COMMUNICATION OF DEPENDENT MESSAGES OVER COMPOUND CHANNELS

Zhong Cheng

A Thesis submitted to the Faculty of Graduate Studies and Postdoctoral Studies in partial fulfillment of the requirements for the degree of Doctorate of Philosophy, Electrical Engineering

July 2015

Ottawa-Carleton Institute for Electrical and Computer Engineering
School of Electrical Engineering and Computer Science
University of Ottawa
Ottawa, Ontario, Canada

To my parents

Communication of Dependent Messages over Compound Channels

Doctorate of Philosophy, Electrical Engineering Thesis
School of Electrical Engineering and Computer Science
University of Ottawa

by Zhong Cheng
July 2015

Abstract

In the communication of multimedia content, certain dependency structure often exists among the source-coded messages by different source coding techniques, where by "dependency" we mean the dependent contributions of the messages to the overall reconstruction quality. Motivated by such notion of dependency, this thesis considers the problem of communicating dependent source-coded messages over compound channels, which include the attractive wireless channels and packet-loss channels. In this thesis we propose a novel framework to model arbitrary dependency structure among source-coded messages from the source-reconstruction perspectives, and formulate the problem of communicating such messages over compound channels as the problem of maximizing the average utility at the receiver. Over discrete memoryless channels (DMC), we derive the expression of maximal achievable utility, which appears to be governed by the channel coding theorem. Over degraded compound channels, we study analytically the maximal achievable utility by superposition codes. To achieve the maximal utility, the encoder chooses the best sub-chain in the utility graph and encodes it using the best superposition code. For the case of two source-coded messages, we show that the maximal utility achieved by superposition codes is the maximum among all coding schemes. Since in practice layered codes (LC) and multiple description codes (MDC) are two most attracted source coding schemes which induce different dependency structures among coded messages, we numerically evaluate the maximal achievable utility for sources coded with those two source coding schemes communicated over DMC and degraded compound channels respectively, and show the impact of communication delay and channel condition on their respective achievable utilities. In addition, for communicating a Gaussian $\mathcal{N}(0,1)$ independent identically distributed (i.i.d.) sequence over degraded compound channels, the joint source channel coding schemes are considered and the minimal achievable distortion is derived and compared for different combinations of source and channel codes. It is shown that the combination of LC and superposition codes outperforms other coding schemes. The comparative behaviour among those techniques is further demonstrated by the experimental results. Practically, we study the performances of various coding schemes for communicating two dependent messages over quasi-static Rayleigh fading channels, which include conventional channel codes, time sharing codes, and a low-density parity-check (LDPC) based coding scheme, termed Bi-LDPC codes. The success rates and the throughput of the considered coding schemes are compared. For communicating a Gaussian $\mathcal{N}(0,1)$ source sequence encoded by LC and MDC respectively over quasi-static Rayleigh fading channels, the average distortion is also compared among different channel coding schemes. It appears that Bi-LDPC codes, conventional channel codes encoding both messages, and time sharing type-2 (TS-2) codes, each has their advantage over different region of signal-to-noise ratio (SNR), compared with each other.

Acknowledgments

First and foremost I would like to express my deep and sincere appreciation to my supervisor, Professor Yongyi Mao, for his open-minded guidance, enthusiastic encouragement and deep understanding on my thesis work. His wide knowledge, profound insight, and logical thinking, especially the excellent mathematical talent have been of precious value for me. His supervising style, on the other hand, full of personality together with a social spirit, makes me enjoy my life of study very much.

Next, I wish to express my deep gratitude to Professor Terence Chan, for his generous sharing of the ideas on research, helpful suggestions, patient communications, and insightful comments on the manuscript. Without his help, I would not be able to fulfill my research work.

Additionally, the rest of our group members are all appreciated, from whom I benefit much through occasional discussions.

Finally, I have my great regard to my family. My loving gratitude goes to my wife Jie Tang, for her loving support both materially and spiritually. I also owe my warm thanks to my parents for their endless encouragement and stimulation. Last, but not least, I would like to thank everyone else who has helped me during my educational life.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation and contribution

Streaming multimedia information, such as video, audio, or image data, over wireless networks is becoming prevalent in the modern world. Wireless channels fundamentally belong to the class of compound channels [1], in which the channel realization is drawn at random from a distribution, and the channel realization is assumed to be unknown to the transmitter, but known to the receiver. Besides that, data transfer over Internet often suffers from unexpected packet loss, and thus data networks can be modelled as packet-loss channels or erasure channels, in which the packet is either successfully delivered to the receiver or lost mainly due to network congestion. In this scenario, the packet is considered to be dropped at a random manner and Internet is essentially deemed as a category of compound channels. Partially motivated by wireless and data network applications, there has been extensive research effort towards developing communication techniques for compound channels [2–15]. In the context of communicating multimedia content, the communication problem is compounded with the complexity of efficiently representing the source in order to accommodate various levels of media fidelity demanded by the heterogeneous receiving devices. To that end, various source coding techniques are employed to compress the multimedia data, where layered coding (LC) [16–30] and multiple description coding (MDC) [31–42] methodologies are the most typical and attractive. One observation made in this thesis is that different source coding techniques induce different "dependency" structures among the source-coded messages. For example, in LC, the source is represented by packets of multiple layers, where the lower-layer packets contain low-fidelity information of the source and the higher-layer packets contain enhancement information; higher layers serve to improve the fidelity in source reconstruction but are only useful when the lower layers have been received and decoded correctly. In MDC, on the other hand, the source is coded into a set of "equal-footing" packets, and every packet serves to improve the overall fidelity in source reconstruction, regardless of whether other packets are decoded. It is clear that LC and MDC induce different dependency structures among the coded packets. However, such a notion of source message dependency, "reconstructive" rather than probabilistic, has hardly been rigorously formulated or carefully examined in the design of communication systems.

Essentially, the problem of efficiently representing a source, known as the source coding problem, is interleaved with a channel coding problem, i.e., how to reliably and efficiently transmit the source-coded messages over compound channels. Research on joint source channel coding over compound channels has then been inspired [43–50] typically for communications over wireless fading channels.

There are several questions of interest along this research. It is necessary to first introduce a notion of message dependency to capture the "reconstructive" dependency of the source-coded messages. For any source under such representation of message dependency, we then investigate how to optimally transmit the source-coded messages over compound channels. Along this direction, we first study the optimal coding schemes for fixed discrete memoryless channels (DMC) as a special case of compound channels, then investigate coding schemes for degraded compound channels. The problem for general compound channels remains open at this time. Furthermore, in terms of the optimal performance, for a source to be transmitted over compound channels, the joint source channel coding problem needs to be taken into consideration under different scenarios. We also investigate practical channel coding schemes for source-coded messages over fading channels and study their performance under various performance metrics.

The contribution of this thesis is summarized as follows.

1) We propose a novel framework to model an arbitrary dependency structure among

source-coded messages.

As the first effort dealing with the fore-mentioned message dependency, this thesis proposes a framework to model the dependency structure among source-coded messages and investigate efficient channel coding strategies for communicating such messages over compound channels.

In this framework, an arbitrary dependency structure of source-coded messages, which may in general be induced by the mixture of LC and MDC, is modelled via a utility function. Such arbitrary dependency structure can be illustrated by a movie process. A movie process is normally formed by the video process, the audio process, and the caption process; the video and audio processes may each be represented by several streams so that various levels of quality may be reconstructed; inside the caption process, there could exist various language tracks. The presented framework associates every possible reconstruction of the movie source with a utility value and uses the utility function to represent the coded source.

2) For a special case of compound channels, namely, fixed channels, we derive the maximal achievable utility and provide the coding scheme for transmitting any given source-coded messages under the proposed model.

   Under the proposed source model, we first consider the scenario of real-time communication of media sources over DMC subject to a "delay factor" constraint. Using the average achieved utility as the performance metric, we derive the maximal achievable utility in the asymptotic regime, which largely follows from the channel coding theorem. To achieve the maximal utility, the encoder selects the best node in the utility graph subject to the channel capacity constraint, treats the set of messages corresponding to that node as a single message, and encodes it using a capacity-achieving channel code. We numerically compare the achievable utilities of an LC-coded source and an MDC-coded source and find that the two source coding schemes each has its own advantage, which depends on the delay-factor constraint as well as the channel statistics.

3) For any given source-coded messages under the proposed model communicated over

degraded compound channels, we derive analytically the maximal utility achieved by superposition codes and provide the optimal superposition coding scheme to achieve that. For the case of two source-coded messages, we show that the maximal utility achieved by superposition codes is the maximum among all coding schemes.

Inspired by the work of [51], we consider to adapt superposition codes to the problem of communicating dependent messages over degraded compound channels. For any source-coded messages with the entropy rates and the utility function defined, we derive analytically the maximal utility achieved by superposition codes and conventional channel codes respectively, where the latter follows from the former as a special case. To achieve the maximal utility, in superposition coding scheme, the encoder codes along the best (in the utility maximizing sense) sub-chain in the utility graph, whereas in conventional channel coding scheme, the encoder codes for the best node. We numerically evaluate the achievable utilities of an LC-coded source and an MDC-coded source by superposition codes. It is shown that for any given source, the performance depends on the distribution of the channel realization and the delay (or channel-code block length) requirement.

4) Considering a Gaussian $\mathcal{N}(0,1)$ source transmitted over degraded compound channels, we find the minimal distortion achieved by different joint source channel coding schemes. For each coding scheme, the optimal solution of code construction is suggested.

   The LC-coded and the MDC-coded sources used as the examples in previous numerical studies are not constructed from a single source, which makes the performance comparison between both source coding techniques infeasible. To make the comparison on their performance, we investigate efficient combination of source coding and channel coding strategies for communicating a Gaussian $\mathcal{N}(0,1)$ source over degraded compound channels for the purpose of demonstration. We derive analytically the minimal achievable distortion of different source coding schemes, namely, rate distortion codes, LC, and MDC at the receiver, by channel codes being conventional channel codes and superposition codes respectively. We find that in the asymptotic regime by conventional channel codes the minimal achievable distortion of different source coding schemes are same, however the minimal achievable distortion of an LC-coded Gaussian source is lower than that of an MDC-coded Gaussian source by superposition codes. Additionally, for either an LC-coded or an MDC-coded Gaussian source, the minimum distortion achieved by superposition codes is not greater than that achieved by conventional channel codes. The optimizing coding strategies under different scenarios are therefore provided. The analytic results are further demonstrated by numerically comparing the minimal achievable distortion of a rate distortion-coded source, an LC-coded source, and an MDC-coded source by both channel codes. The comparative behaviour also depends on the distribution of the channel realization.

5) For two dependent source-coded messages transmitted over quasi-static Rayleigh fading channels, we investigate several coding schemes including conventional channel codes, time sharing codes, and a low-density parity-check (LDPC) based codes, termed Bi-LDPC codes. Performance study based on simulations shows that the considered coding schemes each has its own advantage over different region of channel signal-to-noise ratio (SNR).

   Given the two source-coded dependent messages to be communicated over quasi-static Rayleigh fading channels, as the special case of degraded compound channels, we consider to use several coding schemes, namely, conventional channel codes, time sharing codes, and Bi-LDPC codes which demonstrate the concept of message superposition to some extent. We are to compare the performance among different channel coding schemes. For the purpose of comparison, the conventional channel codes and the time sharing codes are constructed based on LDPC component code. The success rates of both messages and the throughput are compared based on the simulation results. Furthermore, under the assumptions that the two dependent messages are generated by an LC and an MDC respectively to encode a Gaussian $\mathcal{N}(0,1)$ source sequence, the achieved average distortion on source reconstruction at the receiver is also compared among the considered channel codes. In terms of the considered performance metrics, Bi-LDPC codes, conventional channel codes encoding both messages, and time sharing type-2 (TS-2) codes, each has their own advantage over other schemes in different region of channel SNR. Along with the simulation studies and discussions, further optimization of the Bi-LDPC codes and the consideration of high order signaling during transmission are expected to improve their performance.

1.2 Thesis organization

The remaining parts of thesis are organized as follows.

In Chapter 2, we orderly introduce the necessary background around the concepts of source coding, channel coding, coding for compound channels, and joint source channel coding for compound channels, especially wireless fading channels. In the part of source coding, we have the literature review on lossy source coding techniques including rate distortion codes, LC, and MDC, and the applications of modern source coding techniques on multimedia streaming. In the part of channel coding, we briefly review various coding schemes, and treat LDPC codes as an emphasis.

We propose in Chapter 3 the model of dependent source messages with an introduction of the utility function and its graph representation, which we use throughout our discussion. Special examples of utility function are included to better explain the concept.

We derive in Chapter 4 the maximal achievable utility for communicating any given source-coded messages under the proposed model over fixed channels, followed by the numerical study on the example of source-coded messages.

Correspondingly, we derive in Chapter 5 the maximal utility achieved by superposition codes for communicating any given source-coded messages under the proposed model over degraded compound channels, and prove the optimality of superposition codes for the case of two source-coded messages. Numerical study on the example of source-coded messages is also included.

In Chapter 6, we investigate the minimal achievable distortion for communicating a Gaussian $\mathcal{N}(0,1)$ source over degraded compound channels. Different joint source channel coding schemes are analyzed, wherein the source coding scheme is chosen to be rate distortion codes, LC, and MDC respectively, while the channel codes are conventional channel codes and superposition codes respectively. The analytic results are further compared via numerical study.

In Chapter 7 we compare several coding schemes for communicating two dependent source-coded messages over quasi-static Rayleigh fading channels, including conventional channel codes, time sharing codes, and Bi-LDPC codes. Simulation results are also studied.

Chapter 8 briefly concludes the thesis and the future work is to be expected.

Chapter 2

Background

2.1 Source coding

Source coding refers to the techniques that attempt to compress or efficiently represent the source information. Generally, in source coding techniques, the sequence of source information symbols is represented by a sequence of alphabetical symbols. Source coding can be divided into two categories, namely, *lossless* coding and *lossy* coding. The source coding is *lossless* if the information symbols can exactly be reconstructed from the alphabetical symbols otherwise it is *lossy* with distortion or error. This thesis is only concerned with lossy source coding.

2.1.1 Lossy source coding

Compared with lossless source coding, the lossy source coding techniques provide the flexibility on representing the source. They have been widely used in many applications such as audio, video, and image compression. Typically, in multimedia communication, lossy source coding techniques can be applied to not only meet the bandwidth requirements, but also maintain the source fidelity up to some certain criterion. In general, a lossy source coding scheme can be characterized by the rate of compression and the distortion on source reconstruction. Inspired by the introduction of the rate distortion theory [52], the rate distortion theoretic concepts have been tremendously developed and various practical lossy source coding schemes have been presented [53]. In addition to the classic rate distortion coding, or quantization, the concepts of LC and MDC schemes are the most fundamental and typical, for which we now introduce the information theoretic results.

Information theoretic results

Highlighted by the rate distortion theory [52, 54], the rate distortion codes, or simply quantization, work by encoding the source into a single message such that the code rate is the minimum subject to a certain upper limit of distortion between the source and its reconstruction. Typically, we consider a source emitting an independent identically distributed (i.i.d.) sequence $\mathbf{X} := X_1, X_2, \ldots, X_n$, where each X_i is drawn according to the distribution $p(x),\ x \in \mathscr{X}$. By rate distortion code $(2^{nH}, n)$, the source sequence is encoded by a mapping $f:\ \mathscr{X}^n \to \{1, 2, \ldots, 2^{nH}\}$, where H is the rate of the code in bits per source symbol, and the encoded message $f(\mathbf{X})$ is decoded into the reconstruction of $\mathbf{X}$, $\hat{\mathbf{X}}$ by a mapping $g:\ \{1, 2, \ldots, 2^{nH}\} \to \hat{\mathscr{X}}^n$, where $\hat{\mathscr{X}}$ denotes the reproduction alphabet. The cost of source reconstruction is measured by a mapping $d:\ \mathscr{X} \times \hat{\mathscr{X}} \to \mathbb{R}^+$, termed as *distortion function*, where $\mathbb{R}^+$ denotes the set of positive real numbers. The distortion between any source sequence $\mathbf{x}$ and reconstruction sequence $\hat{\mathbf{x}}$ is defined as

$$d(\mathbf{x}, \hat{\mathbf{x}}) := \frac{1}{n} \sum_{i=1}^{n} d(x_i, \hat{x}_i). \tag{2.1}$$

Therefore the distortion D incurred by the rate distortion code is given by

$$D = \mathrm{E} d(X, \hat{X}), \tag{2.2}$$

where the expectation is with respect to the distribution $p(x)$. A rate distortion pair (H, D) is said to be achievable if there exists a sequence of rate distortion codes parametrized by n, such that $\lim_{n\to\infty} \mathrm{E} d(X, \hat{X}) \leq D$ holds. The rate distortion region is the closure of the set of achievable rate distortion pairs (H, D), which can be specified by $H \geq I(X; \hat{X})$, where the joint distribution $p(x, \hat{x})$ satisfies the condition $\mathrm{E} d(X, \hat{X}) \leq D$. The mean-squared error (MSE) is commonly used as the distortion measure in the rate distortion theory.

Being adapted to the varying channel conditions during data transmission, such as channel uncertainty over wireless channels and packet-loss channels, LC as the source coding scheme is a desired candidate in order to obtain different levels of source fidelity at the receiver. LC, theoretically traced back to the idea of "successive refinement of information" [16], represents the source by messages of a base layer and several successive enhancement layers. The base layer provides a coarse reproduction of the source, which can be successively refined on receipt of additional enhancement layers. An enhancement layer is useful only if the base layer and all previous enhancement layers have been correctly received. Therefore, an improvement in source reproduction, namely, reduced distortion, can be achieved. The results of [16] are further generalized and interpreted in [17]. In [17], by LC, the sequence $\mathbf{X}$ is encoded by two mappings $f_1 : \mathscr{X}^n \to \{1, 2, \dots, 2^{nH_1}\}$ and $f_2 : \mathscr{X}^n \to \{1, 2, \dots, 2^{nH_2}\}$, where H_1 and H_2 are rates of the base layer and the enhancement layer respectively. Depending on the availability of the encoded messages $f_1(\mathbf{X})$ and $f_2(\mathbf{X})$, either the mapping $g_1 : \{1, 2, \dots, 2^{nH_1}\} \to \hat{\mathscr{X}}_1^n$ or $g_0 : \{1, 2, \dots, 2^{nH_1}\} \times \{1, 2, \dots, 2^{nH_2}\} \to \hat{\mathscr{X}}_0^n$ is applied to generate the reconstruction of $\mathbf{X}$, $\hat{\mathbf{X}}_1 \in \hat{\mathscr{X}}_1^n$ or $\hat{\mathbf{X}}_0 \in \hat{\mathscr{X}}_0^n$. With the distortion functions defined as the mappings $d_m : \mathscr{X} \times \hat{\mathscr{X}}_m \to \mathbb{R}^+$ for $m = 0, 1$, the distortion between any source sequence $\mathbf{x}$ and the reconstruction sequence is defined as

$$d_m(\mathbf{x}, \hat{\mathbf{x}}_m) := \frac{1}{n} \sum_{i=1}^{n} d_m(x_i, \hat{x}_{m,i}). \tag{2.3}$$

The distortions D_1 and D_0 with $D_1 \geq D_0$ incurred by the LC are given by

$$D_m = \mathrm{E} d_m(X, \hat{X}_m), \; m = 0, 1, \tag{2.4}$$

where the expectation is with respect to the distribution $p(x)$. A rate distortion tuple (H_1, H_2, D_1, D_0) is said to be achievable if there exists a sequence of LC codes parametrized by n, such that $\lim_{n \to \infty} \mathrm{E} d_m(X, \hat{X}_m) \leq D_m$ holds for $m = 0, 1$. The rate distortion region is the closure of the set of achievable rate distortion tuples (H_1, H_2, D_1, D_0), which can be specified by

$$H_1 \geq I(X; \hat{X}_1)$$

and

$$H_1 + H_2 \geq I(X; \hat{X}_0, \hat{X}_1),$$

where the joint distribution $p(x, \hat{x}_1, \hat{x}_0)$ satisfies the conditions $\mathrm{E}d_m(X, \hat{X}_m) \leq D_m$ for $m = 0, 1$.

Alternatively, originated from [31–33], MDC represents the source by a set of messages, namely, descriptions, where all descriptions are equally important. The source can be reproduced by any subset of the descriptions, and the quality of reconstruction improves with the size of the subset. Typically by MDC, the sequence $\mathbf{X}$ is encoded by two mappings $f_1 : \mathscr{X}^n \to \{1, 2, \ldots, 2^{nH_1}\}$ and $f_2 : \mathscr{X}^n \to \{1, 2, \ldots, 2^{nH_2}\}$, where H_1 and H_2 are the rates of description 1 and 2 respectively. Depending on the availability of the encoded messages $f_1(\mathbf{X})$ and $f_2(\mathbf{X})$, one of the three mappings $g_1 : \{1, 2, \ldots, 2^{nH_1}\} \to \hat{\mathscr{X}}_1^n$, $g_2 : \{1, 2, \ldots, 2^{nH_2}\} \to \hat{\mathscr{X}}_2^n$, and $g_0 : \{1, 2, \ldots, 2^{nH_1}\} \times \{1, 2, \ldots, 2^{nH_2}\} \to \hat{\mathscr{X}}_0^n$, is used to generate the reconstruction of $\mathbf{X}$, which are $\hat{\mathbf{X}}_1 \in \hat{\mathscr{X}}_1^n$, $\hat{\mathbf{X}}_2 \in \hat{\mathscr{X}}_2^n$, and $\hat{\mathbf{X}}_0 \in \hat{\mathscr{X}}_0^n$ correspondingly. Associated with the distortion functions defined as the mappings $d_m : \mathscr{X} \times \hat{\mathscr{X}}_m \to \mathbb{R}^+$ for $m = 0, 1, 2$, the distortion between any source sequence $\mathbf{x}$ and the reconstruction sequence is defined as

$$d_m(\mathbf{x}, \hat{\mathbf{x}}_m) := \frac{1}{n} \sum_{i=1}^{n} d_m(x_i, \hat{x}_{m,i}). \tag{2.5}$$

The distortions D_1, D_2, and D_0 with $D_1 \geq D_0$ and $D_2 \geq D_0$ incurred by the MDC are given by

$$D = \mathrm{E}d_m(X, \hat{X}_m),\ m = 0, 1, 2, \tag{2.6}$$

where the expectation is with respect to the distribution $p(x)$. A rate distortion tuple $(H_1, H_2, D_1, D_2, D_0)$ is said to be achievable if there exists a sequence of MDC codes parametrized by n, such that $\lim_{n\to\infty} \mathrm{E}d_m(X, \hat{X}_m) \leq D_m$ holds for $m = 0, 1, 2$. The rate distortion region is the closure of the set of achievable rate distortion tuples

$(H_1, H_2, D_1, D_2, D_0)$, which can be specified by

$$H_1 \geq I(X; \hat{X}_1),$$

$$H_2 \geq I(X; \hat{X}_2),$$

and

$$H_1 + H_2 \geq I(X; \hat{X}_0, \hat{X}_1, \hat{X}_2) + I(\hat{X}_1; \hat{X}_2),$$

where the joint distribution $p(x, \hat{x}_0, \hat{x}_1, \hat{x}_2)$ satisfies the conditions $\mathrm{E}d_m(X, \hat{X}_m) \leq D_m$ for $m = 0, 1, 2$.

Coding techniques

Based on the information theoretic results on different lossy source coding schemes, extensive research has been conducted towards developing various coding techniques. In the area of rate distortion codes, also known as quantization, or scalar quantization, the well known pulse-code modulation (PCM) [55], originally developed for speech applications, is a typical example of scalar quantization. A comprehensive survey of the popular and promising quantization techniques and their variants has been shown in [56].

For LC codes, similar idea on representing the source in the form of several copies with different error levels is presented in [57–59]. Similar to the concept of successive refinement of information, progressive coding scheme, as a means to provide the source reconstruction at different levels of quality, is mainly applied in graphic compression and image coding [18, 60–63]. The well known Joint Photographic Experts Group (JPEG) prescribes the progressive compression scheme after applying the discrete cosine transform (DCT) to the image [18]. The moving picture encoded into a low frequency component and a high frequency component by progressive coding, and more complicated method of visual progressive coding for image based on transformation and weight assigning are patented [62, 63]. Moreover, progressive coding is further used to perform the three-dimensional (3-D) graphic compression [60] and the compression of the medical volumetric data using 3-D integer wavelet packet transforms and set partitioning in hierarchical trees (SPIHT) algorithm [61].

For MDC codes, the history and the development of MDC are reviewed in [34]. The design of practical MDC has been presented in [35–37]. In [35], the design of multiple description scalar quantizer (MDSQ) is formulated as an optimization problem and the design algorithm is developed, while in [36, 37], the design of multiple description lattice vector quantizer (MDLVQ) is considered, which is shown to outperform MDSQ. In the area of image coding, relevant work is presented in [39, 40, 64]. In [39] two MDC schemes on image coding are reported, wherein one scheme is based on the correlating transform, whereas the other uses frame expansions. The design and optimization of high-performance wavelet-based multiple description image coding algorithms are presented in [40]. Furthermore, an MDC image coding scheme is proposed in [64] which is based on the design of pairwise correlating transform (PCT) for the purpose of reducing the distortion when only one description is received.

2.1.2 Source coding in media streaming

With the advancement of multimedia technology, demand for real-time multimedia data transfer has increased, which typically involves large volume of data. Various advanced data compression techniques have been incorporated in order to fulfill the ever increasing bandwidth requirements. In the context of delivering multimedia data stream to heterogeneous receivers, the source is to be efficiently represented. Due to the flexibility provided by LC and MDC for the data transmission over channels with different conditions and the increasing demand for multimedia communication, various state-of-the-art applications of both LC and MDC methodologies have been widely utilized specifically in the area of media communications over years.

Scalable Video Coding (SVC) [19], as the major application of LC, becomes attractive since it is adapted to enable the decoding from partial bit streams depending on the required rate and quality by certain receiving device during video streaming. Extensive research work has been conducted for efficient SVC techniques [20–25]. Notions such as spatial, temporal, and SNR scalability are further defined in Moving Picture Experts Group (MPEG) standard [65], where the SNR scalability is also known as fidelity or quality scalability. The performance of techniques based on different scalability is compared and discussed [25]. The combination of SVC and unequal error protection is investigated for Internet video streaming [20]. An adaptive framework based on SVC is proposed for video delivery over wireless networks [21]. With 3-D SPIHT algorithm originally designed for image coding, a low bit-rate SVC scheme is proposed in [23]. SVC has been standardized as an extension of H. 264/Advanced Video Coding (AVC), commonly referred to as part 10 of the MPEG-4 standard [26]. A typical SVC approach, known as fine granular scalability (FGS) and also adopted in MPEG-4 standard, is described in [22] to show its simplicity and flexibility, and an efficient FGS SVC scheme is also presented [24]. An overview of the concepts on SVC in the extension of H. 264/AVC is provided in [66], and the comprehensive introduction to H. 264 and MPEG-4 standard is contained in [67]. Additionally, the relation between SVC and sorts of mobile delivery methods is introduced [68].

Other related work on LC is included in [27–30]. Earlier work of [27] presents an LC scheme using DCT transform for video transmission over asynchronous transfer mode (ATM) networks. An effective framework for low bit-rate video communications over Internet based on LC is proposed together with an algorithm on the selection of rate distortion optimized mode [28]. Over peer-to-peer (P2P) networks, a distributed protocol using LC approach for video streaming is shown to have promising performance [29]. Recent work of [30] jointly optimizes both the base layer and the enhancement layer for particular cases of LC, namely, MPEG Scalable Advanced Audio Coding (S-AAC) and Scalable-to-Lossless (SLS) coding.

MDC, on the other hand, due to the property that any received subset of data is useful on source reconstruction, is well adapted to media sources. In audio coding, as the original motivation for MDC, an audio stream, is separated into two sub-streams before further encoded, which provides the flexibility to the decoder depending on how many sub-stream it receives [38]. Along with such idea, later work such as [69] proposes an MDC-based packetization scheme for transmitting audio over Internet. For video streaming, due to its requirement on delay constraint, MDC acts as a proper solution [41, 42, 70, 71]. The principles on designing MDC for video streaming are described [41], specifically for codes utilizing prediction and mismatch control. An algorithm for multiple description motion coding (MDMC) is proposed [70], which is shown to have promising performance on video transmission. To achieve good performance over channels with bursty errors, MDC is exploited on designing transmission scheme for low bit-rate video [71]. Furthermore, two schemes of MDC based on motion compensation prediction are proposed for video steaming over Internet [42].

Over packet networks, LC and MDC are further analyzed and compared in different scenarios. In [72], they are compared with or without retransmission in terms of the rate distortion lower bound, while in [73], their performance is compared with or without network path diversity using different transmission schemes. Nevertheless, the combination of LC and MDC is shown to have the advantage of both LC and MDC [74].

2.2 Channel coding

Source coding tries to remove the redundancy present in the source, while channel coding, as a framework to efficiently and reliably transmit the data over channels, adds the redundancy to the information being transmitted. Channel codes, also known as error-control codes, are now widely used not only in communication systems, but also in many applications such as compact disk and hard disk drive. Channel codes are mainly classified into two categories, namely, *block codes* and *convolutional codes*.

2.2.1 Overview

In general, a block code is a set of fixed length sequences, called codewords, where the symbols of a codeword is selected from some finite alphabet, usually the finite field $GF(q)$. Specifically, when $q = 2$, the code is a binary code and the symbols of any codeword are named bits. The most attractive block codes are *linear*, where the sum of any two codewords is a codeword, and any codeword multiplied by an element of $GF(q)$ is a codeword. For instance, Bose-Chaudhuri-Hocquenghem (BCH) codes [75, 76], Reed-Solomon codes [77], and Reed-Muller codes [78, 79] are all well known linear block codes. For illustrative purpose, we consider binary linear block codes.

For positive integers $k < n$, a linear block code, termed (n, k) code, maps a block of k information bits into a length n codeword, chosen from the set of 2^k codewords. The set of all codewords form a k-dimensional subspace of the vector space formed by all length n sequences. The ratio k/n, denoted by R, is defined as the *code rate*. A measure called *Hamming distance* between two codewords is the number of corresponding bits for which they differ. The smallest Hamming distance among the set of all codewords is termed the *minimum distance* of the code, and denoted by $d_{\min}$. Let m denote the sequence of information bits. The encoding of m is performed through a full rank binary *generator matrix* of size $k \times n$, denoted G. We use an example to illustrate the generator matrix.

Example 1 *The generator matrix of a* $(7, 4)$ *Hamming code is as follows.*

$$G = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix} \tag{2.7}$$

The codeword c is generated as $c = mG$. Consider also the dual code of any (n, k) code, which is a linear $(n, n-k)$ code with codewords forming the null space of (n, k) code. Associated with any (n, k) code is the *parity-check matrix* of size $(n-k) \times n$ with binary elements, denoted H, which is the generator matrix of its dual $(n, n-k)$ code. We use an example to illustrate the parity-check matrix as well.

Example 2 *The parity-check matrix of a* $(7, 4)$ *Hamming code is as follows.*

$$H = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} \tag{2.8}$$

For every codeword c of the (n, k) code, it satisfies the condition $cH' = \mathbf{0}$, where $\mathbf{0}$ denotes the all zero sequence. An efficient decoding algorithm based on the parity-check matrix H is called *syndrome decoding*. Given the received word corrupted by the error, a syndrome can be computed. For the computed syndrome, we may find the most likely error pattern based on a syndrome decoding table called standard array, from which the decoded word is yielded. For codes in different classes, various decoding algorithms may exist. For example, BCH codes and Reed-Solomon codes can be decoded by the Berlekamp-Massey algorithm [80, 81], and Reed-Muller codes can be decoded through the notion of majority logic decoding.

Convolutional codes, as another major category of channel codes, generally map an information sequence, of possibly infinite length, into an output sequence. Assume the input data has binary symbols. More precisely, the encoding of a convolutional code is done through a *linear finite-state shift register*, which we now describe. The input sequence is entered into and passed along the shift register k bits at one time. The shift register is also parametrized by the *constraint length* K and n *linear generator* based on the total Kk input bits in the shift register. The code rate R is defined as the ratio k/n. Three different ways, namely, the tree diagram, the trellis diagram, and the state diagram, are used to represent the convolutional code. Based on the trellis representation of the convolutional code, Viterbi algorithm [82] provides the optimum decoding based on maximum likelihood sequence estimation (MLSE) to search through the trellis for the most probable input sequence. The BCJR decoding algorithm [83] is exploited to achieve the optimal bit error by estimating the maximum a posteriori probability (MAP) [84].

With the invention of Tanner graph [85], Tanner-Wiberg-Loeliger graph [86], and later factor graph [87, 88], the traditional code representation has shifted to the paradigm of codes on graph, typically accompanied by the invention of turbo codes [89] and the re-discovery of LDPC codes [90–92]. The state-of-the-art approach of coding and decoding requires both the graph representation of codes and the iterative message passing decoding algorithm. Based on the work of [89], turbo codes are constructed from the parallel concatenation of two elementary recursive systematic convolutional (RSC) codes linked by nonuniform interleaving. The decoder consists of two elementary decoders, each of which is based on a modified BCJR algorithm. The iterative decoding is performed by the decoder, where the MAP information is exchanged between the two elementary decoders. It appears that turbo codes can perform close to Shannon's theoretical limit [93].

Unlike turbo codes which use convolutional codes as the building blocks, LDPC codes are linear block codes, which can be represented by factor graph. In next section we introduce LDPC codes and focus on the factor graph representation of LDPC codes.

2.2.2 Low-density parity-check (LDPC) codes

We first introduce the factor graph representation of linear block codes. A *factor graph* is a bipartite graph consisting of two types of nodes, *variable nodes* and *function (factor) nodes*, where a variable node is connected to a function node by an *edge* if the variable is an argument of the function, and we term the number of the edges connected to a node the *degree* of the node. A linear block code (n, k), completely specified by its generator matrix G or parity-check matrix H, can be represented by the factor graph. By using an example, we explain how to obtain a factor graph from a parity-check matrix.

Example 3 *Consider the parity-check matrix H of a $(7,4)$ Hamming code in (2.8). If we denote the codeword by sequence $(c_1, c_2, \ldots, c_7)$, then in terms of H, the following constraints are satisfied simultaneously.*

$$\begin{aligned} c_1 \oplus c_2 \oplus c_3 \oplus c_5 &= 0, \\ c_2 \oplus c_3 \oplus c_4 \oplus c_6 &= 0, \\ c_1 \oplus c_2 \oplus c_4 \oplus c_7 &= 0, \end{aligned}$$

where $\oplus$ denotes the operation of modulo 2 addition. Let each codeword bit and each constraint be represented by a variable node and a function node respectively. An edge is connected between a variable node and a function node if the variable node is involved in the constraint specified by the function node. Therefore we have the factor graph representation of the $(7,4)$ Hamming code, shown in Figure 2.1.

Experiencing invention in 1960s by Gallager [90] and rediscovery during the recent years [91], LDPC codes become an important class of linear block codes, which are so called due to the fact that in terms of the parity-check matrix, only a small fraction of "1"'s exists in it.

Based on the factor graph representation of LDPC codes, each of the codeword bits represents a *variable node*, while each parity-check constraint represents a *check node*, equivalent to function node in general factor graph. The factor graph of an example of LDPC codes is shown in Figure 2.2. The ensemble of LDPC codes with same length is identified by the degree distributions $\{\lambda_i\}$ and $\{\rho_j\}$, and each of the codes in the

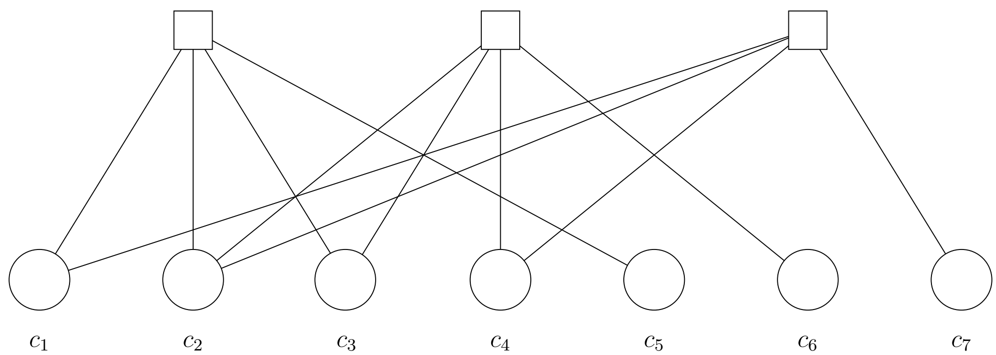

Figure 2.1: Factor graph corresponding to parity-check matrix H in (2.8) (squares represent the function (factor) nodes and circles represent the variable nodes).

ensemble has nearly the same performance. In LDPC codes literature, $\{\lambda_i\}$ is termed *variable (left) degree distribution*, and $\{\rho_j\}$ is termed *check (right) degree distribution*, where λ_i and ρ_j are the probability of an edge being attached to a degree i variable node and a degree j check node respectively. Degree distributions plays an important role on the performance of the codes. If there is only one element involved in $\{\lambda_i\}$ or $\{\rho_j\}$, such a degree distribution is a *regular* distribution, and an *irregular* distribution otherwise. Given those two degree distributions $\{\lambda_i\}$ and $\{\rho_j\}$, the rate of the ensemble can be obtained as

$$R = 1 - \frac{\sum \rho_j/j}{\sum \lambda_i/i}. \tag{2.9}$$

The suboptimal decoding algorithm, sum-product algorithm [88], is widely applied to decode LDPC codes with low complexity. The sum-product algorithm, also terminologically known as belief-propagation (BP) algorithm, passes the information from received noisy codeword symbols in the graph, and the belief of a transmitted codeword symbol becomes more and more precise during iterations. Under the MAP decision rule, the performance of sum-product algorithm is close to optimum, which has been acted as the standard approach for decoding a large family of channel codes with the performance extremely close to Shannon's theoretical limit [91, 92, 94].

With respect to such iterative message-passing decoding algorithm, the asymptotic performance of the ensemble of LDPC codes can be analyzed via a powerful tool, termed

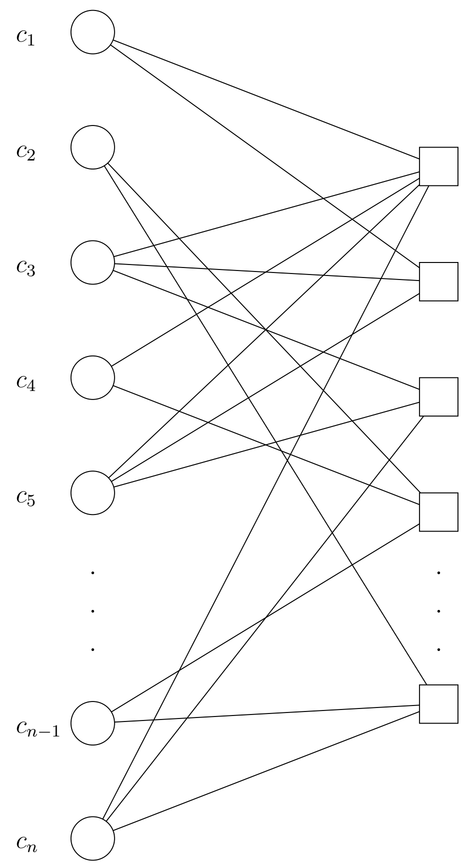

Figure 2.2: The factor graph of an LDPC code (squares represent the check nodes and circles represent the variable nodes).

density evolution (DE) [95]. By DE, the probability density function (pdf) of log-likelihood ratio (LLR) messages keeps evolving during the decoding iterations. Two assumptions need to be imposed essentially in order to validate the density evolution.

On one hand, the pdf $f(x)$ of LLR messages during the iterative decoding keeps satisfying the condition of *symmetric distribution* [95], and hence, the all-zero codeword, or equivalently, the all-one codeword after the binary phase shift keying (BPSK) modulation is considered to be transmitted, where the symmetric distribution refers to the distribution of random variable X with pdf $f(x)$ if and only if $f(x) = e^x f(-x)$. On the other hand, after round l of the iterative decoding process, the paths of the LLR messages passed along are assumed to be strict trees under the factor graph representation. Thus, when l becomes larger, the factor graph is assumed to be a group of trees, without any cycles in it, which guarantees the independence among the LLR messages. This fact is impossible for a finite length code, but only valid for an infinite length code.

As an analytical tool for LDPC codes, DE determines the threshold for the given ensemble parameters $\{\lambda_i\}$ and $\{\rho_j\}$, which equivalently gives the average asymptotic performance of the family of codes over a range of channels. *Threshold*, which is a channel parameter, divides the range of channels into two parts of different code performance. We refer to Gaussian channels as an illustration. If the threshold is the channel variance σ^{*2}, then the error probability after l iterations of decoding algorithm tends to zero with l approaching infinity for all the channels with variance smaller than σ^{*2}, whereas for all the channels with variance greater than σ^{*2}, the decoding error probability is bounded away from zero.

DE is not only good for analysis, but acts as a design tool of LDPC codes as well. Nevertheless, it is such a computation-intensive method that we are interested in other approximation schemes as the replacements. As a result, semi-Gaussian approximation approach based on extrinsic information transfer (EXIT) chart, known as a one-dimensional (1-D) method typical for Gaussian channels, shows its good performance by the gap between the Shannon limit and the design code rate, and its less computational cost with only one parameter evolving [96].

The work of [96] suggests that due to the central limit theorem, the pdf of LLR messages at the output of variable nodes can be well approximated by symmetric Gaussian or a mixture of symmetric Gaussian pdf's, while the density at the output of check nodes keeps its true pdf. EXIT chart provides a way to describe the transition of information on transmitted codeword bits during the decoding iterations, which is always expressed as a function $I_{out} = f(I_{in}, I_0)$, where I_0, I_{in}, and I_{out} are channel information, information from previous iteration, and output information of current iteration, respectively. Based on such semi-Gaussian assumption, the error probability in iterative decoding, which is well approximated by the tail probability of symmetric Gaussian distribution for LLR, can be chosen as the measure of the information. Therefore, the EXIT chart approach, combined with the semi-Gaussian assumption, sufficiently provides a good approximation of DE. On the other hand, for a given Gaussian channel, a good ensemble of LDPC codes can be found by this method together with linear programming, with little computational cost.

2.3 Coding for compound channels

With the rapid development of wireless communication technology, there has been extensive work on channel codes over wireless fading channels, a class of compound channels. In general, under the assumption that both transmitter and receiver are ignorant of the channel realization, the capacity of compound DMC is provided in [2, 3], while the compound channel capacity of a class of finite-state channels (FSC) is derived in [5]. For the compound FSC with time-invariant deterministic feedback, the capacity is provided in [97] when the fixed length block codes are considered. For fading channels, the comprehensive information theoretic results and communication features are described in [6] including the description of several frequently used channel models and the discussion on coding for fading channels as well. Furthermore, the capacity of fading channels with channel side information at the transmitter and receiver, and at the receiver only, is also obtained [4]. In the scenario of multi-user communications, the capacity region of the degraded multiple-input multiple-output (MIMO) compound broadcast channels with two receivers and two messages is characterized in [98], where each receiver is subject to a finite set of possible channel realizations. Another class of compound channels, namely, the compound wiretap channels, have been studied and the upper and lower bounds on the secrecy capacity are derived [99].

In fact, the problems on compound channels can be viewed from their similarity with broadcast channels [51], and theoretically there has been some work on fading channels studying the coding approach applied over the broadcast channels, named broadcast approach terminologically. With such approach, a strategy that adapts the power of transmitted symbols, or equivalently, the transmission rate, to the infinitely many channel realizations based on the concept of superposition coding is shown to have advantage on the average transmission rate over the traditional transmission scheme of fixed rate over slow fading channels with the channel realization known at the receiver only [7]. Such power allocation scheme for finite levels, especially two levels is further studied over quasi-static fading channels, which has also been shown to achieve significant gain on the throughput [8]. As an extension of the work [7], the broadcast transmission approach is investigated with the achievable rate region characterized over slow fading Gaussian MIMO channels as well [9]. The results of [9] are extended in [10] for a suboptimal finite level transmission over the MIMO block fading channels, wherein the information theoretic upper bounds of the achievable rates are derived for MIMO block fading channels.

Other than the broadcast approach applied for the transmission schemes over compound channels, the recently developed polar codes [100], which have been shown to be a family of capacity-achieving codes for a large class of channels, have been considered to be exploited over compound channels as well. The capacity of polar codes under successive cancellation decoding is investigated and shown to be strictly smaller than the compound capacity over the class of compound binary-input memoryless output-symmetric channels [14].

Block fading channels, over which the transmission of a codeword spans a finite number of independent channel realizations, are considered as a case of compound channels, and the coding over block fading channels is also of interest. A novel blockwise concatenation coding strategy is shown to significantly outperform the conventional transmission scheme [11]. The performance of the proposed coding scheme is analyzed in terms of the means of bounds and approximations, which shows that the codes perform close to the information theoretic limit of block fading channels. For block erasure channels, deemed as the simplified block fading channels, the error probability of the coded system is studied and the optimal solution is given in [12], where the diversity-wise maximum-distance separable (MDS) codes are shown to have the lowest error probability. Specifically, if considering full-diversity binary codes over block fading channels, the performance on error probability is further analyzed in [13]. Based on the theoretically fundamental limit of block fading channels, namely, information outage probability, the notion of outage boundary region is introduced into the scenario of communications over block fading channels. It has been shown that the codes are outage achieving if and only if the error probability is independent of the block length. Practically a new family of full-diversity LDPC codes is designed typically for block fading channels, which are shown to perform close to the outage limit [15].

Additionally, the reliable communications over wireless fading channels has been characterized by the tradeoff between diversity and multiplexing gains [101], and in [102], the problem of designing codes that are universally tradeoff-optimal is studied at high SNR, where a precise characterization of codes is provided. Practically, LDPC codes have been a good solution for transmission over fading channels. The performance of LDPC codes over Rayleigh fading channels is studied in [103], and the code optimization based on the technique of differential evolution is shown to be effective. For a two state fading channel model, namely, burst-erasure channels with Gaussian noise, an approach for designing LDPC codes is presented in [104] together with the performance characterization by the maximum resolvable erasure-burst length.

2.4 Joint source channel coding

For the purpose of transmitting a source over a channel, source coding and channel coding can be performed separately and sequentially together with the consideration on their own optimality, provided that the coding block length goes to infinity [93]. In practical, such condition on the block length of the code can not be hold in general, and thus considerable interest has developed on various schemes of joint source channel coding.

Over the decades, plenty of work has been conducted both theoretically and constructively under different scenarios of sources and channels along the research. Theoretically, a coding theorem is established for transmitting a stationary and ergodic source over discrete memoryless noisy channels by joint source and channel trellis coding, which is shown to perform arbitrarily close to the information theoretic limit [105]. For an information stream jointly source channel-coded using variable length codes and transmitted over discrete memoryless noisy channels, it is shown that the probability of decoding error under a convolutional-type decoder goes to zero as the length of codeword goes to infinity [106]. Practically, the design of joint source and channel trellis waveform coders are proposed in [107] to encode the stationary and ergodic source for communication over discrete memoryless noisy channels, which has been shown to outperform the separately optimized source and channel codes. On considering the source codes that are robust against channel errors, an algorithm based on simulated annealing for designing the vector quantizer over noisy channels is developed [108]. In [109], the joint design of a block source encoder and a finite modulation signal set is proposed to minimize the decoding error, where it is shown that the code performance for a correlated source and a sufficiently noisy channel coincides with the theoretically attainable limit. Based on the idea that the residual redundancy at the output of the source encoder can be used to provide error protection in the same way as the insertion of redundancy by the channel encoder provides the error protection against the channel noise, an approach for using such redundancy is proposed in the design of joint source channel codes and has been shown to obtain significant performance gain [110]. The residual correlation left in the bits of audio, image, and video source encoder can be further used in developing a channel decoder, which is based on a modification of the Viterbi algorithm and can be used with all channel codes having binary trellis representation [111]. Additionally, for transmitting nonuniform memoryless sources over Gaussian channels and Rayleigh fading channels by turbo codes, the joint source channel coding problem is investigated in [112], where a recursive nonsystematic convolutional encoder is proposed as the elementary encoder. The performance of the resulted nonsystematic turbo codes is extremely close to Shannon's theoretical limit, and also shows the advantage over the cascade of separated source and channel codes. The rate of source and channel codes can be further adjusted to provide various levels of protection on the source data. By providing the error control protection to the most significant source-encoded bits on source reconstruction, a joint source channel image coding scheme is described to offer the performance approaching the rate distortion bound [113].

Further interest has been focused on the communications over compound channels, which exhibit a varying information theoretic limit, and the joint optimization or the combination of both source and channel codes becomes attractive. Over slow fading wireless channels, in terms of the distortions induced by the source codes and the channel, an adaptive rate allocation scheme is proposed to allocate the rates of the source and channel codes, and shown to be optimal [43]. In the general joint source channel coding scheme, the optimization of source and channel codes over wireless fading channels is further analyzed and a general approach based on a parametric distortion model is developed in [44]. Furthermore, for the image transmission over wireless channels, a fixed length reliable joint source channel coding scheme is proposed in [45], which is shown to achieve the optimal tradeoff between the source and channel codes. In addition, over MIMO block fading channels, the joint source channel coding is considered for communicating a continuous amplitude source [46] and the bounds on the distortion exponent are derived. The scheme of layered source coding with progressive transmission is analyzed for the purpose of minimizing the expected distortion on source reconstruction at the receiver.

The broadcast approach has been shown to be a good solution for transmission over wireless fading channels in Section 2.3, and in fact in joint source channel coding over compound channels, it plays an important role together with the layered source coding. In [47], the problem of transmitting a Gaussian source over Gaussian fading channels is considered, where the source is encoded by LC and the broadcast approach is applied on the transmission. The expected distortion at the receiver is optimized according to the power allocation, where the optimal power allocation algorithms are presented. The properties of such optimal power allocation scheme are further studied in [48], and the minimization of distortion under certain circumstances is shown to be a convex optimization problem. The tradeoff between the power constraint and the distortion is also analyzed in [49], where a joint source channel coding scheme based on the transmission of scaled vector-quantized source can approach the optimal tradeoff between the power constraint and the distortion. The transmission of a Gaussian source through a block fading channel is considered in [50], in which the performance of scalable source coding and the superposition codes are evaluated, and both the rate and the power allocation are optimized to achieve the low distortion on source reconstruction.

The popularity of media streaming, on the other hand, also boosts the research and the applications of joint source channel coding techniques. For video transmission over wireless channels, the relevant research work on joint source channel codes is reviewed in [114], together with the discussion on various video compression and transmission techniques. The joint source channel codes in MPEG-4 standard are discussed as well. The multi-resolution joint source channel codes have been applied into the broadcasting of high definition television (HDTV), and shown to have advantage over the traditional single-resolution schemes by providing various levels of receiving quality [115]. In terms of minimizing the power consumption, a general joint source channel coding scheme with power optimizing framework is discussed in [116] for scalable video streaming over wireless channels. More work has been done towards SVC and transmission over wireless channels. A joint source channel coding approach is presented for motion-compensated scalable video based on DCT transform [117]. The optimization of source and channel codes for the purpose of minimizing the overall distortion on the video reconstruction is taken into the consideration as well. The rate distortion optimization of the joint source channel codes for video streaming has been further considered in [118], where an efficient rate distortion optimization scheme is proposed and shown to have advantage over the existing schemes.

Chapter 3

Source Model

3.1 Model of dependent messages

We first propose a model for dependent source-encoded messages, where the concept of message dependency is from the source reconstruction perspective. We consider that the source coder of a media file (for example, a video) generates N independent discrete stationary processes. Let $\mathcal{N}$ denote $\{1, 2, \ldots, N\}$ and $M_{\mathcal{N}} := \{M_i : i \in \mathcal{N}\}$ be the set of independent random processes, each with entropy rate H_i. Here the process M_i represents the i-th message stream obtained from coding the source. We assume that each M_i is drawn uniformly from $\Omega_i := \{1, 2, \ldots, 2^{n_s H_i}\}$, where n_s is the "emission duration" of the source coder i. For any subset $\alpha \subseteq \mathcal{N}$, H_α denotes the set (or vector) of entropy rates associated with the message set $M_\alpha := \{M_i : i \in \alpha\}$.

A *utility function* κ, which we use to model the dependency structure among the message set $M_{\mathcal{N}}$, is a non-negative function on the power set of $\mathcal{N}$ satisfying the conditions

1) $\kappa(\emptyset) = 0$, and

2) $\kappa(\alpha) \leq \kappa(\beta)$ for any $\alpha \subset \beta \subseteq \mathcal{N}$.

We note that the value $\kappa(\alpha)$ represents the "utility" that a receiver obtains when reconstructing the source using exactly the message set M_α.

The dependency structure of the utility function can be represented by a graph, which we now explain. Let directed graph $\mathcal{G}$ denote the lattice of the power set of $\mathcal{N}$. That is, each vertex of $\mathcal{G}$ is precisely a subset of $\mathcal{N}$ and for any two subsets α and β of $\mathcal{N}$ satisfying $\alpha \subset \beta$ and $|\beta \backslash \alpha| = 1$, there is an edge from α to β. Let $\mathcal{G}'$ denote the subgraph of $\mathcal{G}$ obtained by deleting all nodes η (i.e., subsets of $\mathcal{N}$) with zero utility (i.e., $\kappa(\eta) = 0$). Further, any node α, if $\kappa(\beta) = \kappa(\alpha)$ for every parent β of α, then 1) node α and all edges connecting to it are deleted, and 2) every parent of α is connected to every child of α by a directed edge pointing to the child. This procedure is repeated on the graph $\mathcal{G}'$ until no such node exists. Finally, we mark the utility value of each node beside the node and call the resulting (node weighted) graph the *utility graph* of the message set $M_{\mathcal{N}}$. Figure 3.1 contains four utility graphs, where it is assumed that along every edge, the utility value *strictly* increases. It is easy to see that Figure 3.1 (a) corresponds to LC-coded messages, where message 1 is the based layer and message 2 is the enhancement layer. Figure 3.1 (b) corresponds to two MDC-coded messages. Figure 3.1 (c) corresponds to a mixed structure of LC and MDC schemes: messages 1 and 2 together form the base layer of an LC code, and message 3 forms the enhancement layer; inside the base layer, messages 1 and 2 are MDC-coded.

It is clear that utility function κ induces a dependency structure among the messages, and arbitrary dependency structures among source-coded messages can be induced this way, as shown in Figure 3.1 (d). From the examples in Figure 3.1, it appears that the utility graphs introduced here are a very intuitive tool for representing the dependency structures.

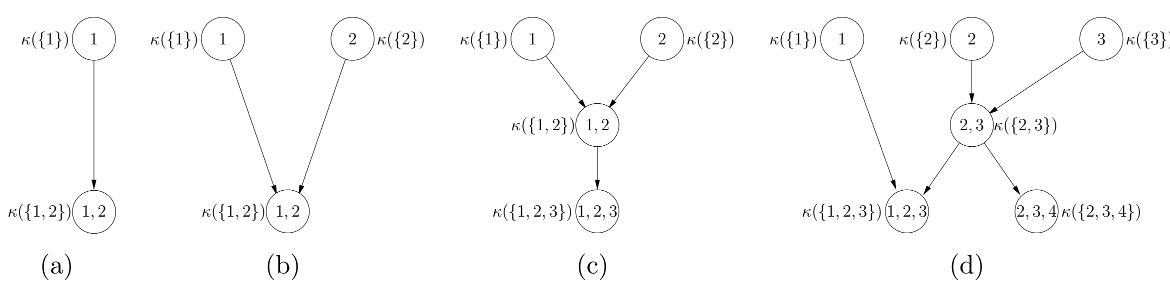

Figure 3.1: Examples of utility graph.

To illustrate the dependency among source-coded messages and the utility graph, we construct examples of an LC-coded and an MDC-coded source based on the work of [73], each generating two output message processes.

Example 4 *Based on the experimental results of [73] on LC-coded and MDC-coded video source, the two coded sources can be constructed as follows.*

The LC-coded source messages are represented by Figure 3.2(a), where message 1 *is the base layer and message* 2 *is the enhancement layer. We choose the entropy rates for both messages* $H_1 = 0.32$ *and* $H_2 = 0.64$*. The average peak signal-to-noise ratio (PSNR) is obtained as* 26.3 *dB and* 29.9 *dB on source reconstruction from message* 1 *only and both messages* 1 *and* 2 *respectively. The utility function is simply taken as the obtained PSNR values in dB when exactly the corresponding message set is available on source reconstruction.On source reconstruction, we obtain utility* 26.3 *with message* 1 *only and utility* 29.9 *with both messages* 1 *and* 2*. However, none utility is obtained with only message* 2 *decoded, meaning that message* 2 *is only useful when message* 1 *is also decoded.*

The MDC-coded source messages are represented by Figure 3.2(b), where message 1 *and message* 2 *are two descriptions. We choose the entropy rates for both messages* $H_1 = 0.48$ *and* $H_2 = 0.48$*. The average PSNR is obtained as* 26.5 *dB,* 26.5 *dB, and* 29.2 *dB on source reconstruction from message* 1 *only, message* 2 *only, and both messages* 1 *and* 2 *respectively. The utility function is taken as the obtained PSNR values in dB when exactly the corresponding message set is available on source reconstruction.It can be seen that on source reconstruction, when either message* 1 *or* 2 *is successfully decoded, utility* 26.5 *can be obtained, meaning that the two descriptions are indeed "equal-footing". When both messages* 1 *and* 2 *are decoded, utility* 29.2 *is obtained.*

We now summarize the source model: the output messages of the source encoder are specified by a set of entropy rates $\{H_i : i \in \mathcal{N}\}$ and a utility function κ, which can be visualized using a utility graph. We note that the source emission time n_s participates in defining the source model and will be used in subsequent analysis, but we choose to ignore n_s from the model, as we are mainly interested in the asymptotic regime where n_s goes to infinity.

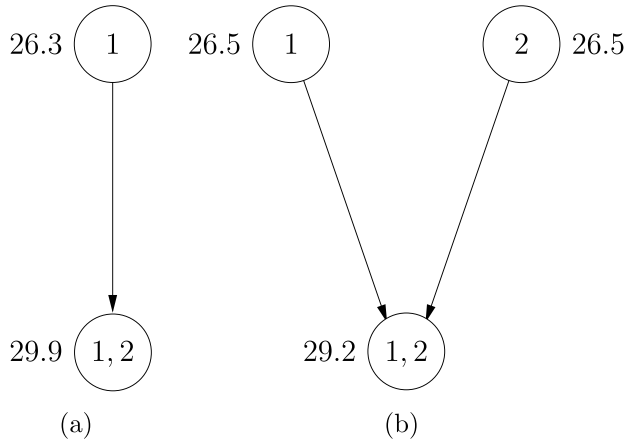

Figure 3.2: Examples of source-coded messages.

3.2 Special examples of utility function

In last section we propose a model to capture the arbitrary dependency structure among source-coded messages. Such dependency structure is modelled via the utility function κ, where it defines the obtained utility when the source is reconstructed at different levels of quality. We note that the notion of utility is in general sense. In order to better understand the implication of the utility function, it needs to be further explained. In this section, we provide several special examples of the utility function.

Various metrics can be used on defining the utility function. Suppose the data rate required to support the transmission of any subset of the source-coded messages is exploited as the utility value of that subset of messages. It can be verified that such defined utility function is non-negative and satisfies the two conditions mentioned in last section. The data rata of the empty set is no doubt zero, and as the subset of source-coded messages becomes enlarged, the data rate of the set is indeed increasing. Furthermore, the successfully transmitted data volume for any subset of source-coded messages is considered the throughput on the transmission of that message set, which serves as an important metric to evaluate the performance of the communication system.

Alternatively, PSNR on source reconstruction from any subset of the source-coded messages is treated as an example of the utility value of that subset of messages. The PSNR is non-negative and can also be verified to meet the two conditions required by a valid utility function. Obviously none of the source information can be decoded, provided that none of the source-coded messages is available, which corresponds to the case that PSNR equals zero for the empty set. With the subset of messages becoming larger, the source is thus reconstructed by using more messages. As a consequence, the quality of source reconstruction becomes better and the PSNR value exhibits the non-decreasing behaviour.

To better understand the meaning of utility function in practical, we further construct an example of utility function as follows.

Example 5 *According to an H.* 264/*AVC based Digital Video Broadcasting - Handheld (DVB-H) service specification on SVC coding scheme [119], there exist four layers of stream in the coding scheme, indexed* 1 *to* 4 *starting from the base layer. The base layer, the second layer, the third layer, and the fourth layer are encoded at the bit rate* 109*,* 140*,* 21*, and* 127 *kbit per second respectively. The base layer supports a low fidelity display of the Quarter Video Graphics Array (QVGA) resolution, which can be extended to a high fidelity of the QVGA resolution by the additional second layer. Furthermore, the combination of the bottom three layers supports a low fidelity display of the Video Graphics Array (VGA) resolution, which can also be extended to a high fidelity of the VGA resolution by the additional fourth layer. The PSNR value, measured on an VGA receiver, is obtained as* 31.22 *dB,* 32.49 *dB,* 33.90 *dB, and* 34.83 *dB, corresponding to the different levels of video quality as discussed above. In this case, we have* $\mathcal{N} = \{1, 2, 3, 4\}$*, and the utility function can be simply taken as the PSNR values such that* $\kappa(\{1\}) = 31.22$*,* $\kappa(\{1, 2\}) = 32.49$*,* $\kappa(\{1, 2, 3\}) = 33.90$*,* $\kappa(\{1, 2, 3, 4\}) = 34.83$*, and for all other* $\alpha \subset \mathcal{N}$*,* $\kappa(\alpha) = 0$*. It is obvious that the constructed utility function satisfies the two necessary conditions.*

Chapter 4

Fixed Channels

4.1 Problem formulation

Under such a source model introduced in Chapter 3, we start to consider a special scenario, real-time communication of media source over DMC subject to a "delay factor" constraint. The media source to be transmitted is the set of message processes $M_{\mathcal{N}}$ specified by entropy rates $\{H_i : i \in \mathcal{N}\}$ and utility function κ. Let the channel conditional distribution be parametrized [1] by θ.

Suppose that each channel use takes one time unit and we consider using the channel rn_s times for communicating the source, where r is a positive number. For real-time communication, such as in a video streaming scenario, the parameter r serves as a "delay factor": when $r < 1$, it is possible to deliver the source process to the receiver without delay; when $r > 1$, the receiver will necessarily suffer from some delay in reconstructing the source process.

For a DMC with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$, an $(H_{\mathcal{N}}, r, n_s)$ code consists of a message space $\prod_{i \in \mathcal{N}} \Omega_i$, an encoding function $f : \prod_{i \in \mathcal{N}} \Omega_i \to \mathcal{X}^{rn_s}$, and a decoding function $g : \mathcal{Y}^{rn_s} \to \prod_{i \in \mathcal{N}} \Omega_i$ attempting to decode $M_{\mathcal{N}}$. Apparently, an $(H_{\mathcal{N}}, r, n_s)$ code is intended for communicating messages $M_{\mathcal{N}}$ with entropy rates $H_{\mathcal{N}}$ over rn_s uses of the channel.

[1] In a general setting, θ could be a vector.

For any $(H_{\mathcal{N}}, r, n_s)$ code C over channel θ, let function $\gamma(\cdot|C,\theta)$ on the power set $2^{\mathcal{N}}$ of $\mathcal{N}$ characterize the probability of successful decoding. More precisely, for every $\alpha \subseteq \mathcal{N}$, $\gamma(\alpha|C,\theta)$ is the probability that the correctly decoded messages over channel θ are precisely those in M_α. We refer to function $\gamma(\cdot|C,\theta)$ as the *decoding profile* of code C over channel θ.

The *average utility* $\mu(C)$ of a code C is defined as

$$\mu(C) := \sum_{\alpha \subseteq \mathcal{N}} \gamma(\alpha|C,\theta)\kappa(\alpha). \tag{4.1}$$

Given a source $(H_{\mathcal{N}}, \kappa)$, over channel θ, the pair (r, Σ), referred to as a *delay-utility pair* is said to be achievable if there exists an $(H_{\mathcal{N}}, r, n_s)$ code whose average utility is Σ. For a given delay factor r, we refer to the supremum of Σ such that delay-utility pair (r, Σ) is achievable as the *maximal achievable utility.*

In this chapter we study the maximal achievable utility for a given source $(H_{\mathcal{N}}, \kappa)$ over a given channel θ.

4.2 Results

As the maximal utility is defined in terms of a supremum, we restrict our attention to $(H_{\mathcal{N}}, r, n_s)$ code sequences (parametrized by n_s) whose decoding profile converges as n_s approaches infinity. We call a code sequence (parametrized by n_s) *simple*, if over any channel θ, its decoding profile approaches a single-point mass distribution, i.e., if as n_s goes to infinity, $\gamma(\alpha|C,\theta)$ approaches 1 for some $\alpha \subseteq \mathcal{N}$ (which depends on θ). Let $\mathrm{Cap}(\theta)$ denote the capacity of channel θ. We have the following results.

Lemma 1 *Given θ, if let the entropy rates $H_{\mathcal{N}}$ and the delay factor r be such that $\sum_{i\in\alpha} H_i/r > \mathrm{Cap}(\theta)$ for some $\alpha \subseteq \mathcal{N}$, then for any sequence of $(H_{\mathcal{N}}, r, n_s)$ codes with decoding profile converging to γ as $n_s \to \infty$, $\gamma(\beta) = 0$ for all $\beta \supseteq \alpha$.*

Proof: For any sequence of $(H_{\mathcal{N}}, r, n_s)$ codes to communicate M_α, the channel code rate of the message M_i is H_i/r. If $\sum_{i\in\alpha} H_i/r > \mathrm{Cap}(\theta)$ holds for some $\alpha \subseteq \mathcal{N}$, then by the strong converse of the channel coding theorem, the probability of error for decoding M_α goes exponentially to one as $n_s \to \infty$. Then the decoding profile γ satisfies $\gamma(\alpha) = 0$.

For any $\beta \supseteq \alpha$, since $\sum\limits_{i\in\beta} H_i/r \geq \sum\limits_{i\in\alpha} H_i/r$, it is easy to verify that $\gamma(\beta) = 0$. □

Lemma 2 *Given θ, if let the entropy rates $H_\mathcal{N}$ and the delay factor r be such that $\sum\limits_{i\in\alpha} H_i/r < \mathrm{Cap}(\theta)$ for some non-empty $\alpha \subseteq \mathcal{N}$, then there exists a sequence of $(H_\mathcal{N}, r, n_s)$ codes whose decoding profile converges to γ satisfying $\gamma(\alpha) = 1$.*

Proof: If $\sum\limits_{i\in\alpha} H_i/r < \mathrm{Cap}(\theta)$ holds for some non-empty $\alpha \subseteq \mathcal{N}$, then by the achievability of the channel coding theorem, there exists a sequence of $(H_\mathcal{N}, r, n_s)$ codes that has the probability of error for decoding M_α approaching zero as $n_s \to \infty$. Thus the decoding profile γ satisfies $\gamma(\alpha) = 1$. □

For any given tuple $(H_\mathcal{N}, \kappa, \theta, r)$, let α^* be defined as

$$\alpha^* := \mathop{\arg\max}_{\alpha: \sum\limits_{i\in\alpha} H_i/r < \mathrm{Cap}(\theta)} \kappa(\alpha), \tag{4.2}$$

and Σ^* be defined as

$$\Sigma^* := \kappa(\alpha^*). \tag{4.3}$$

Then we have the following results.

Theorem 1 *Given the source $(H_\mathcal{N}, \kappa)$ over channel θ and the delay factor r, (r, Σ^*) is achievable.*

Proof: By the definition of α^*, we have $\sum_{i\in\alpha^*} H_i/r < \mathrm{Cap}(\theta)$. By Lemma 2, there exists a sequence of $(H_\mathcal{N}, r, n_s)$ codes C_{n_s} whose decoding profile converges to $\hat{\gamma}$ satisfying $\hat{\gamma}(\alpha^*) = 1$. Therefore, we have

$$\begin{aligned}
\lim_{n_s\to\infty} \mu(C_{n_s}) &= \lim_{n_s\to\infty} \sum_{\alpha\subseteq\mathcal{N}} \gamma(\alpha|C_{n_s}, \theta)\kappa(\alpha) \\
&= \sum_{\alpha\subseteq\mathcal{N}} \hat{\gamma}(\alpha)\kappa(\alpha) \\
&= \kappa(\alpha^*) \\
&= \Sigma^*.
\end{aligned}$$

□

The proof of this theorem suggests that not only Σ^* is achievable, it is in fact achievable by a simple code sequence.

Theorem 2 *Given the source* $(H_{\mathcal{N}}, \kappa)$ *over channel* θ *and the delay factor* r*, suppose (Condition* $(*)$*) that there exists no* $\alpha \subseteq \mathcal{N}$ *that will make both* $\kappa(\alpha) > \kappa(\alpha^*)$ *and* $\sum_{i\in\alpha} H_i/r = \text{Cap}(\theta)$ *hold. Then there exists no* $\Sigma > \Sigma^*$ *such that* (r, Σ) *is achievable.*

Proof: Assume there exists a sequence of $(H_{\mathcal{N}}, r, n_s)$ codes, possibly non-simple code sequence, whose average utility approaches some $\Sigma > \Sigma^*$. Let γ denote the decoding profile when n_s approaches infinity. From the assumption we see that there must exist $\hat{\alpha} \neq \alpha^*$ such that both $\gamma(\hat{\alpha}) > 0$ and $\kappa(\hat{\alpha}) > \kappa(\alpha^*)$ hold. By Lemma 1 and the condition on $\sum_{i\in\hat{\alpha}} H_i/r$, $\sum\limits_{i\in\hat{\alpha}} H_i/r < \text{Cap}(\theta)$ holds. This violates the definition of α^* in (4.2). □

The above theorems, resembling channel coding theorem and its converse, state that Σ^* is the maximal achievable utility for any given source $(H_{\mathcal{N}}, \kappa)$ over channel θ and the delay factor r satisfying Condition $(*)$. Condition $(*)$ essentially excludes the isolated values of delay factor r, such that the overall channel coding rate is exactly at the channel capacity. As the channel coding theorem is inconclusive about the error behaviour when the code rate is equal to capacity, the maximal achievable utilities for these values of r are not determined. From Theorem 1, it can be seen that the optimal coding strategy for a given DMC is to select a message set M_α that has the maximal utility $\kappa(\alpha)$ subject to channel capacity constraint, treat them as a single message and code them using a capacity-achieving channel code.

4.3 Numerical results

For the examples of an LC-coded and an MDC-coded source in Example 4, which are constructed based on [73], we consider communicating such given sources over a binary symmetric channel (BSC) where the crossover probability is θ, and numerically evaluate the maximal achievable utilities.

Figure 4.1 shows the maximal achievable PSNR at varying delay factor r when $\theta = 0.1$ for both sources. It can be seen that the maximal achievable PSNR increases with the

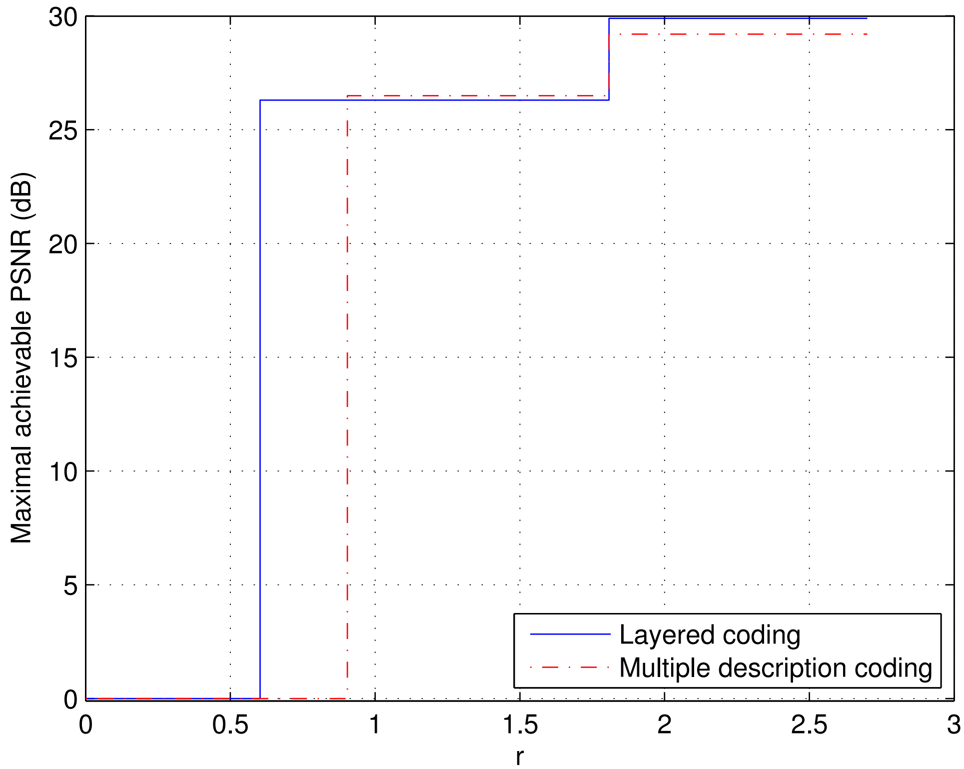

Figure 4.1: Maximal achievable PSNR at various delay factor r over BSC(0.1).

delay factor r. In both cases, the maximal achievable PSNR appears to be a staircase function of delay factor r, where the discontinuity points of the function precisely correspond to the values of r for which Condition $(*)$ in Theorem 2 is violated. For this particular channel parameter $\theta = 0.1$, it appears that the LC achieves higher PSNR than the MDC at low delay; however when allowing longer delay, the MDC results in higher PSNR. It requires even longer delay for both codes to reconstruct the source perfectly (achieving their respective maximal PSNR).

Figure 4.2 shows the maximal achievable PSNR at varying crossover probability θ when the delay factor $r = 2$. It can be seen that the maximal achievable PSNR increases as the channel becomes better. Regarding maximal PSNR as a function of θ, the discontinuities of the function occur precisely at the points of θ for which Condition $(*)$ of Theorem 2 is violated. It appears that for poor channels, the LC performs better

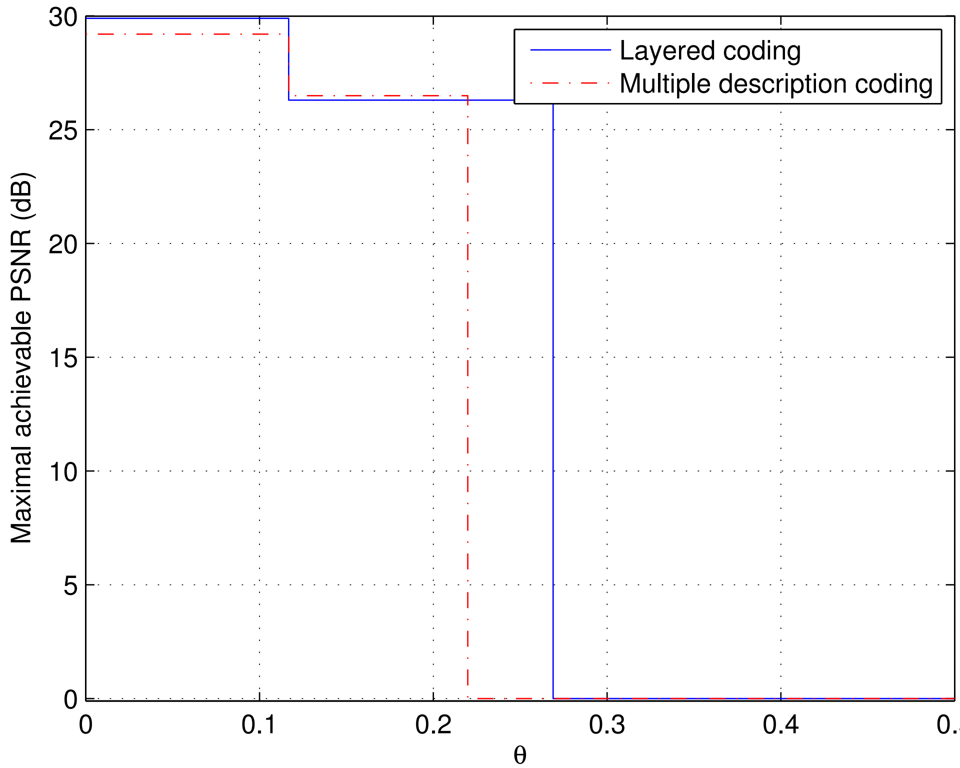

Figure 4.2: Maximal achievable PSNR at different BSC(θ) for $r = 2$.

than the MDC, and as the channel improves, the opposite trend occurs. The crossover behaviour as observed in the figures may not always occur, which depends on the rates and the utility values of the messages.

4.4 Conclusion

In this chapter, we study the problem of communicating source-coded messages under the proposed model over a fixed channel and use the average utility to evaluate the performance of the communication systems. Maximal achievable utilities are derived for any given source over any DMC. Numerical comparisons are made for two specific examples of source models over BSC, where the two models correspond respectively to the messages generated by a layered-coding source encoder and those generated by a multiple-description-coding source encoder. Depending on the tolerable delays and the channel conditions, the two sources each has its own advantage in terms of its maximal achievable PSNR. In practical, when an arbitrary dependency structure exists among the source-coded messages and the utility function is specified by the required data rate for transmission, PSNR on source reconstruction, or even the user satisfaction, we are able to evaluate the throughput, average quality on source reconstruction, or average user satisfaction, and provide the optimal coding scheme to achieve that.

The work of this chapter is still a preliminary study, since the simplicity of DMC models reduces the utility maximization problem completely to a simple mathematical structure that is completely governed by the channel coding theorem. It is our plan to extend this work to the setting of compound channel models, where we expect that the channel coding theorem for DMC will become insufficient for utility maximization and more sophisticated coding schemes are expected later.

Chapter 5

Degraded Compound Channels

5.1 Problem formulation

In last chapter, we consider the scenario of real-time communication of media source over fixed channels where the utility maximization problem is simply governed by the channel coding theorem. In this chapter, we consider communicating the set of messages $M_{\mathcal{N}}$ through a compound channel. Let θ be the parameter of the channel realization, and the compound channel is defined via a distribution q over a space of θ. Let Θ denote the support of q. We assume each $\theta \in \Theta$ parametrizes a DMC with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$. The transition probability function for each θ is denoted by $P_\theta(\mathcal{Y}|\mathcal{X})$, which is completely determined by θ. As is typical, we assume that θ is known to the receiver but not to the transmitter. A channel realization $\theta : X \to Y$ is *degraded* with respect to $\theta' : X \to Y'$ if X, Y', and Y form a Markov chain, denoted by $X - Y' - Y$. We write $\theta \prec \theta'$ or $\theta' \succ \theta$ to denote θ is degraded with respect to θ'. A compound channel is degraded if for any two channel realizations in Θ, one is degraded with respect to another. In the sequel, we consider q being a continuous distribution. In that case, the probability that a specific channel realization θ is chosen becomes zero.

Similar to the case of fixed channels, the communication over compound channels is also subject to a "delay factor" constraint. With the delay factor r, under the assumption that each channel use takes one time unit, we consider using the channel rn_s times for communicating the source. In the case of real-time communication, when $r < 1$, it is possible to deliver the source process to the receiver without any delay, however when $r > 1$, the receiver will necessarily suffer from some delay on source process reconstruction.

An $(H_{\mathcal{N}}, r, n_s)$ code consists of a message space $\prod_{i\in\mathcal{N}} \Omega_i$, an encoding function $f : \prod_{i\in\mathcal{N}} \Omega_i \to \mathcal{X}^{rn_s}$, and a set of decoding functions $\{g_\theta : \mathcal{Y}^{rn_s} \to \prod_{i\in\mathcal{N}} \Omega_i,\ \theta \in \Theta\}$ each attempting to decode $M_{\mathcal{N}}$. To ease our discussion, we also let $n := rn_s$ and R_i denote the channel code rate of M_i, and R_α be the set (or vector) of channel code rates associated with M_α. Thus $R_i = H_i/r$ holds for $i \in \mathcal{N}$, and the $(H_{\mathcal{N}}, r, n_s)$ code is equivalent to the conventional $(R_{\mathcal{N}}, n)$ code.

For any $(H_{\mathcal{N}}, r, n_s)$ code C over channel θ, let the decoding profile $\gamma(\cdot|C,\theta)$ on the power set $2^{\mathcal{N}}$ of $\mathcal{N}$ be well defined to characterize the probability of successful decoding. The *average utility* $\mu(C)$ of a code C is defined as

$$\mu(C) := \int \sum_{\alpha\subseteq\mathcal{N}} \gamma(\alpha|C,\theta)\kappa(\alpha)q(\theta)d\theta. \tag{5.1}$$

Given (κ, q), we use $\mu(C)$ to measure the performance of code C. It is worth noting that if one chooses $\kappa(\alpha) = \sum_{i\in\alpha} H_i/r$, then $\mu(C)$ reduces to the throughput achieved by code C over the compound channel.

Given a source $(H_{\mathcal{N}}, \kappa)$, over compound channel q, the pair (r, Σ), referred to as a *delay-utility pair*, is said to be achievable if there exists an $(H_{\mathcal{N}}, r, n_s)$ code whose average utility is Σ. For a given delay factor r, we refer to the supremum of Σ such that delay-utility pair (r, Σ) is achievable as the *maximal achievable utility*. As the maximal utility is defined in terms of a supremum, we restrict our attention to $(H_{\mathcal{N}}, r, n_s)$ code sequences (parametrized by n_s) whose decoding profile converges as n_s approaches infinity.

Given arbitrary $(\kappa, H_{\mathcal{N}}, q)$, we are interested in finding a coding scheme achieving highest average utility. In general the problem is quite open, and in this chapter we mainly focus on the cases where the channel is degraded compound channel.

5.2 Results

5.2.1 Superposition codes

It is well known since [51], superposition codes are good solution used for degraded broadcast channels [120]. In this chapter, we primarily consider adapting superposition codes to the communication problems over degraded compound channels. The construction of superposition codes in this section is slightly different from the conventional one, which we now describe.

Given a set of messages $M_1, M_2, \ldots, M_N$ with M_i uniformly distributed in Ω_i, a length-n m-layer superposition code for those messages is specified via a Markov chain $X_{\alpha_1} - X_{\alpha_2\backslash\alpha_1} - \cdots - X_{\alpha_m\backslash\alpha_{m-1}} - Y_{\alpha_m\backslash\alpha_{m-1}} - Y_{\alpha_{m-1}\backslash\alpha_{m-2}} - \cdots - Y_{\alpha_1}$ and a rate vector $(R_{\alpha_1}, R_{\alpha_2\backslash\alpha_1}, \ldots, R_{\alpha_m\backslash\alpha_{m-1}})$, where $\alpha_1 \subset \alpha_2 \subset \cdots \subset \alpha_m$ form a chain in the utility graph. We note that a chain refers to a directed path in the graph. For convenience, let $\Delta(1) := \alpha_1$ and $\Delta(i) := \alpha_i \setminus \alpha_{i-1}$ for $i = 2, \ldots, m$. The Markov chain and the rate vector can be rewritten as $X_{\Delta(1)} - X_{\Delta(2)} - \cdots - X_{\Delta(m)} - Y_{\Delta(m)} - \cdots - Y_{\Delta(2)} - Y_{\Delta(1)}$ and $(R_{\Delta(1)}, \ldots, R_{\Delta(m)})$ respectively.

Example 6 *To encode the chain* $(\{1\} \to \{1,2,3\} \to \{1,2,3,4,5\})$*, we have* $m = 3$ *with* $\alpha_1 = \{1\}$*,* $\alpha_2 = \{1,2,3\}$*,* $\alpha_3 = \{1,2,3,4,5\}$*,* $\Delta(1) = \{1\}$*,* $\Delta(2) = \{2,3\}$*, and* $\Delta(3) = \{4,5\}$.

For any given distributions $P_{X_{\Delta(1)}}, P_{X_{\Delta(2)}|X_{\Delta(1)}}, \ldots, P_{X_{\Delta(m)}|X_{\Delta(m-1)}}, P_{Y_{\Delta(m)}|X_{\Delta(m)}}$, $P_{Y_{\Delta(m-1)}|Y_{\Delta(m)}}, \ldots, P_{Y_{\Delta(1)}|Y_{\Delta(2)}}$,

- Map the index set $\{1, \ldots, 2^{n\sum_{l\in\Delta(1)} R_l}\}$ to $\{1, \ldots, 2^{nI(X_{\Delta(1)};Y_{\Delta(1)})}\}$ by h_1. Generate $2^{nI(X_{\Delta(1)};Y_{\Delta(1)})}$ independent length n sequences $X^n_{\Delta(1)}(M_{\alpha_1})$ according to $\prod_{j=1}^n P_{X_{\Delta(1)}}(x_{\Delta(1),j})$.

- For each $i = 2, \ldots, m$ in order, map the index set $\{1, \ldots, 2^{n\sum_{l\in\Delta(i)} R_l}\}$ to $\{1, \ldots, 2^{nI(X_{\Delta(i)};Y_{\Delta(i)}|X_{\Delta(i-1)})}\}$ by h_i. For each sequence $X^n_{\Delta(i-1)}$, generate $2^{nI(X_{\Delta(i)};Y_{\Delta(i)}|X_{\Delta(i-1)})}$ independent length n sequences $X^n_{\Delta(i)}(M_{\alpha_i})$ according to $\prod_{j=1}^n P_{X_{\Delta(i)}|X_{\Delta(i-1)}}(x_{\Delta(i),j}|x_{\Delta(i-1),j})$.

- The process continues until $2^{n(I(X_{\Delta(1)};Y_{\Delta(1)})+I(X_{\Delta(2)};Y_{\Delta(2)}|X_{\Delta(1)})+\cdots+I(X_{\Delta(m)};Y_{\Delta(m)}|X_{\Delta(m-1)}))}$ codewords $X^n_{\Delta(m)}$ have been generated.

- Reveal the codebook to the decoder. The encoder sends the codeword $X^n_{\Delta(m)}(M_{\alpha_m})$ if it sees the message set M_{α_m}.

The mapping $h_i,\ i = 1, \dots, m$ is defined as follows. For any $a \in \{1, \dots, 2^{n\sum_{l\in\Delta(i)} R_l}\}$, $h_i(a) := a \mod 2^{nI(X_{\Delta(i)};Y_{\Delta(i)}|X_{\Delta(i-1)})}$[1], where $\Delta(0)$ is neglected, and $X_{\Delta(0)}$ is the dummy random variable. The rate tuple $(I(X_{\Delta(1)}, Y_{\Delta(1)}), I(X_{\Delta(2)}; Y_{\Delta(2)}|X_{\Delta(1)}), \dots,$ $I(X_{\Delta(m)}; Y_{\Delta(m)}|X_{\Delta(m-1)}))$ has been shown to be achievable over the degraded broadcast channel $X_{\Delta(m)} - Y_{\Delta(m)} - \cdots - Y_{\Delta(1)}$ [120]. We use an example to illustrate the encoding process as follows.

Example 7 *In Example 6, we encode the message set $M_{\{1,2,3,4,5\}}$ as follows. For any distributions $P_{X_{\{1\}}}, P_{X_{\{2,3\}}|X_{\{1\}}}, P_{X_{\{4,5\}}|X_{\{2,3\}}}, P_{Y_{\{4,5\}}|X_{\{4,5\}}}, P_{Y_{\{2,3\}}|Y_{\{4,5\}}}, P_{Y_{\{1\}}|Y_{\{2,3\}}}$, firstly we map the set $\{1, \dots, 2^{nR_1}\}$ to $\{1, \dots, 2^{nI(X_{\{1\}};Y_{\{1\}})}\}$ by h_1 and generate $2^{nI(X_{\{1\}};Y_{\{1\}})}$ independent length n sequences $X^n_{\{1\}}(M_1)$ according to $\prod_{j=1}^n P_{X_{\{1\}}}(x_{\{1\},j})$. Secondly we map the set $\{1, \dots, 2^{n(R_2+R_3)}\}$ to $\{1, \dots, 2^{nI(X_{\{2,3\}};Y_{\{2,3\}}|X_{\{1\}})}\}$ by h_2 and for each sequence $X^n_{\{1\}}(M_1)$, generate $2^{nI(X_{\{2,3\}};Y_{\{2,3\}}|X_{\{1\}})}$ independent length n sequences $X^n_{\{2,3\}}(M_1, M_2, M_3)$ according to $\prod_{j=1}^n P_{X_{\{2,3\}}|X_{\{1\}}}(x_{\{2,3\},j}|x_{\{1\},j})$. Finally we map the set $\{1, \dots, 2^{n(R_4+R_5)}\}$ to $\{1, \dots, 2^{nI(X_{\{4,5\}};Y_{\{4,5\}}|X_{\{2,3\}})}\}$ by h_3 and for each sequence $X^n_{\{2,3\}}(M_1, M_2, M_3)$, generate $2^{nI(X_{\{4,5\}};Y_{\{4,5\}}|X_{\{2,3\}})}$ independent length n codewords $X^n_{\{4,5\}}(M_1, M_2, M_3, M_4, M_5)$ according to $\prod_{j=1}^n P_{X_{\{4,5\}}|X_{\{2,3\}}}(x_{\{4,5\},j}|x_{\{2,3\},j})$.*

We note that slightly different from the superposition codes in [51,120], the mapping h_i is introduced into the encoding of the message set $M_{\Delta(i)}$. The inverse mapping of $h_i,\ i = 1, \dots, m$ exists if and only if h_i is injective. In that case, h_i is called *invertible*. h_i is invertible if and only if

$$\sum_{l\in\Delta(i)} R_l \le I(X_{\Delta(i)}; Y_{\Delta(i)}|X_{\Delta(i-1)}) \tag{5.2}$$

holds. We denote the inequality in (5.2) by $E(i)$. Let $\Lambda_{i|i-1}$ and $\Psi_{i|i+1}$ denote the space of $P_{X_{\Delta(i)}|X_{\Delta(i-1)}}$ and $P_{Y_{\Delta(i)}|Y_{\Delta(i+1)}}$ respectively for $i = 1, \dots, m$, where to ease our discussion, we also let $Y_{\Delta(m+1)}$ denote $X_{\Delta(m)}$. The invertibility of the mapping h_i for $i = 1, \dots, m$

[1]More precisely, $h_i(a) := a \mod 2^{nI(X_{\Delta(i)};Y_{\Delta(i)}|X_{\Delta(i-1)})}$ when $a \neq k2^{nI(X_{\Delta(i)};Y_{\Delta(i)}|X_{\Delta(i-1)})}$ for some integer k, while $h_i(a) := 2^{nI(X_{\Delta(i)};Y_{\Delta(i)}|X_{\Delta(i-1)})}$ when $a = k2^{nI(X_{\Delta(i)};Y_{\Delta(i)}|X_{\Delta(i-1)})}$.

is completely determined by $\lambda \in \prod_i \Lambda_{i|i-1}$, $\psi \in \prod_i \Psi_{i|i+1}$, and R_{α_m}. Thus we consider a sequence of superposition codes C_n parametrized by $(\lambda, \psi, R_{\Delta(1)} \to R_{\Delta(2)} \to \cdots \to R_{\Delta(m)}, n)$, where the arrow indicates the order of superimposed messages from the view of encoder.

5.2.2 Coding theorems

When the channel is a degraded compound channel, the decoder attempts to recover the message set M_{α_m}. We are to find the decoding profile of C_n over any $\theta \in \Theta$ as $n \to \infty$. We note that since the code sequence is specified by the distribution λ, which generalizes a probabilistic ensemble of code sequences, the decoding profile in consideration is indeed the average over the ensemble. Furthermore, let τ_i be defined as

$$\tau_i := \bigcup_{j \leq i \text{ and } E(j) \text{ holds}} \Delta(j). \tag{5.3}$$

Example 8 *In Example 7, suppose the distributions λ and ψ is such that*

$$\begin{aligned} R_1 &< I(X_{\{1\}}; Y_{\{1\}}), \\ R_2 + R_3 &> I(X_{\{2,3\}}; Y_{\{2,3\}} | X_{\{1\}}), \text{ and} \\ R_4 + R_5 &< I(X_{\{4,5\}}; Y_{\{4,5\}} | X_{\{2,3\}}) \end{aligned}$$

hold. We have $\tau_1 = \{1\}$, $\tau_2 = \{1\}$, and $\tau_3 = \{1, 4, 5\}$.

We have the following results, which follow from the work [54, 121].

Lemma 3 *For the sequence of superposition codes C_n parametrized by any given $(\lambda, \psi, R_{\Delta(1)} \to R_{\Delta(2)} \to \cdots \to R_{\Delta(m)})$ over channel $\theta : X_{\Delta(m)} \to Y$, if $\theta^i \prec \theta \prec \theta^{i+1}$, $i = 0, \ldots, m$, the decoding profile γ satisfies $\gamma(\tau_i) = 1$ as $n \to \infty$, where θ^i, $i = 1, \ldots, m$ denotes the channel $X_{\Delta(m)} \to Y_{\Delta(i)}$, and the bounds of θ^0 and θ^{m+1} are neglected.*

The proof of lemma is provided in Appendix I.

Lemma 3 suggests that for any given sequence of $(\lambda, \psi, R_{\Delta(1)} \to R_{\Delta(2)} \to \cdots \to R_{\Delta(m)}, n)$ superposition codes, a degrading sequence of channels $\theta^m \succ \cdots \succ \theta^1$ are induced, which potentially divides the support Θ of degraded compound channel into $m+1$ intervals in order. When the channel realization θ is such that $\theta^i \prec \theta \prec \theta^{i+1}$ for some $i = 0, \ldots, m$, exactly the set of messages M_{τ_i} can be decoded, where τ_i represents the index set of messages on all layers j up to layer i under the view of superposition, such that the mapping h_j is invertible.

In an arbitrary utility graph, let Φ denote the set of all chains and sub-chains, where a sub-chain is a sub-sequence of a chain and may not be a chain in the graph by itself. Here we note that a chain is also a sub-chain by itself, and Φ is the set of all sub-chains. For any sub-chain $\phi \in \Phi$, let m_ϕ be the number of nodes in ϕ. Let $\{1, \ldots, m_\phi\}$ be the index set of the nodes in ϕ, where the index increases with the direction of ϕ, and $\alpha_i \subseteq \mathcal{N}$ be associated to node i for $i \in \{1, \ldots, m_\phi\}$. Let $m = m_\phi$, and we may correspond the m_ϕ messages $M_{\Delta(1)}, M_{\Delta(2)}, \ldots, M_{\Delta(m_\phi)}$ in ϕ to the m messages $M_{\Delta(1)}, M_{\Delta(2)}, \ldots, M_{\Delta(m)}$ discussed above. Pick $\lambda_\phi \in \prod_i \Lambda_{i|i-1}$ and $\psi_\phi \in \prod_i \Psi_{i|i+1}$. Let the sequence of $(\lambda_\phi, \psi_\phi, R_{\Delta(1)} \to R_{\Delta(2)} \to \cdots \to R_{\Delta(m_\phi)}, n)$ superposition codes be applied for communicating messages $M_{\alpha_{m_\phi}}$ over the degraded compound channel q. Our goal is to find the best $\phi \in \Phi$ and best configuration of $(\lambda_\phi, \psi_\phi)$ in terms of achieving the maximal utility. Let Σ^* be defined as

$$\Sigma^* = \max_{\phi \in \Phi} \max_{\lambda_\phi, \psi_\phi} \sum_{j=1}^{m_\phi} \kappa(\tau_j) \Pr(\theta : \theta^j \prec \theta \prec \theta^{j+1}). \tag{5.4}$$

Theorem 3 *Given the source $(H_{\mathcal{N}}, \kappa)$ over degraded compound channel q and the delay factor r, Σ^* is the maximal achievable utility by superposition codes.*

Proof: For the sequence of superposition codes C_n specified by any tuple $(\lambda_\phi, \psi_\phi, R_{\Delta(1)} \to R_{\Delta(2)} \to \cdots \to R_{\Delta(m_\phi)})$, where $R_i = H_i / r$ for $i \in \mathcal{N}$, by equation (5.1), we have

$$\begin{aligned} \lim_{n\to\infty} \mu(C_n) &= \lim_{n\to\infty} \int \sum_{\alpha \subseteq \mathcal{N}} \gamma(\alpha | C_n, \theta) \kappa(\alpha) q(\theta) d\theta \\ &= \int \sum_{\alpha \subseteq \mathcal{N}} \hat{\gamma}(\alpha|\theta) \kappa(\alpha) q(\theta) d\theta, \end{aligned}$$

where $\hat{\gamma}$ denotes the limiting decoding profile. By Lemma 3, we have

$$\begin{aligned}\lim_{n\to\infty}\mu(C_n) &= \int\sum_i \hat{\gamma}(\tau_i)\kappa(\tau_i)q(\theta)d\theta \\ &= \sum_i \kappa(\tau_i)\int\delta(\theta : \theta^i \prec \theta \prec \theta^{i+1})q(\theta)d\theta \\ &= \sum_i \kappa(\tau_i)\Pr(\theta : \theta^i \prec \theta \prec \theta^{i+1}),\end{aligned}$$

where $\delta(\cdot)$ denotes the indicator function valued 1 if the argument is true and 0 otherwise.

Since C_n is specified by different choice of $(\lambda_\phi, \psi_\phi)$ and ϕ, the utility Σ^* is thus achievable by the maximization of $\lim_{n\to\infty}\mu(C_n)$ over all possible sequences of codes for different choices of $(\lambda_\phi, \psi_\phi)$ and $\phi \in \Phi$.

On the other hand, let ϕ' denote the sequence $\alpha'_1 \subset \cdots \subset \alpha'_m$, where $\phi' \notin \Phi$. Suppose the condition $E(i)$ holds for $i = 1, \ldots, m$, as otherwise if $E(i)$ does not hold for any i, we can always achieve better utility by following the proof of Theorem 4 to trivialize or essentially remove $\Delta(i)$ from the chain. Based on ϕ', we can generate a sub-chain $\hat{\phi} \in \Phi$ formed by $\hat{\alpha}_1 \subset \cdots \subset \hat{\alpha}_{\hat{m}}$ as follows, according to the construction of the utility graph. Starting from $i, j = 1$, if α'_i is a node in the utility graph, let $\hat{\alpha}_j = \alpha'_i$; otherwise, if $\kappa(\alpha'_i) > 0$, we find the smallest α in the utility graph such that $\kappa(\alpha'_i) = \kappa(\alpha)$ and $\hat{\alpha}_{j-1} \subseteq \alpha$ hold, and let $\hat{\alpha}_j = \alpha$ if $\hat{\alpha}_{j-1} \subset \alpha$. We note that $\hat{\alpha}_0$ denotes the empty set. Increase j by 1 if $\hat{\alpha}_j$ is updated, as well as i. The process is repeated with increasing i and j until $i = m$. Clearly we have $\hat{m} \leq m$.

We only consider the non-trivial case where $\hat{\phi}$ is not an empty sequence. Let $\Sigma_{\phi'}$ denote the maximal achievable utility for encoding ϕ', achieved by the tuple $(\lambda_{\phi'}, \psi_{\phi'}, R_{\Delta'(1)} \to \cdots \to R_{\Delta'(m)})$. Let $P_{X_{\Delta'(i)}|X_{\Delta'(i-1)}}$ and $P_{Y_{\Delta'(i)}|Y_{\Delta'(i+1)}}$ for $i = 1, \ldots, m$ denote the distributions specifying $\lambda_{\phi'}$ and $\psi_{\phi'}$ respectively, and $P_{X_{\Delta'(i)}}$ and $P_{Y_{\Delta'(i)}}$ be the induced marginal distributions. A degrading sequence of channels $\theta'^m \succ \cdots \succ \theta'^1$ are thus induced. To encode $\hat{\phi}$, generate the distributions $\lambda_{\hat{\phi}}$ and $\psi_{\hat{\phi}}$ as follows. Let $P_{X_{\hat{\Delta}(i)}|X_{\hat{\Delta}(i-1)}}$ and $P_{Y_{\hat{\Delta}(i)}|Y_{\hat{\Delta}(i+1)}}$ for $i = 1, \ldots, \hat{m}$ denote the distributions specifying $\lambda_{\hat{\phi}}$ and $\psi_{\hat{\phi}}$ respectively, and $\theta^{\hat{m}} \succ \cdots \succ \theta^1$ denote the induced sequence of channels. For $j = 1, \ldots, \hat{m}$, let k_j and l_j denote the smallest and the largest i respectively such that $\kappa(\hat{\alpha}_j) = \kappa(\alpha'_i)$ holds. Let

$P_{X_{\hat{\Delta}(1)}} = P_{X_{\Delta'(l_1)}}$,

$$P_{X_{\hat{\Delta}(j)}|X_{\hat{\Delta}(j-1)}} = \sum_{x_{\Delta'(k_j)},\dots,x_{\Delta'(l_j-1)}} \prod_{i=k_j}^{l_j} P_{X_{\Delta'(i)}|X_{\Delta'(i-1)}}$$

for $j > 1$, and

$$P_{Y_{\hat{\Delta}(j)}|Y_{\hat{\Delta}(j+1)}} = \sum_{y_{\Delta'(k_j+1)},\dots,y_{\Delta'(l_j)}} \prod_{i=k_j}^{l_j} P_{Y_{\Delta'(i)}|Y_{\Delta'(i+1)}}$$

for $j \geq 1$. By doing so, we have the marginal distributions $P_{X_{\hat{\Delta}(j)}} = P_{X_{\Delta'(l_j)}}$ and $P_{Y_{\hat{\Delta}(j)}} = P_{Y_{\Delta'(k_j)}}$, and the channel $\theta^j = \theta'^{k_j}$ holds.

Let $\Sigma_{\hat{\phi}}$ denote the utility for encoding $\hat{\phi}$, achieved by the tuple $(\lambda_{\hat{\phi}}, \psi_{\hat{\phi}}, R_{\hat{\Delta}(1)} \to \cdots \to R_{\hat{\Delta}(\hat{m})})$. Based on the code construction, for any $j = 1, \dots, \hat{m}$, we have $\kappa(\tau_j) = \kappa(\tau_i)$ for $i = k_j, \dots, l_j$. Therefore, we have

$$\begin{aligned} \Sigma_{\phi'} &= \sum_{i=1}^{m} \kappa(\tau_i)\Pr(\theta : \theta'^i \prec \theta \prec \theta'^{(i+1)}) \\ &= \sum_{j=1}^{\hat{m}} \sum_{i=k_j}^{l_j} \kappa(\tau_i)\Pr(\theta : \theta'^i \prec \theta \prec \theta'^{(i+1)}) \\ &= \sum_{j=1}^{\hat{m}} \kappa(\tau_j)\Pr(\theta : \theta^j \prec \theta \prec \theta^{j+1}) \\ &= \Sigma_{\hat{\phi}}, \end{aligned}$$

which means that the maximal utility can be obtained by considering to encode an $\phi \in \Phi$, and Σ^* is therefore the maximal achievable utility by superposition codes. □

Theorem 3 states that for any given source $(H_{\mathcal{N}}, \kappa)$ communicated over degraded compound channel q with delay factor r, the maximal utility by superposition codes can be achieved by searching for the best sub-chain ϕ in the utility graph and best superposition code, or equivalently, λ_ϕ and ψ_ϕ, to encode ϕ. For illustrative purpose, in Figure 3.1(d), $(\{1\} \to \{1,2,3\}), (\{2\} \to \{2,3\} \to \{1,2,3\}), (\{2\} \to \{2,3\} \to \{2,3,4\}), (\{3\} \to \{2,3\} \to \{1,2,3\})$, and $(\{3\} \to \{2,3\} \to \{2,3,4\})$ are five maximum chains, where a chain is said to be *maximum* if it is not contained in any other chains except itself. Φ

contains all maximum chains and their sub-chains.

The calculation of (5.4) involves searching over a space of joint probability distribution. However, under some situations the searching space for λ_ϕ can be efficiently reduced for any given ϕ and given distribution ψ_ϕ. We consider the sequence of superposition codes C_n parametrized by any given $(\lambda_\phi, \psi_\phi, R_{\Delta(1)} \to R_{\Delta(2)} \cdots \to R_{\Delta(m_\phi)})$, and let $\hat{\Sigma}$ denote its achievable utility. We have the following result, the proof of which is provided in Appendix II.

Theorem 4 *For the sequence of superposition codes C_n parametrized by any given $(\lambda_\phi, \psi_\phi, R_{\Delta(1)} \to R_{\Delta(2)} \cdots \to R_{\Delta(m_\phi)})$ over degraded compound channel q, if $0 < I(X_{\Delta(i)}; Y_{\Delta(i)}|X_{\Delta(i-1)}) < \sum_{l\in\Delta(i)} R_l$ for any $i \in \{1, \ldots, m_\phi\}$, then there exists other $\lambda'_\phi \in \prod_j \Lambda_{j|j-1}$ such that $I(X_{\Delta(i)}; Y_{\Delta(i)}|X_{\Delta(i-1)}) = 0$ and the utility $\Sigma' \geq \hat{\Sigma}$ achieved by the sequence of superposition codes parametrized by $(\lambda'_\phi, \psi_\phi, R_{\Delta(1)} \to R_{\Delta(2)} \cdots \to R_{\Delta(m_\phi)})$.*

Theorem 4 provides a means to reduce the complexity in the calculation of (5.4). For any given $\phi \in \Phi$, whenever the mapping h_i is not invertible for any $i \in \{1, \ldots, m_\phi\}$, resulting in the fact that the message set $M_{\Delta(i)}$ can not be decoded, the theorem suggests that we can always suppress that message set by letting $I(X_{\Delta(i)}; Y_{\Delta(i)}|X_{\Delta(i-1)}) = 0$ without decreasing the achievable utility. In this case, the mapping h_i and the message set $M_{\Delta(i)}$ are trivialized[2] since none of its information is encoded. Therefore, regarding to (5.4), for any fixed $\phi \in \Phi$, the calculation for any λ_ϕ and ψ_ϕ such that there exists $0 < I(X_{\Delta(i)}; Y_{\Delta(i)}|X_{\Delta(i-1)}) < \sum_{l\in\Delta(i)} R_l$ for any $i \in \{1, \ldots, m_\phi\}$ can be omitted, in terms of achieving the maximal utility.

Superposition codes, originally designed for degraded broadcast channels, have been shown to be a good solution to our problem. Indeed, a correspondence exists between the original problem of degraded broadcast channels and our problem. In degraded broadcast channels with m components, superposition codes are known capacity-achieving codes to transmit m independent source messages to m receivers. In our problem, when the source with arbitrary utility graph is given and the communication is over degraded compound channels, for a fixed sub-chain ϕ of length m_ϕ, $M_{\Delta(1)}, M_{\Delta(2)}, \ldots, M_{\Delta(m_\phi)}$ are the m_ϕ source messages to be encoded. The set of all channel realizations of the degraded

[2]In this case, for any $a \in \{1, \ldots, 2^{n\sum_{l\in\Delta(i)} R_l}\}$, $h_i(a) = 1$.

compound channel is divided into $m_\phi + 1$ subsets. Except the one over which nothing is decodable, each subset contains channel realizations over which exactly the message set M_{τ_i} is decoded for $i = 1, \ldots, m_\phi$, resembling each of m receivers under degraded broadcast channels.

In general, from (5.4), it is difficult to evaluate the maximal achievable utility by superposition codes since it depends on the source $(H_\mathcal{N}, \kappa)$, the compound channel q, and the delay factor r. For a given ϕ, the utility vector $(\kappa(\alpha_1), \ldots, \kappa(\alpha_{m_\phi}))$ for the sequence of nodes along ϕ is given. The maximization taken over all possible λ_ϕ and ψ_ϕ is indeed to find a vector of probability mass with component $\Pr(\theta : \theta^i \prec \theta \prec \theta^{i+1}),\ i = 1, \ldots, m_\phi$ such that its inner product with the vector $(\kappa(\tau_1), \ldots, \kappa(\tau_{m_\phi}))$ is maximized. For given source $(H_\mathcal{N}, \kappa)$ and the delay factor r, the vector of probability mass depends on both the superposition codes and the degraded compound channel q, while the vector of utility depends on the superposition codes. To find the maximum in (5.4), we are to find the best element $\phi \in \Phi$ to encode and the best superposition code to encode that sub-chain, where the best is chosen in terms of maximizing the inner product of the utility vector and the vector of assigned probability mass.

From the practical point of view, due to the complexity of the encoder, if the superposition codes are limited to no more than L layers, the maximization of the achievable utility becomes searching for the best sub-chain up to L nodes and the best possible superposition code to encode it.

5.2.3 The special case of conventional channel codes

When the length of ϕ is reduced to one, i.e., $m_\phi = 1$, superposition coding scheme in fact reduces to the conventional channel coding scheme which encodes a single node of messages α_1 in the utility graph. In this case, the Markov chain specifying the codes becomes $X_{\alpha_1} - Y_{\alpha_1}$, and the distributions λ_ϕ and ψ_ϕ are simplified to $P_{X_{\alpha_1}}$ and $P_{Y_{\alpha_1}|X_{\alpha_1}}$ respectively. Let V denote the set of all nodes in the graph, and Σ^{**} denote the maximal achievable utility by the conventional channel codes. By (5.4), we have

$$\Sigma^{**} = \max_{\phi \in V} \max_{P_{X_{\alpha_1}}, P_{Y_{\alpha_1}|X_{\alpha_1}}} \kappa(\tau_1)\Pr(\theta : \theta \succ \theta^1).$$

In the sequel, we consider the maximal achievable utility for encoding any given $\phi \in V$. We first consider the sequence of C'_n codes where the distributions $P_{X_{\alpha_1}}$ and $P_{Y_{\alpha_1}|X_{\alpha_1}}$ are such that $I(X_{\alpha_1}; Y_{\alpha_1}) = \mathrm{Cap}(\theta^1)$ and $\sum_{l \in \alpha_1} R_l = \mathrm{Cap}(\theta^1)$ hold. Let $P'_{X_{\alpha_1}}$ and $P'_{Y_{\alpha_1}|X_{\alpha_1}}$ denote such $P_{X_{\alpha_1}}$ and $P_{Y_{\alpha_1}|X_{\alpha_1}}$ respectively, and θ' denote such θ^1. Hence $P'_{X_{\alpha_1}}$ is the capacity achieving input distribution for channel θ'. In this case, $\kappa(\tau_1) = \kappa(\alpha_1)$ and we have

$$\lim_{n \to \infty} \mu(C'_n) = \kappa(\alpha_1)\mathrm{Pr}(\theta : \theta \succ \theta'). \tag{5.5}$$

For the sequence of codes C_n where $P_{X_{\alpha_1}} \neq P'_{X_{\alpha_1}}$ and $P_{Y_{\alpha_1}|X_{\alpha_1}} = P'_{Y_{\alpha_1}|X_{\alpha_1}}$, we have $I(X_{\alpha_1}; Y_{\alpha_1}) < \mathrm{Cap}(\theta')$ and therefore $\sum_{l \in \alpha_1} R_l > I(X_{\alpha_1}; Y_{\alpha_1})$. In this case $\kappa(\tau_1) = 0$ and we have

$$\lim_{n \to \infty} \mu(C_n) = 0. \tag{5.6}$$

Let Y' denote the channel output of θ'. For the sequence of codes C_n where $P_{X_{\alpha_1}}$ and $P_{Y_{\alpha_1}|X_{\alpha_1}}$ are such that $\theta^1 \prec \theta'$ holds, by data processing inequality, we have

$$I(X_{\alpha_1}; Y_{\alpha_1}) < I(X_{\alpha_1}; Y') \leq \mathrm{Cap}(\theta'),$$

and therefore $\sum_{l \in \alpha_1} R_l > I(X_{\alpha_1}; Y_{\alpha_1})$. In this case $\kappa(\tau_1) = 0$ and we have

$$\lim_{n \to \infty} \mu(C_n) = 0. \tag{5.7}$$

For the sequence of codes C_n where $P_{X_{\alpha_1}}$ and $P_{Y_{\alpha_1}|X_{\alpha_1}}$ are such that $\theta^1 \succ \theta'$ holds, we have

$$\begin{aligned} \lim_{n \to \infty} \mu(C_n) &= \kappa(\tau_1)\mathrm{Pr}(\theta : \theta \succ \theta^1) \\ &< \kappa(\alpha_1)\mathrm{Pr}(\theta : \theta \succ \theta'). \end{aligned} \tag{5.8}$$

By (5.5)-(5.8), it can be seen that for any given $\phi \in V$, the maximal achievable utility is that in (5.5). Therefore, we have

$$\begin{aligned} \Sigma^{**} &= \max_{\phi \in V} \kappa(\alpha_1)\mathrm{Pr}(\theta : \theta \succ \theta') \\ &= \max_{\alpha_1} \kappa(\alpha_1)\mathrm{Pr}(\theta : \theta \succ \theta'), \end{aligned}$$

where θ' is such that $\sum_{l\in\alpha_1} H_l/r = \text{Cap}(\theta')$ holds. To that end, we have the following results.

Corollary 1 *Given the source $(H_{\mathcal{N}}, \kappa)$ over degraded compound channel q and the delay factor r, Σ^{**} is the maximal achievable utility by conventional channel codes.*

Corollary 1 states that for any given source $(H_{\mathcal{N}}, \kappa)$ communicated over degraded compound channel q with delay factor r, the maximal utility by conventional channel codes can be achieved by choosing the best node in the utility graph and using a capacity-achieving channel code to encode it. Since the conventional channel codes are special case of superposition codes, it is obvious that $\Sigma^{**} \leq \Sigma^{*}$.

5.2.4 Optimality of superposition codes

In Section 5.2.2, we show that Σ^{*} in (5.4) is the maximal utility achieved by superposition codes for any given source $(H_{\mathcal{N}}, \kappa)$ communicated over degraded compound channel q with the delay factor r, and also provide the optimal superposition coding scheme to achieve Σ^{*}. In this section, we investigate the optimality of superposition codes. That is, whether Σ^{*} is indeed the maximal achievable utility among all coding schemes. We have the following results for the case $N = 2$.

Theorem 5 *For $N = 2$, given the source $(H_{\mathcal{N}}, \kappa)$ over degraded compound channel q with the delay factor r, the utility achieved by any code is upper bounded by that achieved by superposition codes.*

Proof: Suppose there exists a sequence of codes $\hat{C}_n$ whose decoding profile converges to $\hat{\gamma}$ as $n \to \infty$. Let $\hat{\Sigma}$ be the achievable utility by $\hat{C}_n$. Let function $\eta(\alpha|C,\theta)$ of any non-empty $\alpha \subseteq \mathcal{N}$ be defined as follows,

$$\eta(\alpha|C,\theta) := \sum_{\beta:\beta\supseteq\alpha \text{ and } \beta\subseteq\mathcal{N}} \gamma(\beta|C,\theta). \tag{5.9}$$

For every non-empty $\alpha \subseteq \mathcal{N}$, $\eta(\alpha|C,\theta)$ is the probability that M_α is decoded over θ. We consider only *normal decoder*, i.e., if $\theta' \succ \theta''$, for any $\alpha \subseteq \mathcal{N}$, $\eta(\alpha|C,\theta') \geq \eta(\alpha|C,\theta'')$ holds. We suppress the role of code C when there exists no ambiguity. Let $\hat{\eta}$ be the

η function induced by $\hat{\gamma}$. Furthermore, let the threshold $\theta_\alpha : X \to Y_\alpha \in \Theta$ denote the channel realization such that $\hat{\eta}(\alpha|\theta) = 0$ for $\theta \prec \theta_\alpha$, and $\hat{\eta}(\alpha|\theta) > 0$ for $\theta \succeq \theta_\alpha$. Clearly, we have $\theta_\alpha \preceq \theta_\beta$ if $\alpha \subset \beta$. For $N = 2$, both $\theta_{\{1\}} \preceq \theta_{\{1,2\}}$ and $\theta_{\{2\}} \preceq \theta_{\{1,2\}}$ hold if all such thresholds exist.

If $\hat{\eta}(\{1,2\}|\theta) = 0$ for any $\theta \in \Theta$, by (5.9), we know the decoding profile satisfies $\hat{\gamma}(\{1,2\}|\theta) = 0$, and $\hat{\gamma}(\{i\}|\theta) = \hat{\eta}(\{i\}|\theta),\ i = 1,2$ for any $\theta \in \Theta$. By the strong converse of the channel coding theorem, we have $R_1 < \mathrm{Cap}(\theta_{\{1\}})$. By the achievability of the channel coding theorem, there exists a sequence of codes $C_{1,n}$ encoding M_1 only which has the probability of error for decoding M_1 under $\theta_{\{1\}}$ approaching zero as $n \to \infty$. Let $C_{1,n}$ be applied over degraded compound channel q, and the decoding profile γ_1 satisfies $\gamma_1(\emptyset|\theta) = 1$ for $\theta \prec \theta_{\{1\}}$ and $\gamma_1(\{1\}|\theta) = 1$ for $\theta \succ \theta_{\{1\}}$. Similarly, there exists a sequence of codes $C_{2,n}$ encoding M_2 only which has the decoding profile γ_2 satisfying $\gamma_2(\emptyset|\theta) = 1$ for $\theta \prec \theta_{\{2\}}$ and $\gamma_2(\{2\}|\theta) = 1$ for $\theta \succ \theta_{\{2\}}$. Let Σ_1 and Σ_2 be the utility achieved by $C_{1,n}$ and $C_{2,n}$ respectively. By (5.1), we have $\Sigma_1 = \kappa(\{1\}) \int_{\theta:\theta\succ\theta_{\{1\}}} q(\theta)d\theta$ and $\Sigma_2 = \kappa(\{2\}) \int_{\theta:\theta\succ\theta_{\{2\}}} q(\theta)d\theta$. Let $a := \sup_{\theta:\theta\in\Theta} \hat{\gamma}(\{1\}|\theta)$ and $b := \sup_{\theta:\theta\in\Theta} \hat{\gamma}(\{2\}|\theta)$. Since $\hat{\gamma}(\{i\}|\theta') \geq \hat{\gamma}(\{i\}|\theta''),\ i = 1,2$ holds for $\theta' \succ \theta''$, and $\hat{\gamma}(\{1\}|\theta) + \hat{\gamma}(\{2\}|\theta) \leq 1$ for any θ, we must have $a + b \leq 1$. It can be shown as follows that $\hat{\Sigma}$ is upper bounded by the utility achieved by time sharing of $C_{1,n}$ and $C_{2,n}$ with fraction a and b respectively, and further by the maximum of Σ_1 and Σ_2,

$$\begin{aligned}
\hat{\Sigma} &= \int\limits_{\theta:\theta\succ\theta_{\{1\}}} \hat{\gamma}(\{1\}|\theta)\kappa(\{1\})q(\theta)d(\theta) + \int\limits_{\theta:\theta\succ\theta_{\{2\}}} \hat{\gamma}(\{2\}|\theta)\kappa(\{2\})q(\theta)d(\theta) \\
&\leq a\kappa(\{1\}) \int\limits_{\theta:\theta\succ\theta_{\{1\}}} q(\theta)d(\theta) + b\kappa(\{2\}) \int\limits_{\theta:\theta\succ\theta_{\{2\}}} q(\theta)d(\theta) \\
&\leq \max(\Sigma_1, \Sigma_2).
\end{aligned}$$

Otherwise, if $\theta_{\{1\}} \preceq \theta_{\{1,2\}}$ and $\theta_{\{2\}} \preceq \theta_{\{1,2\}}$, by (5.9), the decoding profile satisfies $\hat{\gamma}(\{1,2\}|\theta) = 0$, and $\hat{\gamma}(\{i\}|\theta) = \hat{\eta}(\{i\}|\theta),\ i = 1,2$ for any $\theta \prec \theta_{\{1,2\}}$. We consider applying $\hat{C}_n$ over the degraded broadcast channel $X \to Y_{\{1,2\}} \to Y_{\{1\}}$. By the strong converse and the weak converse of the coding theorem for degraded broadcast channels [122], there exists U such that $U \to X \to Y_{\{1,2\}} \to Y_{\{1\}}$ form a Markov chain with $R_1 \le I(U;Y_{\{1\}})$ and $R_2 \le I(X;Y_{\{1,2\}}|U)$. By Lemma 3, we are able to construct a sequence of superposition codes $C_{1,n}$ encoding M_1 first and M_2 superimposed which has the decoding profile γ_1 satisfying $\gamma_1(\emptyset|\theta) = 1$ for $\theta \prec \theta_{\{1\}}$, $\gamma_1(\{1\}|\theta) = 1$ for $\theta_{\{1\}} \prec \theta \prec \theta_{\{1,2\}}$, and $\gamma_1(\{1,2\}|\theta) = 1$ for $\theta \succ \theta_{\{1,2\}}$. Similarly, there exists a sequence of superposition codes $C_{2,n}$ encoding M_2 first and M_1 superimposed which has the decoding profile γ_2 satisfying $\gamma_2(\emptyset|\theta) = 1$ for $\theta \prec \theta_{\{2\}}$, $\gamma_2(\{2\}|\theta) = 1$ for $\theta_{\{2\}} \prec \theta \prec \theta_{\{1,2\}}$, and $\gamma_2(\{1,2\}|\theta) = 1$ for $\theta \succ \theta_{\{1,2\}}$. Besides, there also exists a sequence of codes $C_{3,n}$ encoding both M_1 and M_2 which has the decoding profile γ_3 satisfying $\gamma_3(\emptyset|\theta) = 1$ for $\theta \prec \theta_{\{1,2\}}$ and $\gamma_3(\{1,2\}|\theta) = 1$ for $\theta \succ \theta_{\{1,2\}}$. Let Σ_1, Σ_2, and Σ_3 be the utility achieved by $C_{1,n}$, $C_{2,n}$, and $C_{3,n}$ respectively. By (5.1), we have $\Sigma_1 = \kappa(\{1\}) \int_{\theta:\theta_{\{1\}}\prec\theta\prec\theta_{\{1,2\}}} q(\theta)d\theta + \kappa(\{1,2\}) \int_{\theta:\theta\succ\theta_{\{1,2\}}} q(\theta)d\theta$, $\Sigma_2 = \kappa(\{2\}) \int_{\theta:\theta_{\{2\}}\prec\theta\prec\theta_{\{1,2\}}} q(\theta)d\theta + \kappa(\{1,2\}) \int_{\theta:\theta\succ\theta_{\{1,2\}}} q(\theta)d\theta$, and $\Sigma_3 = \kappa(\{1,2\}) \int_{\theta:\theta\succ\theta_{\{1,2\}}} q(\theta)d\theta$. Let $a := \sup_{\theta:\theta\prec\theta_{\{1,2\}}} \hat{\gamma}(\{1\}|\theta)$ and $b := \sup_{\theta:\theta\prec\theta_{\{1,2\}}} \hat{\gamma}(\{2\}|\theta)$. Since $\hat{\gamma}(\{i\}|\theta') \ge \hat{\gamma}(\{i\}|\theta'')$, $i = 1,2$ holds for $\theta' \succ \theta''$ and $\theta', \theta'' \prec \theta_{\{1,2\}}$, and $\hat{\gamma}(\{1\}|\theta) + \hat{\gamma}(\{2\}|\theta) \le 1$ for any $\theta \prec \theta_{\{1,2\}}$, we must have $a + b \le 1$. It can be shown as follows that $\hat{\Sigma}$ is upper bounded by the utility achieved by time sharing of $C_{1,n}$, $C_{2,n}$, and $C_{3,n}$ with fraction a, b, and $1-a-b$ respectively, and further by the maximum of Σ_1, Σ_2, and Σ_3,

$$\begin{aligned}
\hat{\Sigma} &= \int\limits_{\theta:\theta_{\{1\}}\prec\theta\prec\theta_{\{1,2\}}} \hat{\gamma}(\{1\}|\theta)\kappa(\{1\})q(\theta)d(\theta) + \int\limits_{\theta:\theta_{\{2\}}\prec\theta\prec\theta_{\{1,2\}}} \hat{\gamma}(\{2\}|\theta)\kappa(\{2\})q(\theta)d(\theta) \\
&\quad + \int\limits_{\theta:\theta\succ\theta_{\{1,2\}}} \left(\hat{\gamma}(\{1\}|\theta)\kappa(\{1\}) + \hat{\gamma}(\{2\}|\theta)\kappa(\{2\}) + \hat{\gamma}(\{1,2\}|\theta)\kappa(\{1,2\})\right) q(\theta)d(\theta) \\
&\le a\kappa(\{1\}) \int\limits_{\theta:\theta_{\{1\}}\prec\theta\prec\theta_{\{1,2\}}} q(\theta)d(\theta) + b\kappa(\{2\}) \int\limits_{\theta:\theta_{\{2\}}\prec\theta\prec\theta_{\{1,2\}}} q(\theta)d(\theta) + \kappa(\{1,2\}) \int\limits_{\theta:\theta\succ\theta_{\{1,2\}}} q(\theta)d(\theta) \\
&= a\left(\kappa(\{1\}) \int\limits_{\theta:\theta_{\{1\}}\prec\theta\prec\theta_{\{1,2\}}} q(\theta)d\theta + \kappa(\{1,2\}) \int\limits_{\theta:\theta\succ\theta_{\{1,2\}}} q(\theta)d\theta\right) + b\left(\kappa(\{2\}) \int\limits_{\theta:\theta_{\{2\}}\prec\theta\prec\theta_{\{1,2\}}} q(\theta)d\theta\right. \\
&\quad \left. + \kappa(\{1,2\}) \int\limits_{\theta:\theta\succ\theta_{\{1,2\}}} q(\theta)d\theta\right) + (1-a-b)\,\kappa(\{1,2\}) \int\limits_{\theta:\theta\succ\theta_{\{1,2\}}} q(\theta)d(\theta) \\
&\le \max(\Sigma_1, \Sigma_2, \Sigma_3).
\end{aligned}$$

□

Theorem 5, together with Theorem 3, state that Σ^* in (5.4) is the maximal achievable utility among all coding schemes. For communicating any given source $(H_{\mathcal{N}}, \kappa)$ over degraded compound channel q with delay factor r when $N = 2$, superposition codes are optimal.

5.3 Numerical results

Same as the coded sources used in Chapter 4, for the examples of an LC-coded and an MDC-coded source in Example 4, which are constructed based on [73], we consider communicating such sources over the degraded compound channel with specific q distribution, and numerically evaluate the maximal achievable utilities by superposition codes.

Let θ be the crossover probability of BSC and let distribution q be $q(\theta) = 2\text{Beta}(2\theta; 2, 6)$, where $\text{Beta}(x; \alpha, \beta)$ denotes the pdf of the beta distribution parametrized by α and β for $0 \leq x \leq 1$. Figures 5.1 and 5.2 show the maximal achievable PSNR by superposition codes at varying delay factor r for LC-coded and MDC-coded source respectively, in which the curves of $R_{\{1\}}$ and $R_{\{2\}}$ represent the maximal achievable PSNR for the cases of trivializing message M_2 and M_1 respectively, the one of $R_{\{1,2\}}$ represents the maximal achievable PSNR for conventional channel code encoding both M_1 and M_2, and the one of $R_{\{1\}} \to R_{\{2\}}$ represents that for the non-trivial case. In the curve of $R_{\{1\}} \to R_{\{2\}}$, there exists point of discontinuity which results from the fact that two different configurations of $(P_{X_{\{1\}}}, P_{X_{\{2\}}|X_{\{1\}}}, \theta^1 \prec \theta^2)$ both maximize the achievable PSNR at that point.

It can be seen that for both sources, the maximal achievable PSNR by superposition codes is the envelop of the four curves. This fact coincides with the results of Theorem 3 and 4. For LC-coded source, since $\kappa(\{2\}) = 0$, the maximal achievable PSNR by codes trivializing message M_1 is zero, while for MDC-coded source, since $H_1 = H_2$ and $\kappa(\{1\}) = \kappa(\{2\})$, the maximal achievable PSNR by codes trivializing M_1 and M_2 respectively are equal. For both sources, the maximal achievable PSNR increases with the delay factor r, and becomes to deviate from zero when the delay factor r is such that the channel code rate of a single message becomes supported by some channel realizations. It requires long delay for both codes to reconstruct the source perfectly (approaching their

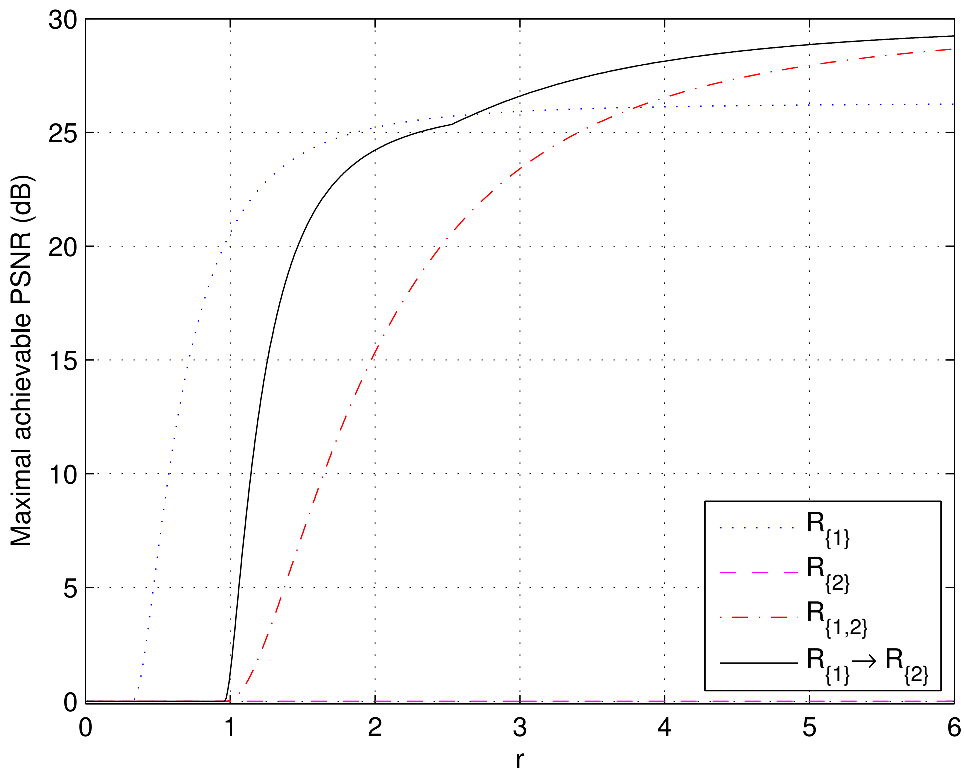

Figure 5.1: Maximal achievable PSNR at various delay factor r over degraded compound channel for LC-coded source.

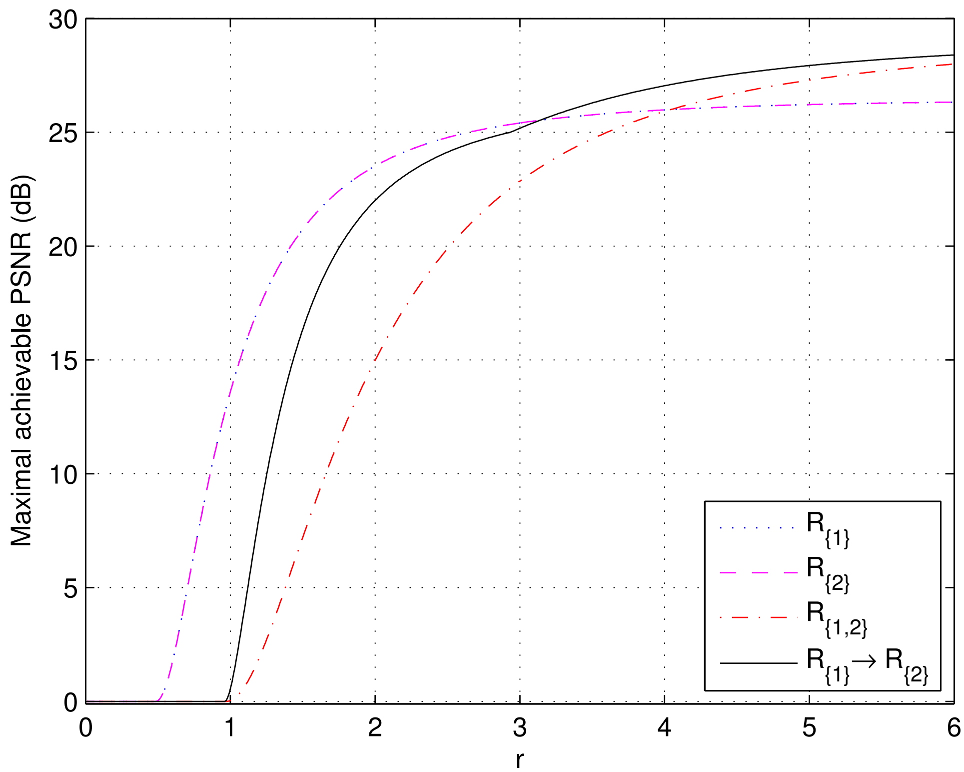

Figure 5.2: Maximal achievable PSNR at various delay factor r over degraded compound channel for MDC-coded source.

respective maximal PSNR).

For both sources, the maximal achievable PSNR by conventional channel codes, as the special case of superposition codes, is the envelop of the curves of $R_{\{1\}}$, $R_{\{2\}}$, and $R_{\{1,2\}}$, which correspond to encoding M_1, M_2, and both M_1 and M_2 together respectively. The fact coincides with the result of Corollary 1. The maximal achievable PSNR by conventional channel codes increases with the delay factor r as well. Apparently, the maximal achievable PSNR of conventional channel codes is dominated by that of superposition codes.

Figure 5.3 shows the comparison of maximal achievable PSNR by superposition codes between the LC-coded and the MDC-coded sources. For this particular case, it can be seen that the LC achieves better quality of source reconstruction than the MDC. However, for other cases of sources and degraded compound channels, it may be different. For any source communicated over any degraded compound channel, the maximal achievable utility by superposition codes can be numerically evaluated.

On the other hand, we are interested in the maximal achievable utility at different degraded compound channels. Figure 5.4 shows the examples of $q(\theta) = 2\text{Beta}(2\theta; 2, \beta)$ for some β. It appears that as β increases, the distribution $q(\theta)$ is shifted to low value of crossover probability θ, i.e., less noisy BSC channels.

Figures 5.5 and 5.6 show the maximal achievable PSNR at degraded compound channel $q(\theta) = 2\text{Beta}(2\theta; 2, \beta)$ with varying β by superposition codes when the delay factor $r = 2$ for LC-coded and MDC-coded source respectively. It can be seen that for both sources, the maximal achievable PSNR by superposition codes is the envelop of the four curves by codes encoding M_1 only, M_2 only, both M_1 and M_2 treated as one message, and M_1 first with M_2 superimposed respectively. It appears that the maximal achievable PSNR increases with β. That is, as the major probability mass of q is shifted to less noisy channel realizations, the maximal achievable PSNR increases.

For both sources, the maximal achievable PSNR by conventional channel codes, as the special case of superposition codes, is the envelop of the curves by codes encoding M_1 only, M_2 only, and both M_1 and M_2 treated as a single message respectively. It appears that the maximal achievable PSNR by conventional channel codes increases with β as well. Apparently, since conventional channel codes are special case of superposition

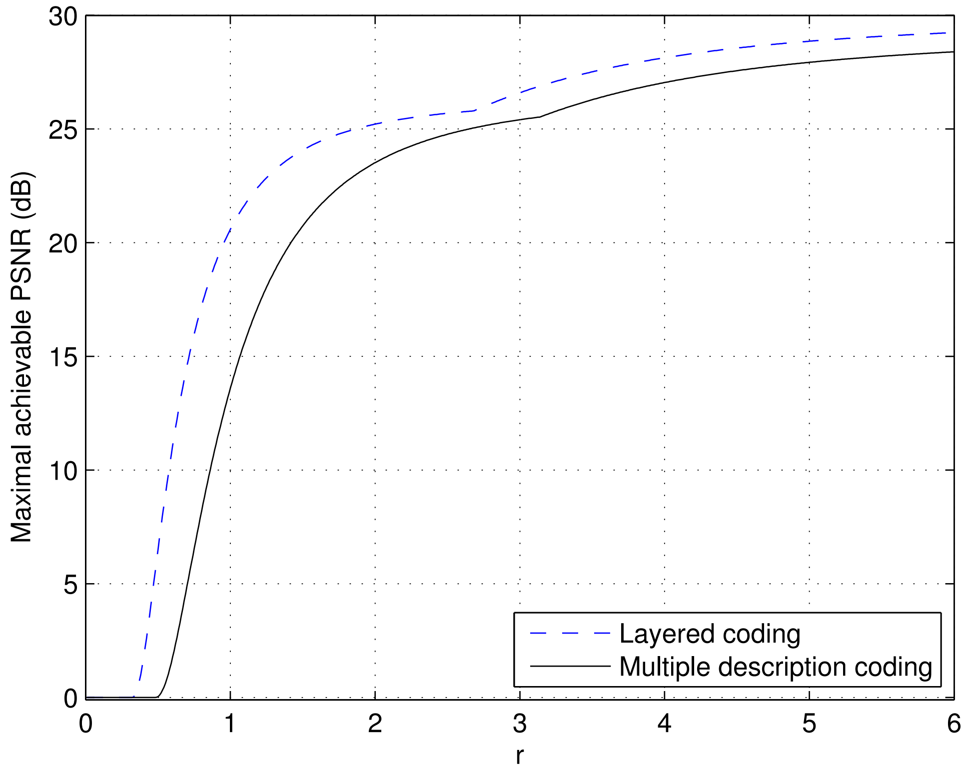

Figure 5.3: Maximal achievable PSNR at various delay factor r over degraded compound channel by superposition codes.

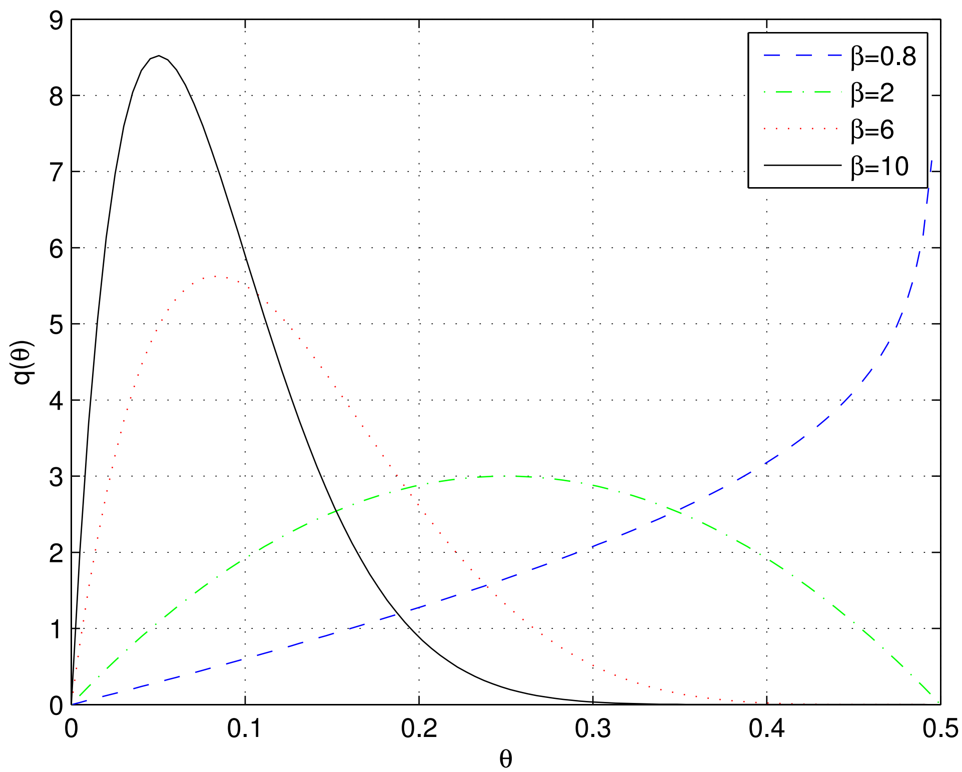

Figure 5.4: Examples of $q(\theta)$ for different β.

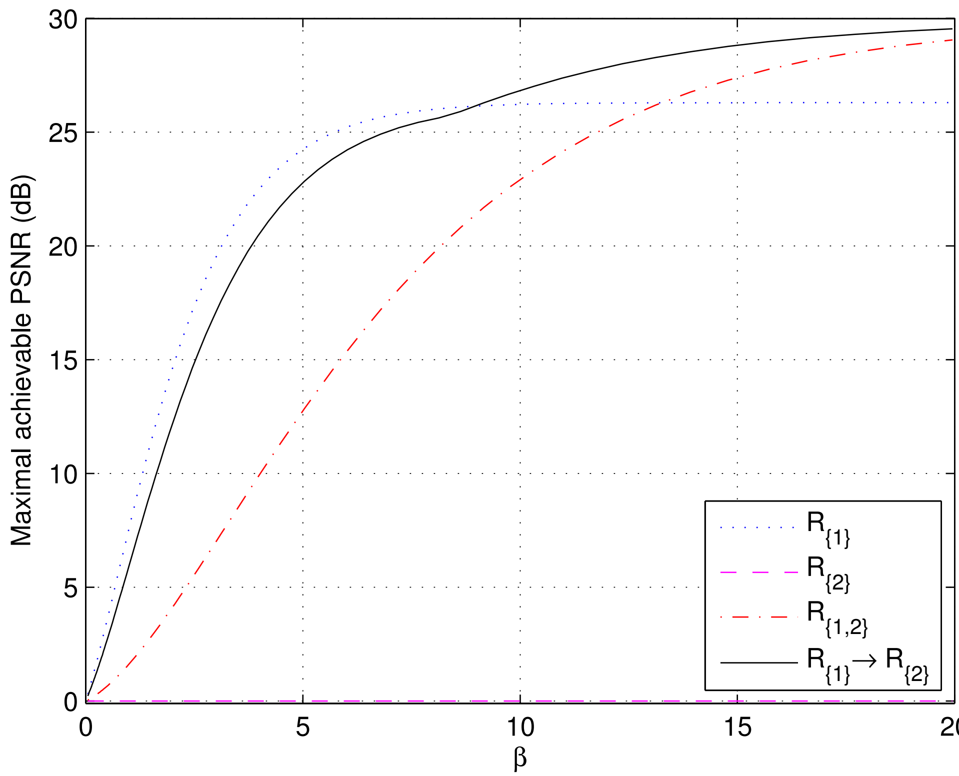

Figure 5.5: Maximal achievable PSNR at different degraded compound channels when $r = 2$ for LC-coded source.

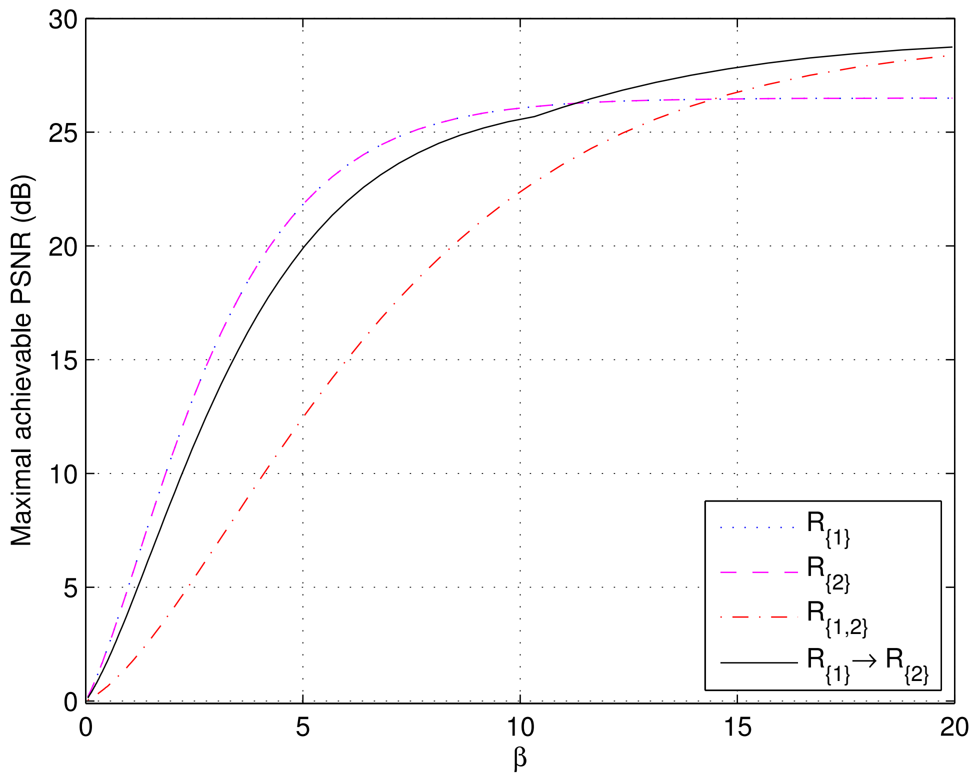

Figure 5.6: Maximal achievable PSNR at different degraded compound channels when $r = 2$ for MDC-coded source.

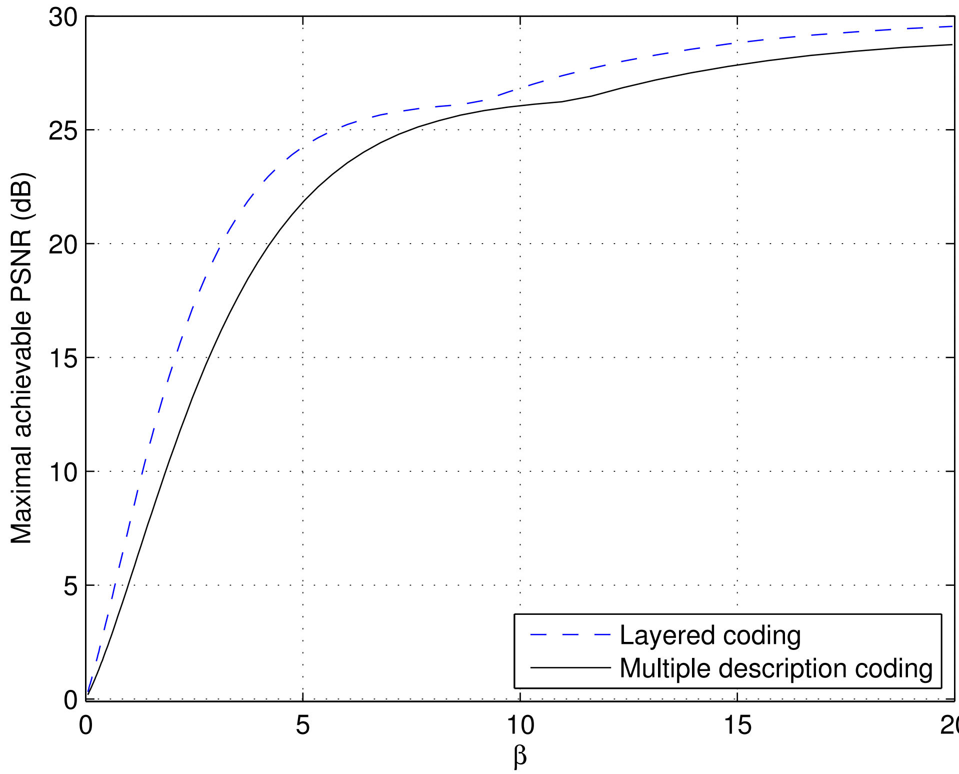

Figure 5.7: Maximal achievable PSNR at different degraded compound channels by superposition codes for $r = 2$.

codes, the maximal achievable utility of conventional channel codes is dominated by that of superposition codes.

Figure 5.7 shows the maximal achievable PSNR by superposition codes for the LC-coded and the MDC-coded sources over different degraded compound channels for $r = 2$. For this particular case, it appears that the LC achieves better quality of source reconstruction than the MDC. However, the comparison behaviour observed in the figure may not occur in other cases of source and degraded compound channel.

5.4 Conclusion

In this chapter we formulate the problem of communicating source-coded messages for multimedia content over degraded compound channels as the problem of maximizing the average utility at the receiver. The performance of superposition coding schemes, which include the conventional channel code special cases, is investigated. We show that with superposition codes, the maximal utility is achieved by coding along the best sub-chain in the utility graph, and when restricting the superposition codes to the conventional channel code family, the utility is maximized by a capacity-achieving channel code over the best node. For the case of two source-coded messages, we show that the maximal utility achievable by superposition codes is the maximum among all coding schemes. We numerically evaluate the maximal achievable PSNR for LC-coded and MDC-coded messages under superposition coding schemes and demonstrate the impact of the channel statistics and delay on the achieved utility. Practically, when a more complicated source-coded media file with an arbitrary dependency structure among messages is to be transmitted over a wireless fading channel or packet-loss channel, we are able to evaluate the maximal throughput, maximal average quality on media reconstruction, or maximal average user satisfaction, and provide the optimal coding scheme to achieve that as well.

In the problem of communicating dependent source messages over degraded compound channels, we conjecture that the maximal utility achievable by superposition codes is in fact the maximum among *all* coding schemes for arbitrary number of source-coded messages. Proving or disproving this conjecture remains as an interesting future work.

Chapter 6

Joint Source Channel Coding of Gaussian Source

In last chapter, we investigate the problem of communicating given source-coded messages over degraded compound channels. Based on the numerical studies, the maximal achievable utility by superposition codes can be evaluated and the optimizing coding scheme is therefore provided. However, the examples of the LC-coded and the MDC-coded source used in Chapter 4 and 5 may not be constructed from a single source, which makes the performance of both source coding techniques not comparable. To better understand the nature of different source coding techniques, and further compare their performance, we heuristically consider communicating a single source over degraded compound channels by joint source channel coding. Specifically, in this chapter we consider the source sequence as a Gaussian $\mathcal{N}(0,1)$ i.i.d. sequence.

6.1 Problem formulation

Instead of considering all families of source and channel coding schemes, on source coding side, we consider rate distortion coding, LC, and MDC schemes respectively, while on channel coding side, we focus on special cases of channel codes, namely, conventional channel codes and superposition codes. We consider communicating the Gaussian $\mathcal{N}(0,1)$

source sequence by the combination of source and channel codes through degraded compound channels described in Chapter 5. Let θ be the parameter of the channel realization, and the degraded compound channel is defined via a distribution q over a space of θ. Let Θ denote the support of q. We assume each $\theta \in \Theta$ parametrizes a DMC. For any two channel realizations in Θ, one is degraded with respect to another. Similar to that discussed in Chapter 5, we consider q being a continuous distribution. In that case, the probability that a specific channel realization θ is chosen becomes zero.

The communication over degraded compound channels is also subject to a "delay factor" constraint. With the delay factor r, we consider using the channel rn_s times for communicating the source sequence. In real-time communication scenario, such as in a video streaming scenario, when $r < 1$, it is possible to deliver the source sequence to the receiver without delay, and otherwise the receiver will necessarily suffer from some delay when reconstructing the source sequence.

With the notations introduced in Chapter 3, the constructions of the encoded source messages by different source codes under consideration are described as follows, together with the characterization of their corresponding information theoretic bounds. Let the source sequence or process $\mathbf{X} := X_1, X_2, \ldots, X_{n_s}$ be i.i.d. Gaussian random variables with mean 0 and variance 1, and $\hat{\mathbf{X}} := \hat{X}_1, \hat{X}_2, \ldots, \hat{X}_{n_s}$ denote the reconstruction of the sequence $\mathbf{X}$. The squared error distortion between sequences $\mathbf{x}$ and $\hat{\mathbf{x}}$ is defined by

$$d(\mathbf{x}, \hat{\mathbf{x}}) := \frac{1}{n_s} \sum_{i=1}^{n_s} (x_i - \hat{x}_i)^2. \tag{6.1}$$

In rate distortion coding scheme, essentially we fix the number of encoded messages $N = 1$. By rate distortion codes, the source sequence $\mathbf{X}$ is encoded into an index $M_1(\mathbf{X}) \in \{1, \ldots, 2^{n_s H_1}\}$, and is reconstructed by function $\hat{\mathbf{X}}(M_1)$, where H_1 is the rate of the code. A rate distortion pair (H_1, D_1) is said to be achievable if there exists a sequence of rate distortion codes, parametrized by n_s, with $\lim_{n_s \to \infty} \mathrm{E} d(X, \hat{X}) \leq D_1$, where the expectation is with respect to the distribution $\mathcal{N}(0,1)$. Based on the rate distortion theory [54], the rate distortion region formed by all achievable pairs (H_1, D_1) is such that

$$H_1 \geq \frac{1}{2} \log \frac{1}{D_1} \tag{6.2}$$

holds for $0 \leq D_1 \leq 1$.

In LC scheme, as that shown in Figure 3.1(a), we choose $N = 2$, i.e., two layers. By LC, $\mathbf{X}$ is encoded into the base layer message $M_1(\mathbf{X}) \in \{1, \ldots, 2^{n_s H_1}\}$ and the enhancement layer message $M_2(\mathbf{X}) \in \{1, \ldots, 2^{n_s H_2}\}$, and is reconstructed by function $\hat{\mathbf{X}}^1(M_1)$ or $\hat{\mathbf{X}}^0(M_1, M_2)$ depending on the availability of M_1 and M_2, where H_1 and H_2 are the rates of based layer and enhancement layer respectively. The tuple (H_1, H_2, D_1, D_0) is said to be achievable if there exists a sequence of LC, parametrized by n_s, with $\lim_{n_s \to \infty} \mathrm{E}d(X, \hat{X}^l) \leq D_l,\ l = 0, 1$, where the expectation is with respect to the distribution $\mathcal{N}(0, 1)$. It has been shown in [16] that the rate distortion region formed by all achievable tuples (H_1, H_2, D_1, D_0) is such that

$$\begin{cases} H_1 \geq \frac{1}{2} \log \frac{1}{D_1} \\ H_1 + H_2 \geq \frac{1}{2} \log \frac{1}{D_0} \end{cases} \tag{6.3}$$

holds for $0 \leq D_0 < D_1 \leq 1$.

In MDC scheme, as that shown in Figure 3.1(b), we choose $N = 2$, i.e., two descriptions. By MDC, $\mathbf{X}$ is encoded into the first description $M_1(\mathbf{X}) \in \{1, \ldots, 2^{n_s H_1}\}$ and the second description $M_2(\mathbf{X}) \in \{1, \ldots, 2^{n_s H_2}\}$, and is reconstructed by function $\hat{\mathbf{X}}^1(M_1)$, $\hat{\mathbf{X}}^2(M_2)$, or $\hat{\mathbf{X}}^0(M_1, M_2)$ depending on the availability of M_1 and M_2, where H_1 and H_2 are the rates of two descriptions. The tuple $(H_1, H_2, D_1, D_2, D_0)$ is said to be achievable if there exists a sequence of MDC, parametrized by n_s, with $\lim_{n_s \to \infty} \mathrm{E}d(X, \hat{X}^l) \leq D_l,\ l = 0, 1, 2$, where the expectation is with respect to the distribution $\mathcal{N}(0, 1)$. Two scenarios, namely, high distortion and low distortion, exist in the MDC of Gaussian $\mathcal{N}(0, 1)$ source [33]. In the case of high distortion where $D_1 + D_2 - D_0 \geq 1$ is required, the rate distortion region formed by all achievable $(H_1, H_2, D_1, D_2, D_0)$ is such that

$$\begin{cases} H_1 \geq \frac{1}{2} \log \frac{1}{D_1} \\ H_2 \geq \frac{1}{2} \log \frac{1}{D_2} \\ H_1 + H_2 \geq \frac{1}{2} \log \frac{1}{D_0} \end{cases} \tag{6.4}$$

holds. In the case of low distortion where $D_1 + D_2 - D_0 < 1$ is required, the rate distortion region is such that

$$\begin{cases} H_1 \geq \frac{1}{2}\log\frac{1}{D_1} \\ H_2 \geq \frac{1}{2}\log\frac{1}{D_2} \\ H_1 + H_2 \geq \frac{1}{2}\log\frac{1}{D_0} + \frac{1}{2}\log\frac{(1-D_0)^2}{(1-D_0)^2-((1-D_1)^{1/2}(1-D_2)^{1/2}-(D_1-D_0)^{1/2}(D_2-D_0)^{1/2})^2} \end{cases} \tag{6.5}$$

holds. In both scenarios, apparently we require $0 \leq D_0 < D_1 \leq 1$ and $0 \leq D_0 < D_2 \leq 1$. For convenience, let

$$s(D_1, D_0) := \log\frac{(1-D_0)^2}{(1-D_0)^2-(1+D_0-2D_1)^2}. \tag{6.6}$$

For any fixed source coding techniques, generally let $(H_{\mathcal{N}}, r, n_s)$ denote a channel code defined in Chapter 5, and let the decoding profile of the channel code C over channel θ, $\gamma(\cdot|C,\theta)$ be well defined on the power set $2^{\mathcal{N}}$ of $\mathcal{N}$ characterizing the probability of successful decoding, as that in Chapter 4. Let $D(\alpha)$ denote the squared error distortion for source sequence reconstruction from M_α for $\alpha \subseteq \mathcal{N}$. For any given source code, i.e., function D, the *average distortion* $\nu(C)$ of a code C is defined as

$$\nu(C) := \int \sum_{\alpha\subseteq\mathcal{N}} \gamma(\alpha|C,\theta)D(\alpha)q(\theta)d\theta. \tag{6.7}$$

Given (D,q), we use $\nu(C)$ to measure the performance of code C.

Given an encoded Gaussian $\mathcal{N}(0,1)$ source, i.e., $(H_{\mathcal{N}}, D)$, over degraded compound channel q, the pair $(r,\mathcal{D})$, referred to as a *delay-distortion pair* is said to be achievable if there exists an $(H_{\mathcal{N}}, r, n_s)$ code whose average distortion is $\mathcal{D}$. For a given delay factor r, we refer to the infimum of $\mathcal{D}$ such that delay-distortion pair $(r,\mathcal{D})$ is achievable as the *minimal achievable distortion*. As the minimal distortion is defined in terms of a infimum, we restrict our attention to $(H_{\mathcal{N}}, r, n_s)$ code sequences (parametrized by n_s) whose decoding profile converges as n_s approaches infinity.

For the Gaussian $\mathcal{N}(0,1)$ source to be communicated over any arbitrary degraded compound channel q with given delay factor r, we are interested in finding the minimal achievable average distortion for different combination of source and channel coding schemes.

6.2 Results

Since in MDC, the source is typically encoded into "equal-footing" descriptions, we naturally choose $H_1 = H_2$ in our discussion and denote it by H.

6.2.1 Conventional channel codes and their achievable distortion

We first consider the case where the channel codes are conventional channel codes. In the conventional channel coding scheme, we select a message set M_α for some non-empty $\alpha \subseteq \mathcal{N}$, treat them as a single message, and code them using a capacity-achieving channel code. Let C_n be the sequence of conventional channel codes for message set M_α, and θ^* be the channel such that $\sum_{i\in\alpha} H_i/r = \mathrm{Cap}(\theta^*)$ holds. Furthermore, let $\mathcal{D}'$ be defined as

$$\mathcal{D}' = \min_{\alpha} D(\emptyset) - (D(\emptyset) - D(\alpha)) \Pr(\theta : \theta \succ \theta^*). \tag{6.8}$$

For Gaussian $\mathcal{N}(0,1)$ source, we have $D(\emptyset) = 1$. We apply C_n on $\theta \in \Theta$ and have the following results, the proof of which is provided in Appendix III.

Theorem 6 *Given the encoded source $(H_{\mathcal{N}}, D)$ over degraded compound channel q with delay factor r, $\mathcal{D}'$ is the minimal achievable distortion by conventional channel codes.*

When conventional channel codes are applied to communicate the rate distortion-coded Gaussian $\mathcal{N}(0,1)$ source, we have the following results. Let $\hat{\mathcal{D}}$ be defined as

$$\hat{\mathcal{D}} = \min_{\theta^*} D(\emptyset) - \left(D(\emptyset) - 2^{-2r\mathrm{Cap}(\theta^*)}\right) \Pr(\theta : \theta \succ \theta^*). \tag{6.9}$$

Theorem 7 *For the rate distortion-coded Gaussian $\mathcal{N}(0,1)$ source over degraded compound channel q with delay factor r, $\hat{\mathcal{D}}$ is the minimal achievable distortion by conventional channel codes.*

Proof: For rate distortion codes, by (6.8), we have the minimal achievable distortion

$$\begin{aligned}\mathcal{D} &= \min_H \min_D \mathcal{D}' \\ &= \min_H \min_D D(\emptyset) - (D(\emptyset) - D(\{1\}))\Pr\left(\theta : \theta \succ \mathrm{Cap}^{-1}\left(\frac{H}{r}\right)\right),\end{aligned}$$

where $\mathrm{Cap}^{-1}(\cdot)$ denotes the inverse capacity function. By (6.2), we have $D(\{1\}) \geq 2^{-2H}$ for any given H. Therefore, we have

$$\begin{aligned}\mathcal{D} &= \min_H D(\emptyset) - \left(D(\emptyset) - 2^{-2H}\right)\Pr\left(\theta : \theta \succ \mathrm{Cap}^{-1}\left(\frac{H}{r}\right)\right) \\ &= \min_{\theta^*} D(\emptyset) - \left(D(\emptyset) - 2^{-2r\mathrm{Cap}(\theta^*)}\right)\Pr(\theta : \theta \succ \theta^*).\end{aligned}$$

□

Theorem 7 suggests that if the channel code is a conventional channel code, the optimal rate distortion-coded Gaussian $\mathcal{N}(0,1)$ source to achieve the minimum distortion can be constructed at the achieving rate H with the minimum distortion 2^{-2H}.

When conventional channel codes are applied to communicate the LC-coded Gaussian $\mathcal{N}(0,1)$ source, we have the following results.

Theorem 8 *For the LC-coded Gaussian $\mathcal{N}(0,1)$ source over degraded compound channel q with delay factor r, $\hat{\mathcal{D}}$ is the minimal achievable distortion by conventional channel codes.*

Proof: For LC, by (6.8), we have the minimal achievable distortion

$$\begin{aligned}\mathcal{D} &= \min_H \min_D \mathcal{D}' \\ &= \min_H \min_D \min\Bigg(D(\emptyset) - (D(\emptyset) - D(\{1\}))\Pr\left(\theta : \theta \succ \mathrm{Cap}^{-1}\left(\frac{H}{r}\right)\right), \\ &\quad D(\emptyset) - (D(\emptyset) - D(\{1,2\}))\Pr\left(\theta : \theta \succ \mathrm{Cap}^{-1}\left(\frac{2H}{r}\right)\right)\Bigg) \\ &= \min_H \min\Bigg(\min_D D(\emptyset) - (D(\emptyset) - D(\{1\}))\Pr\left(\theta : \theta \succ \mathrm{Cap}^{-1}\left(\frac{H}{r}\right)\right), \\ &\quad \min_D D(\emptyset) - (D(\emptyset) - D(\{1,2\}))\Pr\left(\theta : \theta \succ \mathrm{Cap}^{-1}\left(\frac{2H}{r}\right)\right)\Bigg).\end{aligned}$$

By (6.3), we have $H \geq \max(\frac{1}{2}\log\frac{1}{D(\{1\})}, \frac{1}{4}\log\frac{1}{D(\{1,2\})})$, and therefore $D(\{1\}) \geq 2^{-2H}$ and $D(\{1,2\}) \geq 2^{-4H}$ hold for any given H. Thus we have

$$\begin{aligned}
\mathcal{D} &= \min_H \min\Bigg(D(\emptyset) - \big(D(\emptyset) - 2^{-2H}\big)\Pr\left(\theta : \theta \succ \mathrm{Cap}^{-1}\left(\frac{H}{r}\right)\right), \\
&\qquad D(\emptyset) - \big(D(\emptyset) - 2^{-4H}\big)\Pr\left(\theta : \theta \succ \mathrm{Cap}^{-1}\left(\frac{2H}{r}\right)\right)\Bigg) \\
&= \min\Bigg(\min_H D(\emptyset) - \big(D(\emptyset) - 2^{-2H}\big)\Pr\left(\theta : \theta \succ \mathrm{Cap}^{-1}\left(\frac{H}{r}\right)\right), \\
&\qquad \min_H D(\emptyset) - \big(D(\emptyset) - 2^{-4H}\big)\Pr\left(\theta : \theta \succ \mathrm{Cap}^{-1}\left(\frac{2H}{r}\right)\right)\Bigg) \\
&= \min\Big(\min_{\theta^*} D(\emptyset) - \big(D(\emptyset) - 2^{-2r\mathrm{Cap}(\theta^*)}\big)\Pr(\theta : \theta \succ \theta^*), \\
&\qquad \min_{\theta^*} D(\emptyset) - \big(D(\emptyset) - 2^{-2r\mathrm{Cap}(\theta^*)}\big)\Pr(\theta : \theta \succ \theta^*)\Big) \\
&= \min_{\theta^*} D(\emptyset) - \big(D(\emptyset) - 2^{-2r\mathrm{Cap}(\theta^*)}\big)\Pr(\theta : \theta \succ \theta^*).
\end{aligned}$$

□

Theorem 8 suggests that if only the base layer message is chosen to be encoded by conventional channel codes, the optimal LC-coded Gaussian $\mathcal{N}(0,1)$ source can be constructed at the achieving rate H together with the distortion function being specified via $D(\{1\}) = 2^{-2H}$ and $D(\{1,2\}) \geq 2^{-4H}$. On the other hand, if both the base layer and enhancement layer messages are chosen to be encoded by conventional channel codes, the optimal source can be constructed at the achieving rate H with the distortion function being specified via $D(\{1\}) \geq 2^{-2H}$ and $D(\{1,2\}) = 2^{-4H}$.

When conventional channel codes are applied to communicate the MDC-coded Gaussian $\mathcal{N}(0,1)$ source, we have the following results. We note that since $H_1 = H_2$ holds, by symmetry we consider $D(\{1\}) = D(\{2\})$.

Theorem 9 *For the MDC-coded Gaussian $\mathcal{N}(0,1)$ source over degraded compound channel q with delay factor r, $\hat{\mathcal{D}}$ is the minimal achievable distortion by conventional channel codes.*

Proof: For MDC, by (6.8), we have the minimal achievable distortion

$$\begin{aligned}\mathcal{D} &= \min_H \min_D \mathcal{D}' \\ &= \min_H \min_D \min\Bigg(D(\emptyset) - (D(\emptyset) - D(\{1\}))\Pr\left(\theta : \theta \succ \mathrm{Cap}^{-1}\left(\frac{H}{r}\right)\right), \\ &\quad D(\emptyset) - (D(\emptyset) - D(\{1,2\}))\Pr\left(\theta : \theta \succ \mathrm{Cap}^{-1}\left(\frac{2H}{r}\right)\right)\Bigg).\end{aligned}$$

By (6.4) and (6.5), we have $H \geq \max(\frac{1}{2}\log\frac{1}{D(\{1\})}, \frac{1}{4}\log\frac{1}{D(\{1,2\})})$ in the case of high distortion where $2D(\{1\}) - D(\{1,2\}) \geq 1$ holds, and

$$H \geq \max\left(\frac{1}{2}\log\frac{1}{D(\{1\})}, \frac{1}{4}\left(\log\frac{1}{D(\{1,2\})} + s(D(\{1\}), D(\{1,2\}))\right)\right)$$

in the case of low distortion where $2D(\{1\}) - D(\{1,2\}) < 1$ holds. Then, for any given $H > 0$, in the case of high distortion, we have $D(\{1,2\}) \geq 2^{-4H}$ and $D(\{1\}) \geq (1 + D(\{1,2\}))/2 > 2^{-2H}$ such that $(H, D(\{1\}), D(\{1,2\}))$ is achievable, while in the case of low distortion, we have $2^{-2H} \leq D(\{1\}) < (1 + D(\{1,2\}))/2$ and

$$H \geq \frac{1}{4}\left(\log\frac{1}{D(\{1,2\})} + s(D(\{1\}), D(\{1,2\}))\right), \tag{6.10}$$

such that $(H, D(\{1\}), D(\{1,2\}))$ is achievable. When $D(\{1\}) = 2^{-2H}$, we have $D(\{1,2\}) = (2 \cdot 2^{2H} - 1)^{-1}$. Therefore we have

$$\begin{aligned}\mathcal{D} &= \min\Bigg(\min_H \min_D D(\emptyset) - (D(\emptyset) - D(\{1\}))\Pr\left(\theta : \theta \succ \mathrm{Cap}^{-1}\left(\frac{H}{r}\right)\right), \\ &\quad \min_H \min_D D(\emptyset) - (D(\emptyset) - D(\{1,2\}))\Pr\left(\theta : \theta \succ \mathrm{Cap}^{-1}\left(\frac{2H}{r}\right)\right)\Bigg) \\ &= \min\Bigg(\min_H D(\emptyset) - \left(D(\emptyset) - 2^{-2H}\right)\Pr\left(\theta : \theta \succ \mathrm{Cap}^{-1}\left(\frac{H}{r}\right)\right), \\ &\quad \min_H D(\emptyset) - \left(D(\emptyset) - 2^{-4H}\right)\Pr\left(\theta : \theta \succ \mathrm{Cap}^{-1}\left(\frac{2H}{r}\right)\right)\Bigg) \\ &= \min_{\theta^*} D(\emptyset) - \left(D(\emptyset) - 2^{-2r\mathrm{Cap}(\theta^*)}\right)\Pr(\theta : \theta \succ \theta^*).\end{aligned}$$

□

For the MDC-coded Gaussian $\mathcal{N}(0,1)$ source, if a single description is chosen to be encoded by conventional channel codes, the optimal source can be constructed at the achieving rate H together with the distortion function being specified via $D(\{1\}) = 2^{-2H}$ and $D(\{1,2\}) = (2 \cdot 2^{2H} - 1)^{-1}$. The distortion function corresponds to the case of low distortion, where encoding the single description by conventional channel codes gives rise to relatively lower distortion than encoding both descriptions. On the other hand, if both descriptions are chosen to be encoded by conventional channel codes, the optimal source can be constructed at the achieving rate H together with the distortion function being specified via $D(\{1\}) \geq \frac{2^{-4H}+1}{2}$ and $D(\{1,2\}) = 2^{-4H}$. The distortion function corresponds to the case of high distortion, in which the single description results in higher distortion than that in the case of low distortion. Consequently, encoding both descriptions gives rise to lower distortion.

Figure 6.1 shows the achievable region formed by all achievable distortion pair $(D(\{1\}), D(\{1,2\}))$ for the LC-coded and MDC-coded Gaussian $\mathcal{N}(0,1)$ source under given H. For the MDC-coded source, the achievable region of $(D(\{1\}), D(\{1,2\}))$ under given H is the shaded area with both vertical and horizontal lines, while for the LC-coded source, the achievable region of $(D(\{1\}), D(\{1,2\}))$ under given H is all shaded area. For LC-coded source, it can be verified that for any given H, the minimum of $D(\{1\})$ and $D(\{1,2\})$, 2^{-2H} and 2^{-4H} respectively, can be achieved simultaneously. Alternatively, for MDC-coded source, it can be seen that for any given H, the minimum of $D(\{1\})$ and $D(\{1,2\})$ can not be achieved simultaneously. In fact, when the minimum of $D(\{1\}) = 2^{-2H}$ is achieved, $D(\{1,2\}) = (2 \cdot 2^{2H} - 1)^{-1}$ can be achieved; when the minimum of $D(\{1,2\}) = 2^{-4H}$ is achieved, $D(\{1\}) \geq (1+2^{-4H})/2$ can be achieved. Nevertheless, the minimum of $D(\{1\}) = 2^{-2H}$ can be achieved for both LC-coded and MDC-coded sources when a single message is encoded by the conventional channel codes, and the minimum of $D(\{1,2\}) = 2^{-4H}$ can be achieved when both messages are encoded by the conventional channel codes. Consequently, the minimum distortion achieved by conventional channel codes are same for both LC-coded and MDC-coded Gaussian $\mathcal{N}(0,1)$ source.

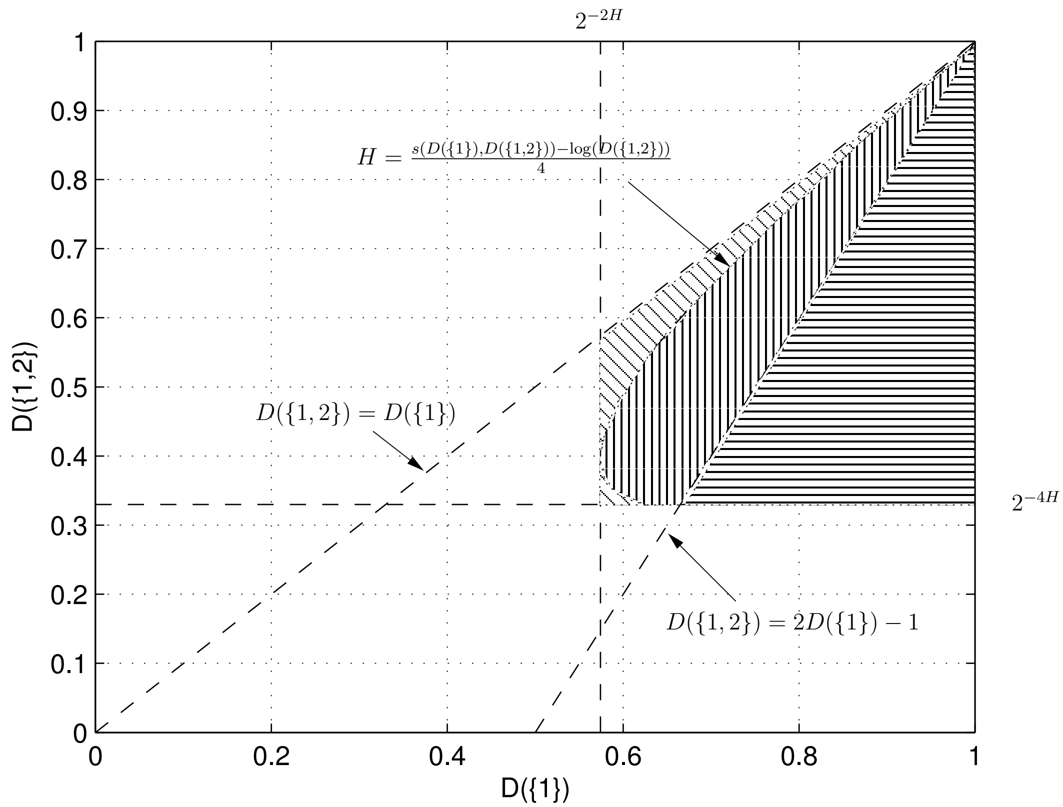

Figure 6.1: The set of all achievable distortion $(D(\{1\}), D(\{1,2\}))$ for an LC-coded and an MDC-coded Gaussian $\mathcal{N}(0,1)$ source under $H = 0.4$.

It can be seen from Theorem 7, 8, and 9 that the minimal distortion achieved by conventional channel codes are same for communicating rate distortion-coded, LC-coded, and MDC-coded Gaussian $\mathcal{N}(0,1)$ source over any degraded compound channel q with delay factor r. In fact conventional channel codes treat either one message or two messages as a single message and encode it, and the minimum distortion on source reconstruction can be achieved.

6.2.2 Superposition codes and their achievable distortion

Next we consider the case where the channel codes are superposition codes. We note that since superposition codes are applied to encode multiple messages, we do not take into account the rate distortion-coded source.

Since $H_1 = H_2$, by symmetry we consider only the superposition codes encoding M_1 first and then M_2 superimposed. A length-n superposition code is specified via a Markov chain $U - X - Y_2 - Y_1$ and a rate pair (R_1, R_2). For any given distributions P_U, $P_{X|U}$, $P_{Y_2|X}$, and $P_{Y_1|Y_2}$ such that $R_1 = I(U;Y_1)$ and $R_2 = I(X;Y_2|U)$ hold, generate 2^{nR_1} independent length n sequences $U^n(M_1)$ according to $\prod_{i=1}^n P_U(u_i)$. For each sequence $U^n(M_1)$, generate 2^{nR_2} independent length n codewords $X^n(M_1, M_2)$ according to $\prod_{i=1}^n P_{X|U}(x_i|u_i)$. Reveal the codebook to the decoder. The encoder sends $X^n(M_1, M_2)$ if it sees the message pair (M_1, M_2). The rate pair (R_1, R_2) has been shown to be achievable over the degraded broadcast channel $X - Y_2 - Y_1$ [120].

Let θ' and θ'' denote the channels $X \to Y_1$ and $X \to Y_2$ respectively. We consider a sequence of superposition codes C_n parametrized by $(P_U, P_{X|U}, \theta', \theta'', R_1, R_2)$. Let $\mathcal{D}^*$ be defined as

$$\mathcal{D}^* = \min_{P_U, P_{X|U}, \theta', \theta''} D(\emptyset) - (D(\emptyset) - D(\{1\}))\Pr(\theta : \theta'' \succ \theta \succ \theta') - (D(\emptyset) - D(\{1,2\}))\Pr(\theta : \theta \succ \theta''). \tag{6.11}$$

Theorem 10 *For LC-coded and MDC-coded source, if given $H_1 = H_2$ and D over degraded compound channel q with delay factor r, $\mathcal{D}^*$ is the minimal achievable distortion by superposition codes.*

The proof of theorem is provided in Appendix IV.

When superposition codes are applied to communicate the LC-coded Gaussian $\mathcal{N}(0,1)$ source, we have the following results. Let $\tilde{\mathcal{D}}_{\text{LC}}$ be defined as

$$\begin{aligned} \tilde{\mathcal{D}}_{\text{LC}} &= \min_{H} \min_{P_U, P_{X|U}, \theta', \theta''} D(\emptyset) - \left(D(\emptyset) - 2^{-2H}\right) \Pr(\theta : \theta'' \succ \theta \succ \theta') \\ &\quad - \left(D(\emptyset) - 2^{-4H}\right) \Pr(\theta : \theta \succ \theta''). \end{aligned} \tag{6.12}$$

Theorem 11 *For LC-coded Gaussian $\mathcal{N}(0,1)$ source over degraded compound channel q with delay factor r, $\tilde{\mathcal{D}}_{\text{LC}}$ is the minimal achievable distortion by superposition codes.*

Proof: By (6.11), we have the minimal achievable distortion

$$\begin{aligned}\mathcal{D} &= \min_H \min_D \mathcal{D}^* \\ &= \min_H \min_D \min_{P_U, P_{X|U}, \theta', \theta''} D(\emptyset) - (D(\emptyset) - D(\{1\}))\Pr(\theta : \theta'' \succ \theta \succ \theta') \\ &\quad -(D(\emptyset) - D(\{1,2\}))\Pr(\theta : \theta \succ \theta'') \\ &= \min_H \min_{P_U, P_{X|U}, \theta', \theta''} \min_D D(\emptyset) - (D(\emptyset) - D(\{1\}))\Pr(\theta : \theta'' \succ \theta \succ \theta') \\ &\quad -(D(\emptyset) - D(\{1,2\}))\Pr(\theta : \theta \succ \theta'').\end{aligned}$$

By (6.3), we have $H \geq \max(\frac{1}{2}\log\frac{1}{D(\{1\})}, \frac{1}{4}\log\frac{1}{D(\{1,2\})})$ and therefore $D(\{1\}) \geq 2^{-2H}$ and $D(\{1,2\}) \geq 2^{-4H}$ hold for any given H. Therefore, we have

$$\begin{aligned}\mathcal{D} &= \min_H \min_{P_U, P_{X|U}, \theta', \theta''} D(\emptyset) - \left(D(\emptyset) - 2^{-2H}\right)\Pr(\theta : \theta'' \succ \theta \succ \theta') \\ &\quad - \left(D(\emptyset) - 2^{-4H}\right)\Pr(\theta : \theta \succ \theta'').\end{aligned}$$

□

Theorem 11 suggests that in order to achieve the minimum distortion by superposition codes, the optimal LC-coded Gaussian $\mathcal{N}(0,1)$ source can be constructed at the achieving rate H together with the distortion function being specified via $D(\{1\}) = 2^{-2H}$ and $D(\{1,2\}) = 2^{-4H}$. Meanwhile, the superposition code is parametrized by the achieving tuple $(P_U, P_{X|U}, \theta', \theta'')$.

When superposition codes are applied to communicate the MDC-coded Gaussian $\mathcal{N}(0,1)$ source, we have the following results. Let $\tilde{\mathcal{D}}_{\text{MDC}}$ be defined as

$$\begin{aligned}\tilde{\mathcal{D}}_{\text{MDC}} &= \min_H \min_{P_U, P_{X|U}, \theta', \theta''} \min_{D(\{1\})} D(\emptyset) - (D(\emptyset) - D(\{1\}))\Pr(\theta : \theta'' \succ \theta \succ \theta') \\ &\quad -(D(\emptyset) - D(\{1,2\}))\Pr(\theta : \theta \succ \theta''), \qquad (6.13)\end{aligned}$$

where $2^{-2H} \leq D(\{1\}) \leq (1 + 2^{-4H})/2$ and $D(\{1,2\})$ is such that

$$H = \frac{1}{4}\left(\log\frac{1}{D(\{1,2\})} + s(D(\{1\}), D(\{1,2\}))\right) \qquad (6.14)$$

holds.

Theorem 12 *For MDC-coded Gaussian* $\mathcal{N}(0,1)$ *source over degraded compound channel* q *with delay factor* r, $\tilde{\mathcal{D}}_{\mathrm{MDC}}$ *is the minimal achievable distortion by superposition codes.*

Proof: By (6.11), we have the minimal achievable distortion

$$\begin{aligned}
\mathcal{D} &= \min_H \min_D \mathcal{D}^* \\
&= \min_H \min_D \min_{P_U, P_{X|U}, \theta', \theta''} D(\emptyset) - (D(\emptyset) - D(\{1\}))\Pr(\theta : \theta'' \succ \theta \succ \theta') \\
&\quad -(D(\emptyset) - D(\{1,2\}))\Pr(\theta : \theta \succ \theta'') \\
&= \min_H \min_{P_U, P_{X|U}, \theta', \theta''} \min_D D(\emptyset) - (D(\emptyset) - D(\{1\}))\Pr(\theta : \theta'' \succ \theta \succ \theta') \\
&\quad -(D(\emptyset) - D(\{1,2\}))\Pr(\theta : \theta \succ \theta'').
\end{aligned}$$

By (6.4) and (6.5), and the definition of function s in (6.6), we have

$$H \geq \max\left(\frac{1}{2}\log\frac{1}{D(\{1\})}, \frac{1}{4}\log\frac{1}{D(\{1,2\})}\right)$$

in the case of high distortion where $2D(\{1\}) - D(\{1,2\}) \geq 1$ holds, and

$$H \geq \max\left(\frac{1}{2}\log\frac{1}{D(\{1\})}, \frac{1}{4}\left(\log\frac{1}{D(\{1,2\})} + s(D(\{1\}), D(\{1,2\}))\right)\right)$$

in the case of low distortion where $2D(\{1\}) - D(\{1,2\}) < 1$ holds. For any given $H > 0$, let $\mathbb{D}$ denote the achievable region of MDC-coded Gaussian $\mathcal{N}(0,1)$ source. Consider the region of distortion $\tilde{\mathbb{D}}$ such that every $\tilde{D} \in \tilde{\mathbb{D}}$ is specified via $2^{-2H} \leq \tilde{D}(\{1\}) \leq (1 + 2^{-4H})/2$ and

$$H = \frac{1}{4}\left(\log\frac{1}{\tilde{D}(\{1,2\})} + s(\tilde{D}(\{1\}), \tilde{D}(\{1,2\}))\right).$$

It can be verified that $(H, \tilde{D})$ is achievable. Furthermore, for any distortion $D \in \mathbb{D}$, there exists $\tilde{D} \in \tilde{\mathbb{D}}$ such that $\tilde{D}(\{1\}) \leq D(\{1\})$ and $\tilde{D}(\{1,2\}) \leq D(\{1,2\})$ hold. Therefore, we have

$$\begin{aligned}\mathcal{D} &= \min_{H} \min_{P_U, P_{X|U}, \theta', \theta''} \min_{D(\{1\})} D(\emptyset) - (D(\emptyset) - D(\{1\}))\Pr(\theta : \theta'' \succ \theta \succ \theta') \\ &\quad -(D(\emptyset) - D(\{1,2\}))\Pr(\theta : \theta \succ \theta'')\end{aligned}$$

for $2^{-2H} \leq D(\{1\}) \leq (1 + 2^{-4H})/2$, where $D(\{1,2\})$ is such that

$$H = \frac{1}{4}\left(\log \frac{1}{D(\{1,2\})} + s(D(\{1\}), D(\{1,2\}))\right)$$

holds. □

Theorem 12 suggests that in order to achieve the minimum distortion by superposition codes, the optimal MDC-coded Gaussian $\mathcal{N}(0,1)$ source can be constructed at the achieving rate H together with the distortion function being specified via some minimizing $D(\{1\})$, $D(\{2\})$, and $D(\{1,2\})$ such that $D(\{1\}) = D(\{2\})$, $2^{-2H} \leq D(\{1\}) \leq (1 + 2^{-4H})/2$, and (6.14) hold. Meanwhile, the superposition code is parametrized by the achieving tuple $(P_U, P_{X|U}, \theta', \theta'')$.

6.2.3 Comparison

By examining $\tilde{\mathcal{D}}_{\text{LC}}$ and $\tilde{\mathcal{D}}_{\text{MDC}}$ in (6.12) and (6.13) respectively, we have the following results.

Corollary 2 $\tilde{\mathcal{D}}_{\text{LC}} < \tilde{\mathcal{D}}_{\text{MDC}}$.

Proof: By contradiction. Assume $\tilde{\mathcal{D}}_{\text{MDC}} \leq \tilde{\mathcal{D}}_{\text{LC}}$ and $\tilde{\mathcal{D}}_{\text{MDC}}$ is achieved by the tuple $(\tilde{H}, \tilde{P}_U, \tilde{P}_{X|U}, \tilde{\theta}', \tilde{\theta}'', \tilde{D}(\{1\}), \tilde{D}(\{1,2\}))$. Since $\tilde{D}(\{1\}) \geq 2^{-2\tilde{H}}$ and $\tilde{D}(\{1,2\}) \geq 2^{-4\tilde{H}}$, and the equalities do not hold simultaneously, with the tuple $(\tilde{H}, \tilde{P}_U, \tilde{P}_{X|U}, \tilde{\theta}', \tilde{\theta}'')$, we can always achieve distortion lower than $\tilde{\mathcal{D}}_{\text{LC}}$ for an LC-coded source by letting $D(\{1\}) = 2^{-2\tilde{H}}$ and $D(\{1,2\}) = 2^{-4\tilde{H}}$. This violates the assumption. □

Corollary 2 states that for any degraded compound channel q and delay factor r, the LC-coded Gaussian $\mathcal{N}(0,1)$ source outperforms the MDC-coded source in terms of achieving the minimal distortion by superposition codes. It can also be justified from

Figure 6.1, in which the achievable region of $(D(\{1\}), D(\{1,2\}))$ for an MDC-coded source is strictly contained in that for an LC-coded source under any given H.

Furthermore, for an LC-coded Gaussian $\mathcal{N}(0,1)$ source over degraded compound channel q with delay factor r, let $\hat{\theta}$ denote the achieving θ^* in (6.9), and $\hat{P}_X$ denote the capacity achieving input distribution for channel $\hat{\theta}$. Consider the sequence of superposition codes C_n parametrized by $(P_U, P_{X|U}, \theta', \theta'')$, where P_U and $P_{X|U}$ induce the marginal distribution $\hat{P}_X$, and θ' and θ'' are such that $I(X;Y_2) = \text{Cap}(\hat{\theta}) - \Delta$ and $I(X;Y_1) = I(X;Y_2) - \Delta$ hold under the channel input distribution $\hat{P}_X$ for any small $\Delta > 0$. Let $\mathcal{D}_\Delta$ denote the minimum distortion achieved by the sequence of C_n codes. By examining $\hat{\mathcal{D}}$ and $\tilde{\mathcal{D}}_{\text{LC}}$ in (6.9) and (6.12) respectively, we have the following results.

Theorem 13 $\lim_{\Delta\to 0} \mathcal{D}_\Delta = \hat{\mathcal{D}}$.

Proof: Since $H_1 = H_2 = H$ holds, for the sequence of C_n codes, we have $I(U;Y_1) = I(X;Y_2|U) = H/r$. On one hand, we have

$$\begin{aligned}
\frac{2H}{r} &= I(U;Y_1) + I(X;Y_2|U) \\
&\overset{(a)}{\leq} I(U;Y_2) + I(X;Y_2|U) \\
&= I(U,X;Y_2) \\
&= I(X;Y_2) + I(U;Y_2|X) \\
&\overset{(b)}{=} I(X;Y_2),
\end{aligned}$$

where (a) follows from the data processing inequality, and (b) follows from the conditional independence of U and Y_2 for given X. On the other hand, we have

$$\begin{aligned}
\frac{2H}{r} &= I(U;Y_1) + I(X;Y_2|U) \\
&\overset{(c)}{\geq} I(U;Y_1) + I(X;Y_1|U) \\
&= I(U,X;Y_1) \\
&= I(X;Y_1) + I(U;Y_1|X) \\
&\overset{(d)}{=} I(X;Y_1),
\end{aligned}$$

where (c) follows from the data processing inequality, and (d) follows from the conditional independence of U and Y_1 for given X. Thus we have $I(X;Y_1) \leq 2H/r \leq I(X;Y_2)$. Since $\lim_{\Delta\to 0} I(X;Y_1) = \lim_{\Delta\to 0} I(X;Y_2) = \text{Cap}(\hat{\theta})$ holds, we have $\lim\limits_{\Delta\to 0} H = r\text{Cap}(\hat{\theta})/2$. As $\Delta \to 0$, we have $\theta' \to \hat{\theta}$ and $\theta'' \to \hat{\theta}$. By (6.12), it follows that

$$\lim_{\Delta\to 0} \mathcal{D}_\Delta = D(\emptyset) - \left(D(\emptyset) - 2^{-4\left(\frac{r\text{Cap}(\hat{\theta})}{2}\right)}\right) \Pr(\theta : \theta \succ \hat{\theta}) = \hat{\mathcal{D}}.$$

□

Theorem 13 states that for LC-coded Gaussian $\mathcal{N}(0,1)$ source over degraded compound channel q with delay factor r, there exists a sequence of superposition codes achieving the distortion that is the minimum distortion achieved by conventional channel codes. By (6.12), it gives rise to the following result, where the proof is obvious.

Corollary 3 $\tilde{\mathcal{D}}_{\text{LC}} \leq \hat{\mathcal{D}}$.

Corollary 3 states that for an LC-coded Gaussian $\mathcal{N}(0,1)$ source over degraded compound channel q with delay factor r, the minimum distortion achieved by superposition codes is not greater than that achieved by conventional channel codes. Along with the similar arguments, we have $\tilde{\mathcal{D}}_{\text{MDC}} \leq \hat{\mathcal{D}}$, which means for an MDC-coded Gaussian $\mathcal{N}(0,1)$ source over degraded compound channel q with delay factor r, the minimum distortion achieved by superposition codes is also not greater than that achieved by conventional channel codes. As a consequence, among all the source coding and channel coding techniques considered in this chapter, the best solution to achieve the minimum distortion when communicating Gaussian $\mathcal{N}(0,1)$ source is to first encode the source with LC and then with superposition codes.

6.3 Numerical results

For specific q distributions, we numerically compute and compare the minimal distortion achieved by superposition codes and by conventional channel codes for communicating different source-encoded messages of the Gaussian $\mathcal{N}(0,1)$ source.

Let θ be the crossover probability of BSC and let distribution q be $q(\theta) = 2\text{Beta}(2\theta; 1, \beta)$.

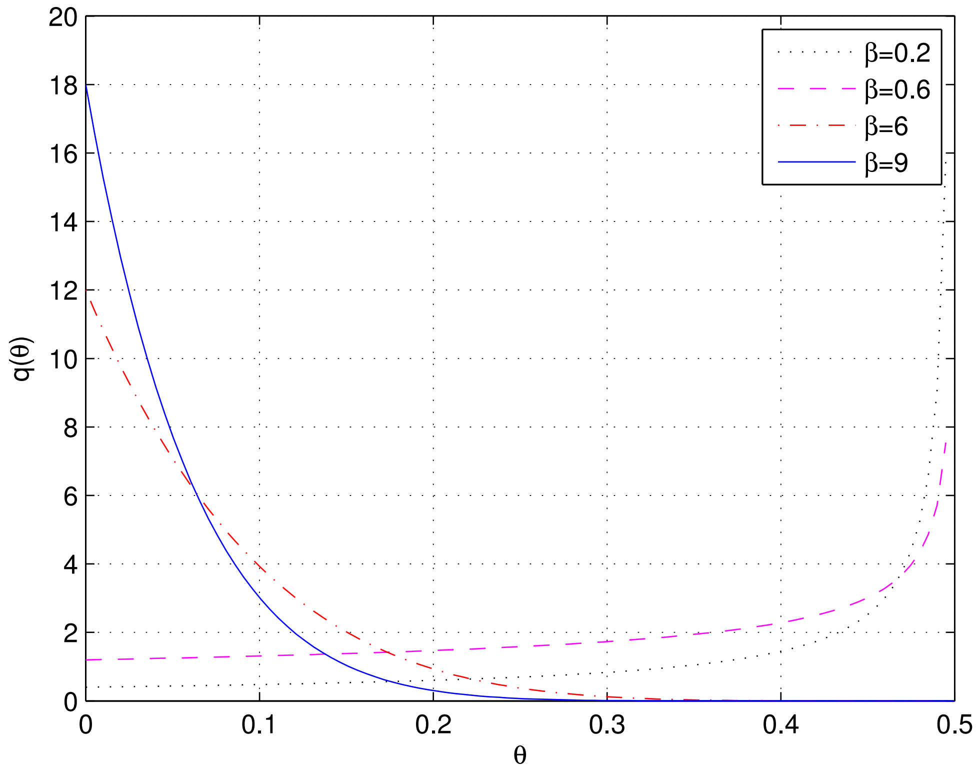

Figure 6.2: Examples of $q(\theta)$ for different β.

Figure 6.2 shows the examples of $q(\theta) = 2\text{Beta}(2\theta; 1, \beta)$ for some β. It appears that as β increases, the distribution $q(\theta)$ is shifted to low value of crossover probability θ, i.e., less noisy BSC channels.

Figure 6.3 shows the signal-to-quantization-noise ratio (SQNR) achieved by the three different combinations of source and channel coding schemes for different q with varying β parameter when the delay factor $r = 1$. In our case, the SQNR value is defined as the ratio of power of Gaussian $\mathcal{N}(0, 1)$ random variable to the achieved minimal distortion for each scheme.

Since the minimal distortion achieved by conventional channel codes are same for rate distortion-coded, LC-coded, and MDC-coded source, only one curve exists for conventional channel codes. It can be seen that the SQNR increases with β for all coding schemes. It coincides with the fact that as the major probability mass of q is shifted

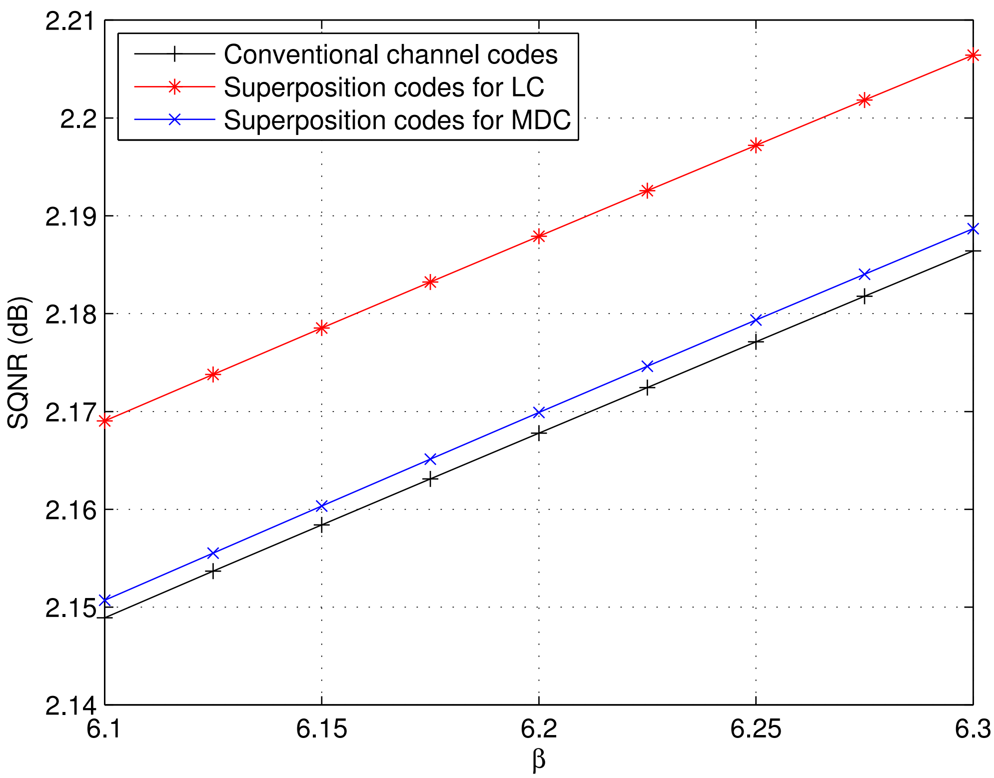

Figure 6.3: SQNR for different coding schemes at various values of β when $r = 1$.

to less noisy channel realizations, the minimal achievable distortion decreases. In Figure 6.3, conventional channel codes perform worse than superposition codes applied on either the LC-coded or the MDC-coded source over the given range of β, which coincides with Corollary 3. In addition, superposition codes for the LC-coded source are shown to achieve higher SQNR value than those for the MDC-coded source, as stated in Corollary 2. The combination of LC and superposition codes is the best to achieve the minimum distortion, or equivalently, highest SQNR. However, due to the small gap at around 0.02 dB between the LC-coded and the MDC-coded source transmitted by superposition codes, in practical MDC with more flexibility is preferred.

6.4 Conclusion

In this chapter we consider the problem of communicating Gaussian $\mathcal{N}(0,1)$ sequence over degraded compound channels. It is formulated as the problem of minimizing the average distortion at the receiver, where the minimization is taken over different schemes of source coding and channel coding. The minimal achievable distortion of rate distortion-coded, LC-coded, and MDC-coded source is derived and compared for the communication by conventional channel codes and superposition codes. By conventional channel codes the minimal achievable distortion of different source coding schemes are same, however by superposition codes the minimal achievable distortion of an LC-coded source is lower than that of an MDC-coded source. Moreover, for either an LC-coded or an MDC-coded source, the minimum distortion achieved by superposition codes outperforms that achieved by conventional channel codes. Out of all the considered schemes, the combination of LC and superposition codes is the best. For LC and MDC, we have focused on $H_1 = H_2$. But our approach extends to more general setting of arbitrary rates.

Chapter 7

Codes for Quasi-Static Rayleigh Fading Channels

In the scenario of multimedia communication over wireless networks, assuming the multimedia source is source-encoded into a set of messages, it is of interest to investigate how to communicate the source-coded messages reliably and efficiently over wireless channels. Fading channels, which model the wireless channels suitably, belong to the class of compound channels. For dependent source messages under the model proposed in Chapter 3, we consider the channel coding problem for communication over fading channels. Specifically, we consider that the source is encoded into two dependent messages and the communication is taken over quasi-static Rayleigh fading channels. The success rates of both messages and the throughput are evaluated and compared for several channel coding schemes including conventional channel codes, time sharing codes, and an LDPC-based Bi-LDPC codes. For an LC-coded and an MDC-coded Gaussian $\mathcal{N}(0,1)$ source transmitted over quasi-static Rayleigh fading channels, the average distortion on source reconstruction is evaluated and compared among different channel coding schemes as well.

7.1 Problem formulation

7.1.1 Problem setup

We consider single-input single-output (SISO) communication over fading channels. The discrete time channel model is given by

$$y_i = ax_i + z_i, \tag{7.1}$$

where at time i, y_i and x_i are the received and transmitted symbols, respectively, and z_i is additive white Gaussian noise with zero mean and variance σ^2. The fading coefficient a is a random variable independent of time, which gives rise to a quasi-static fading model. We further assume the fading coefficient a follows from a Rayleigh distribution $q(a)$, i.e., $q(a) = 2ae^{-a^2}$ for $a \geq 0$, and is available at the receiver but not at the transmitter. The symbol x_i is chosen from a BPSK alphabet $x_i = \pm\sqrt{E_s}$ where E_s is the average energy per transmitted symbol. The *average* SNR of the channel is given by E_s/σ^2, and the SNR of the channel realization, termed as *realized* SNR, is given by a^2E_s/σ^2. For any given average SNR, the set of all realized SNR can be ordered in terms of the fading coefficient a, and therefore quasi-static Rayleigh fading channels belong to the class of degraded compound channels.

In this chapter, we consider communicating two messages M_1 and M_2 over quasi-static Rayleigh fading channels. Suppose we have k_1 and k_2 information bits for message M_1 and M_2 respectively, and we consider the general coding scheme where the encoder maps the two information sequences of length k_1 and k_2 into an n-bit codeword, giving rise to code rates $R_1 := k_1/n$ and $R_2 := k_2/n$, and the decoder attempts to recover both M_1 and M_2 based on the received length-n sequence. Mathematically, a general (R_1, R_2, n) code consists of a message space $\{1, \ldots, 2^{k_1}\} \times \{1, \ldots, 2^{k_2}\}$, an encoding function $f : \{1, \ldots, 2^{k_1}\} \times \{1, \ldots, 2^{k_2}\} \to \mathcal{X}^n$, and a set of decoding functions $\{g_a : \mathcal{Y}^n \to \{1, \ldots, 2^{k_1}\} \times \{1, \ldots, 2^{k_2}\},\ a \sim q(a)\}$ attempting to decode M_1 and M_2, where $\mathcal{X}$ and $\mathcal{Y}$ are channel input and output alphabet respectively, as denoted earlier. In particular, we consider codes with binary codebooks, i.e., $\mathcal{X}$ is binary. Without loss of generality, we assume $k_1 < k_2$, or equivalently, $R_1 < R_2$.

7.1.2 Performance metrics

In such communication scenario, for different coding schemes, success rate of M_1 and M_2 can be used to evaluate the code performance, which is a measure on the robustness of the code. Under any given quasi-static Rayleigh fading channel, the probability that M_1 and M_2 can be decoded individually is termed the *success rate* of M_1 and M_2 respectively, which we denote as P_1^s and P_2^s. Furthermore, let P_1, P_2, and P_{12} denote respectively the probability that only M_1 is decoded in error, only M_2 is decoded in error, and both M_1 and M_2 are decoded in error. The success rates of M_1 and M_2 can be expressed as follows,

$$P_1^s = 1 - P_1 - P_{12} \tag{7.2}$$

and

$$P_2^s = 1 - P_2 - P_{12}. \tag{7.3}$$

In addition to the notion of success rate, throughput plays the role of performance metric to evaluate the efficiency of the coding scheme. On average we use throughput to evaluate the total number of information bits transmitted successfully during one channel use. Let T denote the *throughput* of a code under any given quasi-static Rayleigh fading channel, and it can be calculated as follows,

$$T = R_1 P_1^s + R_2 P_2^s. \tag{7.4}$$

Moreover, suppose the messages M_1 and M_2 are source-coded messages by encoding an i.i.d. Gaussian $\mathcal{N}(0,1)$ sequence using LC and MDC respectively. We consider transmitting the LC-coded and the MDC-coded source messages M_1 and M_2 over given quasi-static Rayleigh fading channel respectively, and evaluate the performance by utilizing the average distortion on source reconstruction.

Based on the discussion in Chapter 6, in which we investigate the achievable rate distortion region of an LC-coded and an MDC-coded Gaussian $\mathcal{N}(0,1)$ source, the LC-coded and the MDC-coded source can be constructed as follows. Assume we pick the delay factor $r = 1$, or equivalently, the length of source symbols $n_s = n$, and thus we have the entropy rates of both messages $H_1 = R_1$ and $H_2 = R_2$. Assume the LC-coded, and MDC-coded source are constructed at their optimized distortions. More precisely, in the LC-coded source, the dependency between M_1 and M_2 is shown as that in Figure 3.1(a), i.e., M_1 is the base layer message and M_2 is the enhancement layer message. We note that the assumption $R_1 < R_2$ guarantees the successful decoding of the base layer when the enhancement layer is decoded, and the enhancement layer can not be decoded by itself. According to (6.3), the distortion can be minimized at $D_1 = 2^{-2R_1}$, $D_2 = 1$, and $D_0 = 2^{-2(R_1+R_2)}$. In the MDC-coded source, the dependency between M_1 and M_2 is shown in Figure 3.1(b). The source messages can be constructed optimally in the sense that all the equalities in (6.5) hold, which corresponds to the case of low distortion. That is, we have $D_1 = 2^{-2R_1}$, $D_2 = 2^{-2R_2}$, and D_0 is such that

$$R_1 + R_2 = \frac{1}{2}\log\frac{1}{D_0} + \frac{1}{2}\log\frac{(1-D_0)^2}{(1-D_0)^2 - ((1-D_1)^{1/2}(1-D_2)^{1/2} - (D_1-D_0)^{1/2}(D_2-D_0)^{1/2})^2}$$

holds. Let $\mathcal{D}$ denote the average distortion of the joint source channel coding scheme. With the above notations, the average distortion is given as

$$\mathcal{D} := D_0(1 - P_1 - P_2 - P_{12}) + D_1P_2 + D_2P_1 + 1 \cdot P_{12}. \tag{7.5}$$

7.2 Coding schemes

From the perspective of practical code construction, instead of considering all possible codes, we restrict our attention to the coding schemes based on the class of well known graph codes, LDPC codes, due to their capacity-achieving performance by BP decoding algorithm. For codes to encode two messages, we consider three different coding schemes, namely, conventional channel codes, time sharing codes, and Bi-LDPC codes.

7.2.1 Conventional channel codes

By conventional channel coding scheme, we have three different strategies to map messages M_1 and M_2 to the length-n codeword. The concatenation of k_1 and k_2 bits can be mapped to the codeword by a single LDPC code, resulting in the code rate $R_1 + R_2$. We denote such scheme by $\mathrm{CC}_{12}(R_1 + R_2)$. Alternatively, we can choose to encode either

M_1 only or M_2 only by a single LDPC code, giving rise to the code rate R_1 or R_2. Such schemes can be denoted by $\mathrm{CC}_1(R_1)$ and $\mathrm{CC}_2(R_2)$ respectively. In that case, the other message is simply dropped by the encoder.

7.2.2 Time sharing codes

Under the concept of time sharing, two categories, namely, type-1 (TS-1) codes and type-2 (TS-2) codes, are taken into consideration. In both categories, the codes are parametrized by two LDPC codes encoding messages M_1 and M_2 separately, which we call component codes, and a time sharing fraction Δ, which dictates the fraction of using one of the two codes in the probabilistic or deterministic sense.

More precisely, the scheme TS-1 is illustrated in Figure 7.1. The two messages M_1 and M_2 are firstly encoded into length-n sequences X_1^n and X_2^n by component codes $\mathcal{C}_1$ and $\mathcal{C}_2$ separately. Secondly, the codeword is chosen to be either X_1^n or X_2^n by a switch parametrized by Δ. Specifically, we choose X_1^n to be the codeword with probability Δ and X_2^n to be the codeword with probability $1-\Delta$. Obviously, the time sharing fraction Δ can vary between 0 and 1. Figure 7.2 shows the encoding process for the scheme TS-2. In TS-2 encoding, the message M_1 is encoded to a length Δn sequence $X_1^{\Delta n}$ by the code $\mathcal{C}_1$, while the message M_2 is encoded to a length $(1-\Delta)n$ sequence $X_2^{(1-\Delta)n}$ by the code $\mathcal{C}_2$. The length n codeword is simply the concatenation of $X_1^{\Delta n}$ and $X_2^{(1-\Delta)n}$. Let r_1 and r_2 denote the rate of the component codes $\mathcal{C}_1$ and $\mathcal{C}_2$ respectively. It can be obtained that

$$r_1 = \frac{k_1}{\Delta n} = \frac{R_1}{\Delta}$$

and

$$r_2 = \frac{k_2}{(1-\Delta)n} = \frac{R_2}{1-\Delta}.$$

Due to the restriction of the binary input over the channel, both r_1 and r_2 must be less than one bit per channel use, which results in the range of Δ is such that $k_1/n < \Delta < 1 - k_2/n$ holds for any given k_1, k_2, and n.

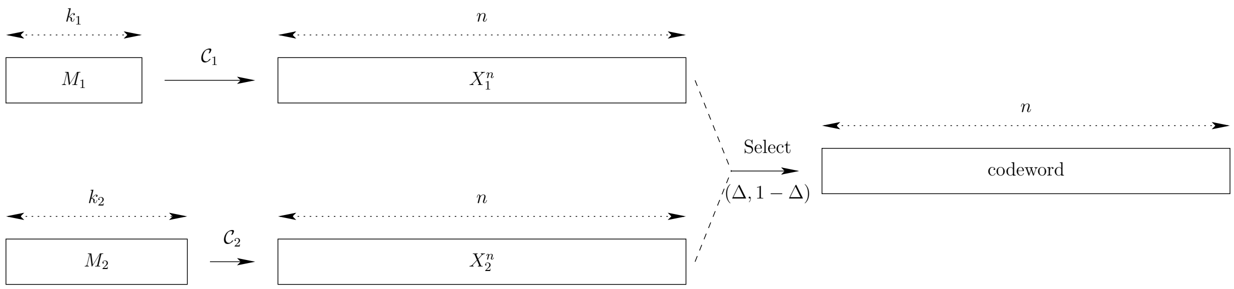

Figure 7.1: TS-1 coding scheme.

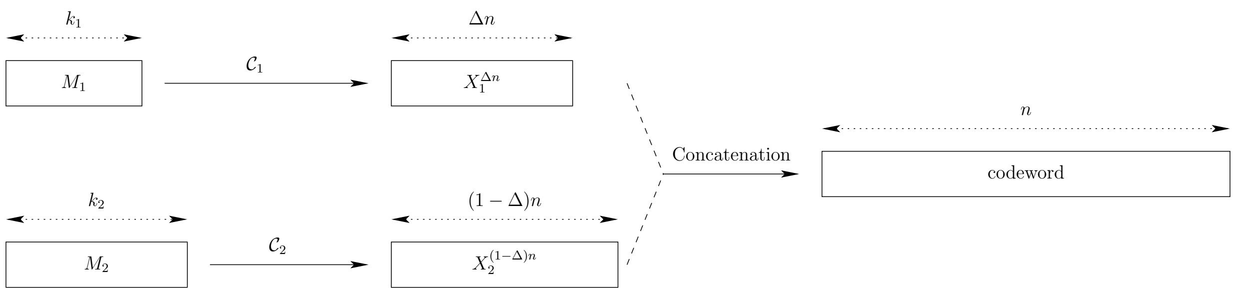

Figure 7.2: TS-2 coding scheme.

7.2.3 Bi-LDPC codes

For channel codes encoding both messages, we consider adapting the information theoretic notion of message superposition into the practical code construction. Inspired by the work of [123], in which the superimposed LDPC codes are investigated over fading Gaussian broadcast channels with the superposition of coded symbols done through the idea of power allocation, we consider a coding scheme, which we call Bi-LDPC, for communicating two messages. Figure 7.3 shows an example of Bi-LDPC code on factor graph representation, where $\mathcal{C}_1$ and $\mathcal{C}_2$ represent two LDPC component codes. During the encoding of M_1 and M_2, the messages M_1 and M_2 are first mapped to an n_1-bit sequence and an n_2-bit sequence by $\mathcal{C}_1$ and $\mathcal{C}_2$ separately. For simplicity, we let $n_1 = n_2$. It can be identified in Figure 7.3 that $n_1 = n_2 = m = n - l + 1$ holds, where l and m are some integers such that $1 < l \leq m < n$ holds. Let the output sequence of $\mathcal{C}_1$ and $\mathcal{C}_2$ be denoted by $(c'_1, c'_2, \ldots, c'_m)$ and $(c''_1, c''_2, \ldots, c''_m)$ respectively. The codeword $(c_1, c_2, \ldots, c_n)$

of the Bi-LDPC code can be generated as follows,

$$c_i := \begin{cases} c'_i, & \text{if } 1 \le i < l; \\ c'_i \oplus c''_{m+l-i} & \text{if } l \le i \le m; \\ c''_{n+1-i} & \text{if } m < i \le n. \end{cases} \tag{7.6}$$

We note that the codeword bits $c_i,\ l \le i \le m$ are generated by the addition of bits from encoding the two messages separately, which resembles the idea of superposition to some extent.

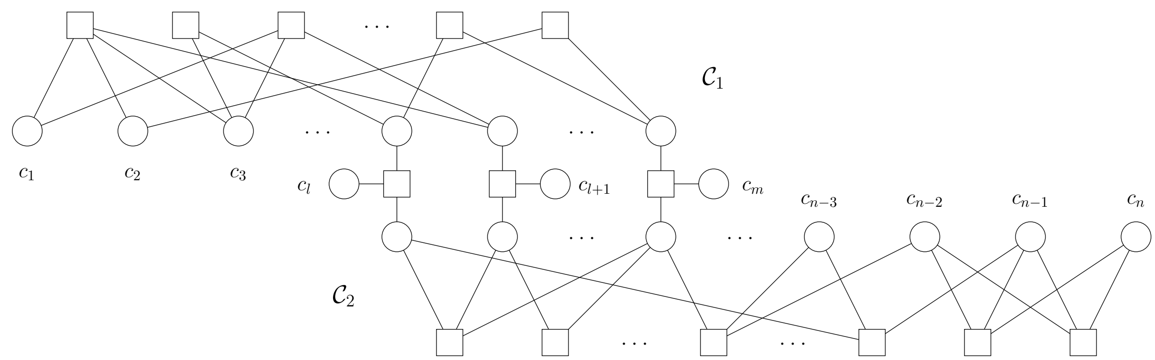

Figure 7.3: The factor graph of a length n Bi-LDPC code (squares represent the check nodes and circles represent the variable nodes).

Let α denote the overlapped fraction of bits between $(c'_1, c'_2, \ldots, c'_m)$ and $(c''_1, c''_2, \ldots, c''_m)$ out of n codeword bits. In Figure 7.3, we have

$$\alpha := \frac{m - l + 1}{n}. \tag{7.7}$$

Furthermore, let (λ_1, ρ_1) and (λ_2, ρ_2) denote the degree distributions parametrizing the codes $\mathcal{C}_1$ and $\mathcal{C}_2$ respectively. Therefore the Bi-LDPC code is parametrized by the tuple $(\lambda_1, \rho_1, \lambda_2, \rho_2, \alpha, n)$. Let r_1 and r_2 denote the rate of codes $\mathcal{C}_1$ and $\mathcal{C}_2$ respectively, i.e.,

$$r_1 = 1 - \frac{\sum \rho_{1,j}/j}{\sum \lambda_{1,i}/i} \tag{7.8}$$

and

$$r_2 = 1 - \frac{\sum \rho_{2,j}/j}{\sum \lambda_{2,i}/i}. \tag{7.9}$$

It can be easily verified that

$$R_1 = r_1 \frac{1+\alpha}{2} \tag{7.10}$$

and

$$R_2 = r_2 \frac{1+\alpha}{2} \tag{7.11}$$

hold.

7.3 Simulation results

We performed Monte Carlo simulations for conventional channel codes, time sharing codes, and Bi-LDPC codes. The success rates of messages M_1 and M_2, and the throughput of the coding schemes are compared for given length of message sequences k_1 and k_2, and length of codeword n. For Gaussian $\mathcal{N}(0,1)$ source sequence encoded by LC and MDC respectively, we evaluate and compare the achieved average distortion for given length of source sequence n, length of encoded message sequences k_1 and k_2, and length of codeword n.

For all coding schemes, we choose $n = 5000$, $k_1 = 625$, and $k_2 = 1875$, giving rise to the rates $R_1 = 0.125$ and $R_2 = 0.375$. Hence in conventional channel codes simulations, all possible codes are specified by $\mathrm{CC}_1(0.125)$, $\mathrm{CC}_2(0.375)$, and $\mathrm{CC}_{12}(0.5)$. In TS-1 codes simulations, the rates of two component codes are 0.125 and 0.375 with the time sharing fraction Δ varying between 0 and 1, while in TS-2 codes simulations, the rates of two component codes are $0.125/\Delta$ and $0.375/(1-\Delta)$ with the time sharing fraction Δ varying between 0.125 and 0.625. In Bi-LDPC codes simulations, α is set to be 0.0999, and thus the rates of two component codes are 0.227 and 0.628. Table 7.1 and 7.2 show the degree distributions of the codes we use in the simulations, which are optimized using the EXIT chart-based design approach [94, 96].

For Gaussian $\mathcal{N}(0,1)$ sequence, provided the parameters k_1, k_2, and n as above, the LC-coded messages can be optimally constructed at the distortions $D_1 = 0.8409$ and

Rate	0.125	0.227	0.375	0.5	0.682
λ_2	0.292350	0.417338	0.368543	0.289956	0.344805
λ_3	0.124880	0.135719	0.230643	0.184750	0.206629
λ_4		0.135881	0.014853		0.223540
λ_5		0.013027		0.222267	0.225027
λ_6	0.121689				
λ_7	0.124470		0.385961		
λ_9		0.298035			
λ_{10}				0.007296	
λ_{11}				0.295732	
λ_{40}	0.336611				
ρ_4	0.100001	1.0			
ρ_5	0.899999		1.0		
ρ_7				0.800001	
ρ_8				0.199999	
ρ_9					0.799999
ρ_{10}					0.200001

Table 7.1: Degree distributions of LDPC codes for $CC_1(0.125)$, $CC_2(0.375)$, $CC_{12}(0.5)$, TS-1, and Bi-LDPC codes.

Rate	0.208	0.313	0.469	0.624	0.625	0.939
λ_2	0.408959	0.241640	0.294023	0.356537	0.320955	0.209468
λ_3	0.137013	0.110291	0.174390	0.237690	0.269766	0.311117
λ_4	0.071145	0.030150	0.025322	0.073345		
λ_5	0.106746	0.127258	0.188011	0.332427	0.005597	
λ_6					0.403683	
λ_7						0.443697
λ_8						0.035719
λ_{10}	0.144494	0.207123				
λ_{11}	0.131644		0.318254			
λ_{40}		0.283537				
ρ_4	1.0					
ρ_6		0.300002	0.200002			
ρ_7		0.699998	0.799998	0.200001		
ρ_8				0.799999	0.599999	
ρ_9					0.400001	
ρ_{59}						1.0

Table 7.2: Degree distributions of LDPC codes for TS-2 codes.

$D_0 = 0.5$. Alternatively, the MDC-coded messages can be optimally constructed at the distortions $D_1 = 0.8409$, $D_2 = 0.5946$ and $D_0 = 0.5345$.

Figures 7.4 and 7.5 show the success rates of messages M_1 and M_2 for different coding schemes. It can be seen that except for the codes $\text{CC}_1(0.125)$ and $\text{CC}_2(0.375)$, the success rates for both messages strictly increase with the channel average SNR. For the code $\text{CC}_1(0.125)$, since only message M_1 is encoded, the success rate of M_2 stays at zero, while for the code $\text{CC}_2(0.375)$, since only message M_2 is encoded, the success rate of M_1 is always zero. As the average SNR becomes high enough, it appears that the success rates of both messages converge. It is shown in Figure 7.4 that for $\text{CC}_{12}(0.5)$ code, TS-2 codes, and the Bi-LDPC code, both success rates converge to one. Compared with the conventional channel code $\text{CC}_{12}(0.5)$ for which $P_1^s = P_2^s$ holds since both messages are treated as a single message to be encoded, the decoder of the Bi-LDPC code is more likely to decode M_1 than M_2. From Figure 7.5, we can see that for TS-1 codes, P_1^s and P_2^s converge to Δ and $1-\Delta$ respectively as the success rate for either message is bounded by the probability that message is transmitted. For the codes $\text{CC}_1(0.125)$ and $\text{CC}_2(0.375)$, P_1^s and P_2^s converge to one respectively. In addition, for both TS-1 and TS-2 codes, as the time fraction Δ increases, for any fixed average SNR, P_1^s increases and P_2^s decreases since in TS-1 coding scheme, M_1 is transmitted with increased probability, and in TS-2 coding scheme, M_1 is in fact encoded by the code of lower rate.

Alternatively, the throughput of different coding schemes is shown and compared in Figures 7.6 and 7.7. For all the coding schemes, the throughput increases with the average SNR and becomes saturated when SNR is large enough. In Figure 7.6, the throughput of $\text{CC}_{12}(0.5)$ code, TS-2 codes, and the Bi-LDPC code goes to $R_1 + R_2 = 0.5$ which is the maximum achieved by any coding scheme for the given selection of k_1, k_2, and n. However, in Figure 7.7, the throughput of the code $\text{CC}_1(0.125)$ and $\text{CC}_2(0.375)$ converges to $R_1 = 0.125$ and $R_2 = 0.375$ respectively since only one message is encoded, while the throughput of TS-1 codes saturates at the value $R_1\Delta + R_2(1-\Delta)$, which is between R_1 and R_2. In terms of the throughput, the performance degradation exists in $\text{CC}_1(0.125)$, $\text{CC}_2(0.375)$, and TS-1 codes especially over channels with large average SNR, compared with other coding schemes. Furthermore, as that can be seen in Figure 7.6, $\text{CC}_{12}(0.5)$ code, TS-2 codes, and the Bi-LDPC code perform closely and each has its advantage over different range of SNR, compared with each other. For example, for SNR less than approximately -3.5 dB, the Bi-LDPC code outperforms $CC_{12}(0.5)$ but underperforms the TS-2 code with $\Delta = 0.6$, while for SNR greater than -2 dB, the Bi-LDPC code outperforms the TS-2 code with $\Delta = 0.6$ but underperforms the $CC_{12}(0.5)$ code.

For the case of communicating an LC-encoded and an MDC-encoded Gaussian $\mathcal{N}(0,1)$ sequence using the above setup, the average distortions for the LC-coded source achieved by different coding schemes are presented in Figures 7.8 and 7.9, while those for the MDC-coded source achieved by different coding schemes are presented in Figures 7.10 and 7.11. Except for the $CC_2(0.375)$ code applied on the LC-coded source, which transmits the message M_2 only and thus gives rise to the distortion one, the average distortion decreases as the channel average SNR increases. For the LC-coded source, as SNR becomes large enough, the distortion of $CC_{12}(0.5)$ code, TS-2 codes, and the Bi-LDPC code converges to $D_0 = 0.5$ when both messages are successfully decoded, while the distortion of $CC_1(0.125)$ code converges to D_1 and that of TS-1 code converges to distortion larger than D_1 depending on the selection of Δ. Similarly, for the MDC-coded source, when SNR becomes large enough, the distortion of $CC_{12}(0.5)$ code, TS-2 codes, and the Bi-LDPC code converges to D_0 due to the successful decoding of both messages, while the distortions of $CC_1(0.125)$ and $CC_2(0.375)$ codes converge to D_1 and D_2 respectively and that of TS-1 code converges to the weighted sum of D_1 and D_2, weighted by Δ. For both source-coded messages, it can be seen that $CC_1(0.125)$, $CC_2(0.375)$, and TS-1 codes perform worse than the other coding schemes especially over channels with large average SNR, in terms of the average achieved distortion. However, $CC_{12}(0.5)$ code, TS-2 codes, and the Bi-LDPC code perform close to each other and each has its own advantage in different range of SNR. For example, for SNR less than approximately -2 dB, the Bi-LDPC code outperforms $CC_{12}(0.5)$ but underperforms the TS-2 code with $\Delta = 0.6$, while for SNR greater than -2 dB, the Bi-LDPC code outperforms the TS-2 code with $\Delta = 0.6$ but underperforms the $CC_{12}(0.5)$ code.

The simulation results of the Bi-LDPC code can be further analyzed. In general, since $R_1 < R_2$ and hence M_2 can not be decoded by itself, for any given average channel SNR, each simulated codeword corresponds to a realized SNR and one of the three decoding outcomes, namely, neither message decoded, only M_1 decoded, and both messages decoded, due to the randomness of the Rayleigh fading coefficient and the Gaussian noise. With the set of realized SNR quantized into intervals of 1 dB, the probability of the event that a simulated codeword corresponds to the realized SNR in a given interval and one certain decoding outcome can be computed based on the simulation results. The calculated results for channels with varying average SNR are shown in Figures 7.12 to 7.17 respectively, where for each average channel SNR, the probability of each decoding outcome is shown as the function of the realized SNR. It can be seen from the figures that for any given average SNR, three peaks of probability mass, indicating the decoding outcomes of neither decoded, only M_1 decoded, and both messages decoded respectively, are ordered with the increasing of the realized SNR. Moreover, as the channel average SNR increases, the probability mass of the decoding outcomes is shifted from the peak of neither decoded to that of both messages decoded.

Figures 7.18 to 7.31 present the simulation results for setup $n = 5000$, $k_1 = 1000$, and $k_2 = 2750$, which give rise to the rates $R_1 = 0.2$ and $R_2 = 0.55$. Thus in conventional channel codes simulations, all possible codes are specified by $\mathrm{CC}_1(0.2)$, $\mathrm{CC}_2(0.55)$, and $\mathrm{CC}_{12}(0.75)$. In TS-1 codes simulations, the rates of two component codes are 0.2 and 0.55 with the time sharing fraction Δ varying between 0 and 1, while in TS-2 codes simulations, the rates of two component codes are $0.2/\Delta$ and $0.55/(1-\Delta)$ with the time sharing fraction Δ varying between 0.2 and 0.45. In Bi-LDPC codes simulations, α is set to be 0.2017, and thus the rates of two component codes are 0.333 and 0.915. Table 7.3 and 7.4 show the degree distributions of the codes used in the simulations, which are also optimized using the EXIT chart-based design approach. For Gaussian $\mathcal{N}(0,1)$ sequence, provided the setup of k_1, k_2, and n, the LC-coded messages can be optimally constructed at the distortions $D_1 = 0.7579$ and $D_0 = 0.3536$, while the MDC-coded messages can be optimally constructed at the distortions $D_1 = 0.7579$, $D_2 = 0.4665$ and $D_0 = 0.4061$.

From Figures 7.18 to 7.31, it can be seen that the comparison behaviour among different coding schemes are similar to that for previous setup of n, k_1, and k_2. In terms of either the throughput, or the average distortion on communicating a Gaussian $\mathcal{N}(0,1)$ source, the performance degradation exists in $\mathrm{CC}_1(0.2)$, $\mathrm{CC}_2(0.55)$, and TS-1 codes especially over channels with large average SNR, compared with other coding

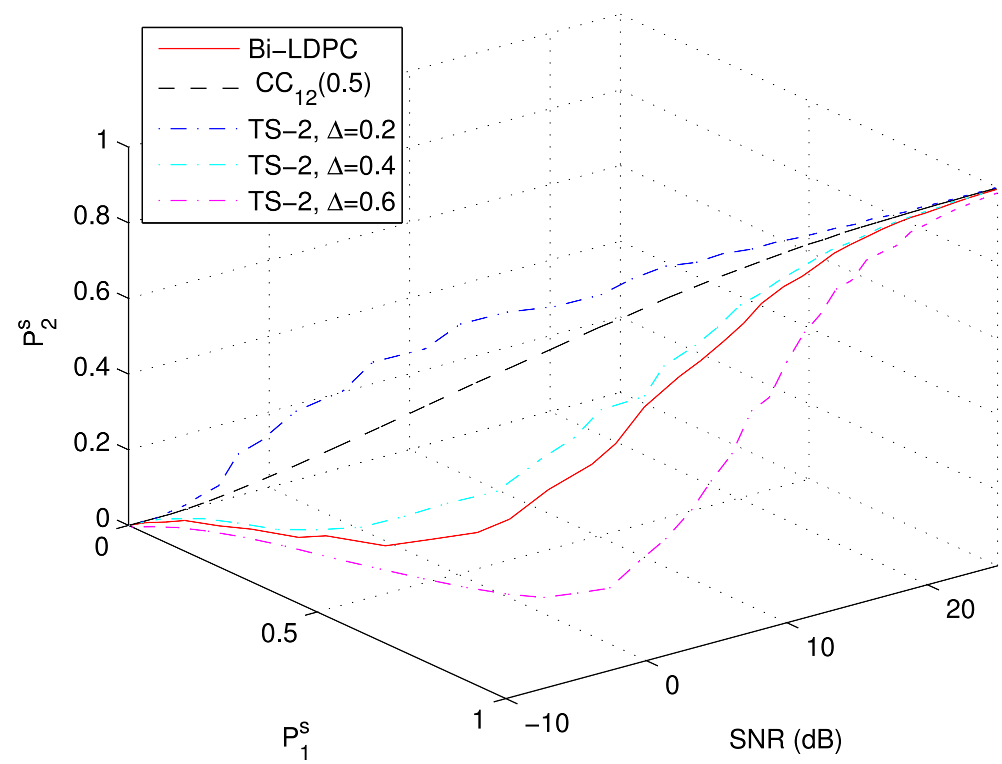

Figure 7.4: Success rates as functions of SNR for $CC_{12}(0.5)$, TS-2, and Bi-LDPC codes.

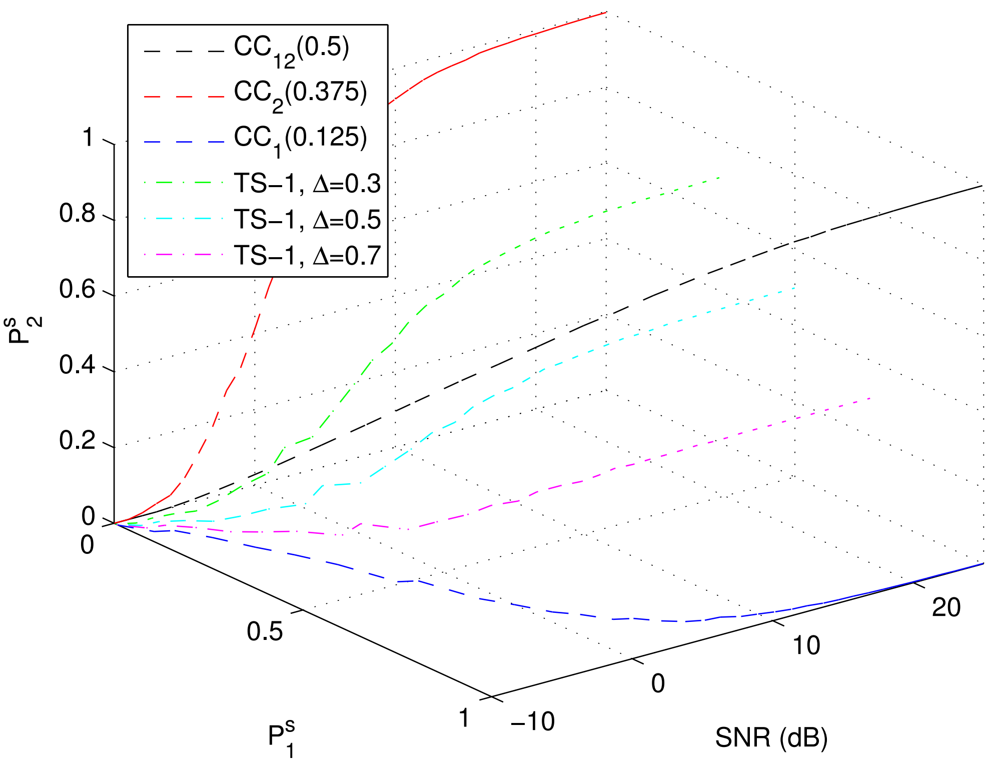

Figure 7.5: Success rates as functions of SNR for $CC_1(0.125)$, $CC_2(0.375)$, and TS-1 codes.

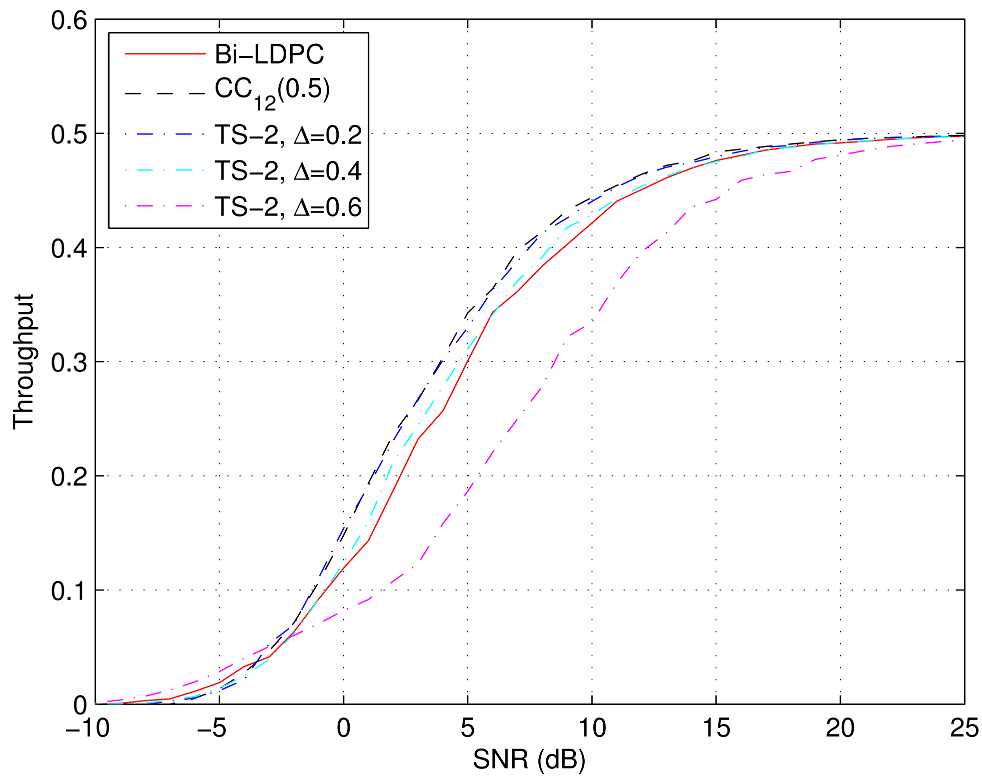

Figure 7.6: Throughput as a function of SNR for $CC_{12}(0.5)$, TS-2, and Bi-LDPC codes.

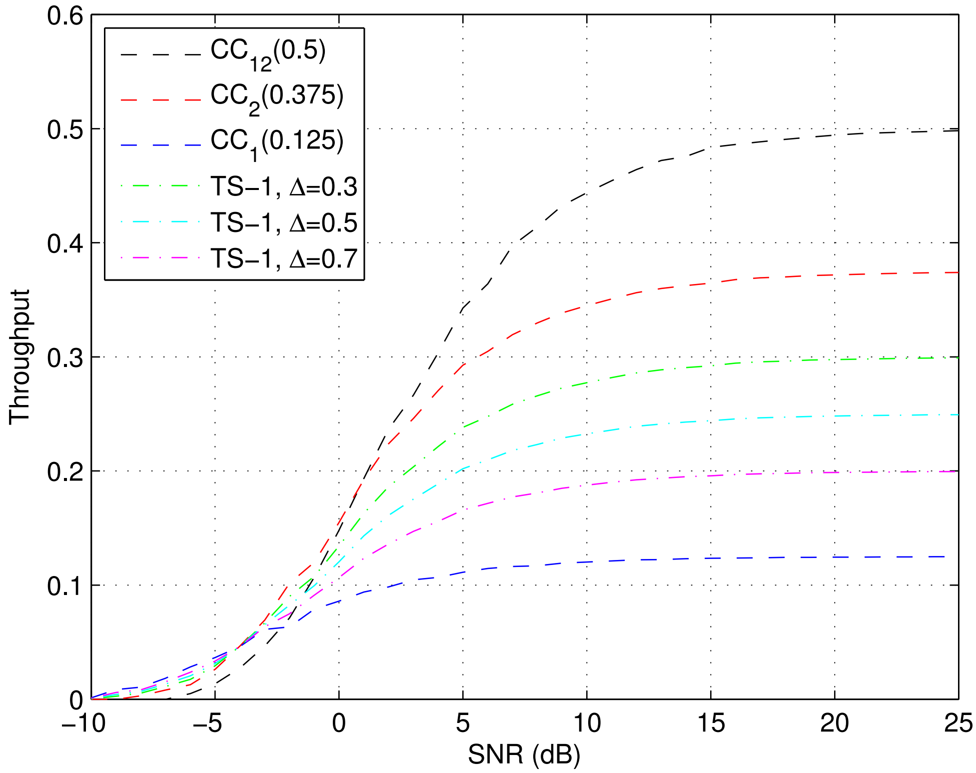

Figure 7.7: Throughput as a function of SNR for $CC_1(0.125)$, $CC_2(0.375)$, and TS-1 codes.

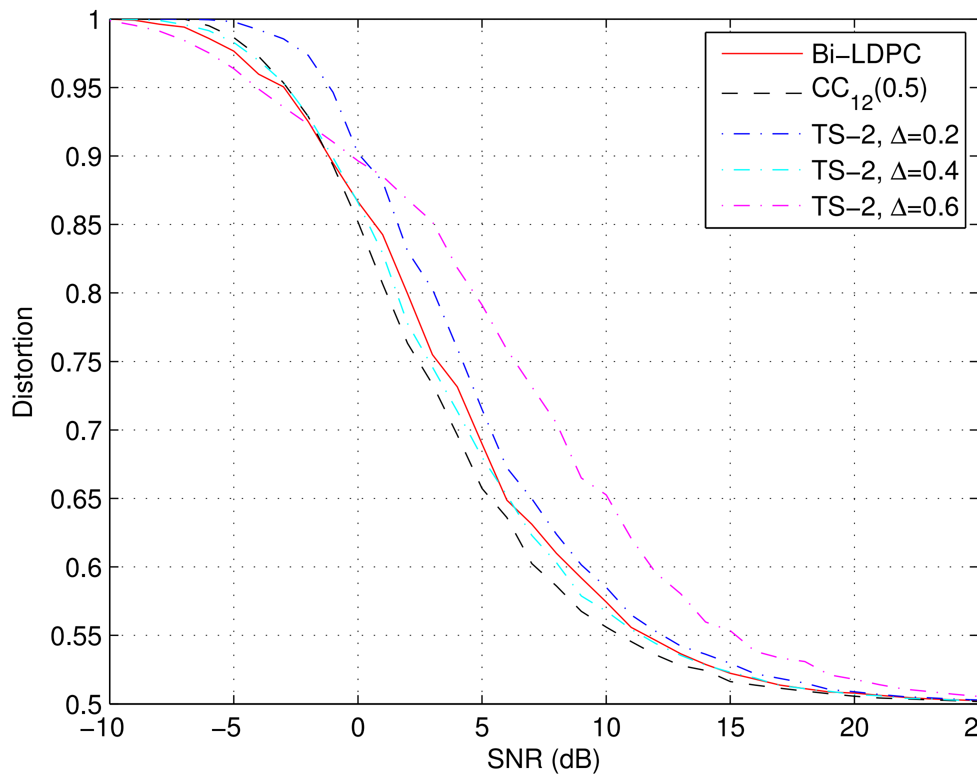

Figure 7.8: Average distortion of an LC-coded Gaussian $\mathcal{N}(0, 1)$ source as a function of SNR for $CC_{12}(0.5)$, TS-2, and Bi-LDPC codes.

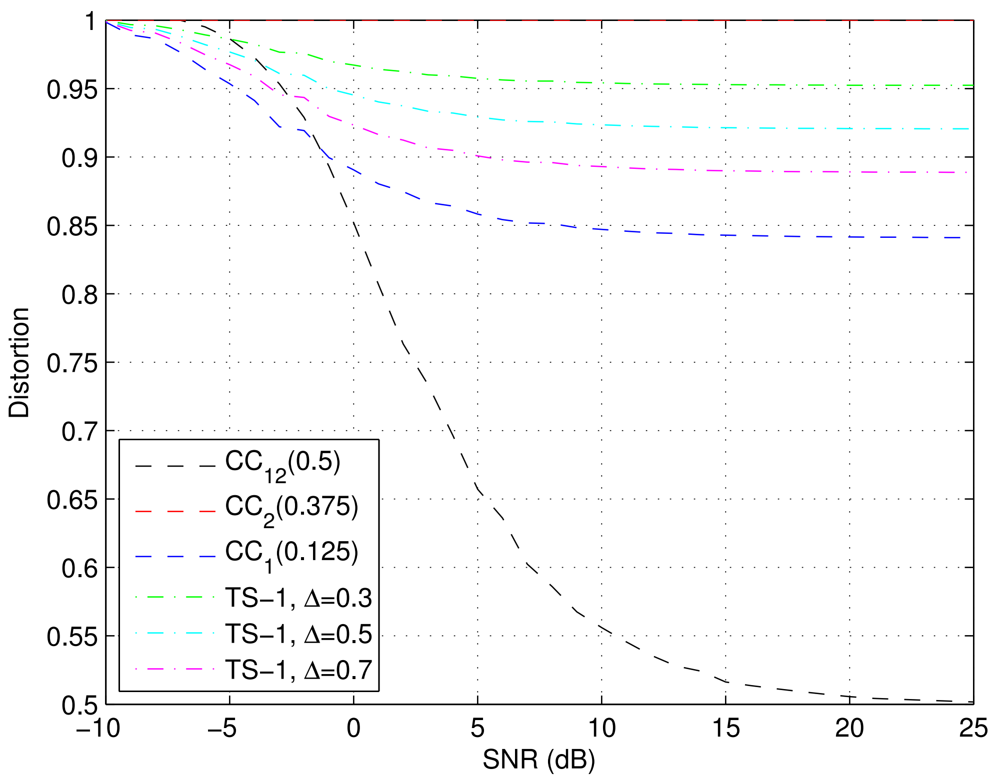

Figure 7.9: Average distortion of an LC-coded Gaussian $\mathcal{N}(0,1)$ source as a function of SNR for $CC_1(0.125)$, $CC_2(0.375)$, and TS-1 codes.

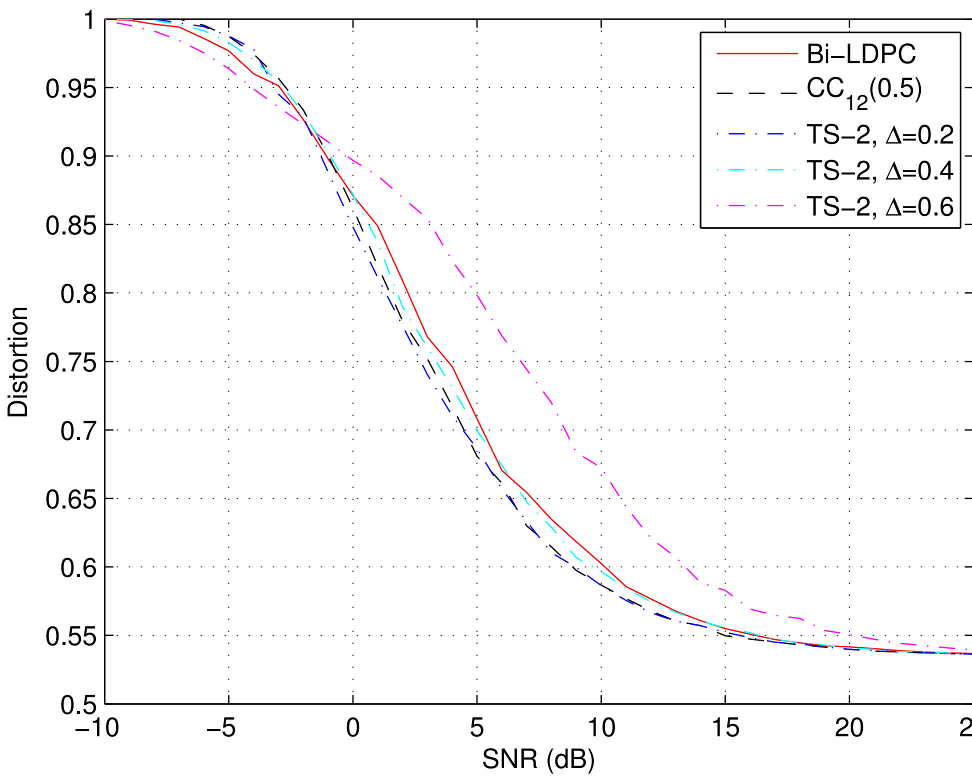

Figure 7.10: Average distortion of an MDC-coded Gaussian $\mathcal{N}(0,1)$ source as a function of SNR for $CC_{12}(0.5)$, TS-2, and Bi-LDPC codes.

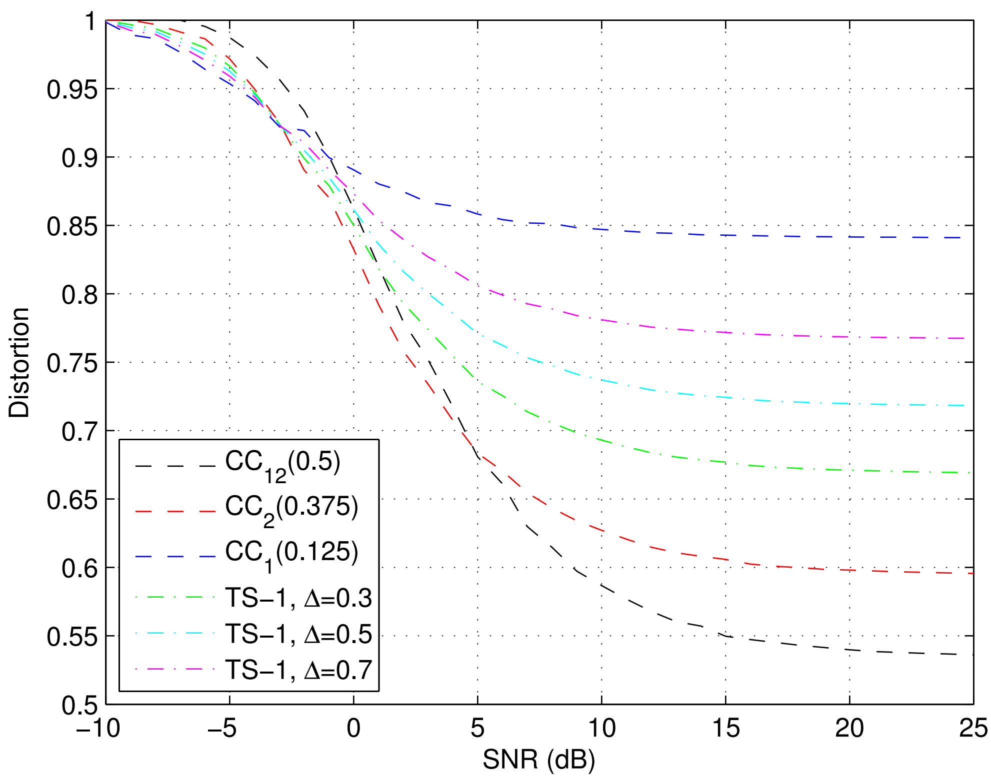

Figure 7.11: Average distortion of an MDC-coded Gaussian $\mathcal{N}(0,1)$ source as a function of SNR for $CC_1(0.125)$, $CC_2(0.375)$, and TS-1 codes.

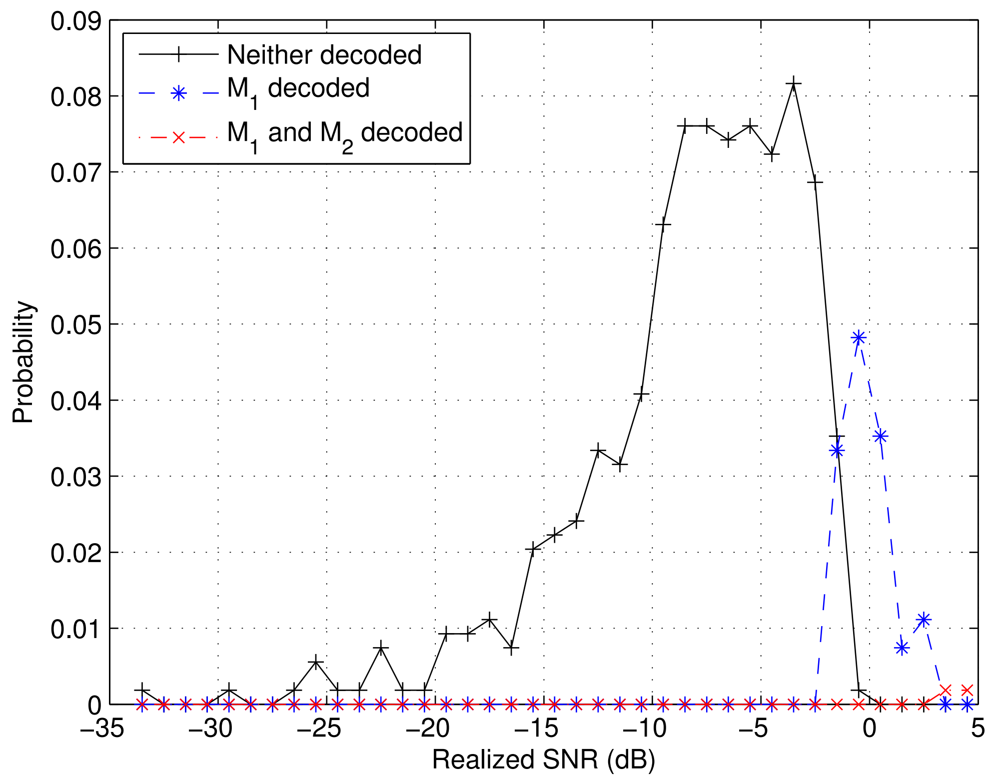

Figure 7.12: The decoding behaviour of Bi-LDPC codes over quasi-static Rayleigh fading channel with average SNR at −5 dB.

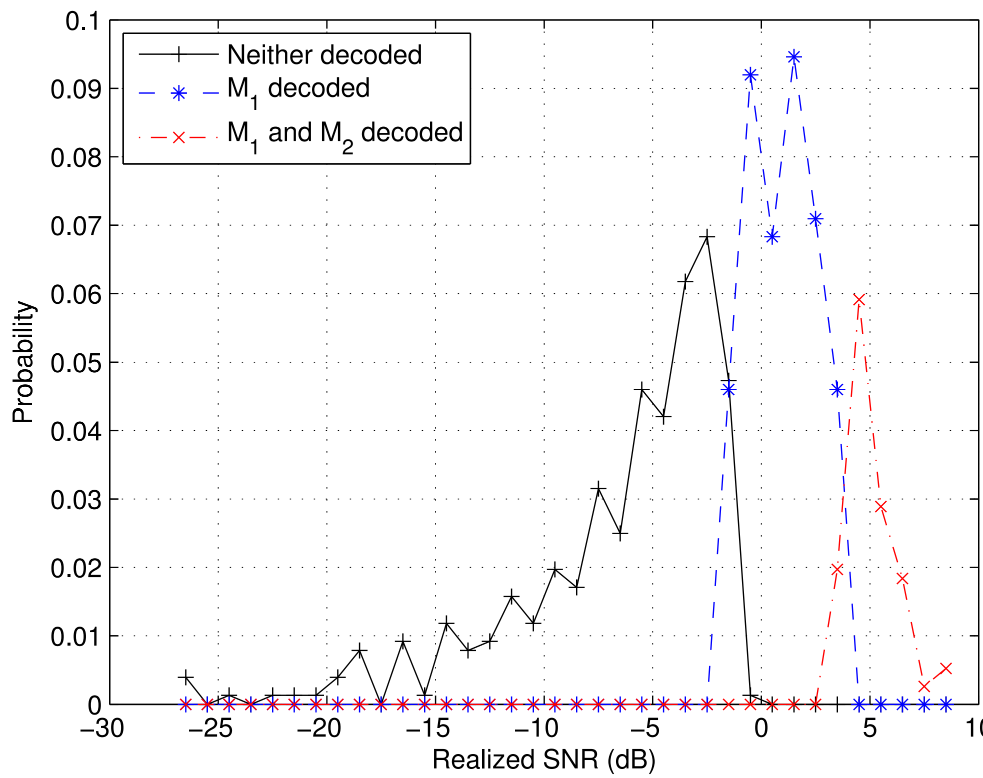

Figure 7.13: The decoding behaviour of Bi-LDPC codes over quasi-static Rayleigh fading channel with average SNR at 0 dB.

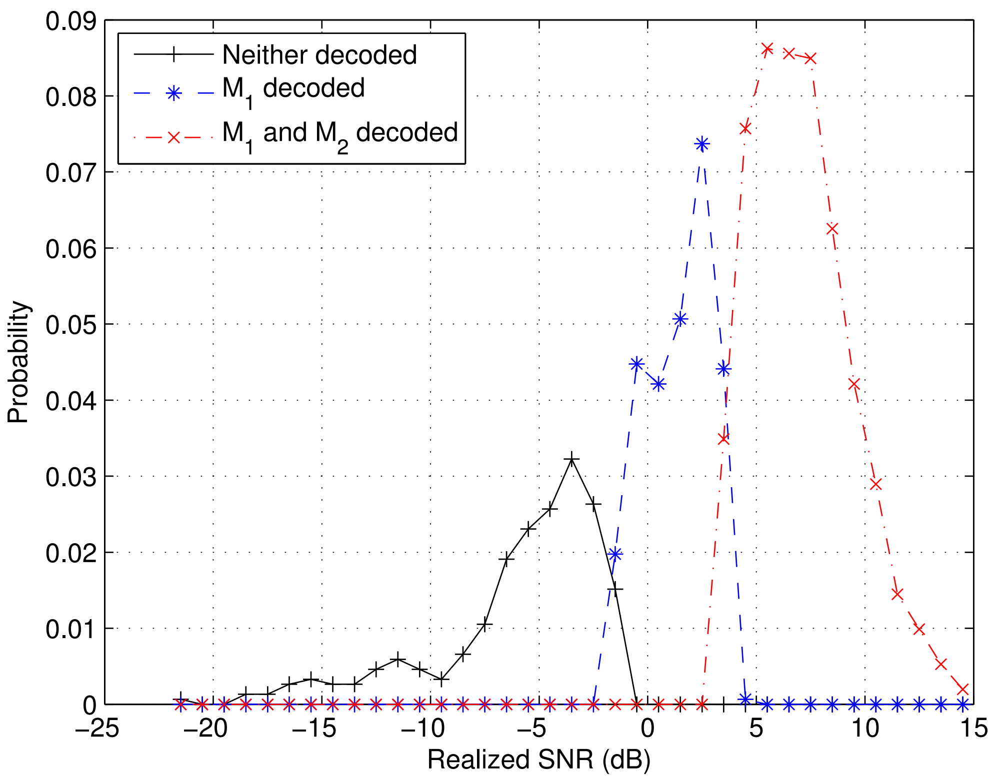

Figure 7.14: The decoding behaviour of Bi-LDPC codes over quasi-static Rayleigh fading channel with average SNR at 5 dB.

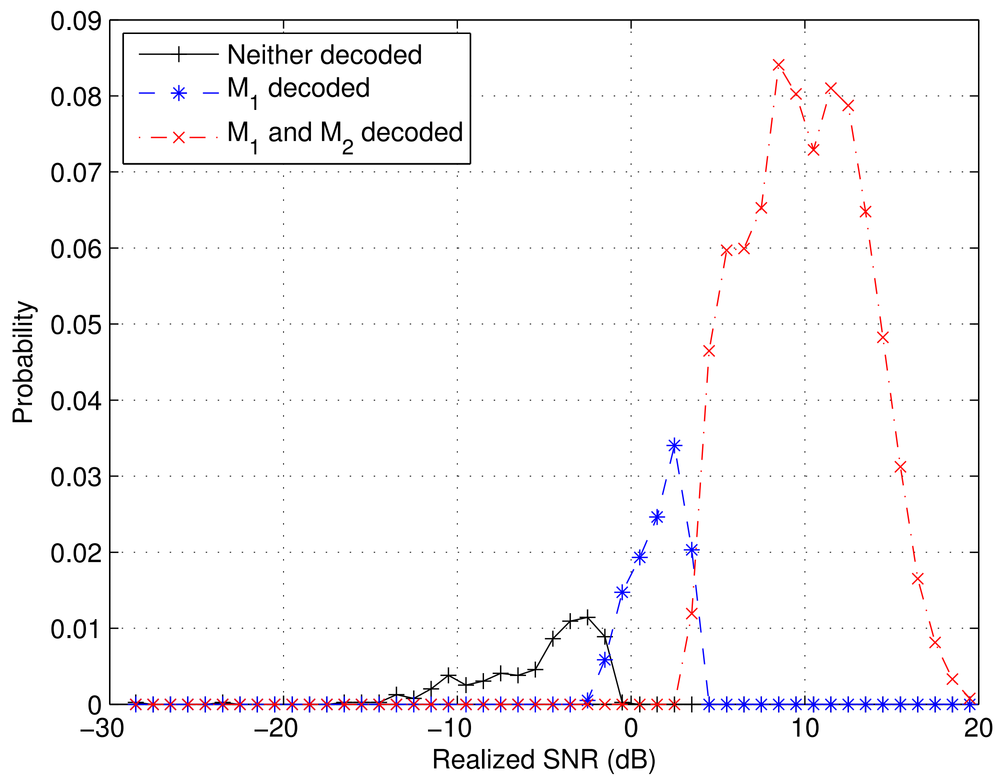

Figure 7.15: The decoding behaviour of Bi-LDPC codes over quasi-static Rayleigh fading channel with average SNR at 10 dB.

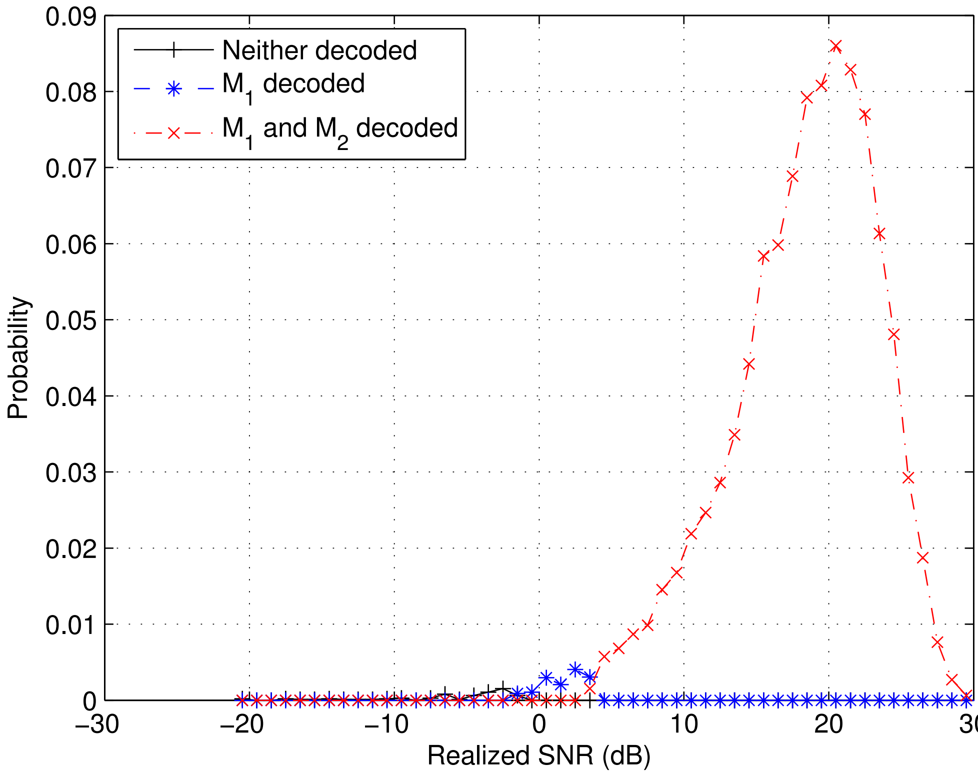

Figure 7.16: The decoding behaviour of Bi-LDPC codes over quasi-static Rayleigh fading channel with average SNR at 20 dB.

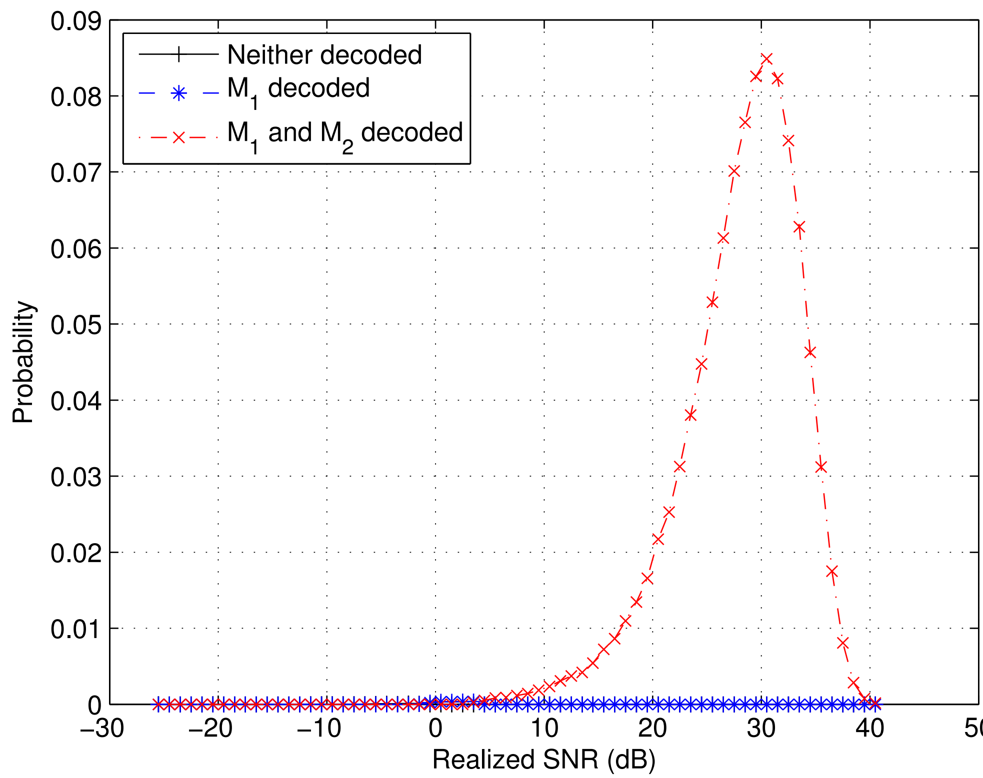

Figure 7.17: The decoding behaviour of Bi-LDPC codes over quasi-static Rayleigh fading channel with average SNR at 30 dB.

Rate	0.200	0.333	0.550	0.750	0.915
λ_2	0.330389	0.319815	0.304852	0.313941	0.199412
λ_3	0.123273	0.163348	0.259917	0.248981	0.251720
λ_4	0.028396			0.057844	0.096172
λ_5	0.124955	0.117181		0.379233	
λ_6		0.145597	0.020007		
λ_7			0.415224		
λ_8	0.160081				0.184646
λ_9					0.268050
λ_{14}		0.195141			
λ_{15}		0.058918			
λ_{22}	0.164466				
λ_{23}	0.068440				
ρ_4	0.300001				
ρ_5	0.699999	0.6			
ρ_6		0.4			
ρ_7			0.600001		
ρ_8			0.399999		
ρ_{12}				0.899996	
ρ_{13}				0.100004	
ρ_{45}					0.799999
ρ_{46}					0.200001

Table 7.3: Degree distributions of LDPC codes for $CC_1(0.2)$, $CC_2(0.55)$, $CC_{12}(0.75)$, TS-1, and Bi-LDPC codes.

Rate	0.500	0.667	0.733	0.783	0.803	0.919
λ_2	0.289956	0.341704	0.323086	0.173846	0.180131	0.201518
λ_3	0.184750	0.240994	0.232337	0.193485	0.205780	0.266573
λ_4		0.068391	0.128518			0.069725
λ_5	0.222267	0.348911	0.316058			
λ_6				0.212200	0.178146	
λ_7					0.105675	
λ_8						0.376488
λ_9						0.085699
λ_{10}	0.007296			0.123475		
λ_{11}	0.295732					
λ_{18}					0.012841	
λ_{19}					0.317427	
λ_{24}				0.275475		
λ_{25}				0.021518		
ρ_7	0.800001					
ρ_8	0.199999	0.099997				
ρ_9		0.900003				
ρ_{11}			0.799995			
ρ_{12}			0.200005			
ρ_{21}				0.199997		
ρ_{22}				0.800003		
ρ_{23}					1.0	
ρ_{47}						1.0

Table 7.4: Degree distributions of LDPC codes for TS-2 codes.

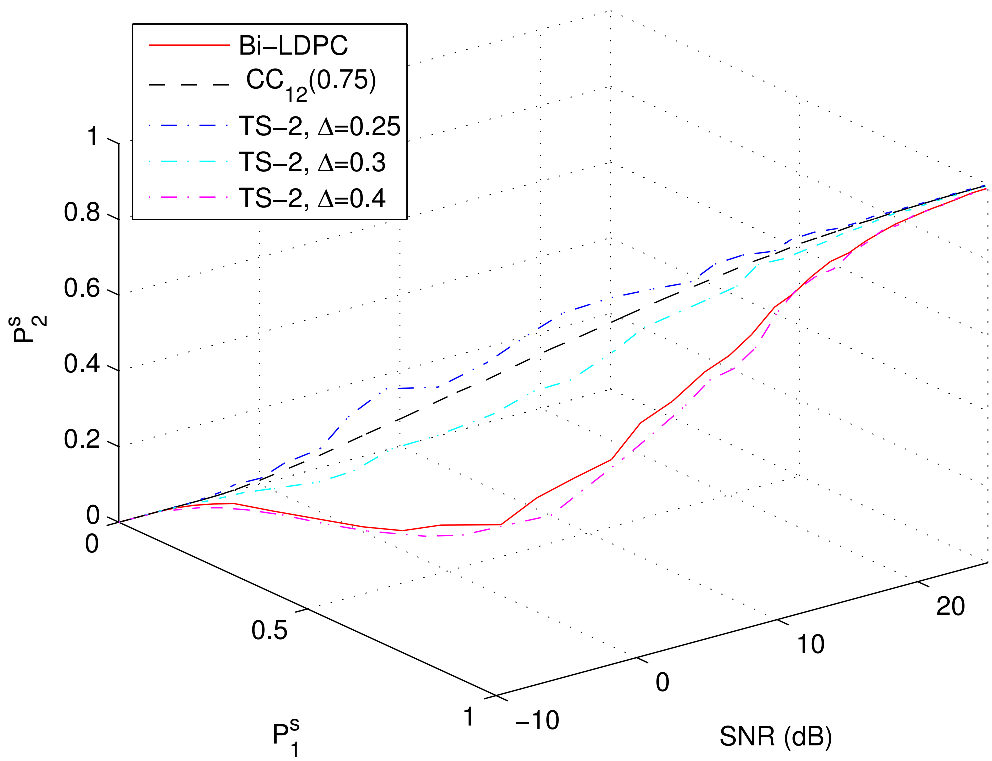

Figure 7.18: Success rates as functions of SNR for $CC_{12}(0.75)$, TS-2, and Bi-LDPC codes.

schemes. However $\mathrm{CC}_{12}(0.75)$ code, TS-2 codes, and the Bi-LDPC code perform closely and each has its own advantage over different range of SNR, compared with each other. The crossover behaviour as observed under previous setup appears to happen at slightly higher channel SNR around 0 to 1 dB. In the simulations, we also note that the TS-2 code with $\Delta = 0.4$ performs very close to the Bi-LDPC code.

7.4 Conclusion and discussion

In this chapter we consider communicating two messages over quasi-static Rayleigh fading channels. The success rates of both messages, and the throughput are evaluated and compared among three different coding schemes, namely, conventional channel codes, time sharing codes, and Bi-LDPC codes. For a Gaussian $\mathcal{N}(0,1)$ source sequence encoded

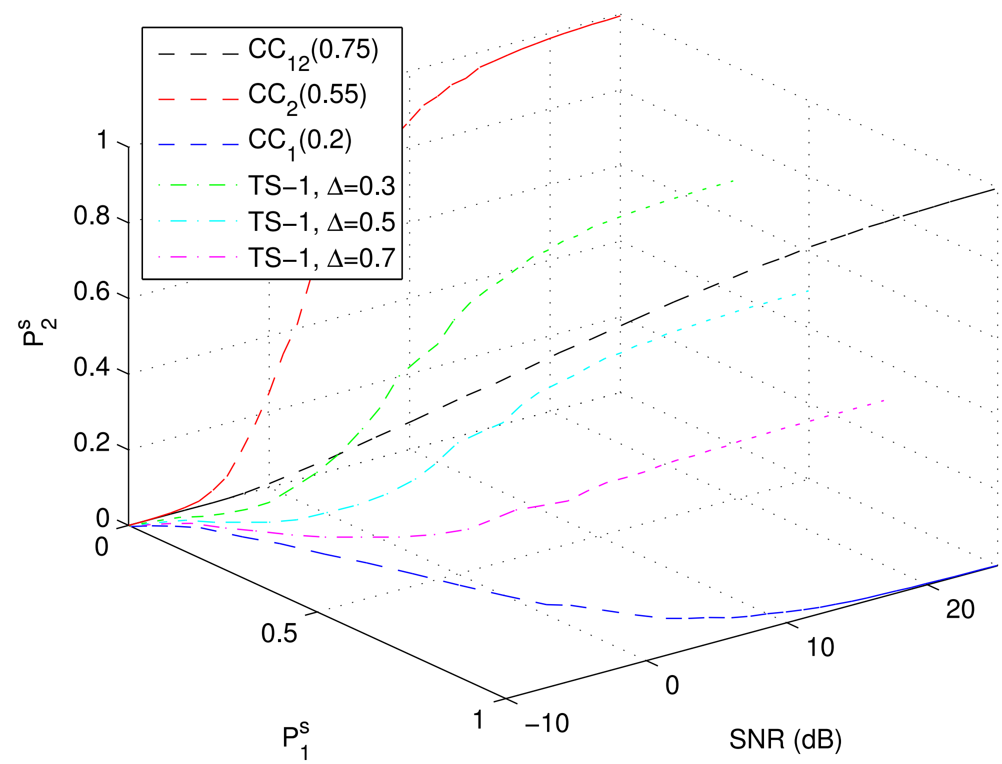

Figure 7.19: Success rates as functions of SNR for $CC_1(0.2)$, $CC_2(0.55)$, and TS-1 codes.

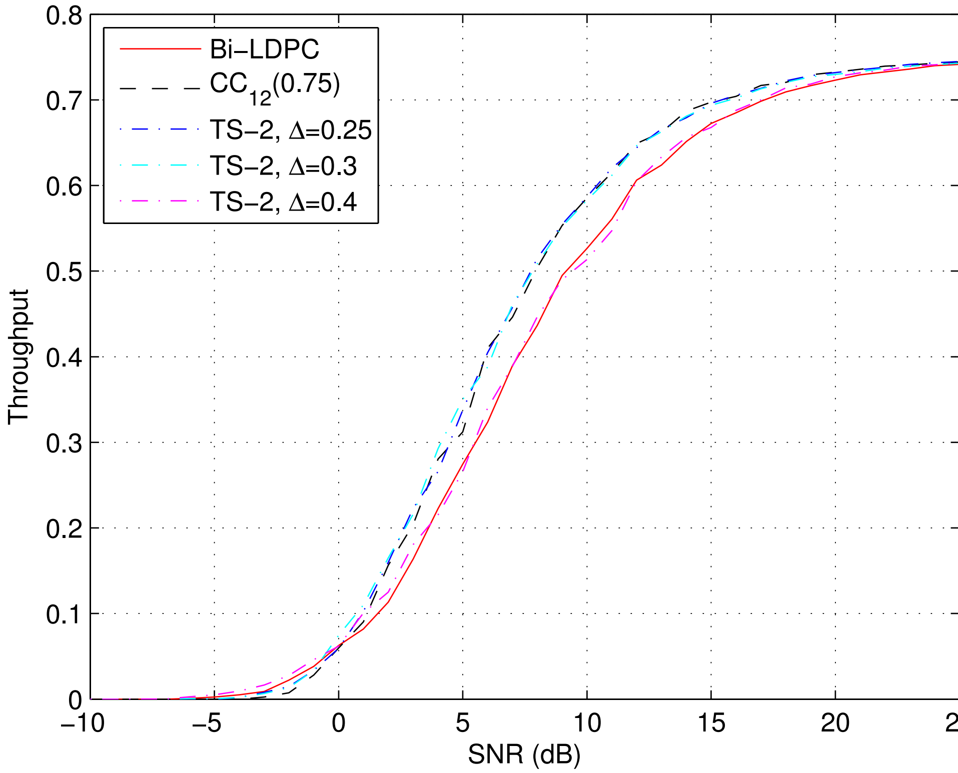

Figure 7.20: Throughput as a function of SNR for $CC_{12}(0.75)$, TS-2, and Bi-LDPC codes.

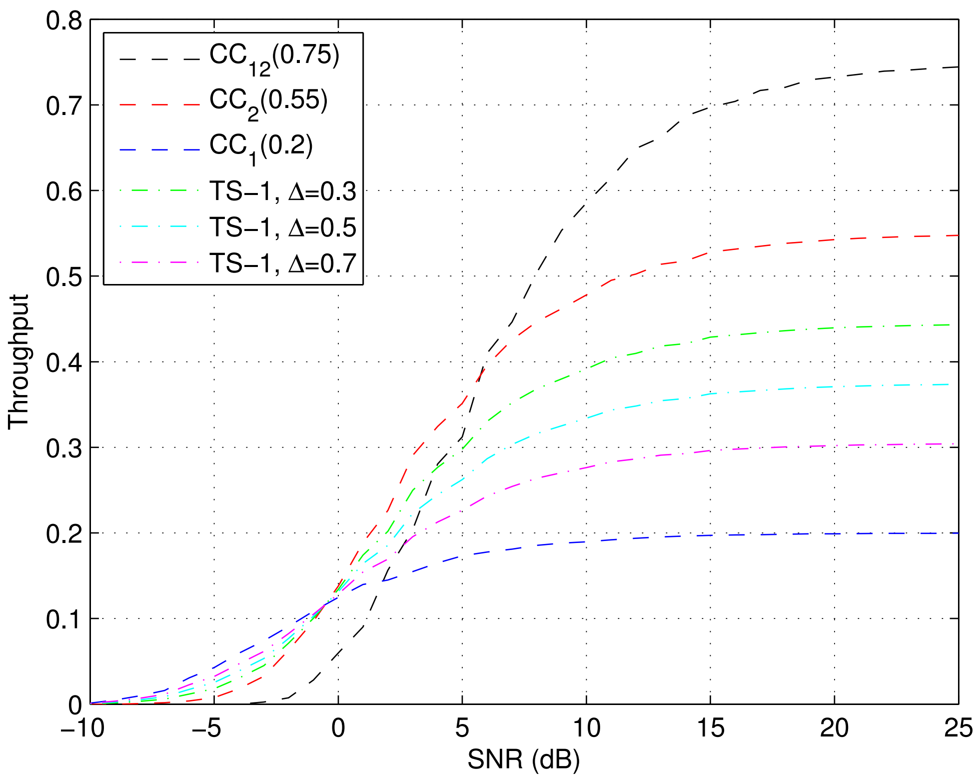

Figure 7.21: Throughput as a function of SNR for $CC_1(0.2)$, $CC_2(0.55)$, and TS-1 codes.

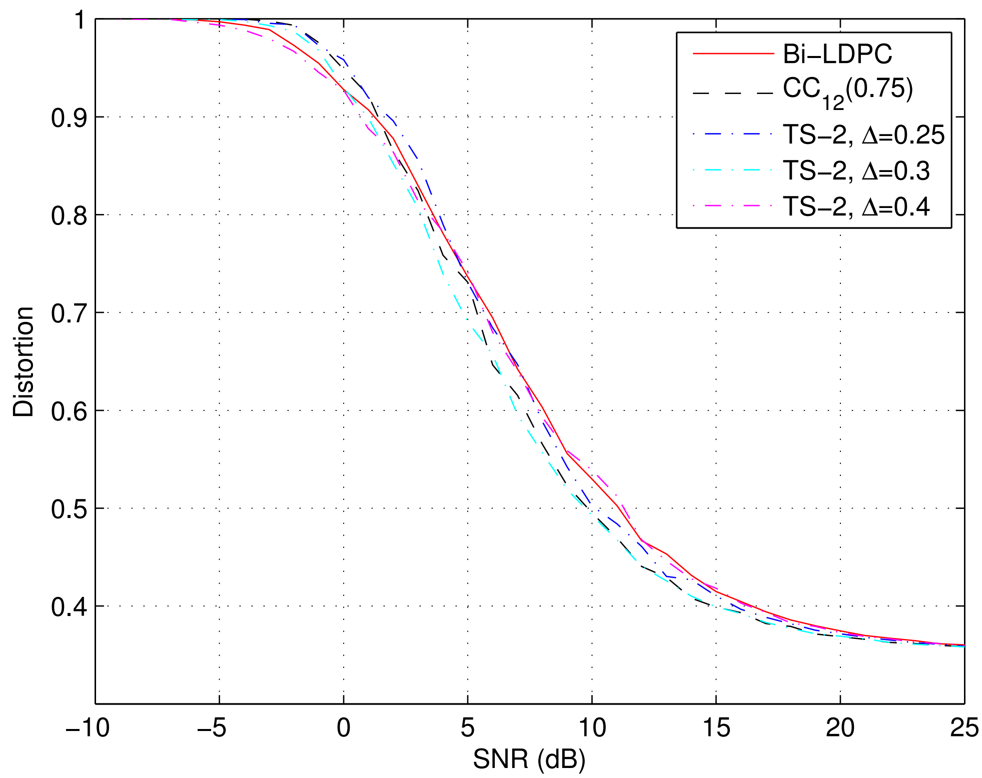

Figure 7.22: Average distortion of an LC-coded Gaussian $\mathcal{N}(0,1)$ source as a function of SNR for $CC_{12}(0.75)$, TS-2, and Bi-LDPC codes.

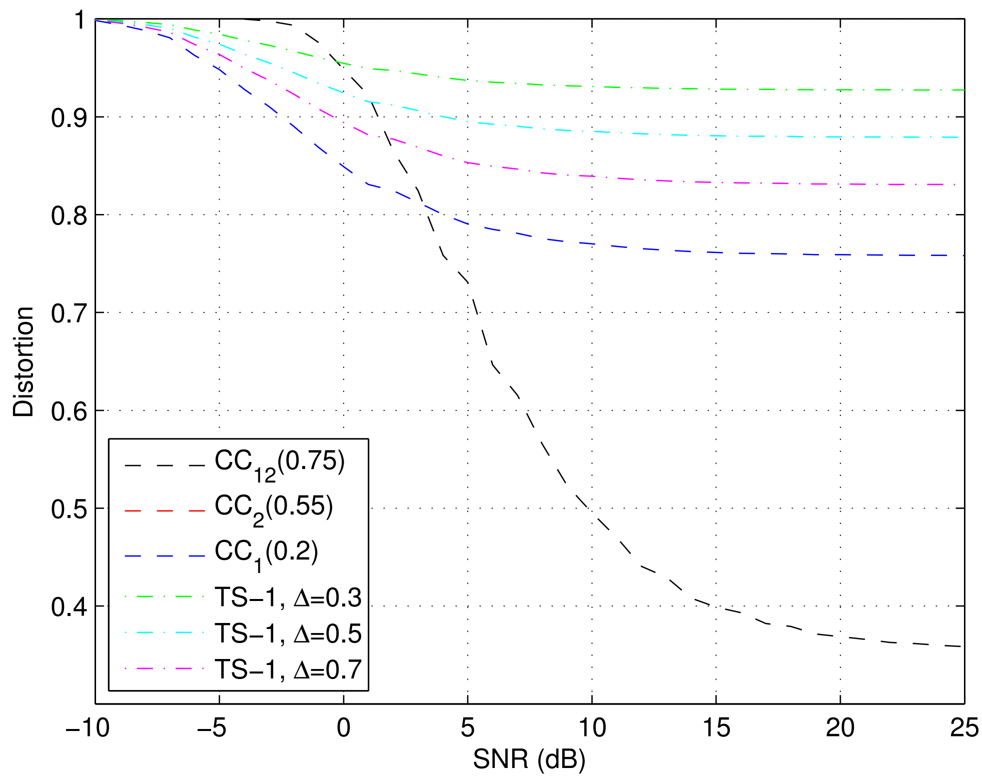

Figure 7.23: Average distortion of an LC-coded Gaussian $\mathcal{N}(0,1)$ source as a function of SNR for $CC_1(0.2)$, $CC_2(0.55)$, and TS-1 codes.

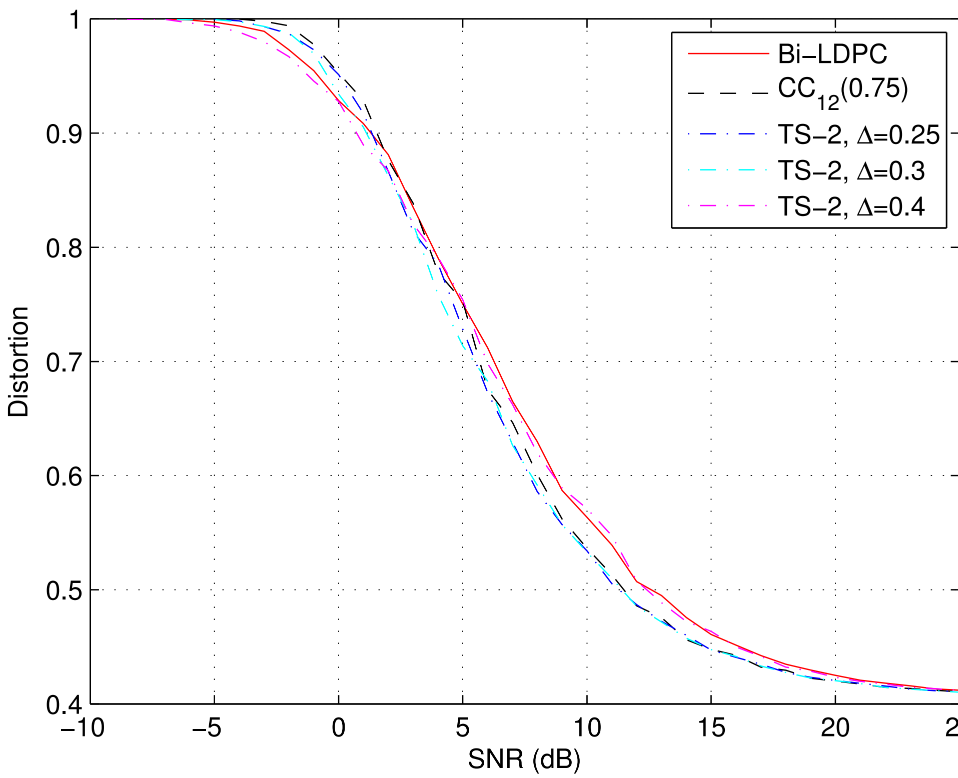

Figure 7.24: Average distortion of an MDC-coded Gaussian $\mathcal{N}(0, 1)$ source as a function of SNR for $CC_{12}(0.75)$, TS-2, and Bi-LDPC codes.

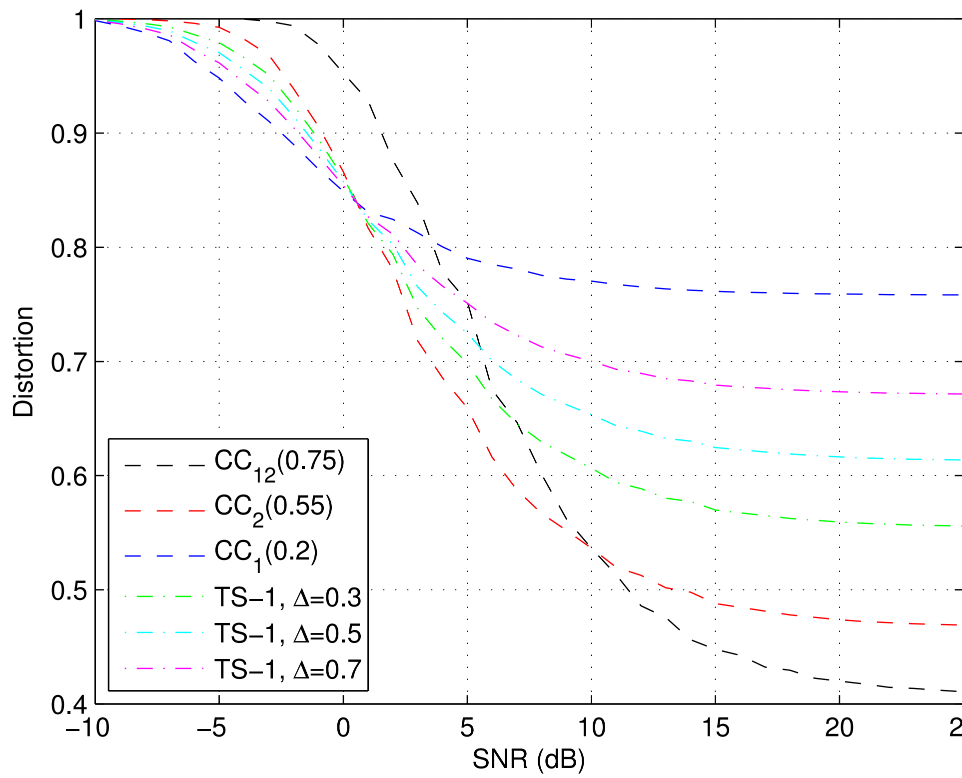

Figure 7.25: Average distortion of an MDC-coded Gaussian $\mathcal{N}(0,1)$ source as a function of SNR for $CC_1(0.2)$, $CC_2(0.55)$, and TS-1 codes.

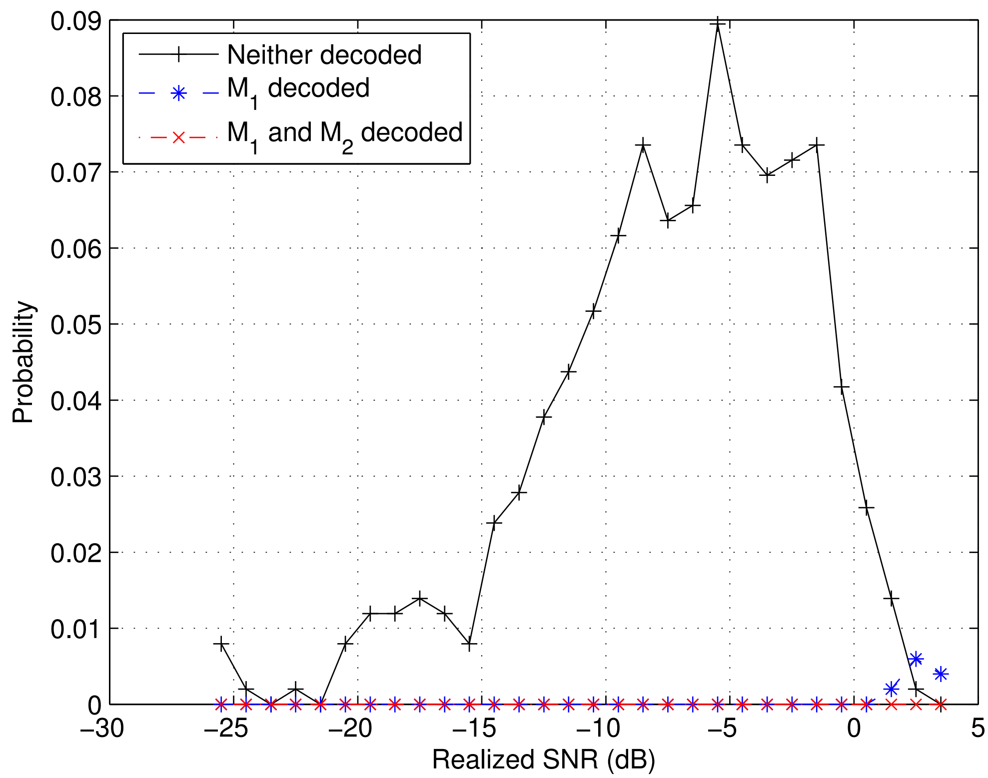

Figure 7.26: The decoding behaviour of Bi-LDPC codes over quasi-static Rayleigh fading channel with average SNR at −5 dB.

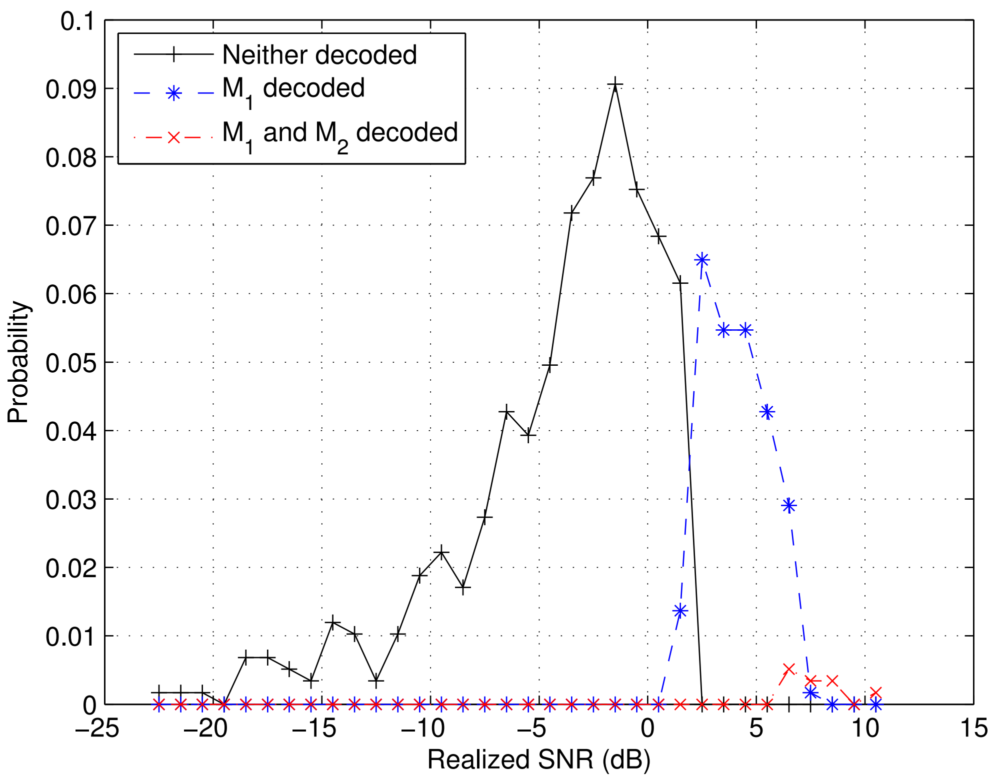

Figure 7.27: The decoding behaviour of Bi-LDPC codes over quasi-static Rayleigh fading channel with average SNR at 0 dB.

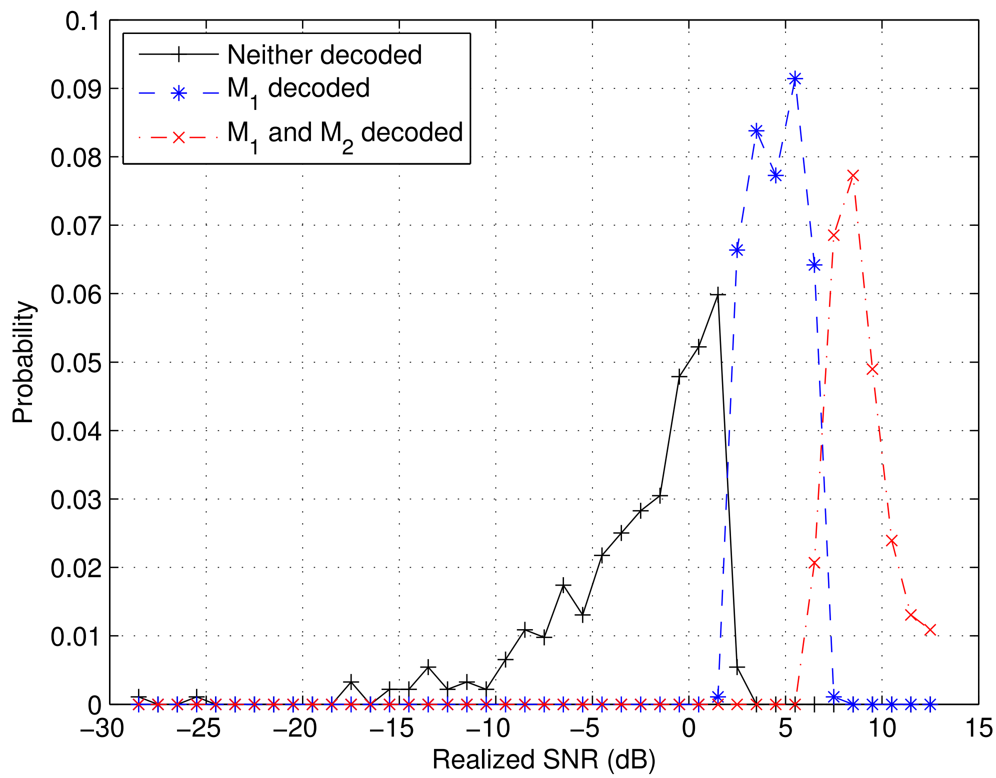

Figure 7.28: The decoding behaviour of Bi-LDPC codes over quasi-static Rayleigh fading channel with average SNR at 5 dB.

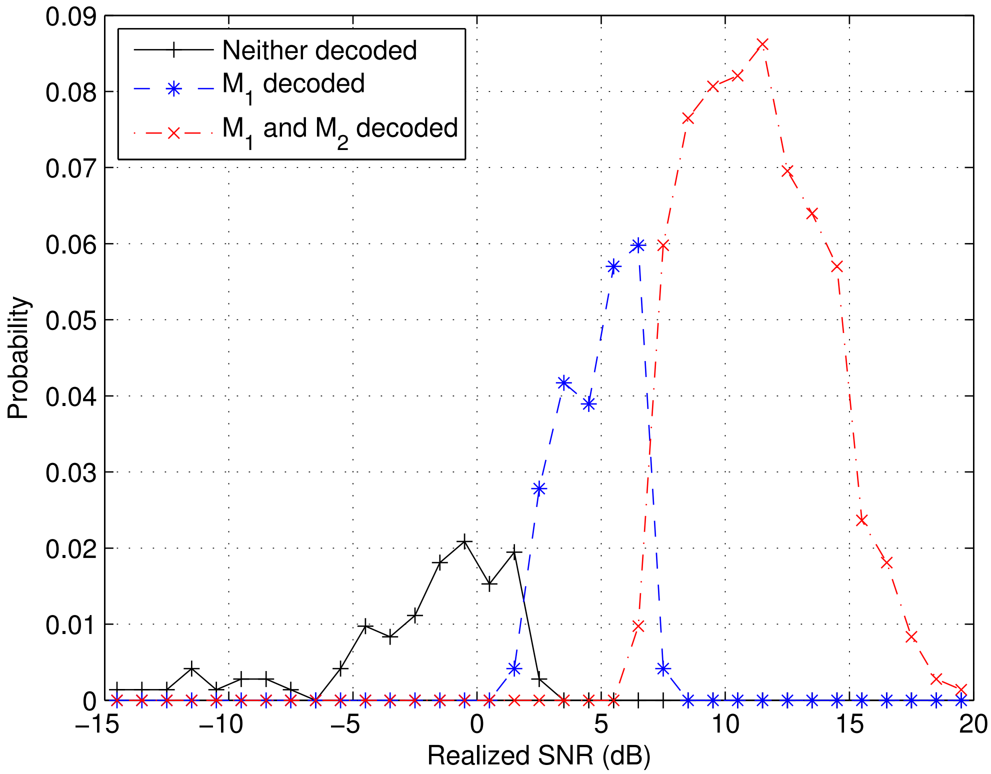

Figure 7.29: The decoding behaviour of Bi-LDPC codes over quasi-static Rayleigh fading channel with average SNR at 10 dB.

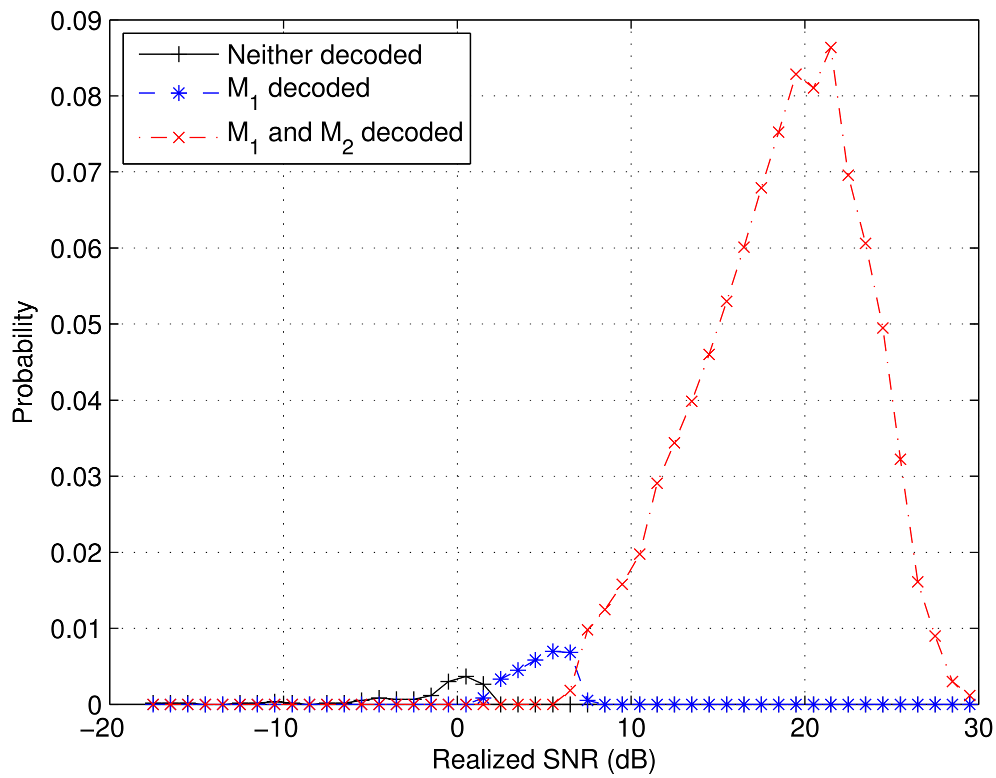

Figure 7.30: The decoding behaviour of Bi-LDPC codes over quasi-static Rayleigh fading channel with average SNR at 20 dB.

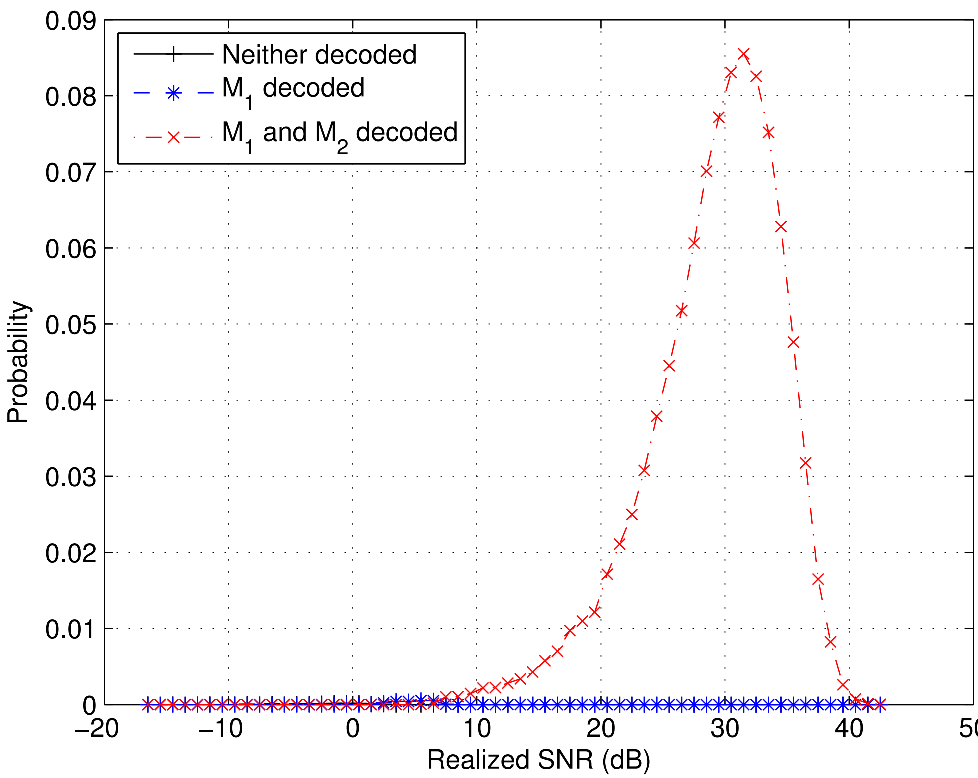

Figure 7.31: The decoding behaviour of Bi-LDPC codes over quasi-static Rayleigh fading channel with average SNR at 31 dB.

by LC and MDC respectively, the average achieved distortion on source reconstruction is further evaluated and compared among different coding schemes. It has been shown by the simulation results that the $\mathrm{CC}_1(R_1)$ code, $\mathrm{CC}_2(R_2)$ code, and TS-1 codes have large performance degradation compared with the other codes especially in high SNR region, in terms of the considered performance metrics, while the $\mathrm{CC}_{12}(R_1 + R_2)$ code, TS-2 codes, and Bi-LDPC each has its own advantage over different range of SNR, compared with each other.

Based on the simulation results, it also appears that the conventional channel codes $\mathrm{CC}_{12}(R_1+R_2)$, TS-2 codes, and Bi-LDPC codes perform similarly in terms of the success rates of both messages, the throughput, and the average distortion on source reconstruction of an LC-coded and an MDC-coded Gaussian $\mathcal{N}(0, 1)$ source. Before making any further conjecture about the properties of the Bi-LDPC code construction, two possible improvements may necessarily be taken into consideration.

On one hand, the Bi-LDPC codes can be optimized at the codeword length n going to infinity, where the optimality depends on the performance metric. Instead of k_1, k_2, and n, we are given the two code rates R_1 and R_2. Suppose the throughput, as a special case of average utility, is the performance metric which we need to maximize. The cost function is thus given as that in (7.4), and the maximization is taken over all possible Bi-LDPC codes of rates R_1 and R_2. Since Bi-LDPC codes are parametrized by $(\lambda_1, \rho_1, \lambda_2, \rho_2, \alpha)$, and the parameters are related according to (7.8) to (7.11). Aware of the subtlety among the fraction α, and the two rates r_1 and r_2, the optimization problem may be solved by either forming it into a convex optimization problem [124] based on DE [94, 95] and the 1-D EXIT chart approach [96], or at least exhaustive search.

On the other hand, it is necessary to investigate whether such performance behaviour is due to the restriction of binary signaling since BPSK modulation is applied during the transmission. Along with such consideration, we are required to determine the theoretical performance limit of Bi-LDPC codes under binary signaling transmission scheme. Instead of the direct study on Bi-LDPC codes, we provide the guidelines to numerically determine the fundamental performance limit of superposition codes. For any given average SNR of the quasi-static Rayleigh fading channel, based on the approach discussed in Chapter 5, we can numerically evaluate the maximal achievable throughput by finding the best superposition code encoding message M_1 first with M_2 superimposed when the channel input is limited to binary. Similarly, for arbitrary distribution on channel input, we are certain to numerically calculate the maximal achievable throughput by superposition codes. Should the performance limitation by binary signaling is not promising enough, we are to consider the high order signaling transmission scheme next. For example, in [123], a 4-ary signaling scheme is studied under the concept of superposition codes.

Chapter 8

Conclusion and Suggestions

Wireless and Internet applications have been stimulated with the increasing demand for multimedia streaming over wireless channels and packet-loss channels. The modern source and channel coding paradigm motivates us to study the communication of dependent messages over compound channels.

This thesis contributes to the subject matter in the following aspects.

- We propose a model for source-coded messages to model arbitrary dependency structure among source-coded messages for the communication of multimedia content. Assigning to each subset of the messages a utility value, this source model has an intuitive graphical representation.

- We study the problem of communicating source-coded messages under such a model and use the average utility to evaluate the performance of the communication systems. Maximal achievable utilities are derived for any given source over any DMC. Numerical comparisons are made for two specific example of source models over BSC, where the two models correspond respectively to the messages generated by a layered-coding source encoder and those generated by a multiple-description-coding source encoder. Depending on the tolerable delays and the channel conditions, the two sources each has its own advantage in terms of its maximal achievable utility.

- We formulate the problem of communicating such messages over degraded compound channels as the problem of maximizing the average utility at the receiver.

The performance of superposition coding schemes, which include the conventional channel code special cases, are investigated. We show that with superposition codes, the maximal utility is achieved by coding along the best sub-chain in the utility graph, and when restricting the superposition codes to the conventional channel code family, the utility is maximized by a capacity-achieving channel code over the best node. For the case of two source-coded messages, we show that the maximal achievable utility by superposition codes is the maximum among all coding schemes. We numerically evaluate the maximal achievable utility for LC-coded and MDC-coded messages under superposition coding schemes and demonstrate the impact of the channel statistics and delay on the achieved utility.

- We consider the problem of communicating Gaussian $\mathcal{N}(0,1)$ sequence over compound channels. It is formulated as the problem of minimizing the average distortion at the receiver, where the minimization is taken over different schemes of source coding and channel coding. For degraded compound channels, the minimal achievable distortion of rate distortion-coded, LC-coded, and MDC-coded source is derived and compared for the communication by conventional channel codes and superposition codes. By conventional channel codes the minimal achievable distortion of different source coding schemes are same, however by superposition codes the minimal achievable distortion of an LC-coded source is lower than that of an MDC-coded source. Moreover, for either an LC-coded or an MDC-coded source, the minimum distortion achieved by superposition codes outperforms that achieved by conventional channel codes. Out of all the considered schemes, the combination of LC and superposition codes is the best.

- We consider communicating two source dependent messages over quasi-static Rayleigh fading channels. The success rates of both messages, and the throughput are evaluated and compared among three different coding schemes, namely, conventional channel codes, time sharing codes, and an LDPC-based Bi-LDPC codes. For a Gaussian $\mathcal{N}(0,1)$ source sequence encoded by LC and MDC respectively, the average achieved distortion on source reconstruction is further evaluated and compared among different coding schemes. We show by simulations that in terms of the considered performance metrics, the conventional channel codes encoding a single message and TS-1 codes have large performance degradation, compared with the other coding schemes in high SNR region, while the conventional channel code encoding both messages, TS-2 codes, and Bi-LDPC code each has its own advantage over different region of SNR, when compared with each other.

At the point, we shall turn to consider far beyond this. Some future work exists in the following areas.

- Under the proposed source model, for any given source-encoded messages, it is of importance to consider the communication of the dependent source messages over general compound channels, possibly formulated as the utility maximization problem. Practically, the block fading channels may be taken into consideration, over which the transmission of a single codeword spans a finite number of independent channel realizations. The state-of-the-art channel coding and transmission schemes are to be investigated on both theoretical and practical aspects.

- Over degraded compound channels, it is also of interest to investigate whether superposition codes are optimal solution for communicating the given source messages within an arbitrary dependency structure for the case of arbitrary number of source messages. Some implications are provided by the work of [7, 9], in which a concept of infinite layer message superposition for transmission over slow fading channels has been shown to be optimal.

- On joint source channel coding of the Gaussian $\mathcal{N}(0, 1)$ sequence, due to the negligible advantage of the two-layer LC codes over the two-description MDC codes, further comparison between LC of more than two layers and MDC of more than two descriptions needs to be considered. The MDC can be constructed by the suboptimal approach proposed in [37].

- As discussed in Chapter 7, attention needs to be given to the optimization of Bi-LDPC codes in terms of some given performance metric. Based on the notion of DE [94, 95] and the 1-D EXIT chart approach [96], the problem is expected to be formulated as a convex optimization problem. Terminologically, in LDPC coding context, an elementary EXIT chart is required to be defined and analyzed potentially.

Appendix

I Proof of Lemma 3

We fist note that a sequence of degraded channels $\theta^1 \prec \cdots \prec \theta^m$ are characterized by the distribution ψ. To prove Lemma 3, we need the results as follows.

Lemma 4 *For the sequence of superposition codes C_n parametrized by any given $(\lambda, \psi, R_{\Delta(1)} \to R_{\Delta(2)} \to \cdots \to R_{\Delta(m)})$ over channel $\theta : X_{\Delta(m)} \to Y$, if $\theta \prec \theta^i$, then the probability of decoding error for $M_{\Delta(i)}$ approaches one as $n \to \infty$.*

Proof: If $\theta \prec \theta^i$, by the chain rule, we have

$$\begin{aligned} I\left(X_{\Delta(i)}; Y, Y_{\Delta(i)} | X_{\Delta(i-1)}\right) &= I\left(X_{\Delta(i)}; Y_{\Delta(i)} | X_{\Delta(i-1)}\right) + I\left(X_{\Delta(i)}; Y | X_{\Delta(i-1)}, Y_{\Delta(i)}\right) \\ &= I\left(X_{\Delta(i)}; Y | X_{\Delta(i-1)}\right) + I\left(X_{\Delta(i)}; Y_{\Delta(i)} | X_{\Delta(i-1)}, Y\right). \end{aligned}$$

Since $X_{\Delta(i)}$ and Y are conditionally independent given $Y_{\Delta(i)}$, we have $I(X_{\Delta(i)}; Y | X_{\Delta(i-1)}, Y_{\Delta(i)}) = 0$. By neglecting the case $I(X_{\Delta(i)}; Y_{\Delta(i)} | X_{\Delta(i-1)}) = 0$ such that the result holds trivially, we have

$$I\left(X_{\Delta(i)}; Y | X_{\Delta(i-1)}\right) < I\left(X_{\Delta(i)}; Y_{\Delta(i)} | X_{\Delta(i-1)}\right).$$

If $i = 1$, directly by the results on Csiszar's sphere packing bound [121], the probability of decoding error for $M_{\Delta(1)}$ goes to one as $n \to \infty$ regardless of the invertibility of the mapping h_1. For $i > 1$, consider an arbitrary decoder $\chi : \mathcal{Y}^n \to \{1, \ldots, 2^{nI(X_{\Delta(i)}; Y_{\Delta(i)} | X_{\Delta(i-1)})}\}$ where $\mathcal{Y}$ is the output alphabet of channel θ. Assume h_i is invertible and the decoder has the knowledge of the vector $X^n_{\Delta(i-1)}(M_{\alpha_{i-1}})$, denoted $x^n_{\Delta(i-1)}$. Then we are given a codeword set

$$\left\{x^n_{\Delta(i),1},\ldots,x^n_{\Delta(i),M}\right\}\subset T^n_{X_{\Delta(i)}|X_{\Delta(i-1)}}\left(x^n_{\Delta(i-1)}\right),$$

where $T^n_{X_{\Delta(i)}|X_{\Delta(i-1)}}(x^n_{\Delta(i-1)})$ is the V-shell of $x^n_{\Delta(i-1)}$ with the conditional type specified by $P_{X_{\Delta(i)}|X_{\Delta(i-1)}}$ and M is a positive integer. The decision region of the decoder χ is

$$D_j=\{y^n:\ \chi(y^n)=j\}$$

for $j=1,\ldots,M$. Let $W^n(y^n|x^n_{\Delta(i)})$ denote the probability of receiving y^n when $x^n_{\Delta(i)}$ is transmitted over the channel $X_{\Delta(i)}\to Y$, and $\mathrm{p}^{(n)}_{e,j}$ denote the probability of decoding error when $x^n_{\Delta(i),j}$ is sent over the channel. For every conditional type V of y^n given $x^n_{\Delta(i)}$ with $P_{X_{\Delta(i)}}$ induced by the joint distribution λ,

$$\begin{aligned}\sum_{j=1}^{M}\left|T^n_{Y|X_{\Delta(i)}}(x^n_{\Delta(i),j})\cap D_j\right| &\le \left|\bigcup_{j=1}^{M}T^n_{Y|X_{\Delta(i)}}\left(x^n_{\Delta(i),j}\right)\right|\\ &\le \left|\{y^n:\ (x^n_{\Delta(i-1)},y^n)\in A^{*(n)}_{\epsilon}\}\right|\\ &\le 2^{n\left(H\left(Y|X_{\Delta(i-1)}\right)+\epsilon'\right)},\end{aligned}$$

where $T^n_{Y|X_{\Delta(i)}}(x^n_{\Delta(i)})$ is the V-shell of $x^n_{\Delta(i)}$ with the conditional type specified by $P_{Y|X_{\Delta(i)}}$, and $A^{*(n)}_{\epsilon}$ denotes the ϵ-strongly jointly typical set. We note that the second inequality follows from Markov lemma and ϵ' can be arbitrarily small with an appropriate choice of ϵ and n.

In the sequel of the proof, we use the notations $\preceq$ and $\approx$ for inequality and equality up to a polynomial factor. Supposing $M\ge 2^{n(I(X_{\Delta(i)};Y_{\Delta(i)}|X_{\Delta(i-1)})+\delta)}$ for arbitrary $\delta\ge 2\epsilon'$, it follows that

$$\begin{aligned}\frac{1}{M}\sum_{j=1}^{M}\frac{\left|T^n_{Y|X_{\Delta(i)}}\left(x^n_{\Delta(i),j}\right)\cap D_j\right|}{\left|T^n_{Y|X_{\Delta(i)}}\left(x^n_{\Delta(i),j}\right)\right|} &\preceq \frac{1}{M}\frac{2^{n\left(H\left(Y|X_{\Delta(i-1)}\right)+\epsilon'\right)}}{2^{nH\left(Y|X_{\Delta(i)}\right)}}\\ &\le 2^{n\left(I\left(X_{\Delta(i)};Y|X_{\Delta(i-1)}\right)-I\left(X_{\Delta(i)};Y_{\Delta(i)}|X_{\Delta(i-1)}\right)+\epsilon'-\delta\right)}.\end{aligned}$$

In particular, if $I(X_{\Delta(i)};Y|X_{\Delta(i-1)}) \leq I(X_{\Delta(i)};Y_{\Delta(i)}|X_{\Delta(i-1)}) + \delta/2$ and $n \geq n_0(\epsilon', \delta)$ (sufficiently large), we have

$$\frac{1}{M}\sum_{j=1}^{M} \frac{\left|T^n_{Y|X_{\Delta(i)}}\left(x^n_{\Delta(i),j}\right) \cap D^c_j\right|}{\left|T^n_{Y|X_{\Delta(i)}}\left(x^n_{\Delta(i),j}\right)\right|} \geq 1 - \Gamma$$

for any $\Gamma > 0$.

Let $\mathrm{p}_e^{(n)}$ be the probability of decoding error for message $M_{\Delta(i)}$. It follows that

$$\begin{aligned}
\mathrm{p}_e^{(n)} &= \frac{1}{M}\sum_{j=1}^{M} \mathrm{p}_{e,j}^{(n)} \\
&= \frac{1}{M}\sum_{j=1}^{M} W^n\left(D^c_j | x^n_{\Delta(i),j}\right) \\
&\geq \frac{1}{M}\sum_{j=1}^{M} W^n\left(T^n_{Y|X_{\Delta(i)}}\left(x^n_{\Delta(i),j}\right) \cap D^c_j | x^n_{\Delta(i),j}\right) \\
&= \frac{1}{M}\sum_{j=1}^{M} \frac{\left|T^n_{Y|X_{\Delta(i)}}\left(x^n_{\Delta(i),j}\right) \cap D^c_j\right|}{\left|T^n_{Y|X_{\Delta(i)}}\left(x^n_{\Delta(i),j}\right)\right|} W^n\left(T^n_{Y|X_{\Delta(i)}}\left(x^n_{\Delta(i),j}\right) | x^n_{\Delta(i),j}\right) \\
&\geq (1-\Gamma) 2^{-nD\left(V||W|P_{X_{\Delta(i)}}\right)}
\end{aligned}$$

for conditional types V with $I(X_{\Delta(i)};Y|X_{\Delta(i-1)}) \leq I(X_{\Delta(i)};Y_{\Delta(i)}|X_{\Delta(i-1)}) + \delta/2$, where $D(V||W|P_{X_{\Delta(i)}})$ denotes the conditional KL divergence of V with respect to W under $P_{X_{\Delta(i)}}$. For sufficiently large n the minimum of $D(V||W|P_{X_{\Delta(i)}})$ approaches zero with arbitrary small δ if $I(X_{\Delta(i)};Y|X_{\Delta(i-1)}) \leq I(X_{\Delta(i)};Y_{\Delta(i)}|X_{\Delta(i-1)})$, and $\mathrm{p}_e^{(n)}$ approaches one. Without the assumptions that the mapping h_i is invertible, and the decoder has the knowledge of the vector $X^n_{\Delta(i-1)}(M_{\alpha_{i-1}})$, the decoder can not perform better and this completes the proof. □

Lemma 5 *For the sequence of superposition codes C_n parametrized by any given $(\lambda, \psi, R_{\Delta(1)} \to \cdots \to R_{\Delta(m)})$ over channel $\theta : X_{\Delta(m)} \to Y$, if $\theta \succ \theta^i$ and $\sum_{l\in\Delta(i)} R_l \leq I(X_{\Delta(i)};Y_{\Delta(i)}|X_{\Delta(i-1)})$, then the probability of decoding error for $M_{\Delta(i)}$ approaches zero as $n \to \infty$.*

Proof: If $\sum_{l\in\Delta(i)} R_l \leq I(X_{\Delta(i)};Y_{\Delta(i)}|X_{\Delta(i-1)})$, meaning h_i is invertible, the probability of decoding error for $M_{\Delta(i)}$ does not depend on which codeword is sent. Since the set of messages $M_{\Delta(i)}$ is treated as a single message by the encoder, without loss of generality, we assume that the message set M_{α_m} such that $M_{\Delta(i)} = 1$ for $i = 1, \ldots, m$ is sent. Let $Q(\cdot)$ denote the conditional probability of an event given the message set M_{α_m} is sent.

If $i = 1$, we define the event

$$E_{Yj} = \left\{ \left(X^n_{\Delta(1)}(j), Y^n \right) \in A^{(n)}_\epsilon \right\},$$

where j denotes the index encoded at layer 1, and $A^{(n)}_\epsilon$ denotes the set of jointly typical sequences. If $\theta \succ \theta^1$, by data processing inequality, we have

$$I\left(X_{\Delta(1)};Y_{\Delta(1)}\right) \leq I\left(X_{\Delta(1)};Y\right).$$

If n is large enough, the probability of error for decoding $M_{\Delta(1)}$ is

$$\begin{aligned} P^{(n)}_e(\Delta(1)) &= Q\left(E^c_{Y1} \cup \bigcup_{j\neq 1} E_{Yj} \right) \\ &\leq Q\left(E^c_{Y1}\right) + \sum_{j\neq 1} Q\left(E_{Yj}\right) \\ &\leq \epsilon + 2^{nI\left(X_{\Delta(1)};Y_{\Delta(1)}\right)} 2^{-n\left(I\left(X_{\Delta(1)};Y\right)-2\epsilon\right)} \\ &\leq 2\epsilon. \end{aligned}$$

If $1 < i \leq m$, we define the events

$$\begin{aligned} E_{Yj_1} &= \left\{ \left(X^n_{\Delta(1)}(j_1), Y^n \right) \in A^{(n)}_\epsilon \right\}, \\ E_{Yj_1j_2} &= \left\{ \left(X^n_{\Delta(1)}(j_1), X^n_{\Delta(2)}(j_1,j_2), Y^n \right) \in A^{(n)}_\epsilon \right\}, \\ &\cdots \\ E_{Yj_1\cdots j_i} &= \left\{ \left(X^n_{\Delta(1)}(j_1), \ldots, X^n_{\Delta(i)}(j_1,\ldots,j_i), Y^n \right) \in A^{(n)}_\epsilon \right\}, \end{aligned}$$

where $j_k,\ k = 1, \dots, i$ denotes the index encoded at layer k. When $\theta \succ \theta^i$, since $\theta^i \succ \cdots \succ \theta^1$, by data processing inequality, we have

$$\begin{aligned} I\left(X_{\Delta(1)}; Y_{\Delta(1)}\right) &\leq I\left(X_{\Delta(1)}; Y\right), \\ I\left(X_{\Delta(2)}; Y_{\Delta(2)}|X_{\Delta(1)}\right) &\leq I\left(X_{\Delta(2)}; Y|X_{\Delta(1)}\right) \\ &\cdots \\ I\left(X_{\Delta(i)}; Y_{\Delta(i)}|X_{\Delta(i-1)}\right) &\leq I\left(X_{\Delta(i)}; Y|X_{\Delta(i-1)}\right). \end{aligned}$$

Then the probability of error for decoding $M_{\Delta(i)}$ is

$$\begin{aligned} P_e^{(n)}(\Delta(i)) &= Q\left(E_{Y1}^c \cup \bigcup_{j_1 \neq 1} E_{Yj_1} \cup E_{Y11}^c \cup \bigcup_{j_2 \neq 1} E_{Y1j_2} \cup \cdots \cup E_{Y1\cdots1}^c \cup \bigcup_{j_i \neq 1} E_{Y1\cdots1j_i}\right) \\ &\leq Q(E_{Y1}^c) + \sum_{j_1 \neq 1} Q(E_{Yj_1}) + Q(E_{Y11}^c) + \sum_{j_2 \neq 1} Q(E_{Y1j_2}) + \cdots + Q(E_{Y1\cdots1}^c) \\ &\quad + \sum_{j_i \neq 1} Q(E_{Y1\cdots1j_i}). \end{aligned}$$

For any $k \leq i$ with $j_1 = \cdots = j_k = 1$, we can bound the probability of error as

$$Q(E_{Y1\cdots1}^c) \leq \epsilon.$$

For any $k \leq i$ with $j_1 = \cdots = j_{k-1} = 1$ and $j_k \neq 1$, we can bound the probability of error as

$$\begin{aligned} Q(E_{Y1\cdots1j_k}) &= \sum_{\left(X_{\Delta(1)}^n, \dots, X_{\Delta(k)}^n, Y^n\right) \in A_\epsilon^{(n)}} Q\left(X_{\Delta(1)}^n(1), \dots, X_{\Delta(k)}^n(1, \dots, j_k), Y\right) \\ &= \sum_{\left(X_{\Delta(1)}^n, \dots, X_{\Delta(k)}^n, Y^n\right) \in A_\epsilon^{(n)}} Q\left(X_{\Delta(1)}^n(1), \dots, X_{\Delta(k)}^n(1, \dots, j_k)\right) Q\left(Y^n|X_{\Delta(1)}^n, \dots, X_{\Delta(k-1)}^n\right) \\ &\leq \sum_{\left(X_{\Delta(1)}^n, \dots, X_{\Delta(k)}^n, Y^n\right) \in A_\epsilon^{(n)}} 2^{-n\left(H\left(X_{\Delta(1)}, \dots, X_{\Delta(k)}\right) - \epsilon\right)} 2^{-n\left(H\left(Y|X_{\Delta(k-1)}\right) - \epsilon\right)} \\ &\leq 2^{n\left(H\left(X_{\Delta(1)}, \dots, X_{\Delta(k)}, Y\right) + \epsilon\right)} 2^{-n\left(H\left(X_{\Delta(1)}, \dots, X_{\Delta(k)}\right) - \epsilon\right)} 2^{-n\left(H\left(Y|X_{\Delta(k-1)}\right) - \epsilon\right)} \\ &= 2^{-n\left(I\left(X_{\Delta(k)}; Y|X_{\Delta(k-1)}\right) - 3\epsilon\right)}, \end{aligned}$$

and hence

$$\begin{aligned}
\sum_{j_k \neq 1} Q(E_{Y1\cdots 1 j_k}) &\leq \sum_{j_k \neq 1} 2^{-n\left(I\left(X_{\Delta(k)};Y|X_{\Delta(k-1)}\right)-3\epsilon\right)} \\
&\leq 2^{nI\left(X_{\Delta(k)};Y_{\Delta(k)}|X_{\Delta(k-1)}\right)} 2^{-n\left(I\left(X_{\Delta(k)};Y|X_{\Delta(k-1)}\right)-3\epsilon\right)} \\
&= 2^{-n\left(I\left(X_{\Delta(k)};Y|X_{\Delta(k-1)}\right)-I\left(X_{\Delta(k)};Y_{\Delta(k)}|X_{\Delta(k-1)}\right)-3\epsilon\right)} \\
&\leq \epsilon.
\end{aligned}$$

Thus, we have

$$P_e^{(n)}\left(\Delta(i)\right) \leq 2i\epsilon$$

if n is large enough. If the mapping h_i is invertible, the above bounds show that we can decode $M_{\Delta(i)}$ with probability of error going to 0 as n goes to infinity. □

Consider the channel θ is such that $\theta^i \prec \theta \prec \theta^{i+1}$ holds for some $i \in \{0, \ldots, m\}$. On one hand, $\theta \prec \theta^{i+1}$, then $\theta \prec \theta^j$ holds for $i+1 \leq j \leq m$. By Lemma 4, the probability of error for decoding $M_{\Delta(j)},\ i+1 \leq j \leq m$ approaches one as $n \to \infty$. On the other hand, $\theta \succ \theta^i$ holds, and then $\theta \succ \theta^j$ holds for $1 \leq j \leq i$. By Lemma 5, the probability of error for decoding $M_{\Delta(j)},\ 1 \leq j \leq i$ approaches zero as $n \to \infty$ if h_j is invertible. For $\Delta(j) \subset \tau_i$, we have $\sum_{l \in \Delta(j)} R_l \leq I(X_{\Delta(j)}; Y_{\Delta(j)}|X_{\Delta(j-1)})$ and hence h_j is invertible. Therefore we have the decoding profile satisfying $\gamma(\tau_i) = 1$ as $n \to \infty$.

II Proof of Theorem 4

For the sequence of C_n codes, since $I(X_{\Delta(i)}; Y_{\Delta(i)}|X_{\Delta(i-1)}) < \sum_{l \in \Delta(i)} R_l$, h_i is not invertible and $M_{\Delta(i)}$ can not be decoded for any $\theta \in \Theta$. Let $P_{X_{\Delta(j)}|X_{\Delta(j-1)}}$ and $P_{X_{\Delta(j)}}$ for $j = 1, \ldots, m_\phi$ denote the distributions specifying λ_ϕ and the marginal distributions induced by λ_ϕ respectively. Generate another $\lambda'_\phi \in \prod_j \Lambda_{j|j-1}$. Let $P'_{X_{\Delta(j)}|X_{\Delta(j-1)}}$ and $P'_{X_{\Delta(j)}}$ for $j = 1, \ldots, m_\phi$ denote the distributions specifying λ'_ϕ and the marginal distributions induced by λ'_ϕ respectively. In the sequel of the proof, we use the subscript λ_ϕ and λ'_ϕ to distinguish the information quantities induced by different distributions.

If $i=1$, i.e., $I(X_{\Delta(1)};Y_{\Delta(1)}) < \sum_{l\in\Delta(1)} R_l$, we generate $P'_{X_{\Delta(j)}|X_{\Delta(j-1)}},\ j=1,\ldots,m_\phi$ as follows. Let $P'_{X_{\Delta(1)}} = P_{X_{\Delta(1)}}$ and $P'_{X_{\Delta(2)}|X_{\Delta(1)}}$ be such that $P'_{X_{\Delta(2)}} = P_{X_{\Delta(2)}}$ and $X_{\Delta(2)}$ is independent of $X_{\Delta(1)}$. For $3 \le j \le m_\phi$ let $P'_{X_{\Delta(j)}|X_{\Delta(j-1)}} = P_{X_{\Delta(j)}|X_{\Delta(j-1)}}$. Then we have $I_{\lambda'_\phi}(X_{\Delta(1)};Y_{\Delta(1)}) = 0$ by data processing inequality, and

$$\begin{aligned}
I_{\lambda'_\phi}\left(X_{\Delta(2)};Y_{\Delta(2)}|X_{\Delta(1)}\right) &= H_{\lambda'_\phi}\left(Y_{\Delta(2)}|X_{\Delta(1)}\right) - H_{\lambda'_\phi}\left(Y_{\Delta(2)}|X_{\Delta(2)}\right)\\
&= H_{\lambda'_\phi}\left(Y_{\Delta(2)}\right) - H_{\lambda'_\phi}\left(Y_{\Delta(2)}|X_{\Delta(2)}\right)\\
&\ge H_{\lambda_\phi}\left(Y_{\Delta(2)}|X_{\Delta(1)}\right) - H_{\lambda_\phi}\left(Y_{\Delta(2)}|X_{\Delta(2)}\right)\\
&= I_{\lambda_\phi}\left(X_{\Delta(2)};Y_{\Delta(2)}|X_{\Delta(1)}\right).
\end{aligned}$$

For $2 < j \le m_\phi$, we have $I_{\lambda'_\phi}(X_{\Delta(j)};Y_{\Delta(j)}|X_{\Delta(j-1)}) = I_{\lambda_\phi}(X_{\Delta(j)};Y_{\Delta(j)}|X_{\Delta(j-1)})$.

If $i \ne 1$, the distributions $P'_{X_{\Delta(j)}|X_{\Delta(j-1)}},\ j=1,\ldots,m_\phi$ can be constructed as follows. Let $X_{\Delta(i-1)}$ take the cardinality of $X_{\Delta(i)}$. Let $P'_{X_{\Delta(i)}|X_{\Delta(i-1)}}$ be such that $P'_{X_{\Delta(i)}|X_{\Delta(i-1)}}(x_{\Delta(i)}|x_{\Delta(i-1)}) = 1$ if $x_{\Delta(i)} = x_{\Delta(i-1)}$, and

$$P'_{X_{\Delta(i-1)}|X_{\Delta(i-2)}} = \sum_{x_{\Delta(i-1)}} P_{X_{\Delta(i)}|X_{\Delta(i-1)}} P_{X_{\Delta(i-1)}|X_{\Delta(i-2)}}.$$

Let $P'_{X_{\Delta(j)}|X_{\Delta(j-1)}} = P_{X_{\Delta(j)}|X_{\Delta(j-1)}}$ for $1 < j < i-1$ and $i \le j \le m_\phi$. With such generated distributions, we have $P'_{X_{\Delta(j)}} = P_{X_{\Delta(j)}}$ for $j \ne i-1$, and $P'_{X_{\Delta(i-1)}} = P'_{X_{\Delta(i)}}$. Then we have

$$\begin{aligned}
I_{\lambda'_\phi}\left(X_{\Delta(i)};Y_{\Delta(i)}|X_{\Delta(i-1)}\right) &= H_{\lambda'_\phi}\left(Y_{\Delta(i)}|X_{\Delta(i-1)}\right) - H_{\lambda'_\phi}\left(Y_{\Delta(i)}|X_{\Delta(i)}\right)\\
&= 0.
\end{aligned}$$

By the chain rule, we have

$$\begin{aligned}
I_{\lambda_\phi}\left(X_{\Delta(i)},X_{\Delta(i-1)};Y_{\Delta(i-1)}|X_{\Delta(i-2)}\right) &= I_{\lambda_\phi}\left(X_{\Delta(i)};Y_{\Delta(i-1)}|X_{\Delta(i-2)}\right)\\
&\quad + I_{\lambda_\phi}\left(X_{\Delta(i-1)};Y_{\Delta(i-1)}|X_{\Delta(i)},X_{\Delta(i-2)}\right)\\
&= I_{\lambda_\phi}\left(X_{\Delta(i-1)};Y_{\Delta(i-1)}|X_{\Delta(i-2)}\right)\\
&\quad + I_{\lambda_\phi}\left(X_{\Delta(i)};Y_{\Delta(i-1)}|X_{\Delta(i-1)},X_{\Delta(i-2)}\right).
\end{aligned}$$

Since $X_{\Delta(i-1)}$ and $Y_{\Delta(i-1)}$ are conditionally independent given $X_{\Delta(i)}$, we have

$$I_{\lambda_\phi}\left(X_{\Delta(i-1)}; Y_{\Delta(i-1)} | X_{\Delta(i)}, X_{\Delta(i-2)}\right) = 0.$$

Thus,

$$I_{\lambda_\phi}\left(X_{\Delta(i)}; Y_{\Delta(i-1)} | X_{\Delta(i-2)}\right) \geq I_{\lambda_\phi}\left(X_{\Delta(i-1)}; Y_{\Delta(i-1)} | X_{\Delta(i-2)}\right)$$

holds. Then we have

$$\begin{aligned} I_{\lambda'_\phi}\left(X_{\Delta(i-1)}; Y_{\Delta(i-1)} | X_{\Delta(i-2)}\right) &= I_{\lambda_\phi}\left(X_{\Delta(i)}; Y_{\Delta(i-1)} | X_{\Delta(i-2)}\right) \\ &\geq I_{\lambda_\phi}\left(X_{\Delta(i-1)}; Y_{\Delta(i-1)} | X_{\Delta(i-2)}\right). \end{aligned}$$

For $j \neq i, i-1$, we have $I_{\lambda'_\phi}(X_{\Delta(j)}; Y_{\Delta(j)} | X_{\Delta(j-1)}) = I_{\lambda_\phi}(X_{\Delta(j)}; Y_{\Delta(j)} | X_{\Delta(j-1)})$.

Let C'_n denote the sequence of superposition codes parametrized by $(\lambda'_\phi, \psi_\phi, R_{\Delta(1)} \to \cdots \to R_{\Delta(m_\phi)})$. Let τ_j and τ'_j for $j = 1, \ldots, m_\phi$ be associated to C_n and C'_n respectively. By the above results, we have $\tau'_j \supseteq \tau_j$, which results in $\kappa(\tau'_j) \geq \kappa(\tau_j)$ for $j = 1, \ldots, m_\phi$. On the other hand, since the distribution ψ_ϕ is fixed, the channels θ^j, $j = 1, \ldots, m_\phi$ are fixed. Consequently, by (5.4), we have $\Sigma' \geq \hat{\Sigma}$.

III Proof of Theorem 6

To prove Theorem 6, we need the results as follows.

Lemma 6 *For the sequence of conventional channel codes C_n specified by any given (H_α, r) over $\theta \in \Theta$, if θ is such that $\theta \succ \theta^*$, the limiting decoding profile γ satisfies $\gamma(\alpha) = 1$; if $\theta \prec \theta^*$, $\gamma(\emptyset) = 1$.*

Proof: First we have the channel code rate $R = \sum_{i \in \alpha} H_i / r$. For the case where $\theta \succ \theta^*$, by data processing inequality, we have $\mathrm{Cap}(\theta) \geq \mathrm{Cap}(\theta^*)$. Since $R = \mathrm{Cap}(\theta^*)$, we have $R < \mathrm{Cap}(\theta)$. By the achievability of the channel coding theorem, the probability of error for decoding M_α approaching zero as $n \to \infty$. Thus the decoding profile γ satisfies $\gamma(\alpha) = 1$.

For the case where $\theta \prec \theta^*$, by data processing inequality, we have $\mathrm{Cap}(\theta) < \mathrm{Cap}(\theta^*)$, and hence $R > \mathrm{Cap}(\theta)$ holds. By the strong converse of the channel coding theorem, the probability of error for decoding M_α goes exponentially to one as $n \to \infty$. Thus the decoding profile γ satisfies $\gamma(\emptyset) = 1$. □

For the sequence of conventional channel codes C_n specified by any tuple (H_α, r), by (6.7), we have

$$\begin{aligned}\lim_{n\to\infty} \nu(C_n) &= \lim_{n\to\infty} \int \sum_{\alpha\subseteq\mathcal{N}} \gamma(\alpha|C_n,\theta) D(\alpha) q(\theta) d\theta \\ &= \int \sum_{\alpha\subseteq\mathcal{N}} \hat{\gamma}(\alpha|\theta) D(\alpha) q(\theta) d\theta,\end{aligned}$$

where $\hat{\gamma}$ denotes the limiting decoding profile. By Lemma 6, we have

$$\begin{aligned}\lim_{n\to\infty} \nu(C_n) &= \int \hat{\gamma}(\{\alpha\}) D(\alpha) q(\theta) d\theta \\ &= D(\alpha) \int \delta(\theta^* \prec \theta) q(\theta) d\theta + D(\emptyset) \int \delta(\theta^* \succ \theta) q(\theta) d\theta \\ &= D(\alpha)\Pr(\theta : \theta^* \prec \theta) + D(\emptyset)\Pr(\theta : \theta^* \succ \theta) \\ &= D(\emptyset) - (D(\emptyset) - D(\alpha)) \Pr(\theta : \theta^* \prec \theta),\end{aligned}$$

where $\delta(\cdot)$ denotes the indicator function valued 1 if the argument is true and 0 otherwise.

Since C_n is specified by different choice of α, the distortion $\mathcal{D}'$ can be achieved by the minimization of $\lim_{n\to\infty} \nu(C_n)$ over all possible sequences of conventional channel codes for different choices of α.

IV Proof of Theorem 10

To prove Theorem 10, we need the results as follows.

Lemma 7 *For the sequence of superposition codes C_n specified by any given $(P_U, P_{X|U}, \theta', \theta'', R_1, R_2)$ over channel $\theta : X \to Y$, if $R_2 > I(X;Y|U)$, then the probability of decoding error for M_2 approaches one as $n \to \infty$.*

Proof: We consider an arbitrary decoder $\chi : \mathcal{Y}^n \to \{1, \ldots, 2^{nR_2}\}$ where $\mathcal{Y}$ is the output alphabet of channel θ, and assume the decoder has the knowledge of the vector $U^n(M_1)$, denoted u^n. Then we are given a codeword set $\{x_1^n, \ldots, x_M^n\} \subset T_{X|U}^n(u^n)$, where $T_{X|U}^n(u^n)$ is the V-shell of u^n with the conditional type specified by $P_{X|U}$ and M is a positive integer. The decision region of the decoder χ is $D_i = \{y^n : \ \chi(y^n) = i\}$ for $i = 1, \ldots, M$. Let $W^n(y^n|x^n)$ denote the probability of receiving y^n when x^n is sent over the channel θ, and $\mathrm{p}_{e,i}^{(n)}$ denote the probability of decoding error when x_i^n is sent over the channel. For every conditional type V of y^n given x^n with P_X induced by P_U and $P_{X|U}$,

$$\begin{aligned}\sum_{i=1}^{M} \left|T_{Y|X}^n\left(x_i^n\right) \cap D_i\right| &\leq \left|\bigcup_{i=1}^{M} T_{Y|X}^n\left(x_i^n\right)\right| \\ &\leq \left|\left\{y^n : \ (u^n, y^n) \in A_\epsilon^{*(n)}\right\}\right| \\ &\leq 2^{n(H(Y|U)+\epsilon')},\end{aligned}$$

where $T_{Y|X}^n(x^n)$ is the V-shell of x^n with the conditional type specified by $P_{Y|X}$, and $A_\epsilon^{*(n)}$ denotes the ϵ-strongly jointly typical set. We note that the second inequality follows from Markov lemma and ϵ' can be arbitrarily small with an appropriate choice of ϵ and n.

In the sequel, we use the notations $\preceq$ and $\approx$ for inequality and equality up to a polynomial factor. Supposing $M \geq 2^{n(R_2+\delta)}$ for arbitrary $\delta \geq 2\epsilon'$, it follows that

$$\begin{aligned}\frac{1}{M}\sum_{i=1}^{M} \frac{\left|T_{Y|X}^n\left(x_i^n\right) \cap D_i\right|}{\left|T_{Y|X}^n\left(x_i^n\right)\right|} &\preceq \frac{1}{M}\frac{2^{n(H(Y|U)+\epsilon')}}{2^{nH(Y|X)}} \\ &\leq 2^{n(I(X;Y|U)-R_2+\epsilon'-\delta)}.\end{aligned}$$

In particular, if $I(X;Y|U) \leq R_2 + \delta/2$ and $n \geq n_0(\epsilon', \delta)$ (sufficiently large), we have

$$\frac{1}{M}\sum_{i=1}^{M} \frac{\left|T_{Y|X}^n\left(x_i^n\right) \cap D_i^c\right|}{\left|T_{Y|X}^n\left(x_i^n\right)\right|} \geq 1 - \Delta$$

for any $\Delta > 0$.

Let $\mathrm{p}_e^{(n)}$ be the probability of decoding error for message M_2. It follows that

$$\begin{aligned}
\mathrm{p}_e^{(n)} &= \frac{1}{M}\sum_{i=1}^{M}\mathrm{p}_{e,i}^{(n)} \\
&= \frac{1}{M}\sum_{i=1}^{M} W^n(D_i^c|x_i^n) \\
&\geq \frac{1}{M}\sum_{i=1}^{M} W^n\left(T^n_{Y|X}(x_i^n)\cap D_i^c|x_i^n\right) \\
&= \frac{1}{M}\sum_{i=1}^{M}\frac{\left|T^n_{Y|X}(x_i^n)\cap D_i^c\right|}{\left|T^n_{Y|X}(x_i^n)\right|}W^n\left(T^n_{Y|X}(x_i^n)|x_i^n\right) \\
&\geq (1-\Delta)2^{-nD(V||W|P_X)}
\end{aligned}$$

for conditional types V with $I(X;Y|U) \leq R_2 + \delta/2$, where $D(V||W|P_X)$ denotes the conditional KL divergence of V with respect to W under P_X. For sufficiently large n the minimum of $D(V||W|P_X)$ approaches zero with arbitrary small δ if $I(X;Y|U) \leq R_2$, and $\mathrm{p}_e^{(n)}$ approaches one. Without the assumption that the decoder has the knowledge of the vector $U^n(M_1)$, the decoder can not perform better and this completes the proof. □

Lemma 8 *For the sequence of superposition codes C_n specified by any given $(P_U, P_{X|U}, \theta', \theta'', R_1, R_2)$ over $\theta \in \Theta$, if θ is such that $\theta \succ \theta''$, the limiting decoding profile γ satisfies $\gamma(\{1,2\}) = 1$; if $\theta'' \succ \theta \succ \theta'$, $\gamma(\{1\}) = 1$; if $\theta \prec \theta'$, $\gamma(\emptyset) = 1$.*

Proof: Let Y denote the output of channel θ. For the case where θ is such that $\theta \prec \theta'$ holds, by data processing inequality, we have $I(X;Y|U) < I(X;Y_2|U)$ and $I(U;Y) < I(U;Y_1)$. Since $R_1 = I(U;Y_1)$ and $R_2 = I(X;Y_2|U)$ hold, we have $R_2 > I(X;Y|U)$ and $R_1 > I(U;Y)$. By Lemma 7, the probability of decoding error for M_2 goes to one as $n \to \infty$. Also, by the results on Csiszar's sphere packing bound [121], the probability of decoding error for M_1 goes to one as well as $n \to \infty$. Thus the limiting decoding profile γ satisfies $\gamma(\emptyset) = 1$.

For the case where θ is such that $\theta'' \succ \theta \succ \theta'$ holds, by data processing inequality, we have $I(X;Y_2|U) > I(X;Y|U)$ and $I(U;Y) \geq I(U;Y_1)$. Thus we have $R_2 >$

$I(X;Y|U)$ and $R_1 \leq I(U;Y)$. Without loss of generality, we assume that the message pair $(M_1, M_2) = (1,1)$ is sent. Let $Q(\cdot)$ denote the conditional probability of an event given the message pair $(1,1)$ is sent. We define the event $E_{Yi} = \{(U^n(i), Y^n) \in A_\epsilon^{(n)}\}$, where $A_\epsilon^{(n)}$ denotes the set of jointly typical sequences. If n is large enough, the probability of error for decoding M_1 is

$$\begin{aligned}
P_e^{(n)}(1) &= Q\left(E_{Y1}^c \cup \bigcup_{i\neq 1} E_{Yi}\right) \\
&\leq Q(E_{Y1}^c) + \sum_{i\neq 1} Q(E_{Yi}) \\
&\leq \epsilon + 2^{nR_1} 2^{-n(I(U;Y)-2\epsilon)} \\
&\leq 2\epsilon.
\end{aligned}$$

On the other hand, by Lemma 7, the probability of decoding error for M_2 goes to one as $n \to \infty$. Thus the limiting decoding profile γ satisfies $\gamma(\{1\}) = 1$.

For the case where θ is such that $\theta \succ \theta''$ holds, by data processing inequality, we have $I(X;Y|U) \geq I(X;Y_2|U)$ and $I(U;Y) \geq I(U;Y_1)$. Thus we have $R_2 \leq I(X;Y|U)$ and $R_1 \leq I(U;Y)$. By the same arguments as above, the probability of decoding error for M_1 goes to zero as $n \to \infty$. Next, we define the event

$$E_{Yij} = \left\{(U^n(i), X^n(i,j), Y^n) \in A_\epsilon^{(n)}\right\}.$$

The probability of error for decoding M_2 is

$$\begin{aligned}
P_e^{(n)}(2) &= Q\left(E_{Y1}^c \cup \bigcup_{i\neq 1} E_{Yi} \cup \bigcup_{j\neq 1} E_{Y1j}\right) \\
&\leq Q(E_{Y1}^c) + \sum_{i\neq 1} Q(E_{Yi}) + \sum_{j\neq 1} Q(E_{Y1j}).
\end{aligned}$$

We can bound the third term as

$$\begin{aligned} Q(E_{Y1j}) &= \sum_{(U^n,\ldots,X^n,Y^n)\in A_\epsilon^{(n)}} Q\left(U^n(1), X^n(1,j), Y^n\right) \\ &= \sum_{(U^n,\ldots,X^n,Y^n)\in A_\epsilon^{(n)}} Q\left(U^n(1)\right) Q\left(X^n(1,j)|U^n(1)\right) Q\left(Y^n|U^n(1)\right) \\ &\leq \sum_{(U^n,\ldots,X^n,Y^n)\in A_\epsilon^{(n)}} 2^{-n(H(U)-\epsilon)} 2^{-n(H(X|U)-\epsilon)} 2^{-n(H(Y|U)-\epsilon)} \\ &\leq 2^{n(H(U,X,Y)+\epsilon)} 2^{-n(H(U)-\epsilon)} 2^{-n(H(X|U)-\epsilon)} 2^{-n(H(Y|U)-\epsilon)} \\ &= 2^{-n(I(X;Y|U)-4\epsilon)}. \end{aligned}$$

Therefore, we have

$$\begin{aligned} P_e^{(n)}(2) &\leq \epsilon + 2^{nR_1} 2^{-n(I(U;Y)-2\epsilon)} + 2^{nR_2} 2^{-n(I(X;Y|U)-4\epsilon)} \\ &\leq 3\epsilon \end{aligned}$$

if n is large enough. Thus the limiting decoding profile γ satisfies $\gamma(\{1,2\}) = 1$. □

For the sequence of superposition codes C_n specified by any tuple $(H_1, H_2, r, P_U, P_{X|U}, \theta', \theta'')$, by (6.7), we have

$$\begin{aligned} \lim_{n\to\infty} \nu(C_n) &= \lim_{n\to\infty} \int \sum_{\alpha\subseteq\mathcal{N}} \gamma(\alpha|C_n,\theta) D(\alpha) q(\theta) d\theta \\ &= \int \sum_{\alpha\subseteq\mathcal{N}} \hat{\gamma}(\alpha|\theta) D(\alpha) q(\theta) d\theta, \end{aligned}$$

where $\hat{\gamma}$ denotes the limiting decoding profile. By Lemma 8, we have

$$\begin{aligned}
\lim_{n\to\infty} \nu(C_n) &= \int \left(\hat{\gamma}(\emptyset)D(\emptyset) + \hat{\gamma}\left(\{1\}\right) D\left(\{1\}\right) + \hat{\gamma}\left(\{1,2\}\right) D\left(\{1,2\}\right)\right) q(\theta) d\theta \\
&= D(\emptyset)\int \delta(\theta \prec \theta')q(\theta)d\theta + D(\{1\})\int \delta(\theta' \prec \theta \prec \theta'')q(\theta)d\theta \\
&\quad + D(\{1,2\})\int \delta(\theta'' \prec \theta)q(\theta)d\theta \\
&= D(\emptyset)\Pr(\theta : \theta \prec \theta') + D(\{1\})\Pr(\theta : \theta' \prec \theta \prec \theta'') + D(\{1,2\})\Pr(\theta : \theta'' \prec \theta) \\
&= D(\emptyset) + (D(\emptyset) - D(\{1\}))\Pr(\theta : \theta' \prec \theta \prec \theta'') + (D(\emptyset) - D(\{1,2\}))\Pr(\theta : \theta'' \prec \theta),
\end{aligned}$$

where $\delta(\cdot)$ denotes the indicator function.

Since C_n is specified by different choice of $(P_U, P_{X|U}, \theta', \theta'')$, the distortion $\mathcal{D}^*$ is thus achievable by the minimization of $\lim_{n\to\infty} \nu(C_n)$ over all possible sequences of superposition codes for different choices of $(P_U, P_{X|U}, \theta', \theta'')$.

References

[1] R. Ahlswede, "Certain results in coding theory for compound channels," in *Proc. Coll. Inf. The. Debrecen 1967,* A. Renyi, Ed. Budapest, Hungary: J. Bolyai Math. Soc., vol. 1, 1968, pp. 35–60.

[2] D. Blackwell, L. Breiman, and A. Thomasian, "The capacity of a class of channels," *The Annals of Mathematical Statistics*, pp. 1229–1241, 1959.

[3] A. Lapidoth and P. Narayan, "Reliable communication under channel uncertainty," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2148–2177, Oct. 1998.

[4] A. J. Goldsmith and P. P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Transactions on Information Theory*, vol. 43, no. 6, pp. 1986–1992, 1997.

[5] A. Lapidoth and I. E. Telatar, "The compound channel capacity of a class of finite-state channels," *Information Theory, IEEE Transactions on*, vol. 44, no. 3, pp. 973–983, 1998.

[6] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: Information-theoretic and communications aspects," *Information Theory, IEEE Transactions on*, vol. 44, no. 6, pp. 2619–2692, 1998.

[7] S. Shamai, "A broadcast strategy for the gaussian slowly fading channel," in *Information Theory. 1997. Proceedings., 1997 IEEE International Symposium on.* IEEE, 1997, p. 150.

[8] Y. Liu, K. N. Lau, O. Y. Takeshita, and M. P. Fitz, "Optimal rate allocation for superposition coding in quasi-static fading channels," in *Information Theory, 2002. Proceedings. 2002 IEEE International Symposium on*. IEEE, 2002, p. 111.

[9] S. S. (Shitz) and A. Steiner, "A broadcast approach for a single-user slowly fading MIMO channel," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2617–2635, Oct. 2003.

[10] A. Steiner and S. Shamai, "Multi-layer broadcasting over a block fading mimo channel," *Wireless Communications, IEEE Transactions on*, vol. 6, no. 11, pp. 3937–3945, 2007.

[11] A. G. i Fabregas and G. Caire, "Coded modulation in the block-fading channel: Coding theorems and code construction," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 91–114, Jan. 2006.

[12] A. G. i Fabregas, "Coding in the block-erasure channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5116–5121, Nov. 2006.

[13] J. J. Boutros, A. Fabregas, and E. C. Strinati, "Binary codes in the block-fading channel," in *Communications Theory Workshop, 2006. Proceedings. 7th Australian*. IEEE, 2006, pp. 7–14.

[14] S. H. Hassani, S. B. Korada, and R. Urbanke, "The compound capacity of polar codes," *arXiv preprint arXiv:0907.3291*, 2009.

[15] J. J. Boutros, A. Guillén i Fàbregas, E. Biglieri, and G. Zémor, "Low-density parity-check codes for nonergodic block-fading channels," *Information Theory, IEEE Transactions on*, vol. 56, no. 9, pp. 4286–4300, 2010.

[16] W. H. R. Equitz and T. M. Cover, "Successive refinement of information," *IEEE Trans. Inf. Theory*, vol. 37, no. 2, pp. 269–275, Mar. 1991.

[17] B. Rimoldi, "Successive refinement of information: Characterization of the achievable rates," *Information Theory, IEEE Transactions on*, vol. 40, no. 1, pp. 253–259, 1994.

[18] W. B. Pennebaker and J. L. Mitchell, *JPEG: Still image data compression standard*. Springer, 1993.

[19] J.-R. Ohm, "Advances in scalable video coding," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 42–56, 2005.

[20] U. Horn, K. Stuhlmüller, M. Link, and B. Girod, "Robust internet video transmission based on scalable coding and unequal error protection," *Signal Processing: Image Communication*, vol. 15, no. 1, pp. 77–94, 1999.

[21] D. Wu, Y. T. Hou, and Y.-Q. Zhang, "Scalable video coding and transport over broadband wireless networks," *Proceedings of the IEEE*, vol. 89, no. 1, pp. 6–20, 2001.

[22] H. M. Radha, M. Van der Schaar, and Y. Chen, "The mpeg-4 fine-grained scalable video coding method for multimedia streaming over ip," *Multimedia, IEEE Transactions on*, vol. 3, no. 1, pp. 53–68, 2001.

[23] B.-J. Kim, Z. Xiong, and W. A. Pearlman, "Low bit-rate scalable video coding with 3-d set partitioning in hierarchical trees (3-d spiht)," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 10, no. 8, pp. 1374–1387, 2000.

[24] Y. He, F. Wu, S. Li, Y. Zhong, and S. Yang, "H. 26l-based fine granularity scalable video coding," in *Circuits and Systems, 2002. ISCAS 2002. IEEE International Symposium on*, vol. 4. IEEE, 2002, pp. IV–548.

[25] R. Aravind, M. R. Civanlar, and A. R. Reibman, "Packet loss resilience of MPEG-2 scalable video coding algorithms," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 5, pp. 426–435, Oct. 1996.

[26] I. Draft, "recommendation and final draft international standard of joint video specification (ITU-T Rec. H. 264— ISO/IEC 14496-10 AVC)," *Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, JVTG050*, vol. 33, 2003.

[27] F. Kishino, K. Manabe, Y. Hayashi, and H. Yasuda, “Variable bit-rate coding of video signals for ATM networks,” *IEEE J. Sel. Areas Commun.*, vol. 7, no. 5, pp. 801–806, June 1989.

[28] M. Gallant and F. Kossentini, “Rate-distortion optimized layered coding with unequal error protection for robust internet video,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 3, pp. 357–372, Mar. 2001.

[29] Z. Liu, Y. Shen, S. S. Panwar, K. W. Ross, and Y. Wang, “Using layered video to provide incentives in p2p live streaming,” in *Proceedings of the 2007 workshop on Peer-to-peer streaming and IP-TV.* ACM, 2007, pp. 311–316.

[30] E. Ravelli, V. Melkote, T. Nanjundaswamy, and K. Rose, “Joint optimization of base and enhancement layers in scalable audio coding,” *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 21, no. 4, pp. 711–724, 2013.

[31] J. Wolf, A. Wyner, and J. Ziv, “Source coding for multiple descriptions,” *Bell System Technical Journal*, vol. 59, no. 8, pp. 1417–1426, 1980.

[32] L. Ozarow, “On a source-coding problem with two channels and three receivers,” *Bell System Technical Journal, The*, vol. 59, no. 10, pp. 1909–1921, 1980.

[33] A. A. E. Gamal and T. M. Cover, “Achievable rates for multiple descriptions,” *IEEE Trans. Inf. Theory*, vol. 28, no. 6, pp. 851–857, Nov. 1982.

[34] V. K. Goyal, “Multiple descriptiion coding: Compression meets the network,” *IEEE Signal Process. Mag.*, vol. 18, pp. 74–93, Sep. 2001.

[35] V. A. Vaishampayan, “Design of multiple description scalar quantizers,” *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 821–834, May 1993.

[36] S. D. Servetto, V. A. Vaishampayan, and N. Sloane, “Multiple description lattice vector quantization,” in *Data Compression Conference, 1999. Proceedings. DCC’99.* IEEE, 1999, pp. 13–22.

[37] V. A. Vaishampayan, N. Sloane, and S. D. Servetto, "Multiple-description vector quantization with lattice codebooks: Design and analysis," *Information Theory, IEEE Transactions on*, vol. 47, no. 5, pp. 1718–1734, 2001.

[38] N. Jayant, "Subsampling of a dpcm speech channel to provide two self-contained half-rate channels," *Bell System Technical Journal*, vol. 60, no. 4, pp. 501–509, 1981.

[39] V. K. Goyal, J. Kovacevic, R. Arean, and M. Vetterli, "Multiple description transform coding of images," in *Image Processing, 1998. ICIP 98. Proceedings. 1998 International Conference on*, vol. 1. IEEE, 1998, pp. 674–678.

[40] S. D. Servetto, K. Ramchandran, V. A. Vaishampayan, and K. Nahrstedt, "Multiple description wavelet based image coding," *Image Processing, IEEE Transactions on*, vol. 9, no. 5, pp. 813–826, 2000.

[41] Y. Wang, A. R. Reibman, and S. Lin, "Multiple description coding for video delivery," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 57–70, Jan. 2005.

[42] N. Franchi, M. Fumagalli, R. Lancini, and S. Tubaro, "Multiple description video coding for scalable and robust transmission over ip," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 15, no. 3, pp. 321–334, 2005.

[43] H. Jafarkhani, P. Ligdas, and N. Farvardin, "Adaptive rate allocation in a joint source/channel coding framework for wireless channels," in *Vehicular Technology Conference, 1996. Mobile Technology for the Human Race., IEEE 46th*, vol. 1. IEEE, 1996, pp. 492–496.

[44] S. Appadwedula, D. L. Jones, K. Ramchandran, and L. Qian, "Joint source channel matching for a wireless image transmission," in *Image Processing, 1998. ICIP 98. Proceedings. 1998 International Conference on*, vol. 2. IEEE, 1998, pp. 137–141.

[45] J. Cai and C. W. Chen, "Robust joint source-channel coding for image transmission over wireless channels," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 10, no. 6, pp. 962–966, 2000.

[46] D. Gunduz and E. Erkip, "Joint source–channel codes for mimo block-fading channels," *Information Theory, IEEE Transactions on*, vol. 54, no. 1, pp. 116–134, 2008.

[47] C. Tian, A. Steiner, S. Shamai, and S. N. Diggavi, "Successive refinement via broadcast: Optimizing expected distortion of a gaussian source over a gaussian fading channel," *Information Theory, IEEE Transactions on*, vol. 54, no. 7, pp. 2903–2918, 2008.

[48] C. T. Ng, D. Gunduz, A. J. Goldsmith, and E. Erkip, "Distortion minimization in gaussian layered broadcast coding with successive refinement," *Information Theory, IEEE Transactions on*, vol. 55, no. 11, pp. 5074–5086, 2009.

[49] H. Behroozi, J. Haghighat, F. Alajaji, and T. Linder, "On the transmission of a memoryless gaussian source over a memoryless fading channel," in *Communications, 2008 24th Biennial Symposium on.* IEEE, 2008, pp. 212–215.

[50] K. Zachariadis, M. L. Honig, and A. K. Katsaggelos, "Source fidelity over fading channels: performance of erasure and scalable codes," *Communications, IEEE Transactions on*, vol. 56, no. 7, pp. 1080–1091, 2008.

[51] T. M. Cover, "Broadcast channels," *IEEE Trans. Inf. Theory*, vol. 18, no. 1, pp. 2–14, Jan. 1972.

[52] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," *IRE Nat. Conv. Rec*, vol. 4, no. 142-163, p. 1, 1959.

[53] T. Berger and J. D. Gibson, "Lossy source coding," *Information Theory, IEEE Transactions on*, vol. 44, no. 6, pp. 2693–2723, 1998.

[54] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* John Wiley and Sons, 1991.

[55] B. Oliver, J. Pierce, and C. Shannon, "The philosophy of pcm," *Proceedings of the IRE*, vol. 36, no. 11, pp. 1324–1331, 1948.

[56] R. M. Gray and D. L. Neuhoff, "Quantization," *Information Theory, IEEE Transactions on*, vol. 44, no. 6, pp. 2325–2383, 1998.

[57] V. Koshelev, "Multilevel source coding and data-transmission theorem," in *Proc. VII All-Union Conf. Theory of Coding and Data Transmission*, 1978, pp. 85–92.

[58] ——, "Hierarchical coding of discrete sources," *Problemy peredachi informatsii*, vol. 16, no. 3, pp. 31–49, 1980.

[59] ——, "An evaluation of the average distortion for discrete scheme of sequential approximation," *Probl. Pered. Inform*, vol. 17, no. 3, pp. 20–33, 1981.

[60] J. Li and C.-C. Kuo, "Progressive coding of 3-d graphic models," *Proceedings of the IEEE*, vol. 86, no. 6, pp. 1052–1063, 1998.

[61] Z. Xiong, X. Wu, D. Y. Yun, and W. A. Pearlman, "Progressive coding of medical volumetric data using three-dimensional integer wavelet packet transform," in *Medical Technology Symposium, 1998. Proceedings. Pacific.* IEEE, 1998, pp. 384–387.

[62] T. Niihara and M. Ueda, "Moving picture signal progressive coding system," Dec. 22 1992, uS Patent 5,173,773.

[63] J. Li, "Method of visual progressive coding," Dec. 4 2001, uS Patent 6,327,392.

[64] Y. Wang, M. T. Orchard, V. Vaishiampayan, and A. R. Reibman, "Multiple description coding using pairwise correlating transforms," *IEEE Trans. Image Process.*, vol. 10, no. 3, pp. 351–366, Mar. 2001.

[65] T. Sikora, "Mpeg digital video-coding standards," *Signal Processing Magazine, IEEE*, vol. 14, no. 5, pp. 82–100, 1997.

[66] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H. 264/AVC standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 17, no. 9, pp. 1103–1120, 2007.

[67] I. E. Richardson, *H. 264 and MPEG-4 video compression: video coding for next-generation multimedia.* John Wiley & Sons, 2004.

[68] T. Schierl, T. Stockhammer, and T. Wiegand, "Mobile video transmission using scalable video coding," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 17, no. 9, pp. 1204–1217, 2007.

[69] W. Jiang and A. Ortega, "Multiple description speech coding for robust communication over lossy packet networks," in *Multimedia and Expo, 2000. ICME 2000. 2000 IEEE International Conference on*, vol. 1. IEEE, 2000, pp. 444–447.

[70] C.-S. Kim and S.-U. Lee, "Multiple description motion coding algorithm for robust video transmission," in *Circuits and Systems, 2000. Proceedings. ISCAS 2000 Geneva. The 2000 IEEE International Symposium on*, vol. 4. IEEE, 2000, pp. 717–720.

[71] W. S. Lee, M. R. Pickering, M. R. Frater, and J. F. Arnold, "A robust codec for transmission of very low bit-rate video over channels with bursty errors," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 10, no. 8, pp. 1403–1412, 2000.

[72] Y. Zhou and W.-Y. Chan, "Performance comparison of layered coding and multiple description coding in packet networks," in *Global Telecommunications Conference, 2005. GLOBECOM'05. IEEE*, vol. 4. IEEE, 2005, pp. 5–pp.

[73] J. Chakareski, S. Han, and B. Girod, "Layered coding vs. multiple descriptions for video streaming over multiple paths," *Multimedia Systems*, vol. 10, no. 4, pp. 275–285, 2005.

[74] P. A. Chou, H. J. Wang, and V. N. Padmanabhan, "Layered multiple description coding," *Proc. PV 13th Int. Packet Video Workshop,* Nantes, France, April 2003.

[75] A. Hocquenghem, "Codes correcteurs derreurs," *Chiffres*, vol. 2, no. 2, pp. 147–56, 1959.

[76] R. C. Bose and D. K. Ray-Chaudhuri, "On a class of error correcting binary group codes," *Information and control*, vol. 3, no. 1, pp. 68–79, 1960.

[77] I. S. Reed and G. Solomon, "Polynomial codes over certain finite fields," *Journal of the Society for Industrial & Applied Mathematics*, vol. 8, no. 2, pp. 300–304, 1960.

[78] I. Reed, "A class of multiple-error-correcting codes and the decoding scheme," *Information Theory, Transactions of the IRE Professional Group on*, vol. 4, no. 4, pp. 38–49, 1954.

[79] D. E. Muller, "Application of boolean algebra to switching circuit design and to error detection," *Electronic Computers, Transactions of the IRE Professional Group on*, no. 3, pp. 6–12, 1954.

[80] E. R. Berlekamp, *Nonbinary BCH decoding.* University of North Carolina. Department of Statistics, 1966.

[81] J. L. Massey, "Shift-register synthesis and bch decoding," *Information Theory, IEEE Transactions on*, vol. 15, no. 1, pp. 122–127, 1969.

[82] G. D. Forney Jr, "The viterbi algorithm," *Proceedings of the IEEE*, vol. 61, no. 3, pp. 268–278, 1973.

[83] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *Information Theory, IEEE Transactions on*, vol. 20, pp. 284–287, 1974.

[84] J. G. Proakis, *Digital Communications, Fourth Edition.* MCGraw-Hill College, Aug. 2000, ch. 5, p. 242.

[85] R. Tanner, "A recursive approach to low complexity codes," *IEEE Trans. Inform. Theory*, vol. 27, no. 5, pp. 533–547, Sep. 1981.

[86] N. Wiberg, H.-A. Loeliger, and R. Kotter, "Codes and iterative decoding on general graphs," *Eur.Trans. Telecomm.*, vol. 6, pp. 513–525, Sep./Oct. 1995.

[87] G. D. Forney, Jr., "Codes on graphs: Normal realizations," *IEEE Trans. Inform. Theory*, vol. 47, pp. 520–548, Feb. 2001.

[88] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.

[89] C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: Turbo-codes," *Communications, IEEE Transactions on*, vol. 44, no. 10, pp. 1261–1271, 1996.

[90] R. G. Gallager, *Low Density Parity Check Codes.* Cambridge, MA: MIT Press, 1963.

[91] D. J. C. Mackay and R. M. Neal, "Near Shannon limit performance of low density parity check codes," *IEE Electron. Lett.*, vol. 33, no. 6, pp. 457–458, Mar. 1997.

[92] T. J. Richardson, M. A. Shokrollahi, and R. L. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 619–637, Feb. 2001.

[93] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423, Jul. and Oct. 1948.

[94] S. Y. Chung, G. D. Forney, T. J. Richardson, and R. L. Urbanke, "On the design of low-density parity-check codes within 0.0045 dB of the Shannon limit," *IEEE Commun. Letter*, vol. 5, pp. 58–60, Feb. 2001.

[95] T. J. Richardson and R. L. Urbanke, "The capacity of low-density parity-check codes under messages-passing decoding," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 599–618, Feb. 2001.

[96] M. Ardakani and F. R. Kschischang, "A more accurate one-dimensional analysis and design of irregular ldpc codes," *Communications, IEEE Transactions on*, vol. 52, no. 12, pp. 2106–2114, 2004.

[97] B. Shrader and H. Permuter, "Feedback capacity of the compound channel," *Information Theory, IEEE Transactions on*, vol. 55, no. 8, pp. 3629–3644, 2009.

[98] H. Weingarten, T. Liu, S. Shamai, Y. Steinberg, and P. Viswanath, "The capacity region of the degraded multiple-input multiple-output compound broadcast channel," *Information Theory, IEEE Transactions on*, vol. 55, no. 11, pp. 5011–5023, 2009.

[99] L. Yingbin, K. Gerhard, and S. Shamai, "Compound wiretap channels," *EURASIP Journal on Wireless Communications and Networking*, vol. 2009, 2009.

[100] E. Arikan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inform. Theory*, vol. 55, no. 7, pp. 3051–3073, Jul. 2009.

[101] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels," *Information Theory, IEEE Transactions on*, vol. 49, no. 5, pp. 1073–1096, 2003.

[102] S. Tavildar and P. Viswanath, "Approximately universal codes over slow-fading channels," *Information Theory, IEEE Transactions on*, vol. 52, no. 7, pp. 3233–3258, 2006.

[103] J. Hou, P. H. Siegel, and L. B. Milstein, "Performance analysis and code optimization of low density parity-check codes on rayleigh fading channels," *Selected Areas in Communications, IEEE Journal on*, vol. 19, no. 5, pp. 924–934, 2001.

[104] M. Yang and W. E. Ryan, "Design of ldpc codes for two-state fading channel models," in *Wireless Personal Multimedia Communications, 2002. The 5th International Symposium on*, vol. 3. IEEE, 2002, pp. 986–990.

[105] J. Dunham and R. M. Gray, "Joint source and noisy channel trellis encoding (corresp.)," *Information Theory, IEEE Transactions on*, vol. 27, no. 4, pp. 516–519, 1981.

[106] V. B. Balakirsky, "Joint source-channel coding with variable length codes," in *Information Theory. 1997. Proceedings., 1997 IEEE International Symposium on.* IEEE, 1997, p. 419.

[107] E. Ayanoglu and R. M. Gray, "The design of joint source and channel trellis waveform coders," *Information Theory, IEEE Transactions on*, vol. 33, no. 6, pp. 855–865, 1987.

[108] N. Farvardin, "A study of vector quantization for noisy channels," *Information Theory, IEEE Transactions on*, vol. 36, no. 4, pp. 799–809, 1990.

[109] V. A. Vaishampayan and N. Farvardin, "Joint design of block source codes and modulation signal sets," *Information Theory, IEEE Transactions on*, vol. 38, no. 4, pp. 1230–1248, 1992.

[110] K. Sayood and J. C. Borkenhagen, "Use of residual redundancy in the design of joint source/channel coders," *Communications, IEEE Transactions on*, vol. 39, no. 6, pp. 838–846, 1991.

[111] J. Hagenauer, "Source-controlled channel decoding," *Communications, IEEE Transactions on*, vol. 43, no. 9, pp. 2449–2457, 1995.

[112] G.-C. Zhu, F. Alajaji, J. Bajcsy, and P. Mitran, "Transmission of nonuniform memoryless sources via nonsystematic turbo codes," *Communications, IEEE Transactions on*, vol. 52, no. 8, pp. 1344–1354, 2004.

[113] J. W. Modestino and D. G. Daut, "Combined source-channel coding of images," *Communications, IEEE Transactions on*, vol. 27, no. 11, pp. 1644–1659, 1979.

[114] R. E. Van Dyck and D. J. Miller, "Transport of wireless video using separate, concatenated, and joint source-channel coding," *Proceedings of the IEEE*, vol. 87, no. 10, pp. 1734–1750, 1999.

[115] K. Ramchandran, A. Ortega, K. M. Uz, and M. Vetterli, "Multiresolution broadcast for digital hdtv using joint source/channel coding," *Selected Areas in Communications, IEEE Journal on*, vol. 11, no. 1, pp. 6–23, 1993.

[116] Q. Zhang, W. Zhu, Z. Ji, and Y.-Q. Zhang, "A power-optimized joint source channel coding for scalable video streaming over wireless channel," in *Circuits and Systems, 2001. ISCAS 2001. The 2001 IEEE International Symposium on*, vol. 5. IEEE, 2001, pp. 137–140.

[117] L. P. Kondi, F. Ishtiaq, and A. K. Katsaggelos, "Joint source-channel coding for motion-compensated dct-based snr scalable video," *Image Processing, IEEE Transactions on*, vol. 11, no. 9, pp. 1043–1052, 2002.

[118] Y. Zhang, W. Gao, Y. Lu, Q. Huang, and D. Zhao, "Joint source-channel rate-distortion optimization for H. 264 video coding over error-prone networks," *Multimedia, IEEE Transactions on*, vol. 9, no. 3, pp. 445–454, 2007.

[119] T. Schierl, C. Hellge, S. Mirta, K. Grüneberg, and T. Wiegand, "Using H. 264/AVC-based scalable video coding (SVC) for real time streaming in wireless ip networks," in *Circuits and Systems, 2007. ISCAS 2007. IEEE International Symposium on*. IEEE, 2007, pp. 3455–3458.

[120] P. P. Bergmans, "Random coding theorem for broadcast channels with degraded components," *IEEE Trans. Inf. Theory*, vol. 19, no. 2, pp. 197–207, Mar. 1973.

[121] I. Csiszar, "The method of types," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2505–2523, Oct. 1998.

[122] R. Ahlswede, P. Gács, and J. Körner, "Bounds on conditional probabilities with applications in multi-user communication," *Probability Theory and Related Fields*, vol. 34, no. 2, pp. 157–177, 1976.

[123] P. Berlin and D. Tuninetti, "LDPC codes for fading gaussian broadcast channels," *Information Theory, IEEE Transactions on*, vol. 51, no. 6, pp. 2173–2182, 2005.

[124] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2009.