copyright/watermark is embedded into the original multimedia data in order to preserve the ownership of intellectual property (see, e.g., [35], [77] and the references therein).

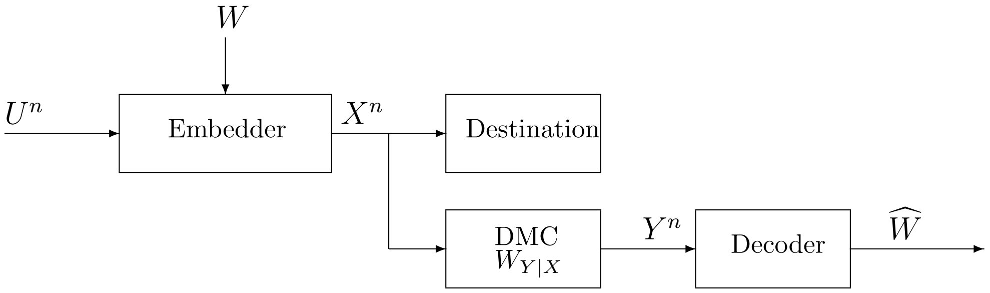


Figure 7.1: A single-user information embedding system.

From an information-theoretic point of view, it is of interest to find the largest achievable embedding rate R (known as the embedding capacity) for which, at the encoder, the distortion between the host message U^n and the stegotext X^n does not exceed a preset threshold, and at the decoder, the secret messages can be reproduced with an arbitrarily small probability of error. It is shown in [88] that for a host source with distribution Q_U, "attack" channel $W_{Y|X}$, and an (average) distortion threshold D, the embedding capacity is given by

$$\max_{P_{TX|U} \in \mathcal{P}_D} [I(T;Y) - I(U;T)]$$

where $\mathcal{P}_D$ is the set of all the conditional distributions $P_{TX|U}$ such that the average distortion between X and U under the distribution $Q_U P_{TX|U}$ is less than D for some auxiliary random variable (RV) T.

In this work we extend the information embedding model for a multi-user setting depicted in Fig. 7.2. Assume that two users separately embed their secret information into two correlated DMSs, U_1 and U_2. Each user can only access one of the two host sources. The stegotexts are sent through a multiple access channel (MAC) to a decoder which tries to reconstruct the secret information. For the two-user information embedding system, we are interested in determining the embedding capacity region, i.e., the two-dimensional set of all the achievable embedding rate pairs. An inner bound for the embedding capacity region is obtained based on a ϵ-strong typicality coding/decoding argument. More specifically, we first map the watermarks w_1 and w_2 as well as the correlated source messages $\mathbf{u}_1$ and $\mathbf{u}_2$ through separate encoders to auxiliary codewords $\mathbf{t}_1$ and $\mathbf{t}_2$, and then we generate two stegotexts $\mathbf{x}_1$ and $\mathbf{x}_2$ which are jointly typical with respect to $(\mathbf{u}_1, \mathbf{u}_2, \mathbf{t}_1, \mathbf{t}_2)$. In the decoding stage, we recover the watermarks by examining the joint typicality of the received sequence $\mathbf{y}$ and all auxiliary codeword pairs $(\mathbf{t}_1, \mathbf{t}_2)$. We employ a generalized rate-distortion encoding scheme to ensure that $(\mathbf{u}_1, \mathbf{u}_2, \mathbf{t}_1, \mathbf{t}_2)$ are jointly typical with high probability. The generalized rate-distortion encoding scheme, introduced in [54] for Gaussian multi-terminal source coding (see also [79], [26]), can be briefly described as follows. One of the encoders, say Encoder 1, generates an auxiliary $\mathbf{t}_1$ such that conditioning on $(\mathbf{u}_1, \mathbf{t}_1)$, $(\mathbf{u}_1, \mathbf{t}_1, U_2^n, T_2^n)$ is ϵ-strongly typical with high probability. The other encoder, Encoder 2,uses the auxiliary codebook of Encoder 1 $(\varphi_1^{(n)}(W_1, U_1^n))$ and generates an auxiliary $\mathbf{t}_2$ such that $(U_1^n, \varphi_1^{(n)}(W_1, U_1^n), \mathbf{u}_2, \mathbf{t}_2)$ is ϵ-strongly typical with high probability. To this end, an extended Markov lemma (see Lemma 5.5) ensures that the auxiliary codewords $\mathbf{t}_1$ and $\mathbf{t}_2$, although generated from separate encoders, are jointly typical with the source sequences with high probability.

We also derive an outer bound for the embedding capacity region with single-letter characterization (see Theorem 7.2), which follows from Fano's inequality and a standard information-theoretical bounding argument. We also study the embedding capacity region when the two host sources are independent, and inner and outer bounds are obtained (see Theorem 7.3). The inner bound is a consequence of Theorem 7.1. To prove the converse part, we sharpen the bound given in Theorem 7.2 by using the independence condition.

In [76], the authors obtained an achievable embedding region for correlated Gaussian host sources and parallel additive Gaussian attack channels. Their proof seems to be incorrect because their encoding approach cannot guarantee the typicality of the output sequences with respect to the host sequences. In fact, our Theorem 7.1 shows that their result (the achievable region) is correct; see the remark after Theorem 7.1. We also point out that a similar setup regarding the multi-user reversible information embedding system was considered in [36] for independent host sources and a MAC. Since in the reversible information embedding problem the secret messages as well as the host sources from both users are reconstructed at the single decoder, the techniques of the generalized rate-distortion coding scheme and the decoding based on the auxiliary codewords are not required in [36] and the coding strategy is fundamentally different from ours.

7.2 Problem Formulation and Main Results

Let the joint distribution of the discrete memoryless host sources U_1 and U_2 with alphabets $\mathcal{U}_1$ and $\mathcal{U}_2$ be $Q_{U_1U_2}$. The secret messages w_1 and w_2 are independently and uniformly chosen from the sets $\mathcal{W}_1 \triangleq \{1, 2, ..., M_1\}$ and $\mathcal{W}_2 \triangleq \{1, 2, ..., M_2\}$, respectively. The attack channel is modeled as a two-sender one-receiver discrete memoryless MAC $W_{Y|X_1X_2}$ having input alphabets $\mathcal{X}_1$ and $\mathcal{X}_2$, output alphabet $\mathcal{Y}$, and a transition probability distribution $W_{Y|X_1X_2}(y|x_1, x_2)$. The probability of receiving $\mathbf{y} \in \mathcal{Y}^n$ conditioned on sending $\mathbf{x}_1 \in \mathcal{X}_1^n$ and $\mathbf{x}_2 \in \mathcal{X}_2^n$ is hence given by $W^{(n)}_{Y|X_1X_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2)$. Let $d_i : \mathcal{U}_i \times \mathcal{X}_i \to [0, \infty)$ be single-letter distortion measures for $i = 1, 2$. For $\mathbf{u}_i \in \mathcal{U}_i^n$ and $\mathbf{x}_i \in \mathcal{X}_i^n$, let $d_i(\mathbf{u}_i, \mathbf{x}_i) = \sum_{j=1}^{n} d_i(u_{ij}, x_{ij})$.

A two-sender one-receiver multiple-access embedding (MAE) code $(f_1^{(n)}, f_2^{(n)}, \psi^{(n)})$

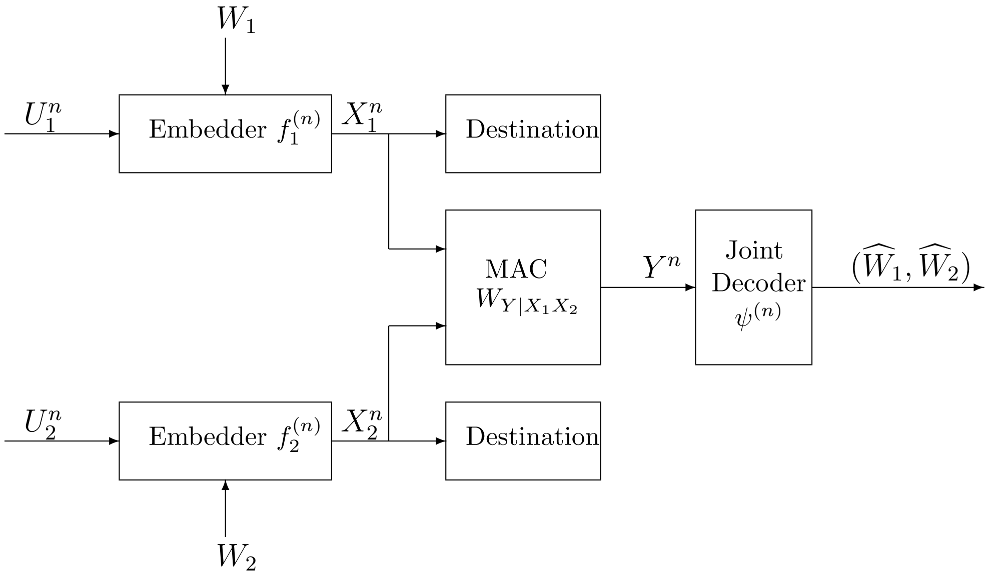


Figure 7.2: A multi-user information embedding system with two embedders.

with block length n consists of (see Fig. 7.2) two encoders (embedders)

$$f_1^{(n)} : \mathcal{W}_1 \times \mathcal{U}_1^n \longrightarrow \mathcal{X}_1^n \qquad \text{and} \qquad f_2^{(n)} : \mathcal{W}_2 \times \mathcal{U}_2^n \longrightarrow \mathcal{X}_2^n$$

with embedding rates $R_1 = \frac{\log_2 M_1}{n}$ and $R_2 = \frac{\log_2 M_2}{n}$, respectively, and decoder

$$\psi^{(n)} : \mathcal{Y}^n \longrightarrow \mathcal{W}_1 \times \mathcal{W}_2.$$

The system depicts a "public" embedding scenario since the host sources are not available at the decoder. The probability of erroneously decoding the secret messages is defined by

$$\begin{aligned}
&P_e^{(n)}(R_1, R_2) \\
&\quad \triangleq \ \Pr\left(\psi^{(n)}(Y^n) \neq (W_1, W_2)\right) \\
&\quad = \ \frac{1}{2^{n(R_1+R_2)}} \sum_{w_1=1}^{M_1} \sum_{w_2=1}^{M_2} \sum_{\mathcal{U}_1^n \times \mathcal{U}_2^n} Q_{U_1U_2}^{(n)}(\mathbf{u}_1, \mathbf{u}_2) W_{Y|X_1X_2}^{(n)}\left(\mathbf{y} : \psi^{(n)}(\mathbf{y}) \neq (w_1, w_2) | \mathbf{x}_1, \mathbf{x}_2\right)
\end{aligned}$$

where $\mathbf{x}_i \triangleq f_i^{(n)}(w_i, \mathbf{u}_i)$ for $i = 1, 2$.

Definition 7.1 Given $Q_{U_1U_2}$, $W_{Y|X_1X_2}$, a rate pair (R_1, R_2) is said to be achievable with respect to distortion levels (D_1, D_2) if there exists a sequence of MAE codes $(f_1^{(n)}, f_2^{(n)}, \psi^{(n)})$ at rates R_1 and R_2 such that

$$\lim_{n\to\infty} P_e^{(n)}(R_1, R_2) = 0$$

and

$$\limsup_{n\to\infty} \frac{1}{n}\mathbb{E}\left[d_i(U_i^n, f_i^{(n)}(W_i, U_i^n))\right] \leq D_i, \quad i = 1, 2.$$

Definition 7.2 The embedding capacity region $\mathcal{R}(D_1, D_2)$ is the closure of the set of achievable rate pairs (R_1, R_2).

Remark: It can be shown by using a time-sharing argument [10] that $\mathcal{R}(D_1, D_2)$ is convex.

Definition 7.3 Given $Q_{U_1U_2}$, $W_{Y|X_1X_2}$, and a pair of distortion levels (D_1, D_2), let $\mathcal{S}_{D_1,D_2}$ be the set of random variables $(U_1, T_1, U_2, T_2, X_1, X_2, Y) \in \mathcal{U}_1 \times \mathcal{T}_1 \times \mathcal{U}_2 \times \mathcal{T}_2 \times \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}$ for some finite alphabets $\mathcal{T}_1$ and $\mathcal{T}_2$ such that the joint distribution $P_{U_1T_1U_2T_2X_1X_2Y}$ satisfies: (1) it can be factorized as

$$P_{U_1T_1U_2T_2X_1X_2Y} = Q_{U_1U_2} P_{T_1X_1|U_1} P_{T_2X_2|U_2} W_{Y|X_1X_2},$$

(2) $I(U_i; T_i) > 0$, and (3) $\mathbb{E}[d_i(U_i, X_i)] \leq D_i$, for $i = 1, 2$.

Definition 7.4 Given $Q_{U_1U_2}$, $W_{Y|X_1X_2}$, and a pair of distortion levels (D_1, D_2), let $\mathcal{P}_{D_1,D_2}$ be the set of random variables $(U_1, T_1, U_2, T_2, X_1, X_2, Y) \in \mathcal{U}_1 \times \mathcal{T}_1 \times \mathcal{U}_2 \times \mathcal{T}_2 \times \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}$ for some finite alphabets $\mathcal{T}_1$ and $\mathcal{T}_2$ such that the joint distribution $P_{U_1T_1U_2T_2X_1X_2Y}$ satisfies: (1) it can be factorized as

$$P_{U_1T_1U_2T_2X_1X_2Y} = Q_{U_1U_2} P_{T_1T_2X_1X_2|U_1U_2} W_{Y|X_1X_2},$$

and (2) $\mathbb{E}[d_i(U_i, X_i)] \leq D_i$, for $i = 1, 2$.

By definition, $\mathcal{S}_{D_1,D_2} \subseteq \mathcal{P}_{D_1,D_2}$. The following are the main results of the paper.

Theorem 7.1 Let $\mathcal{R}_{in}(D_1, D_2)$ be the closure of the convex hull of all (R_1, R_2) satisfying

$$\begin{aligned} R_1 &< I(T_1; T_2, Y) - I(U_1; T_1), && (7.1) \\ R_2 &< I(T_2; T_1, Y) - I(U_2; T_2), && (7.2) \\ R_1 + R_2 &< I(T_1, T_2; Y) - I(U_1, U_2; T_1, T_2), && (7.3) \end{aligned}$$

for some $(U_1, T_1, U_2, T_2, X_1, X_2, Y) \in \mathcal{S}_{D_1,D_2}$. Then $\mathcal{R}_{in}(D_1, D_2) \subseteq \mathcal{R}(D_1, D_2)$.

Remark: Although we only deal with discrete (finite-alphabet) sources and channels, it is not hard to see that, with the appropriate changes in the proof, the achievable region is also valid for a system that incorporates a pair of correlated memoryless Gaussian sources and a Gaussian MAC. In particular, when the MAC is a pair of parallel additive Gaussian channels, $\mathcal{R}_{in}(D_1, D_2)$ reduces to the achievable region obtained in [76], even though the proof provided in [76] is not entirely correct.

Theorem 7.2 Let $\mathcal{R}_{out}(D_1, D_2)$ be the closure of all (R_1, R_2) satisfying (7.1)–(7.3) for some
$(U_1, T_1, U_2, T_2, X_1, X_2, Y) \in \mathcal{P}_{D_1,D_2}$. Then $\mathcal{R}(D_1, D_2) \subseteq \mathcal{R}_{out}(D_1+\delta, D_2+\delta)$ for all $\delta > 0$.

Remark: The theorem states that $\mathcal{R}(D_1, D_2) \subseteq \bigcap_{\delta>0} \mathcal{R}_{out}(D_1 + \delta, D_2 + \delta)$. If we could upper bound the cardinality of the alphabet sizes of the auxiliary RVs T_1 and T_2 in the definition of $\mathcal{R}_{out}(D_1, D_2)$, it could be readily shown that $\bigcap_{\delta>0} \mathcal{R}_{out}(D_1+\delta, D_2+\delta) = \mathcal{R}_{out}(D_1, D_2)$, so that $\mathcal{R}(D_1, D_2) \subseteq \mathcal{R}_{out}(D_1, D_2)$. However, without such an upper bound, we can only state the theorem in the present weaker form. The same remark applies to the outer bound in the next theorem.

We next consider the case when the host sources are independent, i.e., $Q_{U_1U_2} = Q_{U_1}Q_{U_2}$. We then have the following inner and outer bounds.

Theorem 7.3 Let $Q_{U_1U_2} = Q_{U_1}Q_{U_2}$. Let $\mathcal{R}^*_{in}(D_1, D_2)$ be the closure of the convex hull of all (R_1, R_2) satisfying

$$\begin{aligned} R_1 &< I(T_1; Y|T_2) - I(U_1; T_1) && (7.4)\\ R_2 &< I(T_2; Y|T_1) - I(U_2; T_2) && (7.5)\\ R_1 + R_2 &< I(T_1, T_2; Y) - I(U_1; T_1) - I(U_2; T_2) && (7.6) \end{aligned}$$

for some $(U_1, T_1, U_2, T_2, X_1, X_2, Y) \in \mathcal{S}_{D_1,D_2}$, and let $\mathcal{R}^*_{out}(D_1, D_2)$ be the closure of all (R_1, R_2) satisfying (7.4)–(7.6) for some $(U_1, T_1, U_2, T_2, X_1, X_2, Y) \in \mathcal{P}_{D_1,D_2}$. Then

$$\mathcal{R}^*_{in}(D_1, D_2) \subseteq \mathcal{R}(D_1, D_2) \subseteq \mathcal{R}^*_{out}(D_1 + \delta, D_2 + \delta)$$

for all $\delta > 0$.

Remark: In Section 7.3.4 we show that the cardinality of the alphabets of the auxiliary RVs T_1 and T_2 for $\mathcal{R}^*_{in}(D_1, D_2)$ and $\mathcal{R}_{in}(D_1, D_2)$ can be bounded as $|\mathcal{T}_i| \leq |\mathcal{U}_i||\mathcal{X}_i| + 1$ and $|\mathcal{T}_i| \leq |\mathcal{U}_1||\mathcal{U}_2||\mathcal{X}_i| + 1$ $(i = 1, 2)$, respectively.

7.3 Proofs

7.3.1 Proof of Theorem 7.1

We first give an outline of the proof. We need to show that for given $Q_{U_1U_2}$, $W_{Y|X_1X_2}$, and any $(R_1, R_2) \in \mathcal{R}_{in}(D_1, D_2)$, there exists a sequence of codes $(f_1^{(n)}, f_2^{(n)}, \psi^{(n)})$ such that $P_e^{(n)} \to 0$ as $n \to \infty$ and for any $\delta > 0$, $\frac{1}{n}\mathbb{E}[d_i(U_i^n, f_i^{(n)}(W_i, U_i^n))] \leq D_i + \delta$, $i = 1, 2$, for n sufficiently large.

Fix $(P_{T_1|U_1}, P_{X_1|U_1T_1}, P_{T_2|U_2}, P_{X_2|U_2T_2})$ such that $I(U_i; T_i) > 0$ and the following are satisfied for some $\epsilon' > 0$,

$$R_1 < I(T_1; T_2, Y) - I(U_1; T_1) - \epsilon', \tag{7.7}$$

$$R_2 < I(T_2; T_1, Y) - I(U_2; T_2) - \epsilon', \tag{7.8}$$

$$R_1 + R_2 < I(T_1, T_2; Y) - I(U_1, U_2; T_1, T_2) - \epsilon', \tag{7.9}$$

$$\mathbb{E}[d_i(U_i, X_i)] \leq D_i, \ i = 1, 2. \tag{7.10}$$

We will choose $f_1^{(n)}$ and $f_2^{(n)}$ in a random manner. For $\epsilon < \frac{\delta}{2\max\{d_1^{max}, d_2^{max}\}}$, define

$$P_i^{(n)} \triangleq \Pr\Big(\frac{1}{n} d_i\big(U_i^n, f_i^{(n)}(W_i, U_i^n)\big) > D_i + \epsilon d_i^{max}\Big), \ i = 1, 2,$$

where $d_i^{max} \triangleq \max\limits_{u_i, x_i} d_i(u_i, x_i)$, $i = 1, 2$. We will prove that for any $0 < \epsilon_1 \leq \frac{\delta}{6\max\{d_1^{max}, d_2^{max}\}}$, the probabilities $P_e^{(n)}$, $P_1^{(n)}$, and $P_2^{(n)}$, when averaged over the random choice of $f_1^{(n)}$ and $f_2^{(n)}$, satisfy

$$\mathbb{E}[P_e^{(n)}] \leq \epsilon_1, \quad \mathbb{E}[P_1^{(n)}] \leq \epsilon_1, \quad \mathbb{E}[P_2^{(n)}] \leq \epsilon_1$$

for n sufficiently large. Then $\mathbb{E}\{P_e^{(n)} + P_1^{(n)} + P_2^{(n)}\} \leq 3\epsilon_1$, which guarantees that there exists at least one pair of codes $(f_1^{(n)}, f_2^{(n)})$ such that $P_e^{(n)} + P_1^{(n)} + P_2^{(n)} \leq 3\epsilon_1$ and hence $P_e^{(n)} \leq 3\epsilon_1$, $P_1^{(n)} \leq 3\epsilon_1$, $P_2^{(n)} \leq 3\epsilon_1$ are simultaneously satisfied for n sufficiently large. Finally, it can be easily shown that $P_i^{(n)} \leq 3\epsilon_1$ implies for n sufficiently large that

$$\frac{1}{n}\mathbb{E}\left[d_i(U_i^n, f_i^{(n)}(W_i, U_i^n)\right] \leq D_i + \epsilon d_i^{max} + P_i^{(n)} d_i^{max} \leq D_i + \delta.$$

Random Code Design

In what follows, the strongly ϵ-typical set $\mathcal{T}_\epsilon^{(n)}$ is defined under the joint distribution

$$P_{U_1U_2T_1T_2X_1X_2Y} = Q_{U_1U_2} P_{T_1|U_1} P_{X_1|U_1T_1} P_{T_2|U_2} P_{X_2|U_2T_2} W_{Y|X_1X_2} \tag{7.11}$$

and all the marginal and conditional distributions, e.g., $P_{U_2T_2}$, $P_{U_1|U_2T_2}$, etc, are induced by the joint distribution. The parameter ϵ, which is chosen to be sufficiently small, will be specified in the proof.

Generation of codebooks. For $i = 1, 2$ and every $w_i \in \mathcal{W}_i$, generate a codebook

$$\mathcal{C}_{w_i} = \{\mathbf{t}_i(w_i, 1), \mathbf{t}_i(w_i, 2), ..., \mathbf{t}_i(w_i, L_i)\}$$

with $L_i = 2^{n[I(U_i;T_i)+4\epsilon]}$ codewords such that each $\mathbf{t}_i(w_i, l_i)$ is independently selected with uniform distribution from the typical set $\mathcal{T}_\epsilon^{(n)}(T_i)$. Denote the entire codebook for Encoder i by $\mathcal{C}^{(i)} = \{\mathcal{C}_{w_i}\}_{w_i=1}^{M_i}$, where we recall that $M_i = 2^{nR_i}$. For each $\mathbf{u}_i$ and codeword $\mathbf{t}_i(w_i, l_i)$ $(1 \le w_i \le M_i, 1 \le l_i \le L_i)$, generate a codeword $\mathbf{x}_i$ according to $P^{(n)}_{X_i|U_iT_i}(\mathbf{x}_i|\mathbf{u}_i, \mathbf{t}_i)$. Denote the codebook of all the codewords $\mathbf{x}_i$ by $\mathcal{B}^{(i)}$.

Encoder $f_1^{(n)}$: Encoder $f_1^{(n)}$ is the concatenation of a pre-encoder $\varphi_1^{(n)} : \mathcal{W}_1 \times \mathcal{U}_1^n \longrightarrow \mathcal{T}_1^n$ and a mapping $g_1^{(n)} : \mathcal{U}_1^n \times \mathcal{T}_1^n \longrightarrow \mathcal{X}_1^n$.

To define $\varphi_1^{(n)}$, we need the following notation adopted from [54]. We introduce a conditional probability

$$A^{(n)}(\mathbf{u}_1, \mathbf{t}_1) \triangleq P^{(n)}_{U_2T_2|U_1T_1}\left((\mathbf{u}_2, \mathbf{t}_2) : (\mathbf{u}_2, \mathbf{t}_2) \in \mathcal{T}_\epsilon^{(n)}(U_2T_2|\mathbf{u}_1, \mathbf{t}_1) \middle| \, \mathbf{u}_1, \mathbf{t}_1\right).$$

For $\mu \in (0, 1)$, let

$$\mathcal{F}^{(n)}_{\mu,\epsilon}(U_1, T_1) \triangleq \left\{(\mathbf{u}_1, \mathbf{t}_1) : A^{(n)}(\mathbf{u}_1, \mathbf{t}_1) \ge 1 - \mu\right\}.$$

By definition, we have $\mathcal{F}^{(n)}_{\mu,\epsilon}(U_1, T_1) \subseteq \mathcal{T}_\epsilon^{(n)}(U_1, T_1)$.

We now describe the pre-encoding function $\varphi_1^{(n)} = \varphi_1^{(n)}(w_1, \mathbf{u}_1)$ which maps every pair $(w_1, \mathbf{u}_1)$ to a codeword in $\mathcal{C}^{(1)} \subseteq \mathcal{T}_1^n$. Given $w_1 \in \{1, 2, ..., M_1\}$ and $\mathbf{u}_1$, $\varphi_1^{(n)}$ seeks the first codeword $\mathbf{t}_1(w_1, l_1)$ with $l_1 \le L_1 - 1$ in $\mathcal{C}_{w_1}$ such that $(\mathbf{u}_1, \mathbf{t}_1(w_1, l_1)) \in \mathcal{F}^{(n)}_{\mu,\epsilon}(U_1, T_1)$. If there is no such codeword, $\varphi_1^{(n)}$ outputs $\mathbf{t}_1(w_1, L_1)$. Next, for each output $\mathbf{t}_1(w_1, l_1)$ and $\mathbf{u}_1$, $g_1^{(n)}$ sends out the associated codeword $\mathbf{x}_1(w_1, \mathbf{u}_1)$ to the channel. Thus, $f_1^{(n)}(w_1, \mathbf{u}_1) = g_1^{(n)}\left(\mathbf{u}_1, \varphi_1^{(n)}(w_1, \mathbf{u}_1)\right)$.

Encoder $f_2^{(n)}$: Encoder $f_2^{(n)}$ is the concatenation of a pre-encoder $\varphi_2^{(n)} : \mathcal{W}_2 \times \mathcal{U}_2^n \longrightarrow \mathcal{T}_2^n$ and a mapping $g_2^{(n)} : \mathcal{U}_2^n \times \mathcal{T}_2^n \longrightarrow \mathcal{X}_2^n$.

To define $\varphi_2^{(n)}$, let

$$B_{\varphi_1}^{(n)}(\mathbf{u}_2, \mathbf{t}_2) \triangleq \frac{1}{2^{nR_1}} \sum_{w_1=1}^{M_1} P_{U_1|U_2T_2}^{(n)} \left(\mathbf{u}_1 : (\mathbf{u}_1, \varphi_1^{(n)}(w_1, \mathbf{u}_1)) \in \mathcal{T}_\epsilon^{(n)}(U_1T_1|\mathbf{u}_2, \mathbf{t}_2) \middle| \mathbf{u}_2, \mathbf{t}_2 \right).$$

Also, for $\nu \in (0, 1)$, define

$$\mathcal{F}_{\varphi_1,\nu,\epsilon}^{(n)}(U_2, T_2) \triangleq \left\{ (\mathbf{u}_2, \mathbf{t}_2) : B_{\varphi_1}^{(n)}(\mathbf{u}_2, \mathbf{t}_2) \geq 1 - \nu \right\}.$$

By definition, it is seen that $\mathcal{F}_{\varphi_1,\nu,\epsilon}^{(n)}(U_2, T_2) \subseteq \mathcal{T}_\epsilon^{(n)}(U_2, T_2)$.

We now describe the pre-encoding function $\varphi_2^{(n)} = \varphi_2^{(n)}(w_2, \mathbf{u}_2)$ which maps every pair $(w_2, \mathbf{u}_2)$ to a codeword in $\mathcal{C}^{(2)} \subseteq \mathcal{T}_2^n$. Given $w_2 \in \{1, 2, ..., M_2\}$ and $\mathbf{u}_2$, $\varphi_2^{(n)}$ seeks the first codeword $\mathbf{t}_2(w_2, l_2)$ with $l_2 \leq L_2 - 1$ in $\mathcal{C}_{w_2}$ such that $(\mathbf{u}_2, \mathbf{t}_2(w_2, l_2)) \in \mathcal{F}_{\varphi_1,\nu,\epsilon}^{(n)}(U_2, T_2)$. If there is no such codeword, $\varphi_2^{(n)}$ outputs $\mathbf{t}_2(w_2, L_2)$. Next, for each output $\mathbf{t}_2(w_2, l_2)$, $g_2^{(n)}$ sends out the associated codeword $\mathbf{x}_2(w_2, \mathbf{u}_2)$ to the channel. Thus, $f_2^{(n)}(w_2, \mathbf{u}_2) = g_2^{(n)}\left(\mathbf{u}_2, \varphi_2^{(n)}(w_2, \mathbf{u}_2)\right)$.

Decoder $\psi^{(n)}$: Given $\mathbf{y}$, $\psi^{(n)}$ seeks $\mathbf{t}_1(\widehat{w}_1, \widehat{l}_1) \in \mathcal{C}^{(1)}$ and $\mathbf{t}_2(\widehat{w}_2, \widehat{l}_2) \in \mathcal{C}^{(2)}$ such that

$$(\mathbf{t}_1(\widehat{w}_1, \widehat{l}_1), \mathbf{t}_2(\widehat{w}_2, \widehat{l}_2), \mathbf{y}) \in \mathcal{T}_\epsilon^{(n)}(T_1, T_2, Y).$$

If such a pair $(\mathbf{t}_1(\widehat{w}_1, \widehat{l}_1), \mathbf{t}_2(\widehat{w}_2, \widehat{l}_2))$ exists for a unique $(\widehat{w}_1, \widehat{w}_2)$, then $\psi^{(n)}$ outputs $\widehat{w}_1$ and $\widehat{w}_2$ as the decoded messages. If there is no such pair $(\widehat{w}_1, \widehat{w}_2)$, or it is not unique, a decoding error is declared. Letting $\mathbf{t}_i(w_i, l_i) = \varphi_i^{(n)}(w_i, \mathbf{u}_i)$, it is easy to see that if there is a decoding error, then at least one of the following events occurs:

1. E_1: $(\mathbf{t}_1(w_1, l_1), \mathbf{t}_2(w_2, l_2), \mathbf{y}) \notin \mathcal{T}_\epsilon^{(n)}(T_1, T_2, Y)$,

2. E_2: there exist l_1' and $w_1' \neq w_1$ and l_2' (l_2' may or may not be equal to l_2) such that

$$(\mathbf{t}_1(w_1', l_1'), \mathbf{t}_2(w_2, l_2'), \mathbf{y}) \in \mathcal{T}_\epsilon^{(n)}(T_1, T_2, Y),$$

3. E_3: there exist l_2' and $w_2' \neq w_2$ and l_1' (l_1' may or may not be equal to l_1) such that

$$(\mathbf{t}_1(w_1, l_1'), \mathbf{t}_2(w_2', l_2'), \mathbf{y}) \in \mathcal{T}_\epsilon^{(n)}(T_1, T_2, Y),$$

   or

4. E_4: there exist l_1' and $w_1' \neq w_1$ and l_2' and $w_2' \neq w_2$ such that

$$(\mathbf{t}_1(w_1', l_1'), \mathbf{t}_2(w_2', l_2'), \mathbf{y}) \in \mathcal{T}_\epsilon^{(n)}(T_1, T_2, Y).$$

In the following, we will bound the probabilities $P_e^{(n)}$, $P_1^{(n)}$ and $P_2^{(n)}$ averaged over the random choice of all codes $\mathcal{B}^{(1)}$, $\mathcal{B}^{(2)}$, $\mathcal{C}^{(1)}$, and $\mathcal{C}^{(2)}$. To simplify the notation we abbreviate $\mathbb{E}_{\mathcal{B}^{(1)},\mathcal{B}^{(2)},\mathcal{C}^{(1)},\mathcal{C}^{(2)}}[\,\cdot\,]$ as $\mathbb{E}_\Omega[\,\cdot\,]$.

Bounding $\mathbb{E}_\Omega[P_e^{(n)}]$

To analyze the average probability of error, we need the following lemma.

Lemma 7.1 For any discrete RVs (U_1, U_2, T_1, T_2) forming a Markov chain $T_1 \to U_1 \to U_2 \to T_2$ in this order, we have

$$I(U_1, U_2; T_1, T_2) + I(T_1; T_2) = I(U_1; T_1) + I(U_2; T_2).$$

Proof: We first write

$$\begin{aligned} I(U_1, U_2; T_1, T_2) &= H(T_1, T_2) - H(T_1, T_2 | U_1, U_2) \\ &= H(T_1, T_2) - H(T_1 | U_1, U_2) - H(T_2 | U_1, U_2, T_1) \\ &= H(T_1, T_2) - H(T_1 | U_1) - H(T_2 | U_2) \end{aligned}$$

where the last equality follows from the Markov condition. Since $I(T_1; T_2) = H(T_1) + H(T_2) - H(T_1, T_2)$, we have

$$I(U_1, U_2; T_1, T_2) + I(T_1; T_2) = H(T_1) + H(T_2) - H(T_1 | U_1) - H(T_2 | U_2)$$

$$= \quad I(U_1;T_1) + I(U_2;T_2).$$

□

Since the watermarks are independently and uniformly distributed, and by the symmetry of the code construction, we can assume without the loss of generality that some fixed $w_1 \in \mathcal{W}_1$ and $w_2 \in \mathcal{W}_2$ are the transmitted watermarks. Thus we bound the probability of error as

$$\begin{aligned} P_e^{(n)} &= \Pr\left(\left\{\psi^{(n)}(Y^n) \neq (w_1, w_2)\right\}\right) \\ &\leq \Pr(A_1) + \Pr\left(\left\{\psi^{(n)}(Y^n) \neq (w_1, w_2)\right\}\middle|\, A_1^c\right) \end{aligned} \tag{7.12}$$

where A_1 is the event

$$A_1 : (\mathbf{t}_1(w_1, l_1), \mathbf{u}_1, \mathbf{u}_2, \mathbf{t}_2(w_2, l_2), \mathbf{x}_1, \mathbf{x}_2) \notin \mathcal{T}_\epsilon^{(n)}(T_1, U_1, U_2, T_2, X_1, X_2).$$

Recall that $\mathbf{t}_i(w_i, l_i) = \varphi_i^{(n)}(w_i, \mathbf{u}_i)$, $i = 1, 2$. We also let $\mathbf{t}_i(w_i, l_i')$ and $\mathbf{t}_i(w_i', l_i')$ be the l_i'-th codeword in the codebook $\mathcal{C}_{w_i}$ and $\mathcal{C}_{w_i'}$, respectively.

We then introduce the event

$$A_0 : (\mathbf{t}_1(w_1, l_1), \mathbf{u}_1, \mathbf{u}_2, \mathbf{t}_2(w_2, l_2)) \notin \mathcal{T}_\epsilon^{(n)}(T_1, U_1, U_2, T_2).$$

Taking expectation in (7.12) and using the union bound, we have

$$\mathbb{E}_\Omega[P_e^{(n)}] \leq \mathbb{E}_\Omega \Pr(A_0) + \mathbb{E}_\Omega \Pr(A_1|A_0^c) + \mathbb{E}_\Omega \Pr(E_1|A_1^c) + \sum_{k=2}^{4} \mathbb{E}_\Omega \Pr(E_k|A_1^c). \tag{7.13}$$

It immediately follows from Lemma 5.5 that

$$\mathbb{E}_\Omega \Pr(A_0) = \mathbb{E}_{\mathcal{C}^{(1)},\mathcal{C}^{(2)}} \Pr(A_0) \leq \epsilon_0 \tag{7.14}$$

for n sufficiently large, where we set $\epsilon_0 = \epsilon_1/7$ for a given $\epsilon_1 \geq 0$ throughout the proof. When A_0^c holds, since $\mathbf{x}_1$ and $\mathbf{x}_2$ are respectively drawn according to the conditional probabilities $P^{(n)}_{X_1|U_1T_1}(\cdot|\mathbf{u}_1,\mathbf{t}_1)$ and $P^{(n)}_{X_2|U_2T_2}(\cdot|\mathbf{u}_2,\mathbf{t}_2)$, and $\mathbf{y}$ is drawn according to the conditional distribution $W^{(n)}_{Y|X_1X_2}(\cdot|\mathbf{x}_1,\mathbf{x}_2)$, it follows from two successive applications of Lemma 5.2 that

$$\mathbb{E}_\Omega \Pr\left(A_1 \middle| A_0^c \right) \leq \mathbb{E}_\Omega[\epsilon_0] = \epsilon_0 \tag{7.15}$$

and

$$\begin{aligned} &\mathbb{E}_\Omega \Pr\left(E_1 \middle| A_1^c \right) \\ &\leq \mathbb{E}_\Omega \Pr\left(\left\{ \left(\varphi_1^{(n)}(w_1,U_1^n), U_1^n, U_2^n, \varphi_2^{(n)}(w_2,U_2^n), f_1^{(n)}(w_1,U_1^n), f_2^{(n)}(w_2,U_2^n), Y^n \right) \notin \mathcal{T}_\epsilon^{(n)} \right\} \middle| A_1^c \right) \\ &\leq \mathbb{E}_\Omega[\epsilon_0] = \epsilon_0 \end{aligned} \tag{7.16}$$

for n sufficiently large. It remains to bound $\mathbb{E}_\Omega \Pr\left\{ E_k \middle| A_1^c \right\}$ for $k = 2,3,4$. Using the union bound we write

$$\begin{aligned} &\mathbb{E}_\Omega \Pr\left(E_2 \middle| A_1^c \right) \\ &\leq \sum_{w_1' \neq w_1} \sum_{l_1'=1}^{L_1} \Pr\left(\left\{ (T_1^n(w_1',l_1'), Y^n, T_2^n(w_2,l_2')) \in \mathcal{T}_\epsilon^{(n)}(T_1,T_2,Y) \right\} \middle| A_1^c \right), \end{aligned} \tag{7.17}$$

where $T_1^n(w_1',l_1')$ is a RV uniformly drawn from $\mathcal{T}_\epsilon^{(n)}(T_1)$ which is independent of $(T_2^n(w_2,l_2'),Y^n)$ since $w_1' \neq w_1$. Thus we have

$$\begin{aligned} &\Pr\left(\left\{ (T_1^n(w_1',l_1'), Y^n, T_2^n(w_2,l_2')) \in \mathcal{T}_\epsilon^{(n)}(T_1,T_2,Y) \right\} \middle| A_1^c \right) \\ &= \sum_{(\mathbf{t}_2,\mathbf{y}) \in \mathcal{T}_\epsilon^{(n)}(T_2,Y)} \sum_{\mathbf{t}_1 \in \mathcal{T}_\epsilon^{(n)}(T_1|\mathbf{t}_2,\mathbf{y})} \Pr\left(T_2^n(w_2,l_2') = \mathbf{t}_2, Y^n = \mathbf{y} \middle| A_1^c \right) \\ &\qquad \Pr\left(T_1^n(w_1',l_1') = \mathbf{t}_1 \middle| T_2^n(w_2,l_2') = \mathbf{t}_2, Y^n = \mathbf{y}, A_1^c \right) \\ &= \sum_{(\mathbf{t}_2,\mathbf{y}) \in \mathcal{T}_\epsilon^{(n)}(T_2,Y)} \sum_{\mathbf{t}_1 \in \mathcal{T}_\epsilon^{(n)}(T_1|\mathbf{t}_2,\mathbf{y})} \Pr\left(T_2^n(w_2,l_2') = \mathbf{t}_2, Y^n = \mathbf{y} \middle| A_1^c \right) \Pr\left(T_1^n(w_1',l_1') = \mathbf{t}_1 \right) \\ &= \sum_{(\mathbf{t}_2,\mathbf{y}) \in \mathcal{T}_\epsilon^{(n)}(T_2,Y)} \Pr\left(T_2^n(w_2,l_2) = \mathbf{t}_2, Y^n = \mathbf{y} \middle| A_1^c \right) \frac{|\mathcal{T}_\epsilon^{(n)}(T_1|\mathbf{t}_2,\mathbf{y})|}{|\mathcal{T}_\epsilon^{(n)}(T_1)|} \\ &\leq \frac{2^{n[H(T_1|T_2,Y)+\eta]}}{2^{n[H(T_1)-\eta]}} \sum_{(\mathbf{t}_2,\mathbf{y}) \in \mathcal{T}_\epsilon^{(n)}(T_2,Y)} \Pr\left(T_2^n(w_2,l_2') = \mathbf{t}_2, Y^n = \mathbf{y} \middle| A_1^c \right) \end{aligned}$$

$$\leq \ 2^{-n[I(T_1;T_2,Y)-2\eta]}, \tag{7.18}$$

where the first inequality follows from Lemma 5.1. Recalling that $\eta \to 0$ as $n \to \infty$ and $\epsilon \to 0$, we can make sure that $2\eta < \epsilon' - 4\epsilon$ by choosing ϵ small enough and n large enough. Thus from (7.17)

$$\begin{aligned} \mathbb{E}_\Omega \Pr\left(E_2 | \, A_1^c \right) &\leq \ 2^{n[R_1+I(U_1;T_1)+4\epsilon-I(T_1;T_2,Y)+2\eta]} \\ &\leq \ 2^{n[R_1+I(U_1;T_1)-I(T_1;T_2,Y)+\epsilon']} \\ &\leq \ \epsilon_0 \end{aligned} \tag{7.19}$$

for ϵ sufficiently small and n sufficiently large, where (7.19) follows from the assumption (7.7). Similarly we have

$$\mathbb{E}_\Omega \Pr\left(E_3 | \, A_1^c \right) \leq \epsilon_0 \tag{7.20}$$

for ϵ small enough and n sufficiently large. We next bound

$$\begin{aligned} &\mathbb{E}_\Omega \Pr\left(E_4 | \, A_1^c \right) \\ &\leq \ \sum_{w_1' \neq w_1} \sum_{l_1'=1}^{L_1} \sum_{w_2' \neq w_2} \sum_{l_2'=1}^{L_2} \Pr\left(\left\{ (T_1^n(w_1',l_1'), T_2^n(w_2',l_2'), Y^n) \in \mathcal{T}_\epsilon^{(n)}(T_1,T_2,Y) \right\} \middle| \, A_1^c \right), \end{aligned}$$

where $T_1^n(w_1',l_1')$ and $T_2^n(w_2',l_2')$ are RVs independently drawn from $\mathcal{T}_\epsilon^{(n)}(T_1)$ and $\mathcal{T}_\epsilon^{(n)}(T_2)$ according to the uniform distribution, respectively. We have

$$\begin{aligned} &\Pr\left(\left\{ (T_1^n(w_1',l_1'), T_2^n(w_2',l_2'), Y^n) \in \mathcal{T}_\epsilon^{(n)}(T_1,T_2,Y) \right\} \middle| \, A_1^c \right) \\ &= \sum_{\mathbf{y} \in \mathcal{T}_\epsilon^{(n)}(Y)} \ \sum_{(\mathbf{t}_1,\mathbf{t}_2) \in \mathcal{T}_\epsilon^{(n)}(T_1,T_2|\mathbf{y})} \Pr(Y^n = \mathbf{y} | A_1^c) \\ &\qquad \Pr(T_1^n(w_1',l_1') = \mathbf{t}_1, T_2^n(w_2',l_2') = \mathbf{t}_2 | A_1^c, Y^n = \mathbf{y}) \\ &= \sum_{\mathbf{y} \in \mathcal{T}_\epsilon^{(n)}(Y)} \ \sum_{(\mathbf{t}_1,\mathbf{t}_2) \in \mathcal{T}_\epsilon^{(n)}(T_1,T_2|Y)} \Pr(Y^n = \mathbf{y} | A_1^c) \frac{1}{|\mathcal{T}_\epsilon^{(n)}(T_1)|} \frac{1}{|\mathcal{T}_\epsilon^{(n)}(T_2)|} \\ &\leq \sum_{\mathbf{y} \in \mathcal{T}_\epsilon^{(n)}(Y)} \Pr(Y^n = \mathbf{y} | A_1^c) \frac{2^{n[H(T_1,T_2|Y)+\eta]}}{2^{n[H(T_1)-\eta]} 2^{n[H(T_2)-\eta]}} \end{aligned}$$

$$\leq \quad 2^{-n[I(T_1,T_2;Y)+I(T_1;T_2)-3\eta]}$$

and hence

$$\begin{aligned}
&\mathbb{E}_\Omega \Pr\left(E_4 \middle| A_1^c\right)\\
&\quad\leq \quad 2^{n[R_1+R_2+I(U_1;T_1)+I(U_2;T_2)-I(T_1,T_2;Y)-I(T_1;T_2)+8\epsilon+3\eta]}\\
&\quad\leq \quad 2^{n[R_1+I(U_1,U_2;T_1,T_2)-I(T_1,T_2;Y)+\epsilon']}\\
&\quad\leq \quad \epsilon_0 \qquad (7.21)
\end{aligned}$$

for n sufficiently large and ϵ small enough (such that $8\epsilon+3\eta<\epsilon'$), where the second inequality follows from Lemma 7.1 and the last inequality follows from the assumption (7.9). Finally, substituting (7.14)–(7.16), (7.19), (7.20) and (7.21) into (7.13) yields $\mathbb{E}_\Omega[P_e^{(n)}] \leq 7\epsilon_0 = \epsilon_1$ for ϵ sufficiently small and n sufficiently large.

Bounding $\mathbb{E}_\Omega[P_i^{(n)}]$

We only bound $\mathbb{E}_\Omega[P_i^{(n)}]$ for $i=1$, since the case $i=2$ can be dealt with similarly. When $(\mathbf{u}_1, \mathbf{x}_1(w_1,\mathbf{u}_1)) \in \mathcal{T}_\epsilon^{(n)}(U_1,X_1)$,

$$\frac{1}{n} d_1\big(\mathbf{u}_1, \mathbf{x}_1(w_1,\mathbf{u}_1)\big) \leq \mathbb{E}[d_1(U_1,X_1)] + \epsilon d_1^{max} \leq D_1 + \epsilon d_1^{max}$$

for n sufficiently large, where the first inequality follows from the definition of strong typicality and the second inequality follows from (7.10). This means that if $\frac{1}{n}d_1\big(U_1^n, f_1^{(n)}(W_1,U_1^n)\big) > D_1+\epsilon d_1^{max}$, then we must have $\big(U_1^n, f_1^{(n)}(W_1,U_1^n)\big) \notin \mathcal{T}_\epsilon^{(n)}(U_1,X_1)$ for n sufficiently large. Thus, we can bound

$$\begin{aligned}
&\Pr\Big(\frac{1}{n} d_1(U_1^n, f_1^{(n)}(W_1,U_1^n)) > D_1 + \epsilon d_1^{max}\Big)\\
&\quad\leq \quad \Pr\left(\left(U_1^n, f_1^{(n)}(W_1,U_1^n)\right) \notin \mathcal{T}_\epsilon^{(n)}(U_1,X_1)\right)\\
&\quad\leq \quad \Pr\left(\left(U_1^n, \varphi_1^{(n)}(W_1,U_1^n), f_1^{(n)}(W_1,U_1^n)\right) \notin \mathcal{T}_\epsilon^{(n)}(U_1,T_1,X_1)\right)
\end{aligned}$$

$$\begin{aligned}
&\le \Pr\Big(\big(U_1^n, \varphi_1^{(n)}(W_1, U_1^n)\big) \notin \mathcal{T}_\epsilon^{(n)}(U_1, T_1)\Big) \\
&\quad + \Pr\Big(\big(U_1^n, \varphi_1^{(n)}(W_1, U_1^n), f_1^{(n)}(W_1, U_1^n)\big) \notin \mathcal{T}_\epsilon^{(n)}(U_1, T_1, X_1)\Big| \\
&\qquad\qquad\qquad\qquad (U_1^n, \varphi_1^{(n)}(W_1, U_1^n)) \in \mathcal{T}_\epsilon^{(n)}(U_1, T_1)\Big) \\
&\le \Pr\Big(\big(\varphi_1^{(n)}(W_1, U_1^n), U_1^n, U_2^n, \varphi_2^{(n)}(W_2, U_2^n)\big) \notin \mathcal{T}_\epsilon^{(n)}(T_1, U_1, U_2, T_2)\Big) \\
&\quad + \Pr\Big(\big(U_1^n, \varphi_1^{(n)}(W_1, U_1^n), f_1^{(n)}(W_1, U_1^n)\big) \notin \mathcal{T}_\epsilon^{(n)}(U_1, T_1, X_1)\Big| \\
&\qquad\qquad\qquad\qquad (U_1^n, \varphi_1^{(n)}(W_1, U_1^n)) \in \mathcal{T}_\epsilon^{(n)}(U_1, T_1)\Big). \quad (7.22)
\end{aligned}$$

Now taking expectation on both sides, the first term of (7.22) is bounded by $\frac{\epsilon_1}{2}$ by Lemma 5.5, and the second term is bounded by $\frac{\epsilon_1}{2}$ for sufficiently large n by Lemma 5.2. This completes the proof of the bound $\mathbb{E}_\Omega[P_1^{(n)}] \le \epsilon_1$ for n sufficiently large. □

7.3.2 Proof of Theorem 7.2

We will prove the outer bound for the achievable region by using a standard bounding technique based on Fano's inequality. In fact, our proof is a generalization of the proof of the converse in [88] for a single-user embedding system.

We need to show that any MAE code $(f_1^{(n)}, f_2^{(n)}, \psi^{(n)})$ with achievable rate pair (R_1, R_2) must satisfy (7.1)–(7.3) for some auxiliary RVs T_1 and T_2 with joint distribution $P_{U_1U_2T_1T_2X_1X_2Y} \in \mathcal{P}_{D_1,D_2}$. It follows from Fano's inequality that

$$H(W_1, W_2|Y^n) \le n(R_1 + R_2)P_e^{(n)} + H(P_e^{(n)}) \triangleq n\epsilon_n.$$

It is clear that $\epsilon_n \to 0$ if $P_e^{(n)} \to 0$ and

$$\begin{aligned}
H(W_1|Y^n) &\le H(W_1, W_2|Y^n) \le n\epsilon_n, \\
H(W_2|Y^n) &\le H(W_1, W_2|Y^n) \le n\epsilon_n.
\end{aligned}$$

Because W_1 is uniformly drawn from the message set $\{1, 2, ..., 2^{nR_1}\}$ and is independent of U_1^n, we have

$$nR_1 = H(W_1) = I(W_1;Y^n) + H(W_1|Y^n) \leq I(W_1;Y^n) - \underbrace{I(W_1;U_1^n)}_{=0} + n\epsilon_n.$$

In the following, we bound

$$\begin{aligned}
&I(W_1;Y^n) - I(W_1;U_1^n)\\
&\overset{(a)}{=} \sum_{k=1}^{n}\Big[I(W_1;Y_k|Y_1^{k-1}) - I(W_1;U_{1k}|U_{1,k+1}^n)\Big]\\
&= \sum_{k=1}^{n}\Big[H(Y_k|Y_1^{k-1}) - H(Y_k|W_1,Y_1^{k-1},U_{1,k+1}^n) - I(Y_k;U_{1,k+1}^n|W_1,Y_1^{k-1})\\
&\qquad -H(U_{1k}|U_{1,k+1}^n) + H(U_{1k}|W_1,U_{1,k+1}^n)\Big]\\
&\overset{(b)}{=} \sum_{k=1}^{n}\Big[H(Y_k|Y_1^{k-1}) - H(Y_k|W_1,Y_1^{k-1},U_{1,k+1}^n) - I(U_{1k};Y_1^{k-1}|W_1,U_{1,k+1}^n)\\
&\qquad -H(U_{1k}|U_{1,k+1}^n) + H(U_{1k}|W_1,U_{1,k+1}^n)\Big]\\
&\overset{(c)}{=} \sum_{k=1}^{n}\Big[H(Y_k|Y_1^{k-1}) - H(Y_k|W_1,Y_1^{k-1},U_{1,k+1}^n)\\
&\qquad -H(U_{1k}) + H(U_{1k}|W_1,Y_1^{k-1},U_{1,k+1}^n)\Big]\\
&\leq \sum_{k=1}^{n}\Big[H(Y_k) - H(Y_k|W_1,U_{1,k+1}^n,Y_1^{k-1}) - I(U_{1k};W_1,Y_1^{k-1},U_{1,k+1}^n)\Big]\\
&= \sum_{k=1}^{n}\Big[I(Y_k;W_1,U_{1,k+1}^n,Y_1^{k-1}) - I(U_{1k};W_1,Y_1^{k-1},U_{1,k+1}^n)\Big]\\
&\overset{(d)}{\leq} \sum_{k=1}^{n}\Big[I(W_2,U_{2,k+1}^n,Y_1^{k-1},Y_k;W_1,U_{1,k+1}^n,Y_1^{k-1}) - I(U_{1k};W_1,Y_1^{k-1},U_{1,k+1}^n)\Big]\\
&\overset{(e)}{=} \sum_{k=1}^{n}\Big[I(L_{2k},Y_k;L_{1k}) - I(U_{1k};L_{1k})\Big]
\end{aligned}$$

where in (a) $Y_1^{k-1} \triangleq (Y_1,Y_2,...,Y_{k-1})$ and $U_{1,k+1}^n \triangleq (U_{1,k+1},U_{1,k+2},...,U_{1,n})$, (b) follows from the "summation by parts" identity [14, Lemma 7], (c) holds since the source U_1 is memoryless, in (d) $U_{2,k+1}^n \triangleq (U_{2,k+1},U_{2,k+2},...,U_{2,n})$, and in (e) $L_{1k} \triangleq (W_1,Y_1^{k-1},U_{1,k+1}^n)$

and $L_{2k} \triangleq (W_2, Y_1^{k-1}, U_{2,k+1}^n)$. Hence we obtain the bound

$$R_1 \leq \frac{1}{n}\sum_{k=1}^{n}[I(L_{1k}; L_{2k}, Y_k) - I(U_{1k}; L_{1k})] + \epsilon_n. \tag{7.23}$$

Similarly, we can bound

$$R_2 \leq \frac{1}{n}\sum_{k=1}^{n}[I(L_{2k}; L_{1k}, Y_k) - I(U_{2k}; L_{2k})] + \epsilon_n. \tag{7.24}$$

To bound the sum of the rates, we have

$$\begin{aligned} n(R_1 + R_2) &= H(W_1, W_2) = I(W_1, W_2; Y^n) + H(W_1, W_2|Y^n) \\ &\leq I(W_1, W_2; Y^n) - \underbrace{I(W_1, W_2; U_1^n, U_2^n)}_{=0} + n\epsilon_n \end{aligned} \tag{7.25}$$

and

$$\begin{aligned} &I(W_1, W_2; Y^n) - I(W_1, W_2; U_1^n, U_2^n) \\ &= \sum_{k=1}^{n}\Big[I(W_1, W_2; Y_k|Y_1^{k-1}) - I(W_1, W_2; U_{1k}, U_{2k}|U_{1,k+1}^n, U_{2,k+1}^n)\Big] \\ &= \sum_{k=1}^{n}\Big[H(Y_k|Y_1^{k-1}) - H(Y_k|W_1, U_{1,k+1}^n, Y_1^{k-1}, W_2, U_{2,k+1}^n) \\ &\qquad -I(Y_k; U_{1,k+1}^n, U_{2,k+1}^n|W_1, W_2, Y_1^{k-1}) \\ &\qquad -H(U_{1k}, U_{2k}|U_{1,k+1}^n, U_{2,k+1}^n) + H(U_{1k}, U_{2k}|W_1, W_2, U_{1,k+1}^n, U_{2,k+1}^n)\Big] \\ &= \sum_{k=1}^{n}\Big[H(Y_k|Y_1^{k-1}) - H(Y_k|W_1, U_{1,k+1}^n, Y_1^{k-1}, W_2, U_{2,k+1}^n) \\ &\qquad -I(U_{1k}, U_{2k}; Y_1^{k-1}|W_1, W_2, U_{1,k+1}^n, U_{2,k+1}^n) \\ &\qquad -H(U_{1k}, U_{2k}) + H(U_{1k}, U_{2k}|W_1, W_2, U_{1,k+1}^n, U_{2,k+1}^n)\Big] \\ &= \sum_{k=1}^{n}\Big[H(Y_k|Y_1^{k-1}) - H(Y_k|W_1, U_{1,k+1}^n, Y_1^{k-1}, W_2, U_{2,k+1}^n) \\ &\qquad -H(U_{1k}, U_{2k}) + H(U_{1k}, U_{2k}|W_1, W_2, U_{1,k+1}^n, U_{2,k+1}^n, Y_1^{k-1})\Big] \\ &\leq \sum_{k=1}^{n}\Big[H(Y_k) - H(Y_k|W_1, U_{1,k+1}^n, W_2, U_{2,k+1}^n, Y_1^{k-1}) - I(U_{1k}, U_{2k}; L_{1k}, L_{2k})\Big] \end{aligned}$$

$$
\begin{aligned}
&= \sum_{k=1}^{n}\Big[I(Y_k;W_1,U_{1,k+1}^n,W_2,U_{2,k+1}^n,Y_1^{k-1})-I(U_{1k},U_{2k};L_{1k},L_{2k})\Big]\\
&= \sum_{k=1}^{n}\Big[I(Y_k;L_{1k},L_{2k})-I(U_{1k},U_{2k};L_{1k},L_{2k})\Big], \qquad (7.26)
\end{aligned}
$$

which implies

$$
R_1+R_2 \le \frac{1}{n}\sum_{k=1}^{n}[I(L_{1k},L_{2k};Y_k)-I(U_{1k},U_{2k};L_{1k},L_{2k})]+\epsilon_n. \qquad (7.27)
$$

We next introduce a time-sharing RV to simplify the bounds (7.23), (7.24), and (7.27) using a single-letter characterization. Define a RV V with alphabet $\{1,2,...,n\}$ and distribution $P_V(v)=1/n$. We next introduce RVs U_1 and U_2 such that

$$
\Pr(U_1=u_1,U_2=u_2)=\Pr(U_{1k}=u_1,U_{2k}=u_2)=Q_{U_1U_2}(u_1,u_2)
$$

for all $(u_1,u_2)\in\mathcal{U}_1\times\mathcal{U}_2$, which are independent of V. Furthermore, we define new RVs L_1, L_2, X_1, X_2, and Y by

$$
\begin{aligned}
&\Pr(L_1=l_1,L_2=l_2,X_1=x_1,X_2=x_2,Y=y|V=k)\\
&\quad = \Pr(L_{1k}=l_1,L_{2k}=l_2,X_{1k}=x_1,X_{2k}=x_2,Y_k=y)
\end{aligned}
$$

for all $(l_1,l_2,x_1,x_2,y)\in\mathcal{L}_1\times\mathcal{L}_2\times\mathcal{X}_1\times\mathcal{X}_2\times\mathcal{Y}$. It follows that

$$
\begin{aligned}
&\frac{1}{n}\sum_{k=1}^{n}[I(L_{1k};L_{2k},Y_k)-I(U_{1k};L_{1k})]\\
&= I(L_1;L_2,Y|V)-I(U_1;L_1|V)\\
&= H(L_1|V)-H(L_1|L_2,Y,V)-H(U_1|V)+H(U_1|L_1,V)\\
&\overset{(a)}{\le} H(L_1)-H(L_1|L_2,Y,V)-H(U_1)+H(U_1|L_1,V)\\
&= I(L_1;L_2,Y,V)-I(U_1;L_1,V)\\
&\le I(L_1,V;L_2,Y,V)-I(U_1;L_1,V)\\
&\overset{(b)}{=} I(T_1;T_2,Y)-I(T_1;U_1)
\end{aligned}
$$

where (a) holds since conditioning reduces entropy and U_1 is independent of V, and in (b) $T_1 \triangleq (L_1, V)$ and $T_2 \triangleq (L_2, V)$. This shows that

$$R_1 \leq I(T_1; T_2, Y) - I(T_1; U_1) + \epsilon_n. \tag{7.28}$$

By a similar argument, we can show

$$R_2 \leq I(T_2; T_1, Y) - I(T_2; U_2) + \epsilon_n \tag{7.29}$$

and

$$R_1 + R_2 \leq I(T_1, T_2; Y) - I(U_1, U_2; T_1, T_2) + \epsilon_n. \tag{7.30}$$

For such RVs $(U_1, U_2, T_1, T_2, X_1, X_2, Y)$, it can be readily seen that the Markov chain relationship $(U_1, U_2, T_1, T_2) \to (X_1, X_2) \to Y$ holds. In fact,

$$\begin{aligned}
&\Pr(Y = y | U_1 = u_1, U_2 = u_2, T_1 = t_1 = (l_1, k), T_2 = t_2 = (l_2, k), X_1 = x_1, X_2 = x_2) \\
&= \Pr(Y = y | U_1 = u_1, U_2 = u_2, L_1 = l_1, L_2 = l_2, X_1 = x_1, X_2 = x_2, V = k) \\
&= \Pr(Y_k = y | U_{1k} = u_1, U_{2k} = u_2, L_{1k} = l_1, L_{2k} = l_2, X_{1k} = x_1, X_{2k} = x_2) \\
&= \Pr(Y_k = y | X_{1k} = x_1, X_{2k} = x_2) \\
&= W_{Y|X_1X_2}(y | x_1, x_2).
\end{aligned}$$

Next we bound the distortions $\mathbb{E}[d_i(U_i, X_i)]$. Since (R_1, R_2) is achievable under the sequence of codes $(f_1^{(n)}, f_2^{(n)}, \psi^{(n)})$, this implies that for any $\delta > 0$ and all n large enough, we have

$$\begin{aligned}
D_i + \delta &\geq \frac{1}{n} \frac{1}{2^{nR_i}} \sum_{w_i=1}^{M_i} \sum_{\mathcal{U}_i^n} Q_{U_i}^{(n)}(\mathbf{u}_i) d_i \left(\mathbf{u}_i, f_i^{(n)}(w_i, \mathbf{u}_i)\right) \\
&= \frac{1}{n} \sum_{\mathcal{U}_i^n \times \mathcal{X}_i^n} \Pr(U_i^n = \mathbf{u}_i, X_i^n = \mathbf{x}_i) d_i(\mathbf{u}_i, \mathbf{x}_i) \\
&= \frac{1}{n} \sum_{k=1}^{n} \sum_{\mathcal{U}_i^n \times \mathcal{X}_i^n} \Pr(U_i^n = \mathbf{u}_i, X_i^n = \mathbf{x}_i) d_i(u_{ik}, x_{ik})
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{k=1}^{n} P_V(V=k) \sum_{\mathcal{U}_i \times \mathcal{X}_i} \Pr(U_{ik}=u_{ik}, X_{ik}=x_{ik}) d_i(u_{ik}, x_{ik}) \\
&= \sum_{k=1}^{n} P_V(V=k) \sum_{\mathcal{U}_i \times \mathcal{X}_i} \Pr(U_i=u_i, X_i=x_i|V=k) d_i(u_i, x_i) \\
&= \sum_{k=1}^{n} \sum_{\mathcal{U}_i \times \mathcal{X}_i} \Pr(U_i=u_i, X_i=x_i, V=k) d_i(u_i, x_i) \\
&= \sum_{\mathcal{U}_i \times \mathcal{X}_i} P_{U_i X_i}(u_i, x_i) d_i(u_i, x_i).
\end{aligned}
$$

Thus we obtained that $\mathbb{E}[d_i(U_i, X_i)] \leq D_i + \delta$ for $i = 1, 2$. Combined with (7.28)–(7.30) and recalling that $\lim_{n\to\infty} \epsilon_n = 0$ and that $\mathcal{R}(D_1, D_2)$ is closed, we conclude that $\mathcal{R}(D_1, D_2) \subset \mathcal{R}_{out}(D_1 + \delta, D_2 + \delta)$ as claimed. □

7.3.3 Proof of Theorem 7.3

The forward part (achievability) is a consequence of Theorem 7.1 since (U_1, T_1) and (U_2, T_2) are independent and hence $I(T_1; T_2, Y) = I(T_1; Y|T_2)$, $I(T_2; T_1, Y) = I(T_2; Y|T_1)$, and $I(U_1, U_2; T_1, T_2) = I(U_1; T_1) + I(U_2; T_2)$. To prove the converse part, we need to sharpen the bounds in the last proof. We start from

$$
\begin{aligned}
& I(W_1; Y^n) - I(W_1; U_1^n) \\
&= \sum_{k=1}^{n} \Big[I(Y_k; W_1, U_{1,k+1}^n | Y_1^{k-1}) - I(U_{1k}; W_1, Y_1^{k-1}, U_{1,k+1}^n) \Big] \\
&= \sum_{k=1}^{n} \Big[H(W_1, U_{1,k+1}^n | Y_1^{k-1}) - H(W_1, U_{1,k+1}^n | Y_1^{k-1}, Y_k) - I(U_{1k}; W_1, Y_1^{k-1}, U_{1,k+1}^n) \Big] \\
&\overset{(a)}{=} \sum_{k=1}^{n} \Big[H(W_1, U_{1,k+1}^n | W_2, U_{2,k+1}^n, Y_1^{k-1}) - H(W_1, U_{1,k+1}^n | W_2, U_{2,k+1}^n, Y_1^{k-1}, Y_k) \\
&\qquad - I(U_{1k}; W_1, Y_1^{k-1}, U_{1,k+1}^n) \Big] \\
&= \sum_{k=1}^{n} \Big[I(W_1, U_{1,k+1}^n; Y_k | W_2, U_{2,k+1}^n, Y_1^{k-1}) - I(U_{1k}; W_1, Y_1^{k-1}, U_{1,k+1}^n) \Big]
\end{aligned}
$$

$$
\begin{aligned}
&\leq \sum_{k=1}^{n}\Big[I(W_1,U_{1,k+1}^n,Y_1^{k-1};Y_k|W_2,U_{2,k+1}^n,Y_1^{k-1})-I(U_{1k};W_1,Y_1^{k-1},U_{1,k+1}^n)\Big]\\
&= \sum_{k=1}^{n}\Big[I(L_{1k};Y_k|L_{2k})-I(U_{1k};L_{1k})\Big]
\end{aligned}
$$

where (a) follows since $(W_1,U_{1,k+1}^n)$ is now independent of $(W_2,U_{2,k+1}^n)$, and in the last equality we still let $L_{1k}\triangleq(W_1,Y_1^{k-1},U_{1,k+1}^n)$ and $L_{2k}\triangleq(W_2,Y_1^{k-1},U_{2,k+1}^n)$. Thus, using Fano's inequality we have

$$R_1\leq\frac{1}{n}\sum_{k=1}^{n}[I(L_{1k};Y_k|L_{2k})-I(U_{1k};L_{1k})]+\epsilon_n.$$

Similarly we can obtain

$$R_2\leq\frac{1}{n}\sum_{k=1}^{n}[I(L_{2k};Y_k|L_{1k})-I(U_{2k};L_{2k})]+\epsilon_n.$$

To bound the sum of the rates, we have

$$
\begin{aligned}
n(R_1+R_2) &= H(W_1,W_2)=I(W_1,W_2;Y^n)+H(W_1,W_2|Y^n)\\
&\leq I(W_1,W_2;Y^n)-I(W_1;U_1^n)-I(W_2;U_2^n)+n\epsilon_n \qquad (7.31)
\end{aligned}
$$

and

$$
\begin{aligned}
&I(W_1,W_2;Y^n)-I(W_1;U_1^n)-I(W_2;U_2^n)\\
&= \sum_{k=1}^{n}\Big[I(W_1;Y_k|Y_1^{k-1})+I(W_2;Y_k|W_1,Y_1^{k-1})-I(W_1;U_{1k}|U_{1,k+1}^n)-I(W_2;U_{2k}|U_{2,k+1}^n)\Big]\\
&= \sum_{k=1}^{n}\Big[H(Y_k|Y_1^{k-1})-H(Y_k|W_1,Y_1^{k-1},U_{1,k+1}^n)-I(Y_k;U_{1,k+1}^n|W_1,Y_1^{k-1})\\
&\quad +H(Y_k|W_1,Y_1^{k-1})-H(Y_k|W_1,W_2,Y_1^{k-1},U_{2,k+1}^n)-I(Y_k;U_{2,k+1}^n|W_1,W_2,Y_1^{k-1})\\
&\quad -H(U_{1k}|U_{1,k+1}^n)+H(U_{1k}|W_1,U_{1,k+1}^n)-H(U_{2k}|U_{2,k+1}^n)+H(U_{2k}|W_2,U_{2,k+1}^n)\Big]\\
&= \sum_{k=1}^{n}\Big[H(Y_k|Y_1^{k-1})-H(Y_k|W_1,Y_1^{k-1},U_{1,k+1}^n)-I(U_{1k};Y_1^{k-1}|W_1,U_{1,k+1}^n)\\
&\quad +H(Y_k|W_1,Y_1^{k-1})-H(Y_k|W_1,W_2,Y_1^{k-1},U_{2,k+1}^n)-I(U_{2k};Y_1^{k-1}|W_1,W_2,U_{2,k+1}^n)
\end{aligned}
$$

$$
\begin{aligned}
&\quad -H(U_{1k}) + H(U_{1k}|W_1, U_{1,k+1}^n) - H(U_{2k}) + H(U_{2k}|W_1, W_2, U_{2,k+1}^n)\Big] \\
&= \sum_{k=1}^{n} \Big[H(Y_k|Y_1^{k-1}) - H(Y_k|W_1, Y_1^{k-1}, U_{1,k+1}^n) \\
&\quad + H(Y_k|W_1, Y_1^{k-1}) - H(Y_k|W_1, W_2, Y_1^{k-1}, U_{2,k+1}^n) \\
&\quad - H(U_{1k}) + H(U_{1k}|W_1, U_{1,k+1}^n, Y_1^{k-1}) - H(U_{2k}) + H(U_{2k}|W_1, W_2, U_{2,k+1}^n, Y_1^{k-1})\Big] \\
&= \sum_{k=1}^{n} \Big[I(Y_k; W_1, U_{1,k+1}^n|Y_1^{k-1}) + I(Y_k; W_2, U_{2,k+1}^n|W_1, Y_1^{k-1}) \\
&\quad - I(U_{1k}; W_1, U_{1,k+1}^n, Y_1^{k-1}) - I(U_{2k}; W_2, U_{2,k+1}^n, Y_1^{k-1})\Big] \\
&= \sum_{k=1}^{n} \Big[H(W_1, U_{1,k+1}^n|Y_1^{k-1}) - H(W_1, U_{1,k+1}^n|Y_1^{k-1}, Y_k) \\
&\quad + H(W_2, U_{2,k+1}^n|W_1, Y_1^{k-1}) - H(W_2, U_{2,k+1}^n|W_1, Y_1^{k-1}, Y_k) \\
&\quad - I(U_{1k}; W_1, U_{1,k+1}^n, Y_1^{k-1}) - I(U_{2k}; W_2, U_{2,k+1}^n, Y_1^{k-1})\Big] \\
&\overset{(a)}{=} \sum_{k=1}^{n} \Big[H(W_1, U_{1,k+1}^n|Y_1^{k-1}) - H(W_1, U_{1,k+1}^n|Y_1^{k-1}, Y_k) \\
&\quad + H(W_2, U_{2,k+1}^n|W_1, U_{1,k+1}^n, Y_1^{k-1}) - H(W_2, U_{2,k+1}^n|W_1, U_{1,k+1}^n, Y_1^{k-1}, Y_k) \\
&\quad - I(U_{1k}; W_1, U_{1,k+1}^n, Y_1^{k-1}) - I(U_{2k}; W_2, U_{2,k+1}^n, Y_1^{k-1})\Big] \\
&= \sum_{k=1}^{n} \Big[H(W_1, U_{1,k+1}^n, W_2, U_{2,k+1}^n|Y_1^{k-1}) - H(W_1, U_{1,k+1}^n, W_2, U_{2,k+1}^n|Y_1^{k-1}, Y_k) \\
&\quad - I(U_{1k}; W_1, U_{1,k+1}^n, Y_1^{k-1}) - I(U_{2k}; W_2, U_{2,k+1}^n, Y_1^{k-1})\Big] \\
&= \sum_{k=1}^{n} \Big[I(W_1, U_{1,k+1}^n, W_2, U_{2,k+1}^n; Y_k|Y_1^{k-1}) - I(U_{1k}; W_1, U_{1,k+1}^n, Y_1^{k-1}) \\
&\quad - I(U_{2k}; W_2, U_{2,k+1}^n, Y_1^{k-1})\Big] \\
&\leq \sum_{k=1}^{n} \Big[H(Y_k) - H(Y_k|W_1, U_{1,k+1}^n, W_2, U_{2,k+1}^n, Y_1^{k-1}) - I(U_{1k}; W_1, U_{1,k+1}^n, Y_1^{k-1}) \\
&\quad - I(U_{2k}; W_2, U_{2,k+1}^n, Y_1^{k-1})\Big] \\
&= \sum_{k=1}^{n} \Big[I(W_1, U_{1,k+1}^n, W_2, U_{2,k+1}^n, Y_1^{k-1}; Y_k) - I(U_{1k}; W_1, U_{1,k+1}^n, Y_1^{k-1})
\end{aligned}
$$

$$
\begin{aligned}
&\quad -I(U_{2k};W_2,U_{2,k+1}^n,Y_1^{k-1})\Big] \\
&= \sum_{k=1}^{n}\Big[I(L_{1k},L_{2k};Y_k)-I(U_{1k};L_{1k})-I(U_{2k};L_{2k})\Big]
\end{aligned}
$$

where (a) holds since $(W_1, U_{1,k+1}^n)$ is independent of $(W_2, U_{2,k+1}^n)$ and $L_{1k} \triangleq (W_1, Y_1^{k-1}, U_{1,k+1}^n)$ and $L_{2k} \triangleq (W_2, Y_1^{k-1}, U_{2,k+1}^n)$ in the last equality. The above implies

$$R_1 + R_2 \leq \frac{1}{n}\sum_{k=1}^{n}[I(L_{1k},L_{2k};Y_k)-I(U_{1k};L_{1k})-I(U_{2k};L_{2k})]+\epsilon_n.$$

The rest of the proof proceeds the same way as the proof of Theorem 7.2. □

7.3.4 Upper Bounds on $|\mathcal{T}_i|$ for $\mathcal{R}_{in}^*(D_1,D_2)$ and $\mathcal{R}_{in}(D_1,D_2)$

We only bound the cardinality of $\mathcal{T}_1$ and $\mathcal{T}_2$ for the region $\mathcal{R}_{in}(D_1,D_2)$. The bounds for $|\mathcal{T}_1|$ and $|\mathcal{T}_2|$ for the region $\mathcal{R}_{in}^*(D_1,D_2)$ can be conducted in a similar manner. We will need the support lemma (Lemma 5.6), which is based on Carathéodory's theorem.

Using this lemma, we will show that for any given $P_{X_1T_1|U_1}$ and $P_{X_2T_2|U_2}$, there exists a RV $\widehat{T}_1$ with $|\widehat{\mathcal{T}}_1| \leq |\mathcal{U}_1||\mathcal{X}_1|+1$ only depending on U_1 and X_1 such that the following quantities keep invariant,

$$
\begin{aligned}
I(\widehat{T}_1;Y|T_2)-I(U_1;\widehat{T}_1) &= I(T_1;Y|T_2)-I(U_1;T_1) && (7.32)\\
I(T_2;Y|\widehat{T}_1)-I(U_2;T_2) &= I(T_2;Y|T_1)-I(U_2;T_2) && (7.33)\\
I(\widehat{T}_1,T_2;Y)-I(U_1;\widehat{T}_1)-I(U_2;T_2) &= I(T_1,T_2;Y)-I(U_1;T_1)-I(U_2;T_2), && (7.34)
\end{aligned}
$$

and that the expectation of the distortion between U_1 and X_1 is preserved when T_1 is replaced by $\widehat{T}_1$. Note that the upper bound on $|\widehat{\mathcal{T}}_1|$ does not depend on $|\mathcal{T}_2|$.

We first rewrite

$$
\begin{aligned}
I(T_1;Y|T_2)-I(U_1;T_1) &= H(Y|T_2)-H(Y|T_1,T_2)-H(U_1)+H(U_1|T_1),\\
I(T_2;Y|T_1)-I(U_2;T_2) &= H(Y|T_1)-H(Y|T_1,T_2)-I(U_2;T_2),
\end{aligned}
$$

and

$$I(T_1,T_2;Y)-I(U_1;T_1)-I(U_2;T_2)=H(Y)-H(Y|T_1,T_2)-H(U_1)+H(U_1|T_1)-I(U_2;T_2).$$

Recall that the joint distribution of $(U_1,U_2,T_2,T_2,X_1,X_2,Y)$ can be factorized as

$$P_{U_1T_1U_2T_2X_1X_2Y}=Q_{U_1U_2}P_{T_1X_1|U_1}P_{T_2X_2|U_2}W_{Y|X_1X_2}.$$

We note that there exists a Markov chain $(T_1,X_1)\rightarrow U_1\rightarrow U_2\rightarrow(T_2,X_2)$. Writing

$$P_{U_1T_1U_2T_2X_1X_2Y}=P_{T_1}P_{U_1X_1|T_1}P_{U_2|U_1}P_{T_2X_2|U_2}W_{Y|X_1X_2},$$

and noting that $P_{U_2|U_1}$, $P_{T_2X_2|U_2}$ and $W_{Y|X_1X_2}$ are fixed, to apply the support lemma, we need $m-1$ functions to preserve the joint distribution of (U_1,X_1) (see (7.35) below), where $m\triangleq|\mathcal{U}_1||\mathcal{X}_1|$. Specifically, we define the following real-valued continuous functions of distribution $P_{U_1X_1|T_1}(\cdot,\cdot|t_1)$ on $\mathcal{U}_1\times\mathcal{X}_1$ for fixed $t_1\in\mathcal{T}_1$,

$$f_{u_1,x_1}(P_{U_1X_1|T_1}(\cdot,\cdot|t_1))\triangleq P_{U_1X_1|T_1}(u_1,x_1|t_1)$$

for all $(u_1,x_1)\in\mathcal{U}_1\times\mathcal{X}_1$ except one pair (u_1,x_1). Furthermore, we define real-valued continuous functions

$$\begin{aligned} f_m(P_{U_1X_1|T_1}(\cdot,\cdot|t_1)) &\triangleq -H_P(Y|T_1=t_1,T_2)+H_P(U_1|T_1=t_1),\\ f_{m+1}(P_{U_1X_1|T_1}(\cdot,\cdot|t_1)) &\triangleq H_P(Y|T_1=t_1)-H_P(Y|T_1=t_1,T_2),\end{aligned}$$

where the entropies are taken under the joint distribution induced by $P_{U_1X_1|T_1}(\cdot,\cdot|t_1)$. According to the support lemma, there must exist a new RV $\widehat{T_1}$ (jointly distributed with (U_1,X_1)) with alphabet size $|\widehat{\mathcal{T}_1}|=m+1=|\mathcal{U}_1||\mathcal{X}_1|+1$ such that the expectation of f_i, $i=1,2,...,m+1$, with respect to P_{T_1} can be expressed in terms of the convex combination of $m+1$ points, i.e.,

$$P_{U_1X_1}(u_1,x_1) = \sum_{t_1\in\mathcal{T}_1}P_{T_1}(t_1)f_{u_1,x_1}(P_{U_1X_1|T_1}(\cdot,\cdot|t_1))$$

$$= \sum_{\widehat{t}_1 \in \widehat{\mathcal{T}}_1} P_{\widehat{T}_1}(\widehat{t}_1) f_{u_1,x_1}(P_{U_1X_1|\widehat{T}_1}(\cdot,\cdot|\widehat{t}_1)), \tag{7.35}$$

$$\begin{aligned} -H(Y|T_1,T_2) + H(U_1|T_1) &= \sum_{t_1 \in \mathcal{T}_1} P_{T_1}(t_1) f_m(P_{U_1X_1|T_1}(\cdot,\cdot|t_1)) \\ &= \sum_{\widehat{t}_1 \in \widehat{\mathcal{T}}_1} P_{\widehat{T}_1}(\widehat{t}_1) f_m\left(P_{U_1X_1|\widehat{T}_1}(\cdot,\cdot|\widehat{t}_1)\right) \\ &= -H(Y|\widehat{T}_1,T_2) + H(U_1|\widehat{T}_1), \end{aligned}$$

$$\begin{aligned} H(Y|T_1) - H(Y|T_1,T_2) &= \sum_{t_1 \in \mathcal{T}_1} P_{T_1}(t_1) f_{m+1}(P_{U_1X_1|T_1}(\cdot,\cdot|t_1)) \\ &= \sum_{\widehat{t}_1 \in \widehat{\mathcal{T}}_1} P_{\widehat{T}_1}(\widehat{t}_1) f_{m+1}(P_{U_1X_1|\widehat{T}_1}(\cdot,\cdot|\widehat{t}_1)) \\ &= H(Y|\widehat{T}_1) - H(Y|\widehat{T}_1,T_2). \end{aligned}$$

This implies that (7.32)–(7.34) hold. It should be point out that this RV $\widehat{T}_1$ maintains the prescribed distortion level, since $P_{U_1X_1}(u_1,x_1)$ is preserved. Similarly, for any given $P_{X_1T_1|U_1}$ and $P_{X_2T_2|U_2}$, we can show that there exists a RV $\widehat{T}_2$ with $|\widehat{\mathcal{T}}_2| \leq |\mathcal{U}_2||\mathcal{X}_2| + 1$ only depending on U_2 and X_2 such that

$$\begin{aligned} I(T_1;Y|\widehat{T}_2) - I(U_1;T_1) &= I(T_1;Y|T_2) - I(U_1;T_1) && (7.36) \\ I(\widehat{T}_2;Y|T_1) - I(U_2;\widehat{T}_2) &= I(T_2;Y|T_1) - I(U_2;T_2) && (7.37) \\ I(T_1,\widehat{T}_2;Y) - I(U_1;T_1) - I(U_2;\widehat{T}_2) &= I(T_1,T_2;Y) - I(U_1;T_1) - I(U_2;T_2), && (7.38) \end{aligned}$$

and the distortion constraint between U_2 and X_2 is preserved. Thus we conclude that the cardinality of $\mathcal{T}_i$ can be bounded by $|\mathcal{U}_i||\mathcal{X}_i| + 1$, $i = 1, 2$.

Finally, we remark that the support lemma cannot be used to bound the cardinality for $\mathcal{T}_1$ and $\mathcal{T}_2$ for the region $\mathcal{R}_{out}(D_1,D_2)$ and $\mathcal{R}^*_{out}(D_1,D_2)$. For example, to bound the cardinality of $\mathcal{T}_1$ for $\mathcal{R}_{out}(D_1,D_2)$, we need $|\mathcal{U}_1||\mathcal{U}_2||\mathcal{X}_1||\mathcal{X}_2||\mathcal{T}_2| - 1$ real-valued continuous functions to preserve the joint distribution of (U_1,U_2,T_2,X_1,X_2). Therefore, we may need $|\mathcal{U}_1||\mathcal{U}_2||\mathcal{X}_1||\mathcal{X}_2||\mathcal{T}_2| + 1$ letters and this upper bound depends on $|\mathcal{T}_2|$. □

7.4 Conclusions

In this chapter, we studied the public multi-user information embedding system for which two secret information messages are independently embedded into two correlated host sources and are transmitted through a multiple-access attack channel. The tradeoff between the achievable embedding rates and the average distortions for the two embedders is investigated. For given distortion levels, an inner bound and outer bound for the embedding capacity region for the public two-user information embedding system are obtained with single-letter characterization. The bounds are next sharpened when the host sources are independent.

Chapter 8

Summary and Conclusions

In this dissertation, we studied a hybrid digital-analog source-channel coding system for the transmission of a discrete-time memoryless Gaussian source over a discrete-time memoryless Gaussian channel under bandwidth compression. We designed an image communication system based on hybrid digital-analog coding. We then studied the error exponent performance for a joint compression and private information hiding system with Gaussian source-channel pairs under the single-user setting. We next investigated information hiding problems over memoryless multiple access attack channels for multi-user applications. Both private and public information hiding systems are studied. The contributions of this dissertation are summarized in the following.

- For the HDA system, we established information-theoretic upper bounds (under both matched and mismatched channel conditions) on the asymptotically optimal mean squared error distortion. We derived a power allocation scheme for distributing the channel input power between the analog and the digital signals for the mismatched HDA system. We proposed and implemented a low-complexity and low-delay version of the system. An iterative optimization algorithm was next designed and numerical results were presented, which show that the proposed HDA system performs within 0.3 dB of the mismatch distortion upper bound. We designed an image communication system, which combines the proposed bandwidth compression system with the bandwidth expansion system of Skoglund *et al.* [69]. We compared our system to other schemes. Numerical results show that the proposed HDA image coding scheme is robust and superior to purely analog and purely digital systems for a wide range of CSNRs. One direction for future work may include the optimization of the power allocation when the CSNR is governed by a given distribution.

- For a single-user joint compression and private information hiding system with Gaussian source-channel pairs, we derived a random coding error exponent. Our proof methods incorporate a Gallager-type random coding technique, properties of stationary memoryless Gaussian sources and techniques to derive the exponent for Gaussian sources. Numerical results were proposed which show that the error exponent is positive within almost the entire achievable region derived by Karakos and Papamarcou [33].

- We extended the single-user joint compression and private information hiding system to a multi-user setting. In particular, for the case of embedding two independent secret messages into two correlated DMS's and transmitting over MAC, we derived an inner bound and outer bound with single-letter characterization for the achievable compression and watermarking rate region with respect to the distortion levels. Next, we studied a multi-user information hiding model for the transmission of two correlated secret sources over memoryless multiple attack access channel with common host data. We derived a sufficient condition with single-letter characterization for hiding correlated sources against MAC attacks. We also presented an uncomputable (and somewhat trivial) outer bound (converse condition) by applying Fano's inequality in terms of a sequence of n-dimensional joint distributions.

- For the multi-user public information hiding problem, we established inner and outer bounds with single-letter characterization for the embedding capacity region of the two-user information embedding system. The outer bound follows from Fano's inequality and standard information-theoretical bounding arguments. The bounds are tightened in the case that the host sources are independent.

A number of possibly difficult problems have remained open. For example, in Chapter 4, the random coding error exponent was shown to be positive only on a (large) subset of the whole achievable rate region. This might be due to the inefficiency of the proposed bounding techniques, or it could be an inherent trait of the joint compression and watermarking problem. In Chapter 5, we did not provide the conditions where the inner bound and the outer bound are tight. It is still an open problem whether or not a tight bound exists. In Chapter 6, we only have a single-letter sufficient condition for the proposed information hiding problem. The converse condition is in terms of an "infinite dimensiona" characterization, which is uncomputable. A single-letter converse condition seems very difficult for this multi-user joint source-channel coding over a MAC problem. In Chapter 7, we also do not have conditions under which the inner and outer bounds are tight. Furthermore, for the outer bounds, we do not have a bound on the cardinality of the auxiliary RVs T_1 and T_2. Finally, we remark that multi-user information hiding problems tend to inherit the difficulty inherent in multi-user (joint) source/channel coding, where a large number of problems are still open.

Bibliography

[1] R. Ahlswede and T. S. Han, "On source coding with side information via a multiple-access channel and related probalems in multi-user information theory," *IEEE Trans. Inform. Theory*, vol. 29, no. 3, May 1983.

[2] F. Alajaji, N. Phamdo, and T. Fuja, "Channel codes that exploit the residual redundancy in CELP-encoded speech," *IEEE Trans. Speech Audio Processing*, vol. 4, pp. 325-336, Sep. 1996.

[3] R. J. Barron, B. Chen and G. W. Wornell, "The duality between information embedding and source coding with side information and some applications," *IEEE Trans. Inform. Theory*, vol. 49, no. 5, May 2003.

[4] T. Berger, "*Rate Distortion Theory: A Mathematical Basis for Data Compression*," Prentice-Hall, New Jersey, 1971.

[5] B. Chen and G. W. Wornell, "Quantization index modulation: a class of provably good methods for digital watermarking and information embedding," *IEEE Trans. Inform. Theory*, vol. 47, no. 4, pp. 1423-1443, May 2001.

[6] J. Chen, C. Tian, T. Berger, and S. S. Hemami, " Multiple description quantization via Gram-Schmidt orthogonalization," *IEEE Trans. Inform. Theory*, vol. 52, no. 12, pp. 5197-5217, Dec. 2006.

[7] A. S. Cohen and A. Lapidoth, "The Gaussian watermarking game," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1639-1667, Jun. 2002.

[8] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. Inform. Theory*, vol. 29, pp. 439-441, May 1983.

[9] T. M. Cover, A. E. Gamal and M. Salehi, "Multiple access channels with arbitrarily correlated sources," *IEEE Trans. Inform. Theory*, vol. 26, no. 6, Nov. 1980.

[10] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd Edition, Wiley, 2006.

[11] H. Coward, *Joint Source-Channel Coding: Development of Methods and Utilization in Image Communications*, Ph.D thesis, Norwegian University of Science and Engineering (NTNU), 2001.

[12] H. Coward and T. A. Ramstad, "Robust image communication using bandwidth reducing and expanding mappings," in *Proc. 34th Asilomar Conf.*, Pacific Grove, USA, pp. 1384-1388, Oct.-Nov. 2000.

[13] I. J. Cox, M. L. Miller and A. .L. McKellips, "Watermarking as communications with side information," *Proc. IEEE*, vol. 87, no. 7, pp. 1127-1141, Jul. 1999.

[14] I. Csiszár and J. Körner, "Broadcast channels with confidential messages," *IEEE Trans. Inform. Theory*, vol. 24, pp. 339–348, May 1978.

[15] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems.* New York: Academic, 1981.

[16] I. Daubechies and W. Sweldens, "Factoring wavelet transforms into lifting steps," [Technical Report] USA, Princeton University, 1996.

[17] U. Erez, S. Shamai, and R. Zamir, "Capacity and lattice strategies for canceling known interference," *IEEE Trans. Inform. Theory*, vol. 51, no. 11, pp. 3820-3833, Nov. 2005.

[18] U. Erez and R. Zamir, "Achieving $\frac{1}{2}\log(1+\mathrm{SNR})$ on the AWGN channel with lattice encoding and decoding," *IEEE Trans. Inform. Theory*, vol. 50, no. 10, pp. 2293-2314, Oct. 2004.

[19] J. M. Ettinger, "Steganalysis and game equilibria," in *Proc. 1998 Workshop on Information Hiding (Lecture Notes in Computer Sciences)*, Berlin, Germany, Springer-Verlag, 1998.

[20] N. Farvardin, "A study of vector quantization for noisy channels," *IEEE Trans. Information Theory*, vol. 36, pp. 799-809, Jul. 1990.

[21] N. Farvardin and V. Vaishampayan, "On the performance and complexity of channel-optimized vector quantizers," *IEEE Trans. Inform. Theory*, vol. 37, pp. 155-159, Jan. 1991.

[22] G. Fernandez, S. Periaswamy, and W. Sweldens, "LIFTPACK: a software package for wavelet transforms using lifting," in *Proc. SPIE 2825, Wavelet Applications in Signal and Image Processing*, pp. 396-408, 1996.

[23] A. Fuldseth and T. A. Ramstad, "Bandwidth compression for continuous amplitude channels based on vector approximation to a continuous subset of the source signal space," in *Proc. IEEE ICASSP*, pp. 3093-3096, Apr. 1997.

[24] R.G. Gallager, *Information Theory and Reliable Communication*, New York: Wiley, 1968.

[25] S. Gelfand and M. Pinsker, “Coding for a channel with random parameters,” *Problems of Control and Information Theory*, vol. 9, pp. 19–31, 1980.

[26] T. S. Han and K. Kobayashi, “A unified achievable rate region for a general class of multiterminal source coding systems,” *IEEE Trans. Inform. Theory*, vol. 26, no. 3, pp. 396–412, May 1980.

[27] P. Hedelin and J. Skoglund, “Vector quantization based on Gaussian mixture models,” *IEEE Trans. Speech Audio Processing*, vol. 8, pp. 385-401, Jul. 2000.

[28] F. Hekland, “A review of joint source-channel coding,” Technical Report, Norwegian University of Science and Technology, Feb. 2004.

[29] F. Hekland, G. E. Øien, and T. A. Ramstad, “Using 2:1 Shannon mapping for joint source-channel coding,” in *Proc. IEEE Data Compression Conf.*, pp. 223-232, Mar. 2005.

[30] S. Ihara and M. Kubo, “Error exponent for coding of memoryless Gaussian sources with a fidelty criterion,” *IEICE Trans. Fundamentals*, vol. E83-A, no. 10, Oct. 2000.

[31] D. Karakos and A. Papamarcou, “A relationship between quantization and distribution rates of digitally watermarked data,” in *Proc. IEEE Int. Symp. Inform. Theory*, Sorrento, Italy, pp. 47, Jun. 2000.

[32] D. Karakos and A. Papamarcou, “Fingerprinting, watermarking and quantization of Gaussian data,” in *Proc. 39th Allerton Conf. Communication, Control and Computing (Invited Talk)*, Monticello, IL, Oct. 2001.

[33] D. Karakos and A. Papamarcou, “A relationship between quantization and watermarking rates in the presence of additive Gaussian attacks,” *IEEE Trans. Inform. Theory*, vol. 49, no. 8, pp. 1970-1982, Aug. 2003.

[34] D. Karakos, *Digital Watermarking, Fingerprinting and Compression: an Information-Theoretic Perspective*, Ph.D. thesis, University of Maryland, College Park, USA, 2002.

[35] S. Katzenbeisser and F. A. P. Petitcolas, *Information Hiding Techniques for Steganography and Digital Watermarking*, Artech House, 2000.

[36] S. Kotagiri and J. N. Laneman, “Reversible information embedding in multi-user channels,” *Proc. Allerton Conf. Communications, Control, and Computing*, Monticello, IL, Sep. 2005.

[37] I. Kozintsev and K. Ramchandran, “Hybrid compressed–uncompressed framework for wireless image transmission,” in *Proc. IEEE Int. Conf. Commun.*, Montreal, pp. 77-80, Jun. 1997.

[38] H. Kumazawa, M. Kasahara and T. Namekawa, “A construction of vector quantizers for noisy channels,” *Electronics and Engineering in Japan,* vol. 67-B, pp. 39-47, Jan. 1984.

[39] D. Kundur, “Implications for high capacity data hiding in the presence of lossy compression,” *Proc. IEEE Int. Conf. Inform. Tech.: Coding and Computing*, Las Vegas, NV, pp.16-21, Mar. 2000.

[40] A. Kurtenbach and P. Wintz, “Quantizing for noisy channels,” *IEEE Trans. Commun. Technol.*, vol. COM-17, pp. 291-302, Apr. 1969.

[41] A. Lapidoth, “Nearest neighbor decoding for additive non-Gaussian noise channel,” *IEEE Trans. Inform. Theory*, vol. 42, no. 5, pp. 1520-1529, Sep. 1996.

[42] A. Lapidoth, “On the role of mismatch in rate distortion theory,” *IEEE Trans. Inform. Theory*, vol. 43, pp. 38-47, Sep. 1997.

[43] J. M. Lervik, A. Grovlen, and T. A. Ramstad, “Robust digital signal compression and modulation exploiting the advantages of analog communications,” in *Proc. IEEE GLOBECOM*, pp. 1044-1048, Nov. 1995.

[44] J. Lim and D. L. Neuhoff, “Joint and tandem source-channel coding with complexity and delay constraints,” *IEEE Trans. Commun.*, vol. 51, pp. 757-766, May 2003.

[45] Y. Linde, A. Buzo, and R. M. Gray, “An algorithm for vector quantizer design,” *IEEE Trans. Commun.*, vol. 28, pp. 84-95, Dec. 1980.

[46] A. Maor and N. Merhav, “On joint information embedding and lossy compression,” *IEEE Trans. Inform. Theory*, vol. 51, no. 8, pp. 2998-3008, Aug. 2005.

[47] A. Maor and N. Merhav, “On joint information embedding and lossy compression in the presence of a memoryless attack,” *IEEE Trans. Inform. Theory*, vol. 51, no. 9, pp. 3166-3175, Sep. 2005.

[48] N. Merhav, “On random coding error exponent of watermarking systems,”, *IEEE Trans. Inform. Theory*, vol. 46, no. 2, Mar. 2000.

[49] N. Merhav and S. Shamai (Shitz), “On joint source-channel coding for the Wyner-Ziv source and the Gel’fand-Pinsker channel,” *IEEE Trans. Inform. Theory*, vol. 49, no. 11, pp. 2844-2855, Nov. 2003.

[50] N. Merhav and E. Ordentlich, "On causal and semicausal codes for joint information embedding and source coding," *IEEE Trans. Inform. Theory*, vol. 52, no. 1, pp. 213-226, Jan. 2006.

[51] P. Moulin and J. O'Sullivan, "Information-theoretic analysis of information hiding," *IEEE Trans. Inform. Theory*, vol. 49, no. 3, pp. 563-593, Mar. 2003.

[52] P. Moulin and M. K. Mihcak, "The parallel-Gaussian watermarking game," *IEEE Trans. Inform. Theory*, vol. 50, no. 2, Feb. 2004.

[53] U. Mittal and N. Phamdo, "Hybrid digital-analog (HDA) joint source-channel codes for broadcasting and robust communications," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1082-1102, May 2002.

[54] Y. Oohama, "Gaussian multiterminal source coding," *IEEE Trans. Inform. Theory*, vol. 43, no. 6, Nov. 1997.

[55] H. C. Papadopoulos and C. E. W. Sundberg, "Simultaneous broadcasting of analog FM and digital audio signals by means of adaptive precanceling techniques," *IEEE Trans. Commun.*, vol. 46, pp. 1233-1242, Sep. 1998.

[56] F. A. P. Petitcolas, R. J. Anderson and M. G. Kuhn, "Information hiding - a survey," *Proc. IEEE*, vol. 87, no. 7, pp. 1062-1078, Jul. 1999.

[57] N. Phamdo and F. Alajaji, "Soft-decision demodulation design for COVQ over white, colored, and ISI Gaussian channels," *IEEE Trans. Commun.*, vol. 48, no. 9, pp. 1499-1506, Sep. 2000.

[58] N. Phamdo and U. Mittal, "A joint source-channel speech coder using hybrid digital-analog (HDA) modulation," *IEEE Trans. Speech and Audio Processing*, vol. 10, no. 4, pp. 222-231, May 2002.

[59] M. Ramkumar and A. Akansu, "Theoretical capacity measures for data hiding in compressed images," *Proc. SPIE*, vol. 3528, pp. 482-492, Nov. 1998.

[60] T. A. Ramstad, "*Insights into mobile multidedia communications, chapter 26: Combined source coding and modulation for mobile multidedia communication*, pp. 415-430. Academic Press, 1st edition, 1999.

[61] Z. Reznic, M. Feder, and R. Zamir, "Distortion bounds for broadcasting with bandwidth expansion," *IEEE Trans. Inform. Theory*, vol. 52, no. 8, pp. 3778-3788, Aug. 2006.

[62] D. Sakrison, "A geometric treatment of the source encoding of a Gaussian random variable," *IEEE Trans. Inform. Theory*, vol. 14, no. 3, pp. 481-486, May 1968.

[63] S. Sesia, G. Caire, and G. Vivier, "Lossy transmission over slow-fading AWGN channels: a comparison of progressive, superposition and hybrid approaches," in *Proc. IEEE ISIT'05*, Adelaide, Australia, pp. 224-228, Sep. 2005.

[64] S. Shamai, S. Verdú, and R. Zamir, "Systematic lossy source/channel coding," *IEEE Trans. Inform. Theory*, vol. 44, no. 2, pp. 564-579, Mar. 1998.

[65] C. E. Shannon, "A mathematical theory of communication," *Bell Systems Technical Journal*, vol. 27, pp. 379-423 and 623-656, 1948.

[66] C. E. Shannon, "Channels with side information at the transmitter," *IBM J.*, pp. 289-293, Oct. 1958.

[67] P. H. Skinnemoen, *Robust Communication with Modulation Organized Vector Quantization*, Ph.D. thesis, Norwegian Institute of Technology, 1994.

[68] M. Skoglund, “Soft decoding for vector quantization over nosiy channels with memory,” *IEEE Trans.Information Theory*, vol. 45, pp. 1293-1307, May 1999.

[69] M. Skoglund, N. Phamdo and F. Alajaji, “Design and performance of VQ-based hybrid digital-analog joint source-channel codes,” *IEEE Trans. Inform. Theory*, vol. 48, no. 3, Mar. 2002.

[70] M. Skoglund, N. Phamdo, and F. Alajaji, “Hybrid digital-analog source-channel coding for bandwidth compression/expansion,” *IEEE Trans. Inform. Theory*, vol. 52, no. 8, pp. 3757-3763, Aug. 2006.

[71] D. Slepian and J. K. Wolf, “Noiseless coding of correlated information sources,” *IEEE Trans. Inform. Theory*, vol. 19, pp. 471-480, July 1973.

[72] D. Slepian and J. K. Wolf, “A coding theorem for multiple access channels with correlated sources,” *Bell Syst. Tech. Journal*, vol. 52, pp. 1037-1076, 1973.

[73] A. Somekh-Baruch and N. Merhav, “On the error exponent and capacity games of private watermarking systems,” *IEEE Trans. Inform. Theory*, vol. 49, no. 3, pp. 537-563, Mar. 2003.

[74] A. Somekh-Baruch and N. Merhav, “On the capacity game of public watermarking systems,” *IEEE Trans. Inform. Theory*, vol. 50, no. 3, pp. 511-524, Mar. 2004.

[75] A. Somekh-Baruch and N. Merhav, “On the capacity game of private fingerprinting systems under collusion attacks,” *IEEE Trans. Inform. Theory*, vol. 51, no. 3, pp. 884-899, March 2005.

[76] W. Sun and E. H. Yang, “On achievable regions of public multiple-access Gaussian watermarking systems,” *Proc. 6th Int. Information Hiding Workshop*, Toronto, Canada, May 23-25, 2004.

[77] W. Sun, *Joint Compression and Digital Watermarking: Information-Theoretic Study and Algorithms Development*, Ph.D thesis, University of Waterloo, Canada, 2006.

[78] N. Tanabe and N. Farvardin, "Subband image coding using entropy-coded quantization over noisy channels," *IEEE J. Select. Areas Commun.*, vol. 10, pp. 926-943, Jun. 1992.

[79] S. Y. Tung, *Multiterminal Source Coding*, Ph.D dissertation, School of Electrical Engineering, Cornell Univ., Ithaca, NY, May 1978.

[80] V. Vaishampayan, *Combined Source-Channel Coding for Bandlimited Waveform Channels*, Ph.D. thesis, University of Maryland, College Park, USA, 1989.

[81] V. Vaishampayan and S. I. R. Costa, "Curves on a sphere, shift-map dynamics, and error control for continuous alphabet sources," *IEEE Trans. Inform. Theory*, vol. 49, no. 7, pp. 1658-1672, Jul. 2003.

[82] S. Vembu, S. Verdu and Y. Steinberg, "The source-channel separation theorem revisited," *IEEE Trans. Inform. Theory*, vol. 41, pp. 44-54, Jan. 1995.

[83] Y. Wang, F. Alajaji, and T. Linder, "Design of VQ-based hybrid digital-analog joint source-channel codes for image communication," in *Proc. IEEE Data Compression Conf.*, Snowbird, Utah, USA, pp. 193-202, Mar. 2005.

[84] Y. Wang, F. Alajaji and T. Linder,"Hybrid digital-analog coding of memoryless Gaussian sources over AWGN channels with bandwidth compression," in *Proc. IEEE 23nd Biennial Symposium on Communications*, Queen's University, Kingston, Canada, May-June 2006.

[85] Y. Wang, F. Alajaji, and T. Linder, “Hybrid digital-analog coding for memoryless source-channel pairs with applications to image communication,” submitted to *IEEE Trans. Commun.*, May 2007.

[86] Y. Wang, F. Alajaji, and T. Linder, “A random coding error exponent for joint quantization and watermarking of Gaussian sources under memoryless Gaussian attacks,” *Proc. 10th Canadian Workshop on Information Theory* (CWIT’07), Edmonton, AB, Canada, June 6-8, 2007.

[87] Y. Wang, Y. Zhong, F. Alajaji, and T. Linder,“A sufficient condition for private information hiding of two correlated sources under multiple access attacks”, *IEEE Proc. Int. Symp. Inform. Theory* (ISIT’07), Nice, France, June 24-29, 2007.

[88] F. M.J. Willems, “An informationtheoretical approach to information embedding,” Proc. *21st Symposium on Information Theory*, pp. 255–260, Wassenaar, The Netherlands, May 25–26, 2000.

[89] E.-H. Yang and W. Sun, “On watermarking and compression rates of joint compression and private watermarking systems with abstract alphabets,” *Proc. of the 2005 Canadian Workshop on Information Theory*, Montreal, Quebec, Canada, pp. 296-299, Jun. 2005.

[90] S. Zahir, P. Duhamel, and O. Rioul, “Joint source-channel coding: panorama of methods,” *CNES Workshop on Data Compression*, Toulouse, France, Nov. 1996.

[91] R. Zamir and M. Feder, “On universal quantization by randomized uniform/lattice quantizers”, *IEEE Trans. Inform. Theory*, vol. 38, no. 2, pp. 428-436, Mar. 1992.

[92] K. A. Zeger and A. Gersho, “Vector quantizer design for memoryless noisy channels,” in *Proc. IEEE Int. Conf. Commun. (ICC)*, Philadephia, PA, pp. 1593-1597, 1988.

[93] K. A. Zeger and A. Gersho, “Pseudo-Gray coding,” *IEEE Trans. Communications*, vol. 38, pp. 2147-2158, Dec. 1990.