Cross-Layer Communication System Designs Exploiting Statistical Property of Sources

情報源の統計的性質を利用したクロスレイヤ通信システム設計

Yoshito Watanabe

Graduate School of Engineering, Yokohama National University

Advisor: Prof. Hideki Ochiai

Graduate School of Engineering, Yokohama National University

YNU 横浜国立大学
YOKOHAMA National University

This page intentionally left blank

Contents

This page intentionally left blank

Abstract

By Shannon's separation theorem, the combination of independently optimized source coding and channel coding systems can achieve the optimal rate-distortion (R-D) relationship. Most conventional communication systems for multimedia data are modeled based on this theorem, and each technique is developed and optimized independently. However, this theorem is based on the ideal assumption that code length is infinite, and thus a practical system in real life does not conform to it. Joint source-channel coding (JSCC) is a technique to design an optimal system in terms of R-D function or computational complexity considering practical situations, e.g., finite code lengths and bit errors occurring in communications channel. It could be also stated that JSCC may lead to the breakthrough in the development of multimedia data transmission systems considering the fact that the individual studies on source coding and channel coding are reaching maturity.

Unequal error protection (UEP) is a fundamental technique generally implemented in JSCC. The code rates in the framework of UEP can be optimized such that the received source quality is maximized. Most conventional studies on UEP assume that the channel state information is available at the transmitter side such that the code rate can be adjusted depending on the channel. However, this approach increases computational overhead at the transmitter. Furthermore, such an assumption is not applicable to the broadcasting systems where UEP property is most beneficial.

This dissertation tackles the problems involving source and channel coding as described above by designing communication systems over the layers specified by the Open Systems Interconnection (OSI) reference model, where the source coding process corresponds to the application (or upper) layer design whereas the channel coding is implemented in the physical layer. Chapter 1 is dedicated to a literature review on JSCC, and some major issues in the past studies are identified. Chapter 2 describes source coding characteristics observed in modern source compression techniques. An application of image transmission in combination with UEP is also presented as a specific example. Chapter 3 derives the achievable R-D function by JSCC system compared to separately designed source and channel coding system over AWGN channel with a specific channel SNR. The design of JSCC system is applied to the combination of subband coded audio signal and rate compatible punctured convolutional (RCPC) codes as a simple example. In Chapter 4, a novel coded modulation system design that maximizes the average quality of multimedia data, received by a given set of target users, is proposed for compressed video broadcasting based on UEP. To this end, a new measure is introduced that describes the received video quality averaged over the users located within the coverage area of a given broadcasting service. In Chapter 5, from the viewpoint of physical layer implementation, a novel decoding technique is introduced that improves the throughput of the system by exploiting inherent padding bits in the transmitted sequence as a consequence of the mismatch between the length of channel coding frame and that of the source coding. This improvement is achieved by trellis pruning in the decoding process of turbo code. The proposed decoder has the detector of padding bits, and thus the side information about the insertion of padding bits is

unnecessary at the receiver. Finally, the conclusion is given in Chapter 6.

あらまし

シャノンの情報源・通信路分離定理により、それぞれ独立に最適化された情報源符号化と通信路符号化の組み合わせが最適なレート・歪み (R-D: Rate-Distortion) 特性を示すことが証明されている。従来の多くのマルチメディアデータの通信システムはこの理論に則って設計されており、各々の技術は独立に最適化が試みられている。しかしながら、この理論は符号が無限の長さを持つという理想的な仮定に基づいているため、必ずしも現実のシステムの最適性を保証するものではない。そこで近年、情報源・通信路結合符号 (JSCC: Joint Source-Channel Coding) が注目されている。これは、有限長の符号や通信路で生じるビット誤りなど現実的な制約を考慮して、R-D 特性や計算複雑度の観点から最適なシステムを設計する技術である。情報源符号化と通信路符号化の各分野に閉じた研究開発が成熟しつつある現在、JSCC はマルチメディアデータの伝送システムの研究において大きな発展をもたらすものと考えられる。

　不均一誤り訂正 (UEP: Unequal Error Protection) は一般に JSCC で用いられる要素技術である。UEP では、通常符号化率の最適化は受信側での情報源の復号品質が最大になるように行われる。従来の UEP に関する多くの研究においては、送信器側で通信路状態が既知であるという仮定がなされており、通信路に合わせて符号化率を調整する手法が採用されているが、送信器側で必要となる演算量が大きくなるという問題がある。さらにこの仮定は、UEP の応用として有力である放送システムには適用できない。

　本論文では、OSI 参照モデルにおけるレイヤーを縦断した通信システムを設計することにより、上記のような情報源符号化と通信路符号化に関連する問題に取り組む。OSI 参照モデルでは、情報源符号化の過程はアプリケーション層 (または上位層) に相当し、通信路符号化は物理層内で実装される。第 1 章では JSCC に関する研究動向について述べ、過去の研究の問題点を明らかにする。第 2 章では近代のデータ圧縮技術で見られる情報源符号化の特徴を説明する。また、一例として UEP を画像伝送に適用したシステムについての検証も行う。第 3 章では、あるチャネル SNR を持つ AWGN 通信路において JSCC によって達成できる R-D 特性を、情報源符号化と通信路符号化を独立して設計したシステムによる R-D 特性と比較する。簡単な例としてサブバンド符号化されたオーディオ信号とパンクチャド畳み込み符号を組み合わせたシステムを用いて JSCC の設計を行う。第 4 章では、圧縮された動画像を送信する放送システム向けに、サービス対象とするユーザ群が受信する映像の平均品質を最大化する符号化変調システムの設計指針を提案する。ここでは放送サービスのカバレッジ範囲内のユーザの平均的なビデオの受信品質を表す新たな指標を導入する。第 5 章では、情報源符号化と通信路符号化のフレーム長の不一致を補填する目的で一般に広く導入されるパディングビットを受信側での誤り訂正符号の復号に活用することで、従来システムに比べて通信システムのスループットを向上させることが可能であることを示す。この特性向上はターボ符号の復号過程においてトレリスプルーニングを用いることで達成される。なお、提案する復号器にパディングビットの検出器を備えることで、パディングビットの挿入に関するサイド情報の送信が不要となる。最後に第 6 章において結論を述べる。

This page intentionally left blank

Acknowledgements

First of all, I would like to express my deepest gratitude to my supervisor, Prof. Hideki Ochiai, for his excellent guidance. He always gave me his thoughtful and helpful advice both for research and for private life, without which I would not had been able to do this work. I would also like to thank the past and current members of Ochiai laboratory for the opinions and suggenstions.

My special thanks should go to Prof. Ryuji Kohno, Prof. Tomoki Hamagami, Prof. Koichi Ichige, and Prof. Chika Sugimoto. They gave me a lot of stimulating and observant comments, which help to improve this dissertation.

I wish to thank Prof. Rokuya Ishii, who was my former supervisor when I was an undergraduate student, and the members in Ishii laboratory. My fundamental studies on digital signal processing related to multimedia data were initiated there.

I would like to extend my gratitude to my friends. They considerately encouraged and inspired me.

Last but not least, I am thankful to my parents, grandparents, brothers, all relatives for their continuous support and tolerance through these many years of my life, and Akiko for her patience and cheer that eased my pains with this work. These invaluable helps were crucial for me in achieving this work.

This page intentionally left blank

List of Figures

List of Tables

This page intentionally left blank

CHAPTER 1

Introduction

1.1 Background

In a typical communication system, several layers are defined for different functionality. As for networking, the Open Systems Interconnection (OSI) reference model defines the layers ranging from the physical layer, characterized by voltage levels of analog waveforms, to the application layer, with which users interact. The clients can focus on creating or generating multimedia contents, such as text, audio, image and video, without being concerned about how the network is internally constructed. Also, such layered structure enables engineers to maintain their own layers only since the knowledge about other layers is not required.

In terms of quality of service (QoS), the model can be categorized to the two most essential parts, that is, source coding part, i.e., compression technique, and channel coding part, where error correcting codes and modulation are applied (Fig. 1.1). The former corresponds to the application layer, and it generally specifies the type of data and format. The latter is equivalent to the physical layer, where how binary sequences are transmitted over the communication channel.

For example, H.264/Advanced Video Coding (AVC) [1, 2], one of the state-of-the-art video compression techniques, has been developed by the Joint Video Team (JVT) of Video Coding Experts Group (VCEG) in the International Telecommunications Union-Telecommunication Standardization Sector (ITU-T) and Moving Picture Experts Group (MPEG) in the International Organization for Standardization (ISO/IEC) Joint Technical Committee 1 (JTC1) in 2003. It accomplishes the rate range from 10kbps to 240Mbps, corresponding to the compression ratio of 1/20 to 1/100. H.264/AVC has been adopted as a source coding of recent Digital Video Broadcasting-Terrestrial (DVB-T) and DVB-Satellite-Second Generation (DVB-S2) while it is very popular as a codec for storage purpose or video sharing and streaming services such as those represented by YouTube [3]. More recently, H.265/High Efficiency Video Coding (HEVC) [4, 5] has been developed achieving perceptually the same quality as the half data rate of H.264.

As for the field of channel coding, there are error correcting codes that almost achieve channel capacity [6], such as turbo code [7], low-density parity check (LDPC) code [8] and Polar code [9]. Orthogonal frequency division multiplexing (OFDM) is the key technique for high data rate communications due to its high spectral efficiency combining with high order modulation, such as QAM.

By Shannon's source-channel separation theorem [10, 11], the combination of source and channel coding, which are independently optimized, leads to an optimal communications system in terms of rate-distortion (R-D) relationship assuming infinite length of source symbols and channel codewords. Most conventional studies are based on this theory and these two research fields have actually been independently developed as described above. However, this assumption is not realistic, and it is also addressed in [12] that there are cases where this two-stage process is not always appropriate such as:
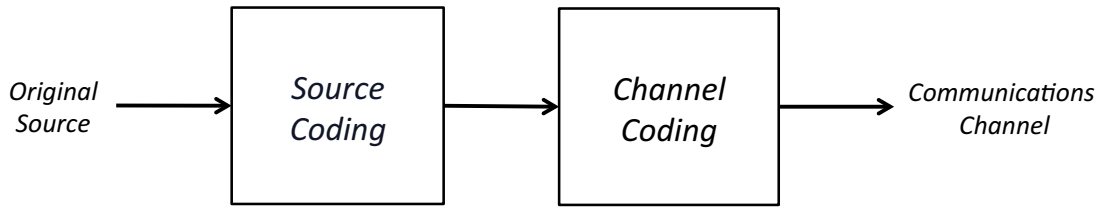
Figure 1.1: A concatenation of source and channel coding system.

- the systems operating with finite length coding,

- multiuser channels where more than two channels are involved in designing communication systems, and

- the systems that take into account human perception of reconstructed data where the moderate number of bit errors is somewhat tolerable.

Moreover, considering the fact that the studies in the respective fields of source coding and channel coding techniques are experiencing maturity, their joint design is the key to breakthrough for the development of multimedia data transmission systems, and this framework is referred to as *joint source-channel coding (JSCC)*.

1.2   Review on JSCC-Related Studies

There have been a number of proposals on JSCC in the past two decades. Based on the concept of simultaneous source and channel coding, where the combination of source coding and channel coding is implemented in one block of system, the joint decoding of entropy code and error correcting code is proposed in [13]. A double LDPC code, where two LDPC codes are utilized for both source and channel coding, is presented in [14]. An analog JSCC [15, 16] is the technique that directly maps source symbols to modulation constellations.

The primary purpose of JSCC is to jointly optimize the rates of both source coding and channel coding such that the source distortion is minimized under a given bandwidth limitation and channel condition, possibly exploiting the unequal error protection (UEP) property of channel coding. In [17, 18], JSCC for general image data has been proposed based on the rate-distortion (R-D) function, where the optimization of the rate allocation for source coding and channel coding is performed by the R-D curve that considers bit error sensitivity (BES) under the constraint of bit error rate (BER) of the channel or a certain channel SNR. In [19], a rate allocation design for source-channel coding with spatial diversity has been proposed over block fading channels. In [20], the variable-rate channel coding is optimized such that the overall distortion specified with peak signal-to-noise ratio (PSNR) of video is minimized. The studies in [21–24] propose UEP-based data protection for audio applications. These approaches design channel codes for output signal from each audio codec using rate compatible codes based on convolutional code or Reed-Solomon code. Most of them calculate coding parameters in every frame of an existing codec according to its BES and channel state information (CSI). These approaches improve system performance under a given bandwidth constraint, but their applicability is limited to the corresponding specific scenarios of source and channel coding. Consequently, they may not be necessarily applicable to a conventional system

based on the concatenation of source and channel coding.

From the viewpoint of physical layer research, a variety of approaches have been proposed for the design of UEP in various levels of applications. Multilevel coded modulation (MLC) [25, 26] is a suitable candidate for efficient implementation of the UEP property under strict bandwidth constraint. Targeting broadcasting systems, several approaches of implementing UEP based on MLC have been studied [27–29]. More recently, in [30], a multiplexed hierarchical modulation (HM) is proposed based on the theoretical BER derived from the symbol constellation with application to progressive transmission. The multilevel spatial hierarchical modulation (MSHM) is proposed in [31] aiming at UEP for line-of-sight (LOS) communications. In view of modern broadcasting systems where capacity-approaching channel codes are adopted, the optimal design of HM for the DVB-S2 system is developed in [32], where the spectral efficiency is considered in order to determine the parameters of the inherent LDPC code. Nevertheless, even though the purpose of UEP is to enhance the quality of the received source data, these studies have not taken into consideration the distortion of specific source data upon their code rate design.

On the other hand, from the perspective of source coding, scalable coding [33] has been developed as a technique to provide various quality from a single source according to the channel condition or the capability of the user equipments; the low but acceptable quality can be obtained by transmitting most essential elements with low data rate, but the full quality is achieved when the total source bits are received, requiring higher data rate. Scalable coding itself can be developed at least independently of channel coding, and thus it has the compatibility with any channel coding. Therefore, the application of scalable coding to UEP might be a more practical JSCC framework considering source quality. In [34], the combination of scalable video coding (SVC) and bit division multiplexing (BDM) is proposed, which achieves high transmission efficiency and low complexity in terms of the baseline layer demapping of SVC compared to the conventional time division multiplexing (TDM). The adaptive JSCC design of scalable video based on the statistics of individual users is proposed in [35], which is designed to guarantee that all the users can receive the data with the minimum quality available. A dynamic joint rate allocation scheme for multiprogram video coding is proposed in [36], where a new complexity measure that incorporates the characteristics of human visual system is introduced. In these studies, optimization of channel coding based on the received source quality can be achieved if the CSI is available at the transmitter. In [37], UEP for the error resilient H.264/Advanced video coding (AVC) is presented under the assumption that channel characteristic is known at the transmitter side, but such an assumption may not be realistic for general broadcasting services.

Some major challenges and limitations associated with the research involving JSCC are summarized as follows:

1. The achievability by JSCC is not clearly identified even though most studies claim that Shannon's separation theorem is unrealistic for practical systems.

2. Even though JSCC systems may achieve better characteristics compared to separately designed source and channel coding (SSCC) systems, its applicability is limited to their specific scenarios.

3. The study of UEP based on the physical layer typically fails to consider the reconstructed source quality.

4. Optimization of JSCC or UEP is generally performed based on the available CSI at the transmitter, and thus feedback process is necessary, which is not practical for broadcasting scenario.

This dissertation tackles these issues. Our approach is based on a cross-layer design through the physical layer to the application layer.

1.3 Thesis Outline

The outline of this thesis is summarized in Fig. 1.2.

Chapter 2 reviews the fundamentals and characteristics of both source and channel coding especially aiming at multimedia data transmission, which will be exploited in the systems in the subsequent chapters. A few UEP frameworks are also introduced as fundamental techniques for cross-layer system designs. Furthermore, a specific example of the combination of existing source and channel coding is analyzed.

In Chapter 3, the numerical R-D model of JSCC is introduced. The technique to optimize the JSCC R-D model is also proposed and the achievability of JSCC R-D model is shown compared to SSCC system, employing the subband-coded audio data and rate compatible punctured convolutional (RCPC) codes.

In Chapter 4, we propose a novel technique of optimizing UEP-based system assuming a broadcasting scenario. To this end, a new measure is introduced considering the average received quality of the users within a targeted coverage area of the transmitter.

In Chapter 5, as a technique in the physical layer, a novel scheme to improve channel decoder performance is proposed exploiting the specific characteristic of source coded data employing turbo code as a channel coding. The proposed technique can be implemented within physical layer, and thus it is applicable to a general system formed by a concatenation of source and channel coding.

Even though the systems in each of the above chapters are independent of the others, there are some connections among them; the system in Chapter 4 solves the problems identified in Chapter 3, and the technique in Chapter 5 is inspired by the system developed in Chapter 4.

Finally, Chapter 6 summarizes this dissertation and describes possible future directions.

Chapter 1. Introduction

Chapter 2. Principles of Source Coding and UEP

Chapter 3.
General R-D Model of
JSCC and Its
Optimization

Chapter 4.
A Novel Design and
Modeling of UEP-
Based Compressed
Video Broadcasting
with Multilevel Coded
Modulation

Chapter 5.
A Novel Channel
Decoder Exploiting
Inherent Padding Bits
for Fixed-Length
Frame Transmission

Chapter 6. Conclusion

Figure 1.2: Outline of this thesis.

This page intentionally left blank

CHAPTER 2

Principles of Source Coding and UEP

This chapter describes properties observed in multimedia source coding, which will be exploited in this thesis. In the framework of channel coding, unequal error protection (UEP) that is essential for implementing joint source-channel coding (JSCC) is also introduced with specific examples. UEP scheme is compared to equal error protection (EEP) considering their application to image transmission systems.

2.1 Source Coding

2.1.1 Sensitivity for Quality

▷ *Bit Significance in Magnitude*

Quantization, which is the process of expressing some real value by discrete values with binary representation, is the most fundamental source coding technique for any multimedia data. Pulse code modulation (PCM) is the technique equivalent to the quantization especially for audio signal. Let us now assume that we use uniform quantizer and two's complement to express a given integer value v with q bits, where the most significant bit (MSB) represents the sign of the number. Each binary value at significance level $\ell \in \{0, \dots, q-1\}$ is denoted as $b_\ell \in \{0, 1\}$. It follows that

$$v = -b_{q-1} 2^{q-1} + \sum_{\ell=0}^{q-2} b_\ell 2^\ell. \tag{2.1}$$

From (2.1), when an error occurs only at the least significant bit (LSB), i.e., b_0, the distortion from original v is just 1 in squared error. On the other hand, the error of the MSB leads to the distortion of $2^{2(q-1)}$, and thus it is obvious that the sensitivity to bit errors is high as the bit significance level increases. Besides, the large value of q leads to high data rate. Therefore, from both perspectives of bit error sensitivity (BES) and data rate, q should be small as much as possible.

A simple solution is to divide v by a given positive integer d such that the magnitude is decreased. In this case, an amplitude that should be quantized is

$$v' = \begin{cases} \lfloor v/d + 0.5 \rfloor & (v \geq 0) \\ \lceil v/d - 0.5 \rceil & (v < 0), \end{cases} \tag{2.2}$$

where $\lfloor x \rfloor$ denotes the maximum integer less than or equal to x and $\lceil x \rceil$ denotes the minimum integer value greater than or equal to x. The decoded value is $v'' = v'd$, and thus the quantization error becomes $|v-v''|$. As d increases, the magnitude of v' decreases, which leads to reducing the data rate but the quantization error increases. Nevertheless, this quantization error does not cause salient distortion if the technique is employed in frequency domain such as that used in JPEG [38], where discrete cosine transform (DCT) is adopted as an orthogonal transform. This property is shown in the following subsection.
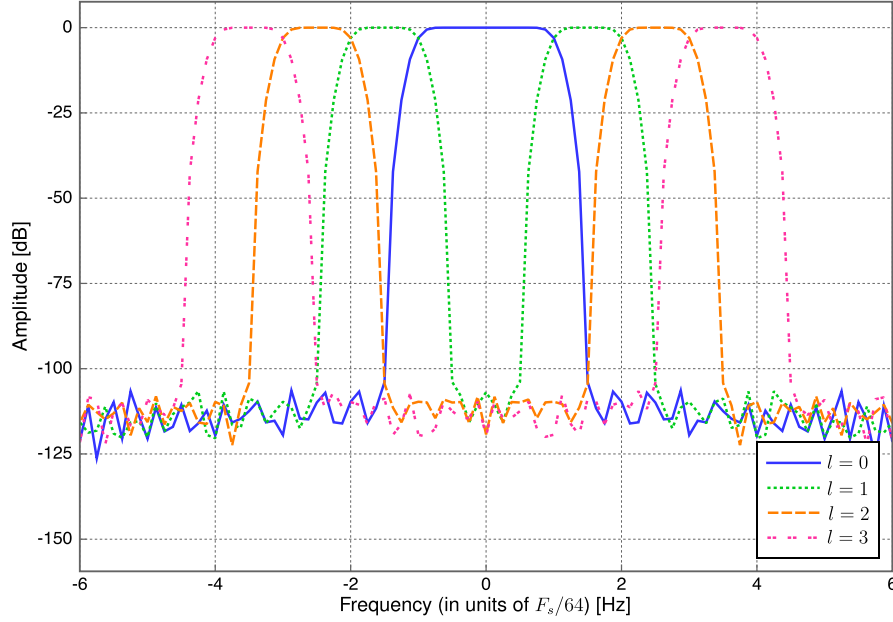
Figure 2.1: Frequency response of the first four bands of the MPEG PQMF.

▷ *Spectral Sensitivity*

There are cases that the decoded multimedia quality does not necessarily degrade perceptually though the signal is somewhat distorted in terms of mean squared error (MSE). Most modern source codecs, such as MPEG-1 Audio Layer-3 (MP3) [39] and MPEG-4 advanced audio coding (AAC) [40] for audio and JPEG [38] and JPEG-2000 [41] for image compression, make use of perceptual characteristic in order to reduce the amount of information that human can hardly perceive.

For simple example, we consider an audio case. The study in [42] proposes the frequency-weighted spectral distortion metric, which considers human audio perception and is defined as

$$SD_{fw} = \sqrt{\frac{1}{W_0} \int |W_B(f)|^2 10 \log \frac{|A_y(f)|^2}{|A_x(f)|^2} df} \qquad (2.3)$$

where $A_x(f)$ is the original spectrum, $A_y(f)$ is the reconstructed spectrum, W_0 is a normalization constant, and $W_B(f)$ is a hearing sensitivity weighting function defined by

$$W_B(f) = \frac{1}{25 + 75(1 + 1.4(f/1000^2))^{0.69}}. \qquad (2.4)$$

We now apply this measure to output of the MPEG pseudo quadrature mirror filter (PQMF) [39, 40, 43], which is the filterbank with 32 subbands. The frequency response of the first 4 filters of the MPEG PQMF is illustrated in Figure 2.1 with the frequency index l and the sampling frequency Fs. Although the prototype filter does not have sharp cut-off and there exists a certain amount of overlapping with adjacent bands, the signal can be reconstructed in the synthesis stage of the decoder.

Figure 2.2 illustrates BES for the signal from each subband of MPEG PQMF quantized with 16 bits, where SD_{fw} is calculated when bit errors occur at rates with 0.01, 0.001,
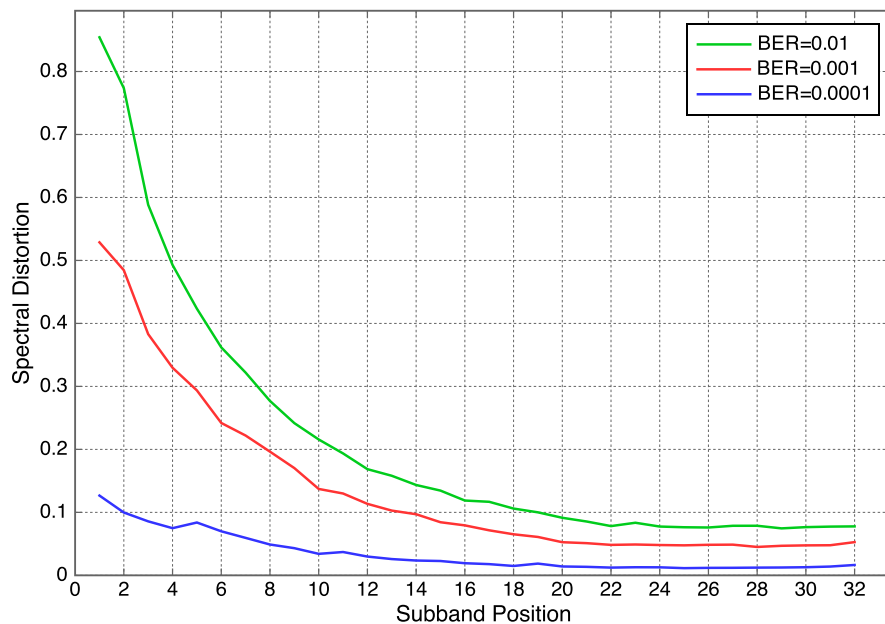
Figure 2.2: Bit error sensitivity for the signal from each subband of MPEG PQMF.

0.0001 individually for each subband. From the figure, it is shown that the lower frequency, i.e., lower subband index, has higher perceptual BES. It follows that bit errors in high frequency signals do not cause severe degradation of audio for human ears.

2.1.2 Scalability

One practical application of two characteristics described above is presented. Figure 2.3 shows the variance σ_i^2 of the DCT coefficient estimated by ensemble images for $i \in \{0,\ldots,63\}$ corresponding to the index in 8×8 block. The index is assigned from the lowest frequency element, i.e., DC coefficient to the highest AC coefficient in zigzag order. We clearly observe that σ_i^2 becomes large when the index i is small, showing that the DCT coefficients with lower indices are dominant for decoding than those with higher indices. In other words, energy of images is concentrated in the low frequency components.

This characteristic can be utilized to construct several multiresolution images from one source. Figure 2.4 shows images of "*Lena*", which are obtained by discarding high frequency coefficients in DCT domain for each 8×8 block. Figure 2.4(a) is decoded from only the lowest frequency, i.e., DC coefficient and the peak signal-to-noise ratio (PSNR)[1]is 23.21 dB, Fig. 2.4(b) is obtained from five lowest DCT coefficients with PSNR=27.82 dB, and Fig. 2.4(c) is the full resolution image. From these images, a few DCT coefficients are sufficient to obtain the abstraction of the original image.

It is apparent that an image with a few DCT coefficients has smaller amount of information than that with the full number of DCT coefficients. The *scalability* [33] is the technique to adapt the quality of one original source according to the channel condition or the user's device; the low but acceptable quality can be obtained by transmitting most essential elements with low data rate, but the full quality is achieved when the total source bits are received, requiring higher data rate.

Progressive image transmission that progressively refines the received image quality

---

[1]PSNR is calculated by $10\log_{10}(255^2)/MSE$.
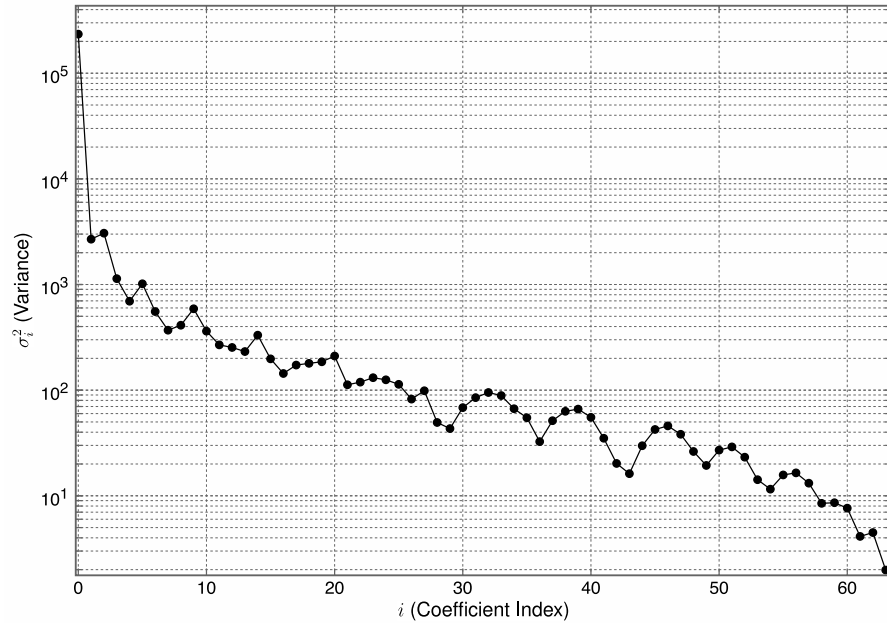
Figure 2.3: Variance of each DCT coefficient estimated by ensemble images.

Figure 2.4: Lena images with multiresolution. (a) From 1 DCT coefficients, PSNR=23.21 dB. (b) From 5 DCT coefficients, PSNR=27.82 dB. (c) Full resolution.

is a practical application of scalability, as represented by the *set partitioning in hierarchical trees (SPIHT)* [41] adopted in JPEG2000 [44]. The *mozjpeg* [45] is the progressive version of JPEG, which originally does not have the progressive capability, and the scheme demonstrated in Fig. 2.4 is employed, where DCT coefficients are reordered over all blocks in an original JPEG image from DC to the highest AC element. The benefit of scalability gets more salient when the original source has large amount of data, and thus the study in the field of video coding is rather active [46–50] than that of image coding.

In scalable coding, source can not be reconstructed without the most essential element. Contrarily, multiple description coding (MDC) [51] is an encoding technique such that multimedia data can be recovered from any of source packets sacrificing some data rate, and thus data becomes more robust against packet loss.

Table 2.1: Example of Huffman table.

$\mathcal{X}$	Prob.	Length	Code
1	0.3	2	00
2	0.25	2	10
3	0.15	3	010
4	0.12	3	011
5	0.1	3	110
6	0.08	3	111

2.1.3 Entropy Codes

Entropy codes are used for lossless coding of discrete random variable. Huffman code [52] is one of the most famous entropy codes with its simple algorithm of constructing optimal codes, where shorter codewords are assigned to the more frequently occurring symbols to achieve an average symbol codeword length that is as close to the symbol source entropy as possible.

Let us now see an example of Huffman codes. Consider a random variable X taking values in the set $\mathcal{X} = \{1, 2, 3, 4, 5, 6\}$ with respective probabilities $\{0.3, 0.25, 0.15, 0.12, 0.1, 0.08\}$. The Huffman table is obtained as shown in Table 2.1. The average code length per symbol is 2.45 bits, which is close to the entropy of 2.42 bits calculated by

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log_2 p(x), \tag{2.5}$$

where $p(x)$ is the probability of the symbol $x \in \mathcal{X}$.

▷ *Variable Length Property*

With the Huffman table above, let us encode a symbol sequence of $\{2, 3, 1, 4, 1\}$. The corresponding code word is 100100001100, which has 12 bits. On the other hand, the codeword of $\{1, 6, 3, 5, 2\}$ is 0011101011010 and has 13 bits. Depending on a sequence to be encoded, the resulting codeword has different code length, and this characteristic will be noticeable as the number of symbols in a sequence increases.

This property sometimes causes problem when encoding several sources with almost the same quality. As for JPEG, Huffman table can be defined in JPEG data, but one defined in Annex K of the standard [38] is usually used as a statistically pseudo-optimal reference since the additional metadata is unnecessary. If the entropy of a given image matches well with the Huffman table, the number of average bits per pixel is expected to be the average code length of the table. However, if the entropy of a given image does not match the Huffman table well, the resulting codeword length will be redundantly large.

▷ *Unique Decoding Property*

In entropy coding, a symbol is coded such that its code is not a prefix of any other symbols or such that any of its prefixes are not used as a code of another symbol. Due to this property, one single error may cause the failure of symbol synchronization and the corruption of an entire image. Therefore, in an entropy coded sequence, symbols after a first bit error should be discarded.

To tackle this problem, several solutions have been proposed. In JPEG, RST (reset) marker, which is a unique binary sequence among all possible codes, is inserted every few DCT blocks. This scheme is also employed in video coding such as MPEG-4 Video [40] in order to prevent error propagation over the adjacent packets. Reversible variable length code (RVLC), which can be decoded both in the forward and backward directions at the cost of reduced coding efficiency compared to entropy codes, is proposed in [53, 54].

2.2   Unequal Error Protection

Contrarily to source coding, which is a process of removing redundancy of original data, channel coding is one that intentionally adds redundancy to source coded data to provide robustness against noise. As discussed above, source has various significance, and thus to equally provide error protection property throughout one source, some redundancy should be introduced. UEP is a method to reduce this redundancy. This section describes examples of UEP schemes, which will be employed in this dissertation.

2.2.1   Rate Compatible Punctured Convolutional (RCPC) Codes

Rate compatible channel codes provide a flexible method to implement UEP, and the most popular one is the rate compatible punctured conventional (RCPC) codes [55, 56], which are based on a special class of punctured conventional codes. One of the main advantages of the RCPC codes is that it can adjust the code rate simply by inserting the dummy bits at the encoder of the transmitter and by eliminating the corresponding bits from actual metric calculation at the decoder of the receiver.

If the rate of mother code is equal to $1/N$, the puncturing period P determines the code rates R as

$$R = \frac{P}{P+l} \qquad\qquad l = 1, \cdots (N-1)P \qquad\qquad (2.6)$$

which can range from $1/N$ up to $P/(P+1)$. The RCPC codes are determined by the mother code and the puncturing matrices, where the element 0 implies the bit position to be punctured and thus not to be actually transmitted.

The upper bound of bit error rate of punctured convolutional code of which code rate is k/n is represented as

$$P_b\left(\frac{k}{n}\right) < \frac{1}{k} \sum_{d=d_{free}}^{\infty} \beta_d P_d \qquad\qquad (2.7)$$

where d_{free} is the minimum free distance of the code and β_d is the number of codewords with Hamming distance d. P_d is expressed as

$$P_d = Q\left(\sqrt{\frac{2dE_s}{N_0}}\right) \qquad\qquad (2.8)$$

where $Q(x)$ is called Q function, which is defined as

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-y^2/2} dy. \qquad\qquad (2.9)$$

Table 2.2 shows parameters of RCPC codes represented by (2.7), in the case of $P = 7$ as an example.

Table 2.2: Parameters of RCPC codes.

Coding Rate	Puncturing Matrix	d_{free}	$\beta_d, d = d_{free}, d_{free}+1, \cdots$
1/2	1111111 1111111	5	1,4,12,32,80,192,448,1024,2304,5120
7/13	1110111 1111111	4	2,17,70,206,528,1419,3934,10346
7/12	1110110 1111111	4	8,48,160,522,1504,4650,13647
7/11	1100110 1111111	4	36,189,340,1866,6416,18464,65900
7/10	1100100 1111111	3	8,130,540,1786,9668,37046,139168
7/9	1100100 1111011	2	4,46,540,1838,14334,53728
7/8	1100100 1011011	2	28,467,4606,39748,323416

2.2.2 Multilevel Coded Modulation (MLC) with Capacity-Approaching Codes

There are powerful error correcting codes approaching channel capacity [6, 10] such as low-density parity-check (LDPC) codes [57] and turbo code [7, 58], and these codes usually take block structure, where a codeword is produced in block unit. Due to this constraint, encoding of multimedia source by such capacity-approaching codes in a straightforward manner will cause "*cliff effect*" of decoded quality at the decoder side. MLC [25, 26, 28, 29, 59] is one of the channel coding schemes suitable for UEP, especially with block codes.

For example, let us consider the MLC system with LDPC codes as its component code. The LDPC encoder at the ℓ-th level in MLC with L levels, denoted as $\mathcal{C}^{(\ell)}$, has an $M^{(\ell)} \times N$ sparse parity check matrix $\mathbf{H}^{(\ell)}$ and a $K^{(\ell)} \times N$ generator matrix $\mathbf{G}^{(\ell)}$ which can be obtained by transforming $\mathbf{H}^{(\ell)}$. An N-dimensional codeword $\mathbf{c}^{(\ell)}$ is produced as follows:

$$\mathbf{c}^{(\ell)} = \mathbf{m}^{(\ell)} \mathbf{G}^{(\ell)}, \tag{2.10}$$

where $\mathbf{m}^{(\ell)}$ consists of $K^{(\ell)}$ information bits. Therefore, the code rate of $\mathcal{C}^{(\ell)}$ results in $r^{(\ell)} = K^{(\ell)}/N$ ($0 < r^{(\ell)} \leq 1$).

In MLC, encoding is implemented in each level individually and the codeword of each level is expressed as

$$\mathbf{c}^{(1)} = (c_0^{(1)}, c_1^{(1)}, \cdots, c_{N-1}^{(1)})$$
$$\mathbf{c}^{(2)} = (c_0^{(2)}, c_1^{(2)}, \cdots, c_{N-1}^{(2)})$$
$$\vdots$$
$$\mathbf{c}^{(L)} = (c_0^{(L)}, c_1^{(L)}, \cdots, c_{N-1}^{(L)}),$$

where $c_t^{(\ell)} \in \{0, 1\}$ and each of them is the element of $\mathcal{C}^{(1)}, \mathcal{C}^{(2)}, \cdots, \mathcal{C}^{(L)}$, respectively. Based on the sequence $(c_t^{(1)} c_t^{(2)} \cdots c_t^{(L)})$, 2^L-ary modulation is implemented for all $0 \leq t \leq N-1$

and then mapped into the signal set $\mathcal{M}$. Finally, we can obtain the symbol sequence

$$\mathcal{M}(\mathbf{c}^{(1)} * \mathbf{c}^{(2)} * \cdots * \mathbf{c}^{(L)})$$
$$=(\mathcal{M}(c_0^{(1)}c_0^{(2)} \cdots c_0^{(L)}),\ \mathcal{M}(c_1^{(1)}c_1^{(2)} \cdots c_1^{(L)}),$$
$$\cdots,\ \mathcal{M}(c_{N-1}^{(1)}c_{N-1}^{(2)} \cdots c_{N-1}^{(L)})).$$

The total rate of the system can be written as $r = \sum_{\ell=1}^{M} r^{(\ell)}$.

At the receiver side, multi-stage decoding (MSD) is employed as the decoding method of MLC. In all levels, sum-product decoding is employed as a decoding method of LDPC code, which is based on iterative decoding. In each iteration of decoding, until the maximum iteration number χ is reached, the parity condition is checked according to the following equation:

$$\hat{\mathbf{c}}^{(\ell)}\mathbf{H}^{(\ell)^{T}} = \mathbf{0}, \tag{2.11}$$

where $\hat{\mathbf{c}}^{(\ell)}$ is a temporarily estimated codeword in the decoding process. If the equation (2.11) is satisfied by χ iterations, $\hat{\mathbf{c}}^{(\ell)}$ is output as a final codeword, which is correctly decoded. Since LDPC code has the error detection capability by this process, an additional cyclic redundancy check (CRC) is not necessary to this system.

2.3 Application to JPEG Transmission

As an example of the combination of source and channel coding, we consider a transmission of JPEG image by the MLC system with UEP property. Even though JPEG is not designed for transmission over noisy channels and thus is not optimized for graceful degradation, it is shown that adapting its binary frame structure to the MLC system can provide JPEG with quality scalability.

2.3.1 Adaptation of JPEG Bitstream to the MLC Frame

Variable length source coding such as Huffman and run-length codes is applied in the standard JPEG, combined with DCT within an 8×8 pixel block. Let us denote the number of bits assigned to the ith DCT coefficient by v_i after entropy coding, where $i \in \{0,\ldots,63\}$ is arranged in the zigzag order and $i = 0$ corresponds to the index of the DC coefficient. Then the ith coefficient of a pixel block will be expressed by a binary sequence $\mathbf{b}_i = (b_{i,0}, b_{i,1}, \cdots, b_{i,k_i-1})$, and the entire pixel block after this entropy coding is given by $\mathbf{B} = (\mathbf{b}_0, \mathbf{b}_1, \cdots, \mathbf{b}_{63})$. The number of bits per block is expressed by

$$K = \sum_{i=0}^{63} k_i. \tag{2.12}$$

Upon adapting the source bits to the MLC frame, the binary sequence $\mathbf{B}$ is divided into the sub-sequences that correspond to L levels. Given the fact that the element of the vector $\mathbf{B}$ is ordered such that $\mathbf{b}_0$ is most important and $\mathbf{b}_{63}$ is least important, and entropy encoded data can only be decoded from the first bits, assignment of a given sequence $\mathbf{B}$ to the MLC codewords can be performed in a straightforward manner. Denoting the total rate of L-level coded modulation as $r_{\text{total}} = \sum_{\ell=1}^{L} r^{(\ell)}$, simple decomposition of the source bit sequence such that the number of source coded bits in the ℓth level meets

$$K^{(\ell)} = \left\lceil K \frac{r^{(\ell)}}{r_{\text{total}}} \right\rceil, \tag{2.13}$$
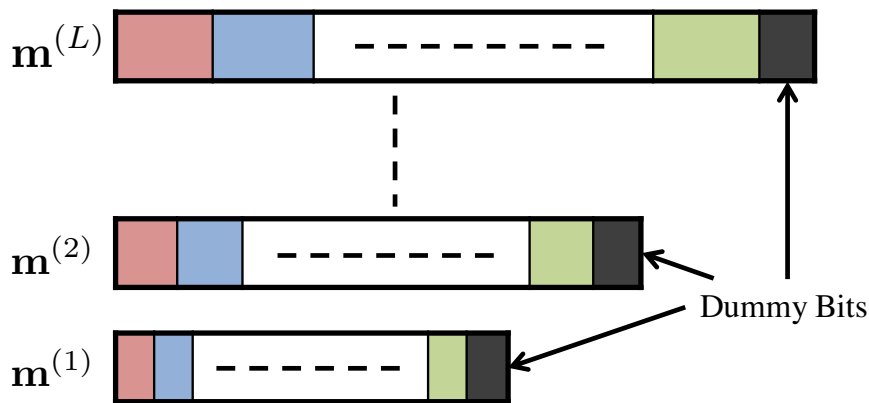
Figure 2.5: The source frame structure in MLC with LDPC codes system. The same tile color indicates the same block element of JPEG.
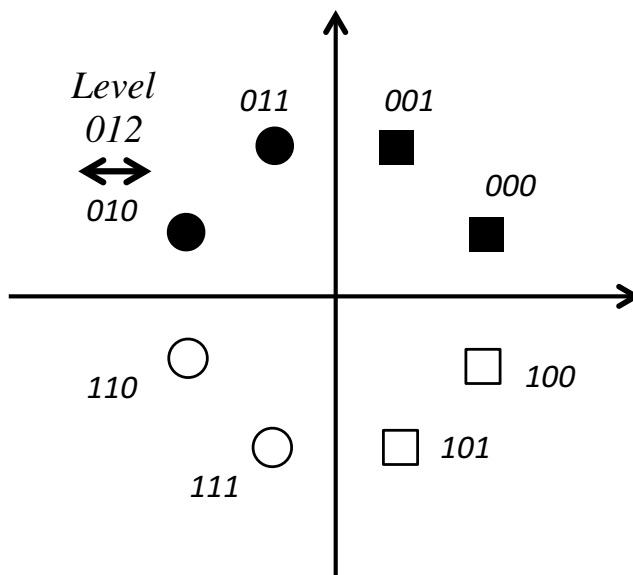
Figure 2.6: 8-PSK with BP labeling.

where $\lceil x \rceil$ denotes the minimum integer greater than or equal to x, can guarantee that the more significant bits are assigned to the lower levels. This bit assignment enables the receiver to display an image of given quality from $\sum_{\ell=1}^{\ell'} K^{(\ell)}$ available bits even if a decoding error occurs at the $(\ell'+1)$th level for $0 \le \ell' \le L-1$ in MSD. Since K is itself a random variable, $K^{(\ell)}$ varies as well, depending on the image.

Upon channel encoding the source coded bits, a multiple number of blocks, say L, are put together and they form one LDPC-coded MLC frame as shown in Fig. 2.5. In practice, it is difficult to match the length of the source coded bits of each level to exactly the same number as the corresponding channel coding frame, and it is necessary to add some dummy bits (i.e., the padding bits) per codeword, as shown in Fig. 2.5. These dummy bits can be further utilized to reduce the frame error rate or improve error rate performance. This topic is discussed in Chapter 5.

Table 2.3: Simulation Setup of MLC system with UEP and CM systems.

	MLC w/ UEP-1	MLC w/ UEP-2	CM
Code Rate	$r^{(1)} = 0.5$ $r^{(2)} = 0.75$ $r^{(3)} = 0.85$	$r^{(1)} = 0.67$ $r^{(2)} = 0.75$ $r^{(1)} = 0.67$	0.7
Total Rate	2.1	2.08	2.1
Code Length N	9000	9000	18000
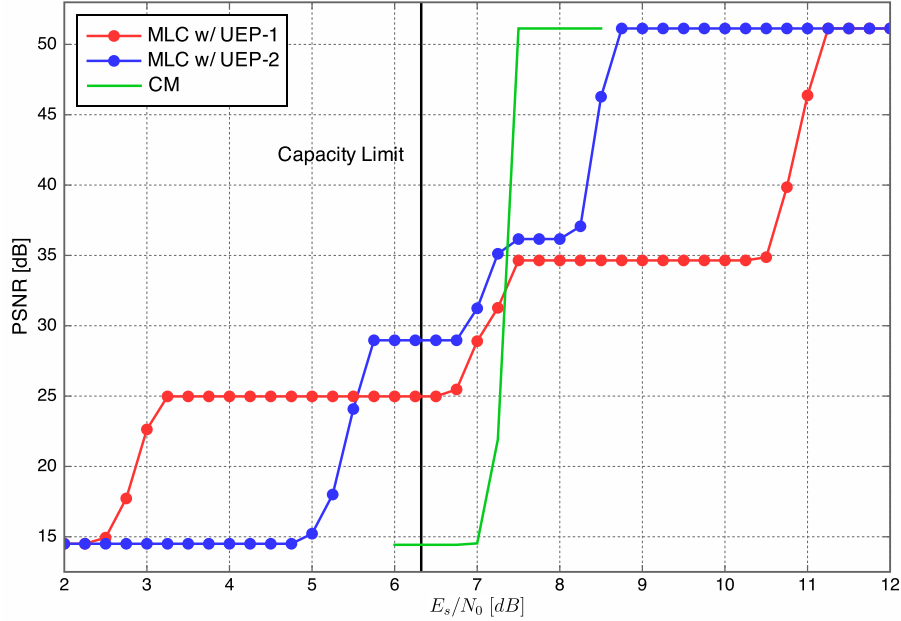
Figure 2.7: PSNR for various E_s/N_0 with MLC and CM systems.

2.3.2 Computer Simulation

We evaluate the MLC system comparing other modulation schemes in the case of $L = 3$, namely, 8-PSK with block partitioning (BP) labeling [26], shown in Fig. 2.6, which has UEP property.

We employ two UEP schemes by MLC and a conventional LDPC coded modulation (CM) system with EEP property. They are designed such that their bandwidth efficiency is similar and the detailed parameters are listed in Table.2.3. Simulations are executed for the Lena's image.

Figure 2.7 shows PSNR with respect to various channel SNR, E_s/N_0 for the proposed UEP-1 and UEP-2 and CM systems. In the figure, the capacity limit when $r = 2.1$ is also plotted by black line. As one can see, UEP-1 can achieve the PSNR with about 25 dB at the range of very low channel SNR. We can also see that UEP-2 can achieve higher quality than UEP-1 at $\ell = 1$, but the start-up E_s/N_0 of the graph is higher than that of UEP-1. This result implies that the trade-off between the channel SNR and the decoded image quality can be flexibly adjusted.

In comparison with EEP, the required E_s/N_0 of the UEP systems to achieve the highest quality image is higher than EEP. In other words, the full quality is achieved at the lowest E_s/N_0 by the EEP system. In the absence of the channel state fluctuation with high E_s/N_0,

one can see that EEP is the best solution to transmit images. However, if E_s/N_0 falls into the lower region, cliff effect is caused, while UEP schemes achieve graceful degradation. It follows that there is a trade-off between UEP and EEP, and if UEP is used, code rate design determines how graceful degradation is achieved. The study in [60] and Chapter 4 of this dissertation attempts to optimize this trade-off assuming a broadcasting scenario.

This page intentionally left blank

CHAPTER 3

General R-D Model of JSCC and Its Optimization

In this chapter, an optimized rate-distortion (R-D) function for multimedia transmission is estimated from a perspective of joint source-channel coding (JSCC). Multimedia data such as audio, image and video is distorted by quantization and compression process. Furthermore, bit errors caused by the noise in communication channels impose additional distortion. In our R-D model, the original multimedia data is divided into some layers according to bit error sensitivities (BES), and an optimal R-D function is calculated by searching the combination of a quantizer and an error correcting code for each layer to minimize both distortion and rate. This process is implemented by an exhaustive search for all possible quantizers and channel coders over AWGN channel. Through our computer simulation, this algorithm is applied to audio signal as a simple example, and the performance of JSCC scheme is compared with that of a separated source-channel coding (SSCC) as a conventional coding scheme.

3.1 Generalizing Rate-Distortion Model

3.1.1 Source Coding

Source data can be divided into L layers according to its BES. The lth layer has a quantizer of which the number of bits is q_l, and distortion caused in the lth layer and the corresponding rate are denoted by $d_{S,l}(q_l)$ and $r_{S,l}(q_l)$, respectively. Therefore, the total distortion and the total rate for source coding can be expressed, respectively, as

$$D_S(\mathbf{q}) = \sum_{l=1}^{L} d_{S,l}(q_l) \tag{3.1}$$

$$R_S(\mathbf{q}) = \sum_{l=1}^{L} r_{S,l}(q_l) \tag{3.2}$$

where $\mathbf{q} = \{q_1, q_2, \cdots, q_L\}$. The above equations show the relationship between the quantizers and R-D function, that is, the dependence of the distortion and the rate on a quantizer in each sub-band. In general, mean squared error (MSE) and signal-to-noise ratio (SNR) are used as a distortion measure, and a rate corresponds to the number of quantization bits.

3.1.2 Joint Source-Channel Coding

The bit errors through communication channels deteriorate the source coded data in bitstream. Channel error correction codes are implemented in order to protect the data from these errors. The distortion to original data through a communication channel depends not only on the source quantizers but also on the bit error probabilities. This relationship can be represented as

$$D_{S+C}(\mathbf{q}, \mathbf{p}) = D_S(\mathbf{q}) + D_C(\mathbf{q}, \mathbf{p}) \tag{3.3}$$

where

$$D_C(\mathbf{q},\mathbf{p}) = \sum_{l=1}^{L} d_{C,l}(q_l,p_l). \tag{3.4}$$

The bit error probability vector $\mathbf{p} = \{p_1, p_2, \cdots, p_L\}$ depends on the channel code vector $\mathbf{c} = \{c_1, c_2, \cdots, c_L\}$ in a practical transmission system. Therefore, the alternative representation of (3.3) can be obtained as

$$D_{S+C}(\mathbf{q},\mathbf{c}) = D_S(\mathbf{q}) + D_C(\mathbf{q},\mathbf{c}) \tag{3.5}$$

where

$$D_C(\mathbf{q},\mathbf{c}) = \sum_{l=1}^{L} d_{C,l}(q_l,c_l). \tag{3.6}$$

Note that $D_C(\mathbf{q},\mathbf{c})$ is affected not only by $\mathbf{c}$, but also by $\mathbf{q}$. This is because the larger the number of quantization bits for each layer gets, the higher the bit error sensitivity generally becomes. It also follows that the distortion through a communication channel can be reduced by choosing an optimal combination of $\mathbf{q}$ and $\mathbf{c}$.

Similarly to the distortion, the rate of the channel coded data depends not only on the number of quantization bits, but also on channel code rates in a JSCC model. Assuming $R_C(\mathbf{c})$ is the average of channel code rates for all layers, a total rate is represented as

$$R_{S+C}(\mathbf{q},\mathbf{c}) = \frac{R_S(\mathbf{q})}{R_C(\mathbf{c})} \tag{3.7}$$

where

$$R_C(\mathbf{c}) = \frac{1}{L} \sum_{l=1}^{L} r_{C,l}(c_l). \tag{3.8}$$

Since $R_C(\mathbf{c}) \leq 1$ is generally satisfied, (3.7) means that the redundancy is added by the channel encoder.

As further additional expressions, we mention that (3.5) and (3.7) can be transformed to the following equations:

$$D_{S+C}(\mathbf{q},\mathbf{c}) = \sum_{l=1}^{L} \left[d_{S,l}(q_l) + d_{C,l}(q_l,c_l) \right]$$

$$= \sum_{l=1}^{L} d_{S+C}(q_l,c_l) \tag{3.9}$$

$$R_{S+C}(\mathbf{q},\mathbf{c}) = \frac{1}{L} \sum_{l=1}^{L} \frac{r_{S,l}(q_l)}{r_{C,l}(c_l)} = \sum_{l=1}^{L} r_{S+C}(q_l,c_l) \tag{3.10}$$

The above equations mean that the overall distortion and rate are equivalent to sums of those in each layer for JSCC scenario in this chapter.

In the R-D model of (3.5) and (3.7), we can also treat the case that $\mathbf{q}$ is fixed, that is, a conventional multimedia transmission model, where the original data is optimally compressed beforehand using, e.g., a perceptual coding or an entropy code without
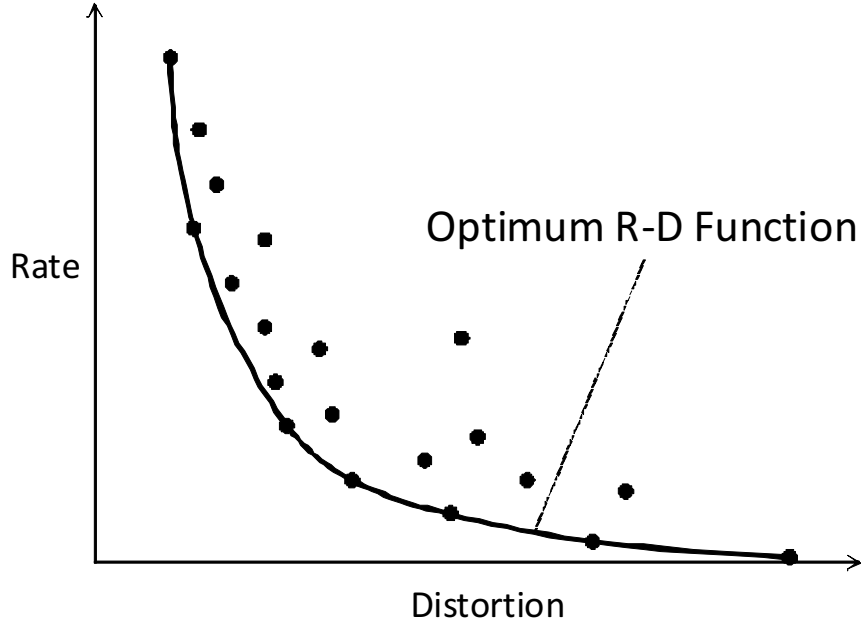
Figure 3.1: An example of relationship between rate and distortion. Each dot is a value of R-D with one combination of **q** and **c**.

considering channel error effect. In this case, the additional distortion and the rate depend only on **c**, and thus finding the optimum vector **c** results in optimizing R-D function. We call this case as SSCC.

On the other hand, equations (3.9) and (3.10) show that we have only to find the optimal combination of the quantizer and the channel coder in each layer respectively in order to optimize the overall R-D function. This characteristic is important for the estimation algorithm of the optimum R-D function described in the following section.

3.2   Estimation of An Optimum R-D Function

Figure 3.1 illustrates an example of relationship between rate and distortion, where each dot represents a value of R-D with one combination of **q** and **c**. To estimate the optimum R-D function in this figure, we adopt the algorithm derived in [61]. However, since the algorithm is developed only for source coding, it must be modified for JSCC system considering bit error influences and channel code rates.

The algorithm is implemented with 5 steps as follows:

1. Determine the initial vector $\mathbf{q}^{(0)}$ and $\mathbf{c}^{(0)}$. In this chapter, they are selected such that $R_{S+C}(\mathbf{q}^{(0)}, \mathbf{c}^{(0)}) = 0$ is satisfied.

2. Giving a certain channel SNR, calculate slopes s_l of R-D function, which is defined as (3.11), under the condition of $d_{S+C,l}(q'_l, c'_l) < d_{S+C,l}(q_l^{(m)}, c_l^{(m)})$, and find the combination of q'_l and c'_l such that s_l is maximal among all possible s_l for each l respectively.

$$s_l = \frac{r_{S+C,l}(q'_l, c'_l) - r_{S+C,l}(q_l, c_l)}{d_{S+C,l}(q'_l, c'_l) - d_{S+C,l}(q_l, c_l)} \tag{3.11}$$

3. Determine l for which s_l is largest among all l and assign $q_l^{(m+1)} = q'_l$ and $c_l^{(m+1)} = c'_l$
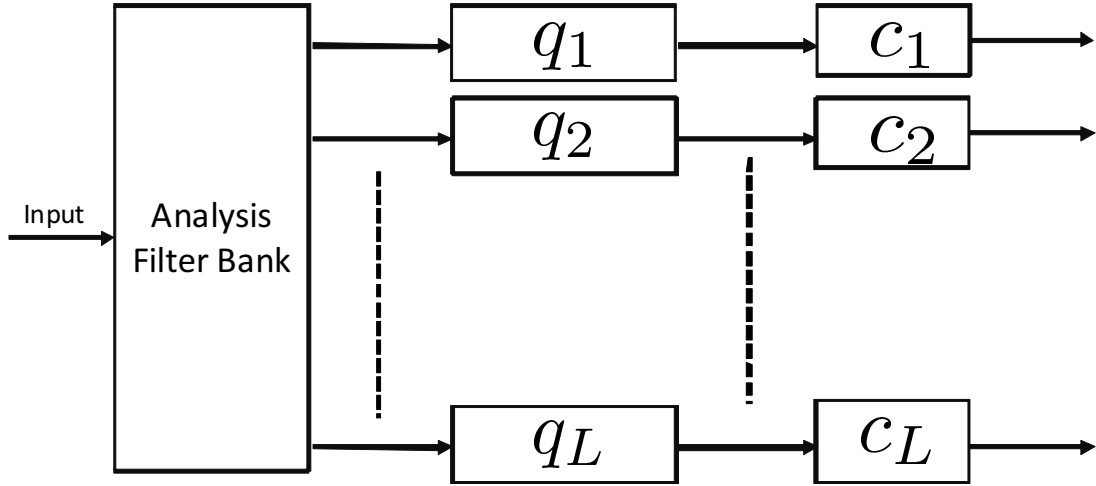
Figure 3.2: Block diagram of our transmission model. An input signal is divided into 32 sub-bands by PQMF, quantized with $r_{S,l}(q_l)$ bits, and then channel encoded with the code rate $r_{C,l}(c_l)$

to this l, and $q_l^{(m+1)} = q_l^{(m)}$ and $c_l^{(m+1)} = c_l^{(m)}$ for other l.

4. For $\mathbf{q}^{(m+1)}$ and $\mathbf{c}^{(m+1)}$, calculate $R_{S+C}(\mathbf{q}^{(m+1)}, \mathbf{c}^{(m+1)})$ and $D_{S+C}(\mathbf{q}^{(m+1)}, \mathbf{c}^{(m+1)})$ and check if the desired bit rate is obtained. If so, stop the algorithm.

5. Repeat steps 2, 3 and 4.

Contrary to [17], where the algorithm is implemented with two stages, our algorithm decides both q_l and c_l at the same time.

3.3   Computer Simulation

3.3.1   Simulation Model

To evaluate the proposed estimation algorithm, we utilize a simple audio transmission model shown in Figure 3.2. The input signal is divided into $L = 32$ sub-bands using pseudo quadrature mirror filter (PQMF) and the lth sub-band signal is linearly quantized with $r_{S,l}(q_l)$ bits where the maximum number is 16 for all l. Unequal error protection (UEP) is implemented with changing the code rate $r_C(c_l)$ for each sub-band using rate compatible punctured convolutional (RCPC) code [55, 56] with the puncturing period 7, where the base convolutional code has 2 registers, 1/2 code rate and the generator polynomial (5,7) in an octal form. Binary codes is modulated by BPSK modulation. MSE is used as a distortion measure defined by

$$\text{MSE} = \frac{1}{N} \sum_{n=0}^{N-1} (x[n] - \hat{x}[n])^2 \tag{3.12}$$

where N is the number of samples in each frame, $x[n]$ is an original signal and $\hat{x}[n]$ is a distorted signal.

3.3.2   Optimum R-D Functions for Various Channel SNR

Figure 3.3 illustrates R-D functions of JSCC with channel SNR at $E_s/N_0 =5$, 6, 7 and 8 dB. The R-D function in which the estimation algorithm is applied only for source coding
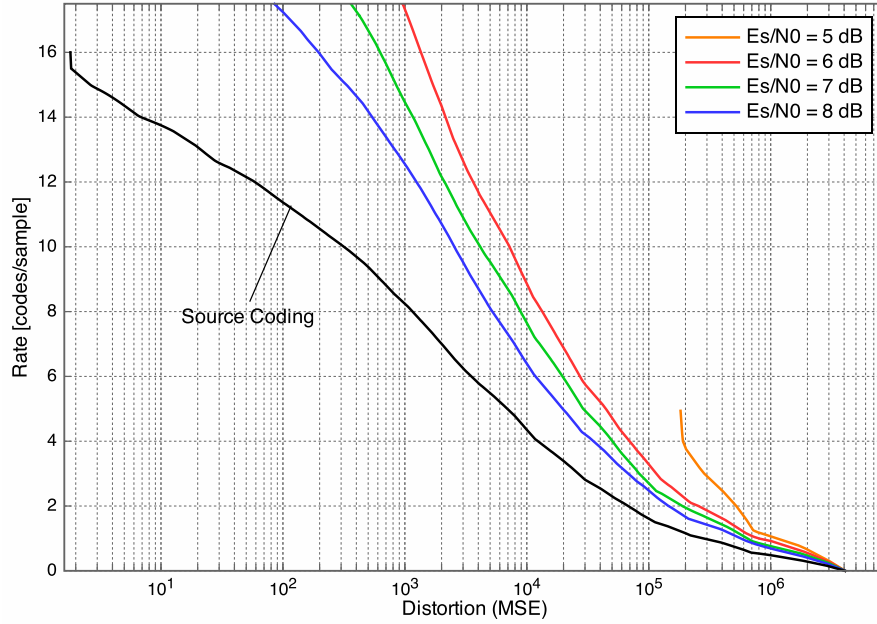
Figure 3.3: R-D functions of JSCC with channel SNR at E_s/N_0 =5, 6, 7 and 8 dB. The R-D function only for source coding is also drawn.

is also plotted as a solid line, which corresponds to the case of $E_s/N_0 \rightarrow \infty$. The higher the channel SNR gets, the lower the R-D function becomes, and it approaches the lower bound of the solid line. Note that in the case of $E_s/N_0 = 5$ dB, the curve terminates at a certain point of MSE. This is because the original signal is highly distorted, and thus s_l in (3.11) has become the negative infinite value. From this result, it follows that the channel SNR in the estimation algorithm of an R-D function is an important factor.

3.3.3 Comparison with SSCC

We compare the R-D function of a JSCC system with that of a SSCC system, which is an example of a conventional transmission system. The latter case is estimated by applying the R-D estimation algorithm with two stages. First, $\mathbf{q}$ is estimated so that both $R_S(\mathbf{q})$ and $D_S(\mathbf{q})$ are minimum, where the algorithm in [61] can be applied without modification. Second, $\mathbf{c}$ is estimated so that both $R_{S+C}(\mathbf{q},\mathbf{c})$ and $D_{S+C}(\mathbf{q},\mathbf{c})$ are minimum.

Figure 3.4 illustrates the result of applying the estimation algorithm at $E_s/N_0 = 6$ dB for JSCC with the solid line and SSCC plotted as the cross marks. The result of SSCC shown in the figure is a part of all possible patterns because there are a number of combinations of $\mathbf{c}$ for each $\mathbf{q}$. As shown in Fig.3.4, SSCC can hardly achieve the performance of JSCC and this characteristic is remarkable as the rate increases. It is shown that JSCC is effective for a code design.

3.4 Summary

In this chapter, we have proposed a general R-D model of JSCC together with the estimation algorithm of the optimum R-D function. Utilizing that JSCC can be expressed by some forms, the algorithm derived in [61] is applied to JSCC by adding AWGN with a certain channel SNR. Our R-D model can also be applied to SSCC by assuming a quantizer is estimated separately from a channel coder.
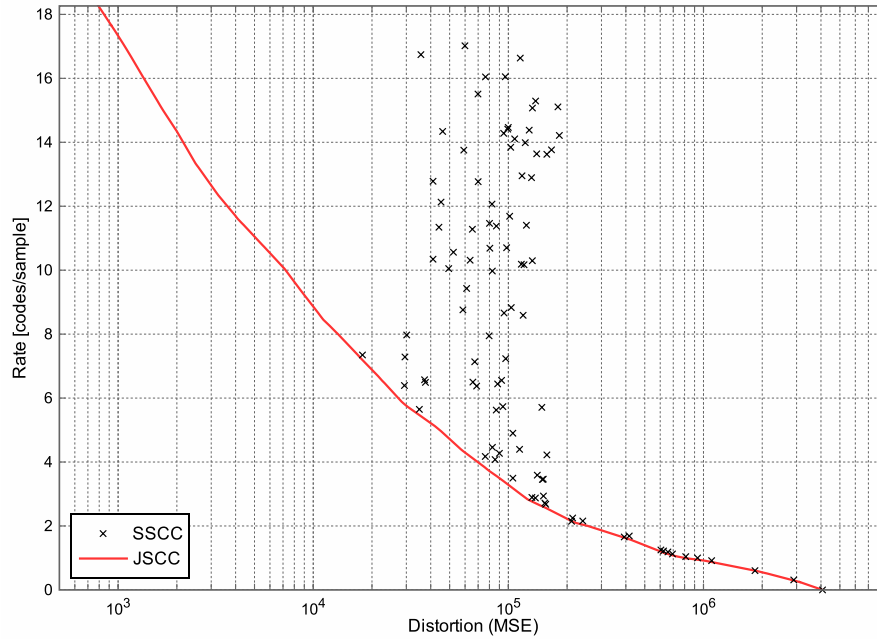
Figure 3.4: R-D function of JSCC with the solid line and SSCC with cross marks at $E_s/N_0 = 6$ dB.

In the simulation result section, we have shown that JSCC scheme can achieve lower R-D function than SSCC under the assumption that the channel state information (CSI) is available at the transmitter side. It is shown that the system design with considering the source distortion caused by bit errors in communications channel is effective. However, it is impossible to continue estimating the R-D function until the termination condition if the too strong noise is added in the algorithm.

Though it has been found that the JSCC system achieves the superior R-D performance compared to SSCC systems, the approach in this chapter can not be applied to the system with entropy coding and capacity-approaching codes. The following chapter tackles this limitation assuming a broadcasting scenario.

A Novel Design and Modeling of UEP-Based Compressed Video Broadcasting with Multilevel Coded Modulation

4.1 Introduction

Future broadcasting services through wireless channels should support transmission of multimedia data with even higher quality under a strict limitation of spectral resources. In wireless broadcasting, however, the signal-to-noise ratio (SNR) of the channel varies depending on the location of each user and thus it may not be necessarily feasible to support all the users with the same best quality of multimedia data. A more practical approach is to provide each user with its best quality achievable by the channel SNR of the corresponding user based on *graceful degradation*. Such an approach can be realized by exploiting the scalability of multimedia data in the application layer and the unequal error protection (UEP) property of coding and modulation in the physical layer. In order to achieve good graceful degradation property in the broadcasting framework, the quality of decoded multimedia data as well as the randomness of the user locations should be taken into consideration.

In this chapter, a novel coded modulation system design that maximizes the average quality of multimedia data, received by a given set of target users, is proposed for compressed video broadcasting based on unequal error protection (UEP). An integrated source and channel coding system with graceful degradation is introduced, where the multilevel coded modulation (MLC) with capacity-approaching codes is combined with an existing source compression technique. By assigning critical information that is required for decoding compressed data to the levels of MLC with higher reliability, we demonstrate that graceful degradation can be naturally incorporated into practical video broadcasting systems. Furthermore, a new measure of achievable average receiver quality, i.e., the quality of the received videos averaged over the users located within the coverage area of a given broadcasting service, is introduced. Based on the proposed measure, an optimal rate allocation algorithm for MLC, utilizing the code rate design based on channel capacity rule, is developed. In the case of MPEG-4 video coding, our numerical results reveal that a well-designed MLC system can achieve higher average quality experienced by target users than the conventional coded modulation as well as hierarchical modulation systems for a given coverage area. This chapter is based on the journal paper published in [62].

4.1.1 Contributions

Several studies on UEP systems that consider the quality of the received video are found in the literature, but most of them do not sufficiently address the design optimization

of the physical layer aspects such as coding and modulation. Also, broadcasting nature of the system with variable received SNR is not fully taken into consideration for the quality evaluation in the preceding work. The main contributions of this chapter are summarized as follows:

1. We consider a UEP-based broadcasting system that takes into account the quality of received multimedia data.

2. Statistical distortion model of MPEG-4 is developed, and the trade-off relationship between the rate of channel codes and received source distortion is derived.

3. A new measure that takes into account the system suitability in terms of average user satisfaction is introduced, where it is calculated based on the distance of the user locations from the transmitter considering their statistical variations in terms of path loss and the resulting quality of received multimedia data of individual users.

4. The optimization problem is formulated such that the average quality experienced by the target users within a specific coverage area is maximized for a given transmission rate (i.e., maximum spectral efficiency).

To demonstrate the effectiveness of our proposed approach, MLC, two-level HM, and conventional coded modulation (CM) systems with and without UEP are compared using several actual video data encoded by MPEG-4. It is shown that a carefully designed MLC system based on an appropriate code rate optimization can outperform the other UEP-based systems compared here in terms of achievable average quality.

We note that this work is an extension of our initial work [60] which has focused on a simplified scenario where JPEG images were chosen as a source codec and a simplified path-loss model (SPLM) was considered as a channel model; this chapter explores a more practical broadcasting scenario where the video transmission based on the compressed MPEG-4 packets is adopted as a source codec and the effect of shadowing is taken into account for a propagation model.

4.1.2   Organization

This chapter is organized as follows. Section 4.2 gives an overview of a general UEP-based system model considered in this chapter. Section 4.3 describes the structure of compressed MPEG-4 packets and develops an R-D model based on statistical distortion. Section 4.4 summarizes a UEP design of coded modulation together with capacity-approaching binary channel codes as its component codes. Based on the integration of the R-D model developed in Section 4.3 and the coded modulation design described in Section 4.4, a new quality measure is introduced in Section 4.5, where the rates of channel codes and the resulting quality of the received source data are jointly taken into consideration. Numerical evaluation of the effectiveness of our R-D model and comparisons of various coded modulation systems based on the proposed measure are given in Section 4.6. Section 4.7 concludes this work.

(a)

(b)
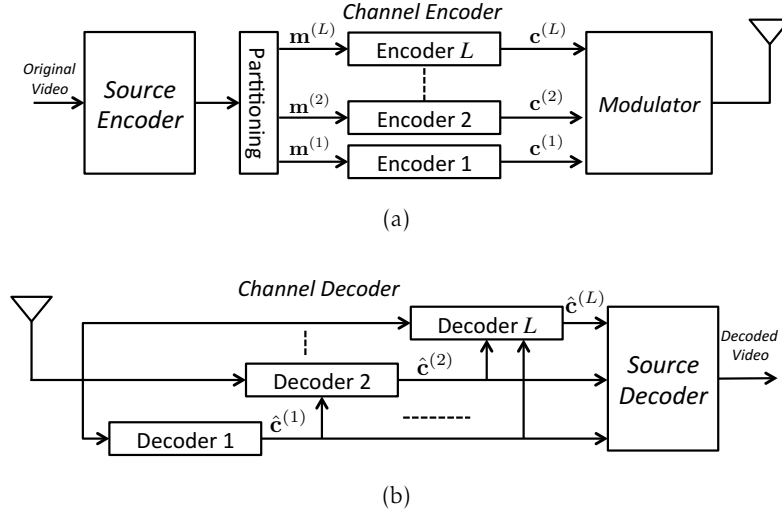
Figure 4.1: A general source and channel coded transmission system model with UEP. (a) Transmitter and (b) receiver where each decoder performs near optimal decoding based on the received QAM symbols.

4.2 A General Coded Modulation System Model

Figure 5.3 describes a basic transmitter and receiver model of our UEP-based coded modulation system considered in this work. The binary sequence representing source-coded data is first divided into *L partitions* depending on its importance, and each partition is individually encoded by its corresponding binary channel code. Each channel-coded bit corresponds to each *level* of an M-ary quadrature amplitude modulation (QAM) symbol to be transmitted over a wireless broadcasting channel. The detailed description of each component is given in what follows.

4.2.1 A General UEP-based Source Coding

We assume that the transmitter broadcasts a hierarchical video stream of *L UEP partitions*; the first (or the lowest) partition corresponds to the most essential data that is required for decoding of any information (such as a packet header), whereas the subsequent UEP partitions serve as an increment of quality, thus achieving graceful degradation. Specifically, let $\mathbf{m} \in \{0,1\}^K$ denote the binary sequence of length K which corresponds to a video-encoded packet. (Note that MPEG-4 is assumed as a specific example of our video encoder in this work and its packet structure is described in Section 4.3.) This is then partitioned into the L binary sub-sequences of distinct length, each of which is denoted by $\mathbf{m}^{(\ell)}$, where $\ell \in \{1, 2, \cdots, L\}$. Let $K^{(\ell)}$ denote the length of the ℓth binary sub-sequence. It then follows that

$$K = \sum_{\ell=1}^{L} K^{(\ell)}. \tag{4.1}$$

Since the subsequent partitions serve as an increment of quality to the former partitions, the sequence $\mathbf{m}^{(\ell)}$ can be decoded only if all its prior partitions, i.e., $\mathbf{m}^{(\ell')}$ with $\ell' = 1, 2, \ldots, \ell - 1$, can be decoded successfully at the receiver. The quality of the multimedia data successfully decoded up to the ℓth partition is then specified by some

non-negative metric $q^{(\ell)}$, which should satisfy

$$q^{(L)} > q^{(L-1)} > \cdots > q^{(1)} > q^{(0)} = 0. \tag{4.2}$$

Here, $q^{(0)} = 0$ corresponds to the case where no information can be retrieved at the source decoder.

4.2.2   A General Coded QAM Transmission

In order to achieve higher bandwidth efficiency, the use of high-order modulations such as M-ary QAM is essential. Let $\mathcal{X}$ denote the set of the corresponding constellations, with its cardinality given by $|\mathcal{X}| = M$. In general, the channel composed by M-ary modulation can be decomposed into $\log_2 M$ binary channels, but they have different resilience to channel noise. It is also essential that the source-coded data should be encoded by channel coding to enhance the robustness against channel impairments. In what follows, we assume that the constellation size M is chosen such that the number of the partitions of the video stream is equal to that of the binary channels, i.e., $L = \log_2 M$ *levels*. More general cases are discussed in Section 4.4.

Let us denote the total number of QAM symbols that are transmitted as a single *frame* (i.e., a transmission unit) by N. Since one QAM symbol consists of L bits, the total number of channel coded bits (i.e., codewords) necessary to construct one transmission frame is equal to LN for general CM systems. We assume that the length of codewords is equal to N for all the partitions, but possibly with different code rates. Recall that the partitions of the source-coded sequence are sorted based on the priority and importance of the data and each partition $\mathbf{m}^{(\ell)}$ has $K^{(\ell)}$ bits. Let $\mathcal{C}^{(\ell)}$ denote the binary channel code for the ℓth partition with its code rate given by $r^{(\ell)} = K^{(\ell)}/N$. The codeword for the ℓth partition, $\mathbf{c}^{(\ell)} = (c_0^{(\ell)}, c_1^{(\ell)}, \ldots, c_{N-1}^{(\ell)})$, is generated by

$$\mathbf{c}^{(\ell)} = \mathbf{m}^{(\ell)} \mathbf{G}^{(\ell)}, \tag{4.3}$$

where $\mathbf{G}^{(\ell)}$ represents the generator matrix for $\mathcal{C}^{(\ell)}$. The total code rate (or transmission rate) of the system is defined as

$$R_{\text{total}} = \sum_{\ell=1}^{L} r^{(\ell)} = \frac{1}{N} \sum_{\ell=1}^{L} K^{(\ell)} = \frac{K}{N}, \tag{4.4}$$

which determines the spectral efficiency of the overall system from the physical layer perspective.

As representative examples of possible coded modulation schemes, we introduce two such frameworks with UEP capability, i.e., MLC and general CM based on time sharing. In MLC, the L-tuples composed of the nth codeword bits $(c_n^{(1)}, c_n^{(2)}, \ldots, c_n^{(L)})$ is mapped onto the nth QAM symbol $X_n \in \mathcal{X}$ according to some bit labeling strategy. On the other hand, in the case of time sharing, each of L codewords separately forms a UEP partition consisting of N/L QAM symbols. According to our notations, for $n \in \{0, 1, \cdots, N-1\}$, the nth QAM symbol is composed of $(c_{n'}^{(\ell)}, c_{n'+1}^{(\ell)}, \ldots, c_{n'+L-1}^{(\ell)})$, where $\ell = 1 + \lfloor nL/N \rfloor$ and $n' = nL$ mod N, with $\lfloor x \rfloor$ denoting the integer part of x. For example, the initial QAM symbol ($n = 0$) of a transmission frame is composed of the L-tuples $(c_0^{(1)}, c_1^{(1)}, \ldots, c_{L-1}^{(1)})$, and the set of the first N/L QAM symbols forms the first UEP partition. One transmission frame is thus composed of all the L UEP partitions formed by the total of N QAM symbols allocated in a time division multiplexing manner.

Consequently, in either of the above coded modulation approaches, the entire set of the codewords $(\mathbf{c}^{(1)}, \ldots, \mathbf{c}^{(L)})$ forms the aforementioned frame consisting of N QAM symbols denoted by $\mathbf{X} = (X_0, X_1, \ldots, X_{N-1})$. Note that $\ell = 1$ corresponds to the lowest partition that is usually designed to have the maximum protection against noise, whereas $\ell = L$ corresponds to the highest partition that has the least protection.

Without loss of generality, we have so far assumed that the source-coded data in the leading partition is assigned to the codeword with the highest error protection. If the source is encoded by variable length coding (VLC) and once a bit error is found at some partition, the rest of the source information may not be correctly retrieved from the subsequent partitions in general. Therefore, if the codeword corresponding to a given UEP partition ℓ' is not available at the receiver due to the decoding error at this partition, the received codeword corresponding to its higher UEP partitions (i.e., the partitions with $\ell > \ell'$) will be discarded even if they could be decoded correctly.

4.2.3 Receiver

The receiver performs decoding for $\mathcal{C}^{(\ell)}$ starting from the lowest UEP partition successively after the detection of QAM symbols. The corresponding channel decoding structure is described in Fig. 5.3. For MLC, multi-stage decoding (MSD) [25], which will be described in Section 4.4.2, is employed as the corresponding decoding framework. For the conventional CM, each channel decoder independently performs near optimal decoding based on the received QAM symbols of length N/L.

By assumption, the decoder terminates its operation at the level (or partition) where the decoding error is detected, and thus the decoding for $\mathcal{C}^{(\ell)}$ is performed given that the codewords corresponding to its lower partitions are correctly decoded. Note that the detection of any binary error in each codeword can be performed by additional standard error detection codes such as cyclic redundancy check (CRC) codes, or if the code itself has a structure of block codes such as LDPC codes, the error detection can be performed by its inherent parity-check matrix. As a result, only the correctly decoded source-coded bits are passed to the source decoder. For example, if the channel decoder is able to decode correctly up to the ℓth UEP partition, the source decoder will receive only the source-coded bits corresponding to the binary sequence $(\mathbf{m}^{(1)}, \mathbf{m}^{(2)}, \ldots, \mathbf{m}^{(\ell)})$.

4.2.4 Channel Model

In a typical terrestrial broadcasting scenario, the users in different locations experience different SNR realizations due to shadowing. Therefore, the achievable SNR for each user can be modeled as a random variable, and its distribution is specified by the channel modeling. In this work, for the purpose of demonstration of our proposed approach and analytical tractability, the SPLM with log-normal shadowing [63, 64] will be adopted as our channel model. Extensions to more complicated fading channels are mathematically involved but straightforward.

Denoting the radiation power from the transmit antenna by P_T, the received signal power at distance ρ in the case of the SPLM is represented as

$$P_R(\rho) = P_T L_0 \left(\frac{\rho_0}{\rho}\right)^{\alpha}, \tag{4.5}$$

where ρ_0 is a reference distance and α is the path-loss exponent. The case with $\alpha = 2$

is equivalent to the free space path-loss model and that with $\alpha = 4$ corresponds to the two-ray model [63]. The reference distance ρ_0 should be selected in accordance with the assumed environment. The reference path-loss L_0 is derived from the free-space path-loss model and expressed as

$$L_0 = \left(\frac{\lambda}{4\pi\rho_0} \right)^2,\tag{4.6}$$

where $\lambda = c/f_c$ is the wavelength of the transmitted radio frequency (RF) signal, f_c is its carrier frequency, and c is the speed of light (3×10^8 m/s).

For the combined model of the SPLM and *shadowing*, the *instantaneous* SNR at the distance ρ is given by $\gamma_\rho = p_{R,\rho}/P_N$, where $p_{R,\rho}$ is the received *instantaneous* signal power and P_N represents the average power of additive white Gaussian noise (AWGN). The former follows the log-normal distribution with its pdf given by

$$f_{\gamma_{\mathrm{dB}},\rho}(\gamma_{\mathrm{dB}}) = \frac{1}{\sqrt{2\pi}\sigma_{\gamma_{\mathrm{dB}}}} \exp\left[-\frac{(\gamma_{\mathrm{dB}} - \mu_{\gamma_{\mathrm{dB}},\rho})^2}{2\sigma_{\gamma_{\mathrm{dB}}}^2} \right],\tag{4.7}$$

where $\mu_{\gamma_{\mathrm{dB}},\rho} = (P_R(\rho)/P_N)_{\mathrm{dB}}$ is the average channel SNR according to the SPLM (4.5) and $\sigma_{\gamma_{\mathrm{dB}}}^2$ is the corresponding variance [63].

4.3   Partition and Distortion Modeling based on MPEG-4 Encoder

In order to exploit scalability for efficient protection (or retrieval) of multimedia data, in this section we describe our proposed design to partition the source-coded binary sequence based on MPEG-4 standard. Our method is to allocate the compressed video packets such that the most important elements are assigned to the leading partitions, which will then be mapped onto the lowest modulation levels with highest robustness against bit errors caused by channel noise.

We also propose a method to quantify the distortion associated with loss of the partitioned data due to the error caused by the channel decoder. This can be used for modeling the quality measure introduced in Section 4.5.1.

Even though the MPEG-4 standard is not originally designed to support scalability, judicious partitioning of its binary frame structure can naturally provide MPEG-4 with quality scalability. The selection of MPEG-4 as our source coding is based on the following practical reasons:

- Establishing the R-D model is essential for design of our UEP-based coded modulation systems. In the case of MPEG-4 video, there exists solid work [65] that is useful for developing our R-D model. This is a challenging issue for newer standards such as H.264/SVC.

- Compared to MPEG-4, the newer standards such as H.264/AVC introduce bi-directional predicted picture (B-picture), which makes use of the pictures encoded in the past and future. The resulting encoding process is thus significantly complicated. The simplicity of MPEG-4 video encoding process makes it feasible to design and analyze our R-D curves in a straightforward manner.

Since we analyze the effect of only the distortion associated with bit errors caused by channel, the effect of quantization noise is not considered in this chapter. It is thus
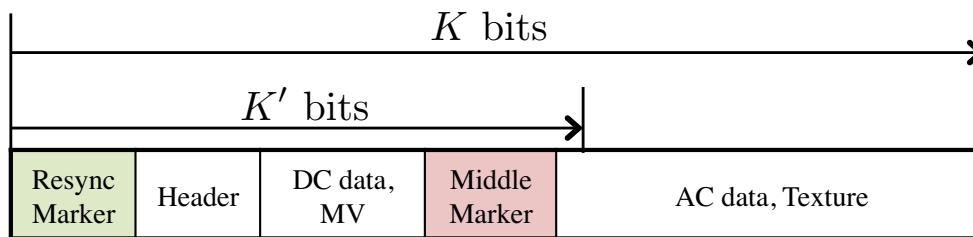
Figure 4.2: Generalized packet structure of MPEG-4.

assumed that the original video data is encoded according to the MPEG-4 standard beforehand at the transmitter with the commonly-used color component format, where the ratio of 4:2:0 for luma (Y) and chroma (Cb, Cr) components is employed [66].

4.3.1 Packet Structure with Error Resilience Tools

First, we give a brief overview of the MPEG-4 data structure employed in this work. Video coded data consists of intra- (I) and predictive- (P) pictures. Encoding process of an I-picture is described as follows: 1) discrete cosine transform (DCT) is applied to every 8×8 macroblock, 2) DCT coefficients are quantized, and 3) variable length coding (VLC) is employed. We adopt the error resilience tools in MPEG-4, namely, data partitioning together with packetization, and thus the VLC of DC coefficients is independent of that of AC coefficients to avoid error propagation. A P-picture is encoded using inter prediction referring to the previous I-picture or P-picture, such that 1) motion estimation is exploited in a unit of 16×16 macroblocks, which produces motion vectors (MV) between the previous picture and residual data (texture), 2) DCT is applied to every 8×8 macroblock in texture, and 3) VLC is independently applied to MV and texture. We assume that there are no B-pictures in this work since they do not support the error resilience tools in MPEG-4.

In data streams, one picture consists of a series of several packets by packetization. Figure 4.2 illustrates a generalized model for the packet structure of MPEG-4. Each packet starts with the resynchronization marker, which consists of more than 16 bits of '0' that are followed by a single bit of '1', for preventing error propagation over the adjacent packets in conjunction with VLC. By data partitioning, important information such as DC coefficients for I-picture and MV for P-picture is located in the first segment following the header information, and then less important information such as AC data and texture is placed in the second segment divided by unique binary sequences as the second marker, which is referred to as the middle marker in what follows.

At the receiver side, our MPEG-4 decoder is implemented independently of the channel decoder. We employ the standard error concealment technique of MPEG-4 for erroneous data; the source decoder declares that the packet is erroneous if 1) the source decoder receives an illegal entropy code, 2) a semantic error is detected, 3) the number of decoded DCT coefficients disagrees with 64, or 4) inconsistent header information is received. If an error is detected before the middle marker, all information in that packet will be considered as unavailable, and the same pixel values are copied from the corresponding macro blocks in the previous picture. When an error is detected after or at the beginning of the middle marker, decoding is executed using only the first segment.

4.3.2   Partitioning of MPEG-4 Video Packet

Following the notations introduced in Section 4.2.1, a packet sequence $\mathbf{m} \in \{0,1\}^K$ is first divided into the sub-sequences $\mathbf{m}^{(\ell)}$ with L partitions, the former partitions representing data of higher importance. As described in the previous subsection, since significant information such as header and fundamental data is gathered in the beginning part of a MPEG-4 packet, this importance-based partitioning can be performed in a straightforward manner. Recall that $K^{(\ell)}$ denotes the number of bits in the ℓth binary vector $\mathbf{m}^{(\ell)}$. From (4.1), if the channel decoder was able to correct up to the ℓth partition, the length of the available source-coded information sequence can be expressed as

$$K' = \sum_{m=1}^{\ell} K^{(m)}. \tag{4.8}$$

4.3.3   Packet-wise R-D Model

For the channel decoding process at the receiver prior to the source decoding, we assume that any decoding error caused by the channel noise is detectable as described in Section 5.2.2, and the corresponding binary sequence will be discarded. Replacing the discarded sequence by the all-zero sequence seems to be an appropriate and simple method. However, the resynchronization marker must have a unique length of consecutive zeros, which is defined in the first header of each picture, and thus the zero replacement will cause the resynchronization error of each packet. To avoid this, we may use the sequence of ones as the replacement bits of erroneous sequence.

Referring to Fig. 4.2, let $K' \in \{0,\dots,K\}$ denote the number of the total bits within successfully decoded partitions defined in (4.8), where we assume that $K' = 0$ corresponds to the case where none of the packet bits is correctly decoded, whereas $K' = K$ indicates that all the packet bits are correctly decoded. Let $R = K'/K \in [0,1]$ denote the ratio of the correctly received source-coded bits in a packet. Employing the mean square error (MSE), the average distortion of this packet *per pixel block* is defined as

$$D_{\mathrm{p}}(R) = \frac{1}{3N_{\mathrm{h}}N_{\mathrm{v}}} \sum_{c \in \{\mathrm{R,G,B}\}} \sum_{i=0}^{N_{\mathrm{h}}-1} \sum_{j=0}^{N_{\mathrm{v}}-1} \left(V_c(i,j) - \tilde{V}_c(i,j|R) \right)^2, \tag{4.9}$$

where $V_c(i,j)$ is the quantized value of the pixel position (i,j) within the domain represented by the corresponding packet for the color component $c \in \{\mathrm{R,G,B}\}$, $\tilde{V}_c(i,j|R)$ is the decoded pixel value at the corresponding position from $K' = RK$ source bits for a given value of R, and N_{h} and N_{v} are the numbers of horizontal and vertical pixels in the packet, respectively.

Note that this model is based on a quasi-static channel state, namely, time fluctuation of channel SNR is negligible for each receiver, which is the main assumption of our broadcasting scenario.

4.3.4   Numerical Example

In practice, the packet length K depends on the specific source data encoded by MPEG-4, and thus it should be modeled as a random variable. Consequently, K' is also a random variable conditioned that R is fixed. Furthermore, the distortion statistics in (4.9) should
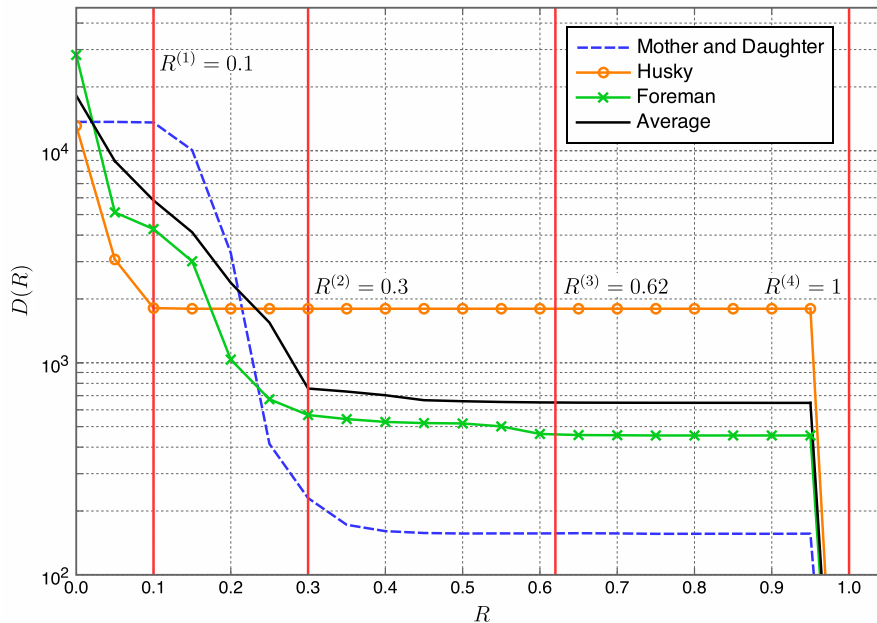
Figure 4.3: The R-D curves obtained from the three specific test sequences of respective video sample as well as the ensemble data of 30 video samples. The normalized cumulative rate $R^{(\ell)}$ that corresponds to an example combination of code rates $\mathbf{r} = (r^{(1)}, r^{(2)}, r^{(3)}, r^{(4)}) = (0.25, 0.50, 0.80, 0.95)$ are separately indicated by the vertical lines.

strongly depend on a specific video scene. In what follows, we develop the average distortion *per video* conditioned that the rate R is achieved by a given channel (i.e., no error is detected up to the ℓth partition) as $D_v(R) = E\left\{D_p(R)\right\}$, with $E\{\cdot\}$ representing an ensemble averaging.

The following video test sequences in a common intermediate format (CIF) size (352×288 at 30 Hz) are employed for our numerical evaluation [67]:

- *Mother and Daughter*: Motion is observed only in the foreground and the background is stationary. (PSNR: 42.93 dB and $\bar{K} = 212$.)

- *Husky*: The entire screen is subject to constant movement. (PSNR: 38.87 dB and $\bar{K} = 3686$.)

- *Foreman*: Large and small motions are mixed. (PSNR: 40.05 dB and $\bar{K} = 686$.)

These samples are encoded by the tool FFmpeg [68] with the corresponding peak signal-to-noise-ratio (PSNR) and the average packet length specified by $\bar{K}$.

For the above video samples, example $D_v(R)$ values are calculated as a function of R (i.e., *the $R - D$ curves*) in Fig. 4.3. In order to average out the dependence on particular video properties, we also consider the distortion averaged over a number of sample videos, which we denote by $D_{av}(R) = E\{D_v(R)\}$. In Fig. 4.3, we also plot the corresponding R-D curve of the ensemble average based on 30 video samples with average PSNR of 40.20 dB and average packet length of 834 bits. It is observed that as R increases, the corresponding distortion $D(R)$ decreases since the number of available bits K' increases on average. For relatively high R, a constant quality is achieved since K' tends to fall

within the second segment for all the tested packets. At $R = 1$, where all the packets are successfully decoded, the original MPEG-4 data are obtained with $D_v(R) = 0$. Several interesting observations depending on specific video types can be made from this result. For example, in the case of *Husky*, the sequence is dominated by texture data rather than MV, and thus the first drop of the R-D curve is observed at lower R than those of the other samples. However, since texture contains significant information for such a large motion video, the *floor* of distortion level with an increase of R is highest among all the sequences tested here. To the contrary, since *Mother and Daughter* has less amount of texture than the others, the reduction of distortion is observed at relatively high R but the resulting distortion floor is lowest. Finally, the R-D curve of *Foreman* shows similar behavior to that of the ensemble average as expected.

As pointed out in [65], the distortion estimated *per video*, i.e., $D_v(R)$, may not be necessarily suitable for real-time transmission due to the requirement of overhead in terms of computation as well as side information to obtain sufficient statistics and share them with the source decoder at the receiver. Therefore, the use of the average values of the ensemble videos, i.e., $D_{av}(R)$, which can be precalculated and shared by the receiver beforehand, should be most practical approach in our communication scenario.

4.3.5   The Relationship Between R-D Model and Code Rate

Suppose that the decoding process succeeds up to the ℓth partition and the decoding error takes place at the partition $\ell + 1$ of the UEP system. In this case, binary sequence $\left(\mathbf{m}^{(1)}, \mathbf{m}^{(2)}, \ldots, \mathbf{m}^{(\ell)}\right)$ is passed to the source decoder as described in Section 5.2.2. At the channel decoder, the number of the correctly decoded bits $K^{(\ell)}$ is then expressed by $K^{(\ell)} = KR^{(\ell)}$, where $R^{(\ell)} \in [0, 1]$ denotes the normalized cumulative rate up to the ℓth partition defined as

$$R^{(\ell)} = \sum_{m=1}^{\ell} \frac{r^{(m)}}{R_{\text{total}}} \tag{4.10}$$

with R_{total} representing the total transmission rate defined in (4.4). Note that $R^{(\ell)}$ is a non-decreasing function with ℓ and $R^{(L)} = 1$. Given that the first decoding error occurs at the ℓth partition, the resulting distortion is given by $D(R)$ in (4.9) with R replaced by $R^{(\ell)}$ of (4.10). Consequently, since $D(R)$ is a non-increasing function with R, increasing $R^{(\ell)}$ results in lower distortion at the decoding stage corresponding to the ℓth partition. In other words, $R^{(\ell)}$ determines the quality of the decoded video at the ℓth partition of UEP.

As an example, let us consider the combination of the code rates $\mathbf{r} = (r^{(1)}, r^{(2)}, r^{(3)}, r^{(4)}) = (0.25, 0.50, 0.80, 0.95)$. The corresponding cumulative rates after normalization are given by $(R^{(1)}, R^{(2)}, R^{(3)}, R^{(4)}) = (0.10, 0.30, 0.62, 1)$ and they are plotted as the vertical lines in Fig. 4.3. From the numerical R-D curve, it is observed that as the number of successfully decodable partitions increases, the corresponding distortion decreases. It follows that by adjusting the code rate for each UEP partition, the achievable quality can be controlled.
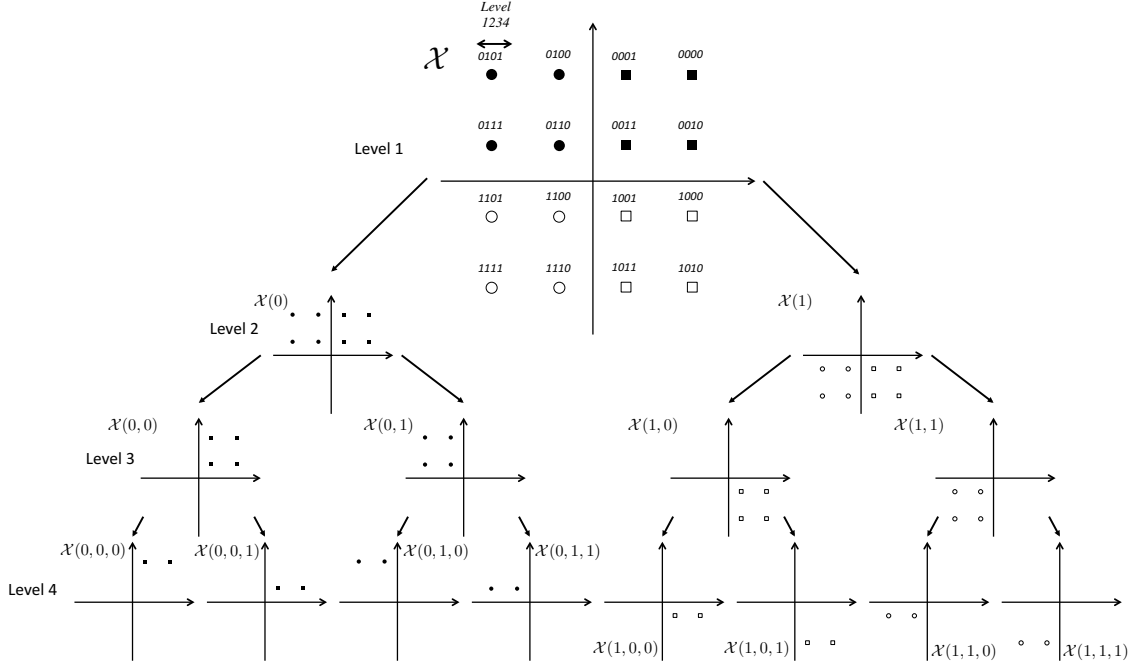
Figure 4.4: The symbol mappings of 16-QAM with the BP labeling and the subsets for each level.

4.4 Coded Modulation Design for UEP Property

In this section, we describe how to match the UEP partitions of source coded sequence to a QAM symbol frame. In particular, we address an appropriate code rate design for each binary channel of MLC/MSD system. We also describe the equivalent code rate design of a general CM based on time sharing.

4.4.1 Mapping of Partitions onto UEP

Up to the previous section, we have assumed that the number of the UEP partitions of source coded sequence and that of the codewords that form one transmission frame are equally given by $L = \log_2 M$ for M-ary QAM. In what follows, we relax this model and assume that the number of the UEP partitions is given by J where $J \leq L$. We then define the following notation that describes the mapping of J partitions onto the set of L codewords:

$$\phi(\mathbf{v}_1; \mathbf{v}_2; \ldots; \mathbf{v}_J), \tag{4.11}$$

where the vector $\mathbf{v}_j$ represents the set of the codewords that form the jth UEP partition of the source coded sequence. For example, $\phi(1, 2; 3; 4)$ indicates that $J = 3$ and $L = 4$, and that the first UEP partition of the source coded bits is divided into the two codewords possibly with the highest error protection, whereas the second and third UEP partitions are mapped onto the third and fourth codewords with decreasing error protection, respectively.

4.4.2   Code Rate Design Rule

▷ *MLC/MSD System*

As described in Section 4.2.2, the information in each UEP partition ℓ is protected by an independent channel code of rate $r^{(\ell)}$. In our source coding framework, the subsequent partitions are useful only if all its prior partitions are correctly retrieved at the source decoder. Since each partition is matched to each binary channel (or multiple binary channels) of QAM symbols in the case of MLC, the codeword of the upper level is useful only if all the codewords assigned to its lower levels are correctly decoded. For this reason, MLC with MSD, which makes use of the correct decisions of lower levels for those of higher levels, is a natural and effective encoding and decoding approach.

For a given M-ary QAM constellation set $\mathcal{X}$, let $\mathcal{X}(b_1, b_2, \ldots, b_m)$ denote its subset with the corresponding first m bits specified by $(b_1, b_2, \ldots, b_m) \in \{0,1\}^m$. For example, $\mathcal{X}(0,1)$ indicates the subset of $\mathcal{X}$ with its left-most bit given by 0 and its second left-most bit given by 1. Since $|\mathcal{X}| = M$, it follows that $|\mathcal{X}(0,1)| = M/4$. Now the process of MSD is as follows: at the first level, the decoding for the codeword $\mathbf{c}^{(1)}$ is performed and we obtain the decoded (hard-decision) codeword $\hat{\mathbf{c}}^{(1)} = (\hat{c}_0^{(1)}, \ldots, \hat{c}_{N-1}^{(1)}) \in \{0,1\}^N$. For every symbol index $n \in \{0, \ldots, N-1\}$, the corresponding signal subset $\mathcal{X}(\hat{c}_n^{(1)})$ is selected. At the second decoding level, the decoding for $\mathbf{c}^{(2)}$ is performed based on $\mathcal{X}(\hat{c}_n^{(1)})$ and the subset $\mathcal{X}(\hat{c}_n^{(1)}, \hat{c}_n^{(2)})$ is determined for each n. This decoding process continues until the codeword of the highest level L is decoded, or decoding is terminated at some level due to the decoding error.

When capacity-approaching codes are employed as a component channel code for each level, the channel capacity is a suitable measure for estimating its performance [26]. Let $C(\mathcal{X})$ denote the constellation constrained channel capacity for a given constellation $\mathcal{X}$. The constrained channel capacity given that the first m bits, $m \in \{1, 2, \ldots, L\}$, are determined by its lower levels *a priori* is expressed as

$$C_m = E_{b_1, b_2, \ldots, b_m} \{ C\left(\mathcal{X}(b_1, b_2, \ldots, b_m)\right) \}. \tag{4.12}$$

Note that C_m monotonically decreases with m and $C_0 = C(\mathcal{X})$.

The capacity of the ℓth level (equivalent binary channel) is then given by

$$C_b^{(\ell)} = C_{\ell-1} - C_\ell, \quad \text{for} \quad \ell = 1, 2, \cdots, L. \tag{4.13}$$

Consequently, the constellation labeling considerably affects the achievable capacity for each level since the equivalent channel capacity at each level is determined by its Euclidean distance property within the subset.

Note that in the case of the square-type QAM where its in-phase and quadrature channels are considered as two independent PAM constellations, L levels can be divided into two equivalent $L/2$ levels and each equivalent level has the same binary capacity. For each PAM constellation, a suitable labeling strategy that naturally provides UEP capability is known as block partitioning (BP) labeling [26, 28]. An example 16-QAM constellation with BP labeling applied to each 4-PAM is shown at the top of Fig. 4.4, where levels 1 and 3 form the PAM in the quadrature channel and levels 2 and 4 form the PAM in the in-phase channel. Also, the subsets of BP labeling for each level are shown in the figure. In contrast to the Ungerboeck set-partitioning that maximizes the minimum intra-subset Euclidean distance, the BP strategy is to make the partitioned
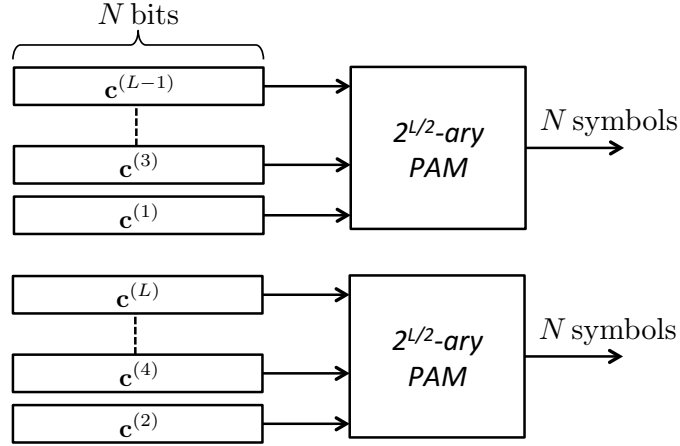
Figure 4.5: An equivalent L-level MLC system with BP labeling for 2^L-ary QAM constellations with even L.

subsets form the dense clusters of the signal points [26]. Figure 4.5 illustrates the equivalent structure of a transmitter side of the L-level MLC system with the BP labeling, where the group of even and odd levels can be considered as independent PAM symbols. Also, a specific channel capacity example in the case of 16-QAM with the BP labeling is given in Fig. 4.6, where $C_b^{(1)}$ and $C_b^{(2)}$ are identical because of the orthogonality of the in-phase and quadrature channels. Likewise, $C_b^{(3)}$ and $C_b^{(4)}$ are identical.

From the equivalent capacities shown in Fig. 4.6, we can design various UEP properties by setting the target channel SNR for each UEP stage. Several assignments of the code rates are possible. For a transmission rate of $R_{\text{total}} = 2.5$, the following two specific examples are considered; $\mathbf{r} = (r^{(1)}, r^{(2)}, r^{(3)}, r^{(4)}) = (0.25, 0.50, 0.80, 0.95)$ with $\phi(1; 2; 3; 4)$, and $\mathbf{r} = (0.625, 0.625, 0.625, 0.625)$ with $\phi(1, 2; 3, 4)$. They are plotted in Fig. 4.6 as the staircase lines that indicate the cumulative rates up to R_{total}. The vertical lines of both the staircases specify the target channel SNR over which error-free decoding can be achieved at each level by MSD, provided that the capacity achieving binary channel codes are employed. Furthermore, the same qualities will be obtained when the channel SNRs indicate the same horizontal line of the corresponding staircase. It thus follows that the system can provide graceful degradation property. The latter case with $J = 2$ is a special case of MLC and equivalent to the two-level HM formed by superposition of two quadrature phase shift keying (QPSK) constellations with different radii, where the first two levels form a "larger" QPSK and the remaining two levels form a "smaller" QPSK. This property should be apparent by observing the four constellations of "Level 3" in Fig. 4.4.

Since the BP labeling strategy satisfies [26]

$$C_b^{(\ell)} \geq C_b^{(\ell+1)} \tag{4.14}$$

for any SNR and $\ell \in \{1, 2, \ldots, L - 1\}$, the BP labeling is suitable for designing UEP based on equivalent channel capacities.

▷ *A General CM with UEP Based on Time Sharing*

For the conventional CM with UEP based on time sharing, each codeword is decoded directly from the received QAM symbols. Therefore, the equivalent channel capacities
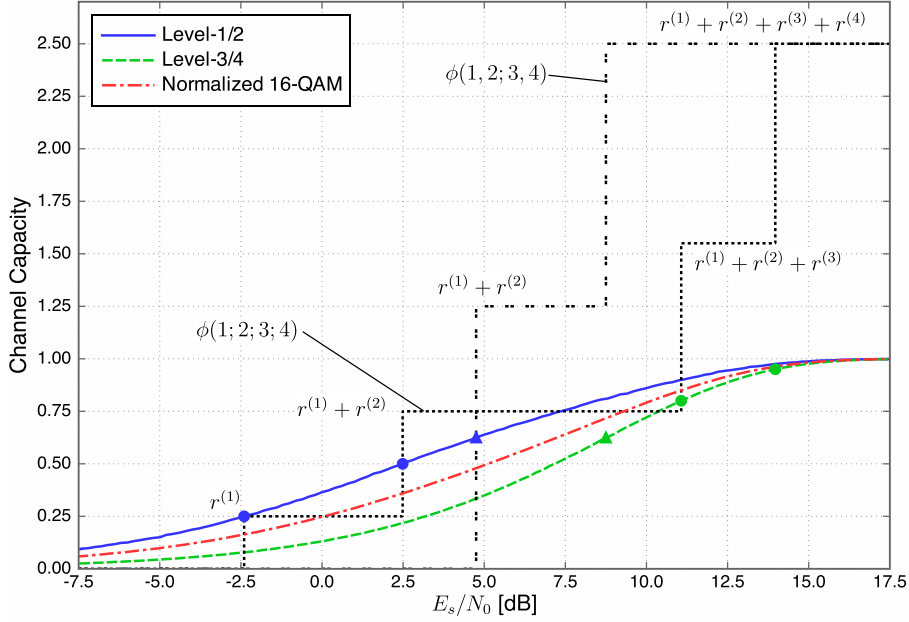
Figure 4.6: Equivalent channel capacities of 16-QAM with the BP labeling defined in Fig. 4.4. The equivalent channel capacities of levels 1 and 2 are identical, and so are those of levels 3 and 4. The two staircase (dotted and dashed) lines correspond to the example rate designs of 1) $(r^{(1)}, r^{(2)}, r^{(3)}, r^{(4)}) = (0.25, 0.50, 0.80, 0.95)$ for $J = 4$ with $\phi(1;2;3;4)$ and 2) $(r^{(1)}, r^{(2)}, r^{(3)}, r^{(4)}) = (0.625, 0.625, 0.625, 0.625)$ for $J = 2$ with $\phi(1,2;3,4)$, and their sums of the rates indicate the total achievable rate for a given SNR.

are inappropriate for its code rate design. Instead, the capacity of the original M-QAM constellation $\mathcal{X}$ should be used for a given transmission rate and the normalized capacity $C(\mathcal{X})/L$ serves as a reference for the corresponding code rate design. The normalized capacity is also plotted in Fig. 4.6.

Note that the two-level HM and conventional CM are introduced here for the purpose of comparisons with the proposed MLC approach in the subsequent simulation result.

4.5   A New Figure of Merit for a UEP-Based Approach in Multimedia Broadcasting

In Section 4.3, from the viewpoint of source coding, it has been shown that the normalized cumulative rate $R^{(\ell)}$ should be increased if one wishes to achieve higher quality at the ℓth partition. On the other hand, in the previous section, from the viewpoint of channel coding it has been demonstrated that by decreasing the code rate $r^{(\ell)}$ at the ℓth level, the corresponding frame can be made robust against noisy channels. There is thus a fundamental trade-off between the channel SNR and the achievable quality.

In this section, we address a method to analyze the effectiveness of a general UEP approach for broadcasting applications. The primary objective of introducing graceful degradation in broadcasting systems is to provide as many users as possible with best-quality services based on their respective SNR. In this sense, it is reasonable to consider the *average quality experienced by the users* served by the same broadcasting transmitter
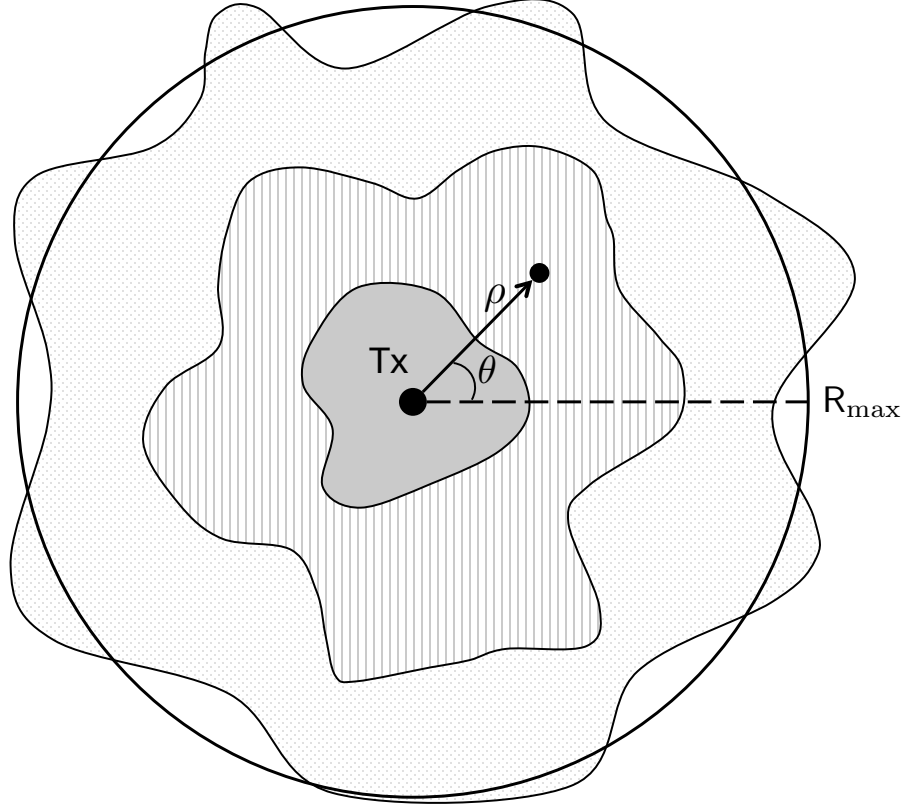
Figure 4.7: Coverage area model and the qualities provided by UEP. The users located in the region with the same filled pattern can achieve the same quality.

as a possible figure of merit. This section thus develops a novel measure that takes into account the effect of the overall quality experienced by the users within a given coverage area in the multimedia broadcasting framework.

4.5.1 Quality Measure

Let $D(R)$ denote the distortion model introduced in Section 4.3; $D(R)$ can be averaged over a single video steam, i.e., $D_v(R)$, or a large number of videos, i.e., $D_{av}(R)$, as discussed in Section 4.3.4. As a quality measure of video sequences, PSNR is commonly adopted. However, its value depends on specific source data and thus is not suitable for our unified analysis. Instead, we express the quality measure as

$$q(R) = 1 - \frac{D(R)}{D(0)}. \tag{4.15}$$

It is easy to observe that $0 \leq q(R) \leq 1$ due to the normalization by $D(0)$, and $q(R) = 1$ indicates the highest quality achievable. We define the quality achieved by up to the ℓth partitions of the source-coded bits as

$$q^{(\ell)} = q(R^{(\ell)}), \tag{4.16}$$

which should satisfy the property of incremental quality defined by (4.2).

We now consider the broadcasting service scenario illustrated in Fig. 4.7, where a single transmitter has an isotropic antenna with the coverage area of radius R_{max} in two dimensional plane. Without loss of generality, we assume that the transmitter is located

at the origin, and the users are randomly located at the point (ρ, θ) in polar coordinates, where ρ and θ follow a joint probability density function (pdf) $f_{\rho,\theta}(\rho, \theta)$.

We denote the achievable quality of multimedia data with the instantaneous channel SNR γ_ρ at the distance ρ through the system with UEP property. For a given realization of γ_ρ and code rate set $\mathbf{r} = (r^{(1)}, \ldots, r^{(L)})$, the maximum number of the correctly decodable partitions ℓ can be uniquely determined. In other words, the quality of the decoded video is given by $q^{(\ell)}$ of (4.2), where ℓ is uniquely determined by the received SNR γ_ρ, given $\mathbf{r} = (r^{(1)}, \ldots, r^{(L)})$. We thus denote this quality measure by $q^{\{\ell(\gamma_\rho|\mathbf{r})\}}$, where $\ell(\gamma_\rho|\mathbf{r}) \in \{0, 1, 2, \ldots, L\}$. We note that since γ_ρ is a random variable, so is $q^{\{\ell(\gamma_\rho|\mathbf{r})\}}$.

Let $\gamma^{(\ell)}(\mathbf{r})$ denote the required channel SNR in order to achieve the normalized cumulative rate $R^{(\ell)}$ that can be calculated from $\mathbf{r}$ according to (4.10). Since $\gamma^{(\ell)}(\mathbf{r})$ can be determined from the channel capacity curve plotted in Fig. 4.6, the corresponding maximum level $\ell(\gamma|\mathbf{r})$ that can be decoded correctly is expressed as

$$\ell(\gamma|\mathbf{r}) = \max_{\ell \in \{0,1,\ldots,L\}} \left\{ \ell : \gamma \geq \gamma^{(\ell)}(\mathbf{r}) \right\}, \tag{4.17}$$

with $\gamma^{(0)}(\mathbf{r}) = 0$.

4.5.2 Average Quality

As illustrated in Fig. 4.7, the users located even at the same distance ρ may experience different instantaneous SNR in the presence of shadowing. The average quality experienced by the users located at the distance ρ can be expressed as

$$Q(\rho|\mathbf{r}) = \int_0^\infty p_\gamma(\gamma|\rho) q^{\{\ell(\gamma|\mathbf{r})\}} d\gamma, \tag{4.18}$$

where $p_\gamma(\gamma|\rho)$ is the pdf of the channel SNR γ at the distance ρ of which its dB value, $\gamma_{\mathrm{dB}} = 10 \log_{10} \gamma$, follows (4.7) in the presence of shadowing. Otherwise, without fading and shadowing we simply have $Q(\rho|\mathbf{r}) = q^{\{\ell(\gamma_\rho|\mathbf{r})\}}$ with $\gamma_\rho = P_R(\rho)/P_N$.

The quality averaged over the location of the receiver can be expressed as

$$A(\mathbf{r}, R_{\max}) = \int_0^{2\pi} \int_0^{R_{\max}} f_{\rho,\theta}(\rho, \theta) Q(\rho|\mathbf{r}) d\rho d\theta. \tag{4.19}$$

In summary, given a coverage distance $R_{\max}$, (4.19) represents the average quality experienced by the target users which is achieved by a specific UEP broadcasting system with code rate set $\mathbf{r}$. We further note that if the users are uniformly distributed in the coverage area such that

$$f_{\rho,\theta}(\rho, \theta) = f_\theta(\theta) f_\rho(\rho)$$

with

$$f_\theta(\theta) = 1/2\pi \quad \text{and} \quad f_\rho(\rho) = 2\rho/R_{\max}^2,$$

then (4.19) reduces to

$$A(\mathbf{r}, R_{\max}) = \frac{2}{R_{\max}^2} \int_0^{R_{\max}} \rho Q(\rho|\mathbf{r}) d\rho. \tag{4.20}$$

Due to the definition of our quality measure (4.15), its average (4.20) should also satisfy $0 \leq A(\mathbf{r}, R_{\max}) \leq 1$.

4.5.3 Optimal Rate Design

Since the primary purpose of UEP in the broadcasting service is to satisfy as many users as possible without over-compromising the quality of each user, it is necessary to find the balancing point between the rate of channel codes and resulting quality of service (QoS). Using the proposed measure introduced in this section, we attempt to find the balancing point between the transmission rate and achievable *average* quality experienced by the users located within a given coverage area.

We are interested in the code rate vector $\mathbf{r}$ that maximizes the average quality $\mathsf{A}(\mathbf{r}, \mathsf{R}_{\max})$ defined in (4.20), which depends on the coverage distance $\mathsf{R}_{\max}$ for a specified modulation format. The code rate vector $\mathbf{r}$ itself is subject to the transmission rate constraint defined in (4.4). Therefore, the rate optimization problem can be formulated as

$$\mathbf{r}_{\mathrm{opt}} = \arg \max_{\mathbf{r} \in \mathcal{R}^L} \quad \mathsf{A}(\mathbf{r}, \mathsf{R}_{\max})$$

$$\text{subject to} \quad \sum_{\ell=1}^{L} r^{(\ell)} = R_{\mathrm{total}},$$

where $\mathcal{R}$ is a finite set of the code rates that are practically realizable by a given channel code of interest. Apparently, the above optimization problem depends on the two parameters; the coverage distance $\mathsf{R}_{\max}$ and the transmission rate R_{total}. The optimal rate allocation vector $\mathbf{r}_{\mathrm{opt}}$ that maximizes $\mathsf{A}(\mathbf{r}, \mathsf{R}_{\max})$ should be thus a function of a given pair $(\mathsf{R}_{\max}, R_{\mathrm{total}})$.

4.6 Numerical Results

In this section, we compare the proposed average quality measure for the following four scenarios; 1) the proposed MLC system with UEP, 2) the two-level HM system with UEP, 3) the conventional CM system with UEP based on time sharing, and 4) the conventional CM system with equal error protection (EEP). The fourth system with EEP is designed under the constraint of $J = 1$ in the code rate design, i.e., without any consideration of the inherent UEP applicability of MPEG-4 frame structure.

Table 4.1 shows the parameters used for our numerical evaluations. The average power of AWGN is given by $P_N = N_0 B$ with its power spectral density $N_0 = k_B T_0 = 4.14 \times 10^{-21}$ W/Hz, where $k_B = 1.38 \times 10^{-23}$ W-s/K is the Boltzmann's constant, $T_0 = 300$ K is the noise temperature, and the signal bandwidth is given by $B = 6$ MHz.

In our numerical results, we fix the coverage distance as $\mathsf{R}_{\max} = 20$ km and the average qualities are derived for various values of R_{total}. The resulting curve that represents the relationship between the average quality and the total transmission rate is referred to as the *R-A curve* in what follows. As binary component codes for error correction of MLC, the regular LDPC codes are employed. For conventional CM systems, each transmission frame is assumed to employ LDPC coded modulation with 16-QAM for channel coding and modulation. The number of the codewords that form a transmission frame is set as $L = 4$ for all UEP systems. The set of the candidate vectors $\mathcal{R}$ thus consists of the code rates that can be generated by the regular LDPC codes [8] in a straightforward manner. As described in Section 5.2.2, MSD generally requires the error detection capability in order to avoid error propagation, but since the LDPC codes inherently have an error

Table 4.1: Simulation parameters.

P_T	100 W
ρ_0	1 km
f_c	900 MHz
P_N	2.48×10^{-14} W
α	5.0
$\sigma^2_{\gamma_{\mathrm{dB}}}$	6.0
$R_{\max}$	20 km
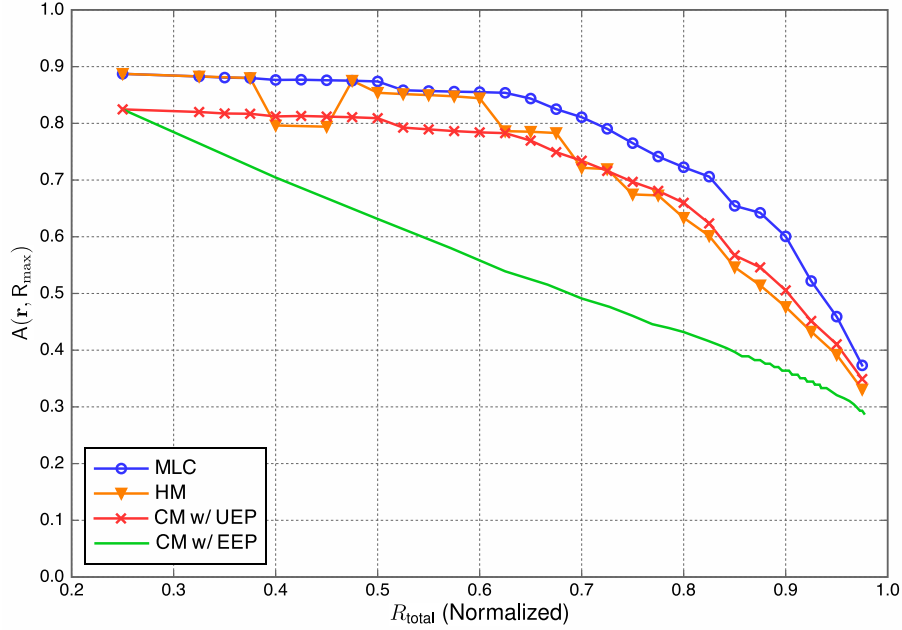
Figure 4.8: Numerical R-A functions with $\sigma^2_{\gamma_{\mathrm{dB}}} = 6$ and $R_{\max} = 20$ km.

detection capability after iterative decoding through the use of parity-check matrix, additional error detection codes such as CRC codes may not be necessary.

4.6.1   Results of Quality Improvement for Ensemble Videos

We first show the R-A curves based on the R-D model of the ensemble video average. In other words, $D_{\mathrm{av}}(R)$ is applied to (4.15) for its quality calculation. The R-A curves are derived by finding $\mathbf{r}_{\mathrm{opt}}$ for each R_{total}. The results are shown in Fig. 4.8. First, we observe that all the systems that exploit UEP property, namely, MLC, HM, and CM, are superior to the CM with EEP for all the values of R_{total} evaluated here, which clearly demonstrates the benefit of graceful degradation in our multimedia broadcasting framework. Furthermore, it is apparent that MLC outperforms the two-level HM and CM even with UEP, since the incremental quality can be efficiently implemented up to the levels where the decoding error first takes place. This is due to the capacity property of the BP labeling characterized in (4.14), which indicates that a larger number of source coded bits in the lowest UEP partitions can be supported, compared to that of the conventional CM, under the same channel SNR condition. This result clearly demonstrates that the MLC is more suitable for multimedia transmission with graceful degradation than the conventional CM system, even with UEP capability. Since the

number of available code rate sets is strictly limited, the two-level HM system exhibits performance inferior to the CM system with UEP for several cases of R_{total}.

4.6.2 Results of Quality Improvement for Individual Video Streams

Finally, we verify the effectiveness of employing the code rate design rule based on the distortion model using ensemble videos, i.e., $D_{\text{av}}(R)$, as well as the video-specific distortion model $D_{\text{v}}(R)$ when it is applied to a particular video stream.

The R-A curves of the three specific video streams considered in Section 4.3.4 are compared in Fig. 4.9 (a), (b), and (c). In each figure, the rate designs employing MLC, two-level HM, and CM systems are compared, all employing the video-specific distortion model $D_{\text{v}}(R)$. In addition, the MLC with the rate design based on the ensemble videos $D_{\text{av}}(R)$ is plotted for comparison. From these results, it is observed that the UEP systems optimized for individual videos outperform the EEP system, and MLC exhibits the best performance among all the systems.

In the case of the source videos of *Foreman* and *Husky*, the results of MLC designed based on $D_{\text{v}}(R)$ and $D_{\text{av}}(R)$ are similar, justifying the use of distortion model based on the ensemble videos instead of a single video stream. On the other hand, in the case of *Mother and Daughter*, a gap between the two cases becomes apparent as R_{total} increases, and this stems from the fact that there is a gap in the R-D models between the individual video and ensemble average as observed in Fig. 4.3 due to the peculiar motion in the video stream. Nevertheless, it still outperforms the CM with EEP and justifies the use of MLC-based system design using the distortion model based on average ensemble.

It can be concluded that the code rate design approach developed based on the R-D model of ensemble videos is applicable to most scenes in digital broadcasting.

4.7 Conclusion

Under the scenario of multimedia broadcasting services, we have proposed a new approach for designing the UEP system with graceful degradation. A new measure that quantifies the average quality of the videos received by uniformly distributed users has been introduced. As a specific practical approach, the code rate design of MLC with capacity-approaching codes combined with the distortion model based on the MPEG-4 video coding has been developed.

The numerical results have shown that compared to the conventional CM systems with and without UEP, the proposed MLC-based approach can achieve highest user satisfaction (or QoS) for given transmission rate and specified coverage area. It has been revealed that achievable average quality experienced by the target users within a given coverage area depends on the characteristic of videos and the code rate design. It also has turned out that the code rate design based on the ensemble videos can achieve good performance in most videos, while that based on the distortion model of a specific video stream always provides best graceful degradation property at the cost of an extra overhead.

The major constraints of the proposed approach are that the code rate design rule should be established depending on the employed channel coding technique, and the distortion model of a given source coding has to be developed beforehand. Nevertheless, the proposed design concept and optimization approach are applicable not only to the
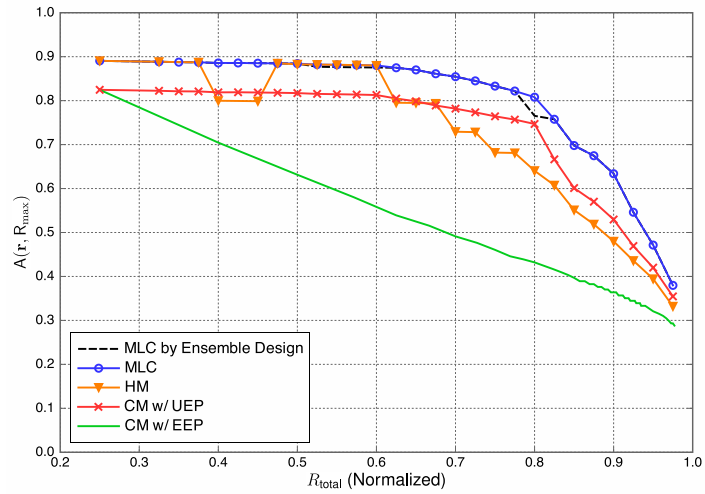
(a)
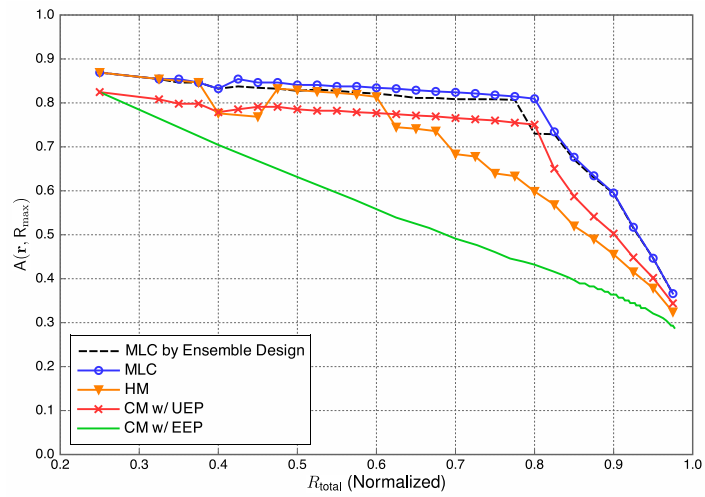
(b)
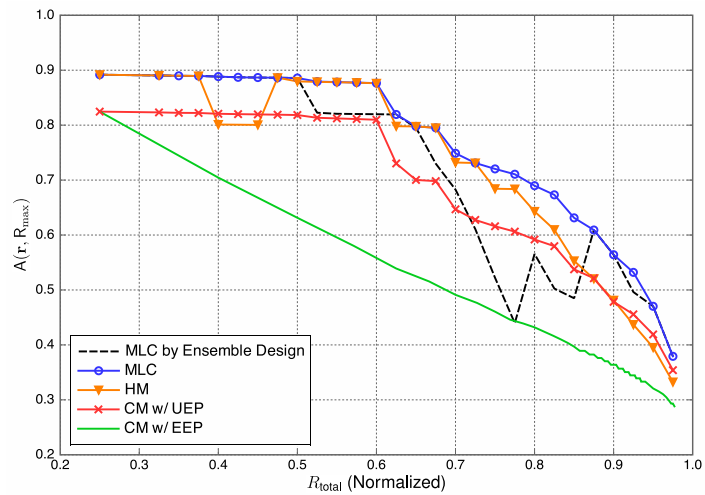
(c)

Figure 4.9: The R-A curves for specific sources. (a) *Foreman*. (b) *Husky*. (c) *Mother and Daughter*.

system considered in this chapter, but also to more general systems that combine source and channel coding. Moreover, its extension to other propagation channel models is straightforward. Further investigation in terms of its applicability to the most recent video coding standards is left as future work.

This page intentionally left blank

CHAPTER 5

A Novel Channel Decoder Exploiting Inherent Padding Bits for Fixed-Length Frame Transmission

5.1 Introduction

The multimedia source encoders such as video, image, and audio compression usually produce coded bits of variable length as a result of entropy coding as described in Chapter 2. On the other hand, the data in the physical (PHY) layer is usually transmitted block-wise as a unit such as a *frame*, and the length of the frame is fixed *a priori*. Therefore, the length of the codeword after the channel coding should be subject to this frame length constraint. Since the rate of a channel code is usually determined *a priori*, the number of information bits to be encoded by this code is fixed as well. As a consequence, there should be a mismatch between the length of the source coded bits and that of the information bits to be encoded by a channel code. A practical and straightforward approach is to simply insert some dummy bits, which will be called *padding bits* in what follows, such that the length of the source coded sequence matches that of the information bits processed by the fixed-rate channel encoder. This scenario can be applied not only to a simple source-channel coding system, but also to a general network framework. Classic Ethernet, for example, defines the minimum frame length in order to enable collision detection, and if the actual data frame is shorter than the minimum length required, padding bytes are appended to the data field [69]. More recent wireless LAN standards such as IEEE 802.11n and 11ac introduce frame aggregation [70], where more than a single packet may be contained in one transmission frame. In such systems, each packet may have a few inherent null bytes to form a single transmit frame in the medium access control (MAC) layer. Furthermore, null data packets (NDP) may be inserted when the number of the packets generated by the upper layer is not sufficient to fill the entire frame. In this case, these NDP can be also considered as a collection of padding bits.

In the conventional system, since the padding bits are simply inserted prior to the channel encoder at the transmitter and then automatically removed *after* the channel decoder at the receiver, it neither improves throughput nor enhances reliability, thus resulting in a reduction in terms of bandwidth efficiency or effective throughput. A straightforward approach that exploits these redundant padding bits from the viewpoint of improving reliability is to replace the redundant null bits in the data frame with the parity bits of the systematic codes [71], but such an approach not only lacks flexibility in code design but also makes the soft decision decoding challenging even if such information is available from the output of the detector. In this work, we take an alternative approach inspired by *code doping* [72] for the iterative decoding of turbo code; Code doping is a technique to improve the performance of iterative decoder and accomplished by inserting the pilot bits to the original information bits or by replacing

the output of the outer coded bits with the pilot bits, provided that their positions and values are known at the receiver side. In [73], code doping is employed for removing the error floor of bit-interleaved coded modulation with iterative detection (BICM-ID) system. The *trellis pruning* is also an equivalent technique, and such an approach is recently studied in the framework of the cooperative communications [74]. Trellis pruning is applicable not only to the block code, but also to the convolutional code [75]. As related work, the unequal error protection (UEP) using *path-pruned* convolutional code is proposed in [76] and [77], where the latter study combines the use of rate-compatible punctured code [55].

This chapter proposes an approach that makes effective use of the redundant padding bits based on trellis pruning such that the error rate performance is improved. We start by assuming that the padding event at the transmitter conforms to the Bernoulli distribution, and thus the fixed length padding bits are generated with a certain probability. Turbo code is employed as our example channel code, and the decoder incorporates the maximum *a posteriori* (MAP) decoding based on the BCJR algorithm. Without loss of generality, we assume that "zero" bits are always chosen as dummy bits, and depending on how this zero padding (ZP) is performed, we define the following two approaches. The first approach is to put the zero bits consecutively in a suffix of an information sequence similar to those introduced in a conventional system. We refer to this approach as a consecutive zero padding (CZP). The second, rather unconventional approach is to insert the dummy zeros sparsely in the information sequence, which will be referred to as a sparse zero insertion (SZI). Unlike the above-mentioned conventional studies where the receivers are assumed to have the knowledge of dummy bit positions, our proposed padding event detector eliminates the necessity of transmitting side information that specifies the existence of padding bits. Based on the acquired knowledge of padding bit positions according to the result of the proposed detector at the receiver, the turbo decoder with trellis pruning is employed, which can be implemented by an additional low complexity process without any additional redundancy.

While SZI has no major difference from CZP in terms of spectral efficiency and implementation cost, we show that SZI together with the BCJR algorithm results in better error correction performance than CZP through an extensive analysis based on the extrinsic information transfer (EXIT) chart [78]. In addition, it is shown that the proposed integrated decoder contributes to an improvement of the system throughput over the conventional system that simply discards the padding bits after channel decoding. Our numerical results are also confirmed by the theoretical analysis based on a finite blocklength channel coding regime [79, 80].

Our main contributions in this chapter are summarized as follows:

1. In order to facilitate the analyses and simulations, a suitable padding event model at the transmitter side, based on the Bernoulli distribution with fixed-length padding bits, is introduced. This model is also extended to more general variable-length cases.

2. A novel zero padding approach based on SZI for effectively utilizing the trellis pruning is proposed and its effectiveness over the conventional CZP is analyzed through the EXIT chart.
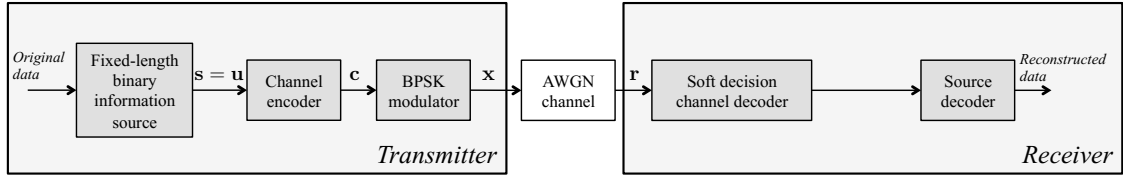
Figure 5.1: A ZP-unaware transmission system.

3. We develop a detecting technique of padding bits based on hypothesis testing at the receiver side.

4. Through computer simulations and based on a finite blocklength analysis, we demonstrate that the proposed system leads to the enhancement in terms of the system throughput.

This chapter is a substantial extension of our initial work presented in [81] and organized as follows. In Section 5.2, we give an overview of conventional ZP-unaware as well as proposed ZP-aware transmitter models together with the proposed decoder. Section 5.3 describes the conventional and proposed zero padding approaches at the transmitter and the proposed ZP-aware iterative decoder based on the BCJR algorithm at the receiver. In Section 5.4, the detection of padding bit event at the receiver based on hypothesis testing is described. Section 5.5 derives throughput expressions for the systems with and without padding bits. Section 5.6 compares the performance of CZP and SZI in terms of FER and evaluates their achievable throughputs as well as those based on the finite blocklength analysis. Section 5.7 concludes this work.

Throughout this chapter, we employ BPSK as our modulation for simplicity, but our approach is also applicable to more general systems operating with high-order QAM.

5.2 System Description

In this section, we briefly review a ZP-unaware system model based on the conventional assumption that the length of data bits exactly matches that of the information bits input to the channel encoder, thus requiring no padding bit. We then develop our encoder model that has the knowledge of the padding bit event followed by the description of our proposed integrated decoder.

5.2.1 ZP-Unaware Model

Figure 5.1 shows a ZP-unaware transmission system where after the source encoding of the original data, the binary information sequence of length K', denoted by $\mathbf{s} = (s_1, s_2, \ldots, s_{K'}) \in \{0,1\}^{K'}$, is generated. This is input to the channel encoder. Let $\mathbf{u} = (u_1, u_2, \ldots, u_K) \in \{0,1\}^K$ denote a binary information sequence of length K to be processed by the channel encoder. In a conventional system, we assume that the length of the data bits or packet bits (received from higher layer) to be transmitted and that of the information bits to be encoded by channel coding are perfectly matched or the residual mismatch is not taken into consideration upon performance evaluation. In other words, it is simply assumed that $K' = K$ which indicates that $\mathbf{u} = \mathbf{s}$.

The subsequent channel encoder with its code rate R generates the codeword of N bits denoted by $\mathbf{c} = (c_1, c_2, \ldots, c_N) \in \{0,1\}^N$. The entire codeword is assumed to be transmitted
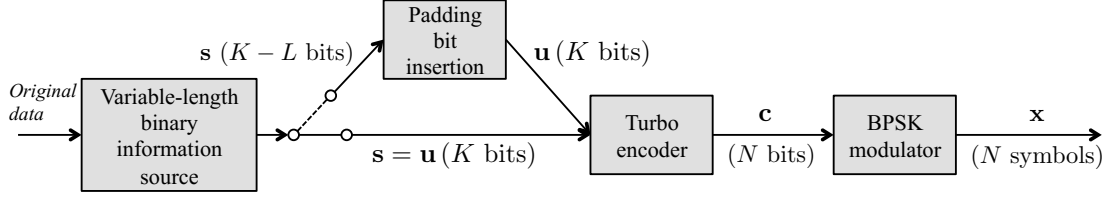
Figure 5.2: The proposed ZP-aware transmitter model.

as a single frame. It then follows that the relationship between the information bits and the channel coded bits is given by $K = NR$. Each element of the vector $\mathbf{c}$ is then modulated by BPSK, and the symbol vector $\mathbf{x} = (x_1, x_2, \ldots, x_N)$, where $x_i \in \mathcal{X} = \{1, -1\}$ and $x_i = 2c_i - 1$ for $i \in \{1, \ldots, N\}$, is transmitted over an AWGN channel. The received complex symbol vector $\mathbf{r} = (r_1, r_2, \ldots, r_N) \in \mathbb{C}^N$ is expressed as

$$\mathbf{r} = \sqrt{E_s}\mathbf{x} + \mathbf{n}, \tag{5.1}$$

where $\mathbf{n} = (n_1, n_2, \ldots, n_N) \in \mathbb{C}^N$ is an AWGN term with each element being an independent and identically distributed (i.i.d.) complex Gaussian random variable with zero mean and variance $N_0/2$ per dimension, and E_s is the average energy of the received symbol.

5.2.2   ZP-Aware Transmitter Model and the Proposed Decoder

▷ *Transmitter*

Figure 5.2 shows our ZP-aware system model at the transmitter side. As is often the case with systems in practice, we assume that there is a gap between the length K' of the binary data sequence received from the higher layer and that of the binary information K that forms a fixed-length frame after channel encoding with a fixed code rate R. Without loss of generality, we assume that $K' \leq K$. In order to meet the constraint on the frame length of N, the insertion of the padding bits of length $L' = K - K'$ is necessary.

For simplicity in terms of our subsequent analysis and implementation, let a positive integer L denote the predetermined threshold of the padding bit length and let us consider only the following two cases; if the length of the actual padding bits L' is larger than or equal to L, the length of the padding bits is assumed to be L. On the other hand, if $L' < L$, we assume that the padding bit length is 0. With this assumption, we define the following two events at the transmitter side:

$$\begin{aligned}
&H_0 : \text{padding bits are not inserted.} \\
&H_1 : \text{padding bits are inserted.}
\end{aligned} \tag{5.2}$$

The probability of the padding bit insertion at the transmitter side is denoted by $p(H_1)$, whereas that of the other case is given by $p(H_0) = 1 - p(H_1)$. The gap of L bits should be filled by the padding bits, and these dummy bits do not carry any information. Therefore, the maximum length of the useful information sequence K' is reduced to $K - L$ in this case. After this probabilistic padding event, the input to the channel encoder $\mathbf{u}$ is classified as either the sequence $\mathbf{s}$ plus the dummy bits of unknown length (but less than L), corresponding to the event H_0, or that with the padding bits of length at least L, corresponding to the event H_1.
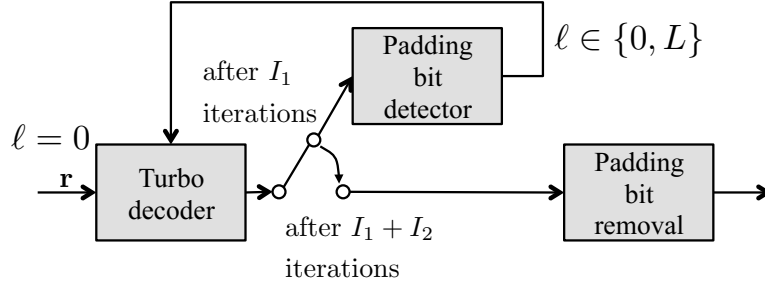
Figure 5.3: The proposed decoding structure with the detector of padding bits.

▷ *Receiver*

Direct transmission of the information that identifies the existence of padding bits, though feasible, requires additional signal transmission and thus reduces spectral efficiency as well as power efficiency, especially considering the fact that this information is required for each codeword (or frame). Therefore, it would be desirable if the system can be designed such that this implicit side information can be easily retrieved at the receiver side through blind estimation. For the ZP-aware transmitter model defined in Fig. 5.2, the corresponding integrated decoder we propose is shown in Fig. 5.3, where the padding bit event detection is incorporated into our channel decoder described in the next section. By assumption, the padding bit event follows the Bernoulli distribution and the length of the padding bits ℓ can be chosen from 0 or L, i.e., $\ell \in \{0, L\}$. The corresponding proposed integrated decoder is described as follows:

Step 1) Perform I_1 iterations of the turbo decoding under the assumption that $\ell = 0$.

Step 2) Perform detection of the padding bits. If the padding bits are detected, set $\ell = L$. Otherwise, set $\ell = 0$.

Step 3) Perform I_2 iterations of the turbo decoding with the number of padding bits ℓ.

The above algorithm performs $I_1 + I_2$ iterations in total, and it aims to gradually improve the performance of the decoder. The detail on the padding bit event detector will be discussed in Section 5.4. The proposed detecting algorithm is also applicable to the cases operating with multiple padding bit lengths in a straightforward manner, which will be also discussed in Section 5.4.

5.3 Design of ZP-Aware Systems for Turbo Decoding

This section describes the two representative approaches for insertion of padding bits in Fig. 5.2, followed by the proposed turbo decoding that takes into account the padding bit event.

5.3.1 Encoding Process with ZP

Figure 5.4(a) shows a general turbo encoder, i.e., parallel concatenated convolutional codes (PCCC) [7], with the code rate 1/3, where the two recursive systematic convolutional (RSC) encoders produce the parity bits. A binary information vector of length K, denoted by $\mathbf{u} = (u_1, u_2, \ldots, u_K)$, will be permuted by the pseudo-random bit interleaver of size K, denoted by π in the figure, before it is input to the second

(a)

(b)
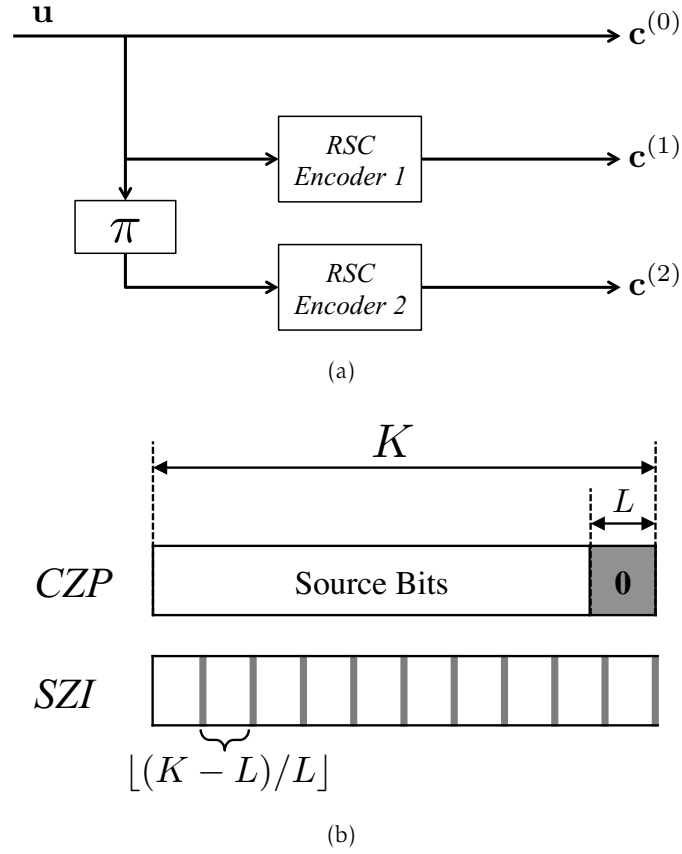
Figure 5.4: The proposed turbo encoder structure. (a) A general turbo encoder with the code rate 1/3. (b) The structures of **u** after ZP with L padding bits. In the case of SZI, the null bits are sparsely inserted with the period of $\lfloor (K-L)/L \rfloor$.

RSC encoder. Due to the systematic structure of the turbo code, **u** is also considered as an output sequence denoted by $\mathbf{c}^{(0)} = \left(c_1^{(0)}, c_2^{(0)}, \ldots, c_K^{(0)} \right)$ in Fig. 5.4(a), and the two parity sequences $\mathbf{c}^{(1)} = \left(c_1^{(1)}, c_2^{(1)} \ldots, c_K^{(1)} \right)$ and $\mathbf{c}^{(2)} = \left(c_1^{(2)}, c_2^{(2)}, \ldots, c_K^{(2)} \right)$ are generated accordingly. It follows that the resulting codeword is expressed as $\mathbf{c} = \left(\mathbf{c}^{(0)}, \mathbf{c}^{(1)}, \mathbf{c}^{(2)} \right)$. This codeword is modulated by BPSK and transmitted over an AWGN channel as described in Section 5.2.1. The vector of the corresponding received BPSK symbols is expressed as $\mathbf{r} = \left(\mathbf{r}^{(0)}, \mathbf{r}^{(1)}, \mathbf{r}^{(2)} \right)$, where $\mathbf{r}^{(k)} = \left(r_1^{(k)}, r_2^{(k)}, \ldots, r_K^{(k)} \right) \in \mathbb{C}^K$ corresponds to the transmitted binary vector $\mathbf{c}^{(k)}$ for $k \in \{0, 1, 2\}$.

Figure 5.4(b) illustrates how zero padding is performed in order to generate **u** from the binary data sequence **s**. Assuming that exactly L zero bits are to be inserted, i.e., $\mathbf{s} \in \{0, 1\}^{K-L}$ and $\mathbf{u} \in \{0, 1\}^K$, they can be in general related by

$$\mathbf{u} = \mathbf{s} \mathbf{P}_{K,L}, \qquad (5.3)$$

where $\mathbf{P}_{K,L}$ represents a $(K-L) \times K$ transformation matrix. We consider the two cases shown in Fig. 5.4(b), which will be described in what follows.

▷ *Consecutive Zero Padding (CZP)*

A straightforward approach is to pad L dummy bits as a postfix of $\mathbf{s}$. In this case, the transformation matrix can be expressed as

$$\mathbf{P}_{K,L} = [\mathbf{I}_{K-L}|\mathbf{O}_{K-L,L}], \qquad (5.4)$$

where $\mathbf{I}_N$ is the identity matrix of size N and $\mathbf{O}_{N,M}$ is the $N \times M$ zero matrix. By (5.3), we have $u_i = 0$ for $K - L + 1 \leq i \leq K$. For convenience of our subsequent analysis, we define the set of the indices of the padding bit locations in $\mathbf{u}$ as

$$\mathcal{I} = \mathcal{I}_{\text{CZP}} = \{K - L + 1, \ldots, K\}. \qquad (5.5)$$

It should be note that this approach may be equivalent to the conventional *shortened* block code where the information bits are shortened by the length of L bits. In general, the shortened part of the information bits will not be transmitted in order to enhance the code rate, but in our application they are transmitted as dummy bits for the purpose of detecting the padding event at the receiver.

▷ *Sparse Zero Insertion (SZI)*

In this case, we periodically insert a single null bit such that all null bits are separately distributed, each adjacent pair with the maximum distance. In order to distribute L zero bits over $(K - L)$ information bits in the aforementioned manner, every two adjacent bits should be separated by $\lfloor (K-L)/L \rfloor$, where $\lfloor x \rfloor$ is the largest integer not greater than x. The resulting transformation matrix $\mathbf{P}_{K,L}$ can be expressed as

$$\mathbf{P}_{K,L} = \begin{bmatrix} \mathbf{I}_P & 0 & \mathbf{O}_{P \times P} & 0 & \cdots & 0 & \mathbf{O}_{P \times P} & 0 \\ \mathbf{O}_{P \times P} & 0 & \mathbf{I}_P & 0 & \cdots & 0 & \mathbf{O}_{P \times P} & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \mathbf{O}_{P \times P} & 0 & \mathbf{O}_{P \times P} & 0 & \cdots & 0 & \mathbf{I}_P & 0 \end{bmatrix}, \qquad (5.6)$$

where $P = \lfloor (K-L)/L \rfloor = \lfloor K/L \rfloor - 1$. It is easy to observe that for a given information vector $\mathbf{u} = (u_1, u_2, \cdots, u_K)$, $u_i = 0$ whenever i is a multiple of $\lfloor K/L \rfloor$. Subsequently, the set of the indices of the padding bit locations in $\mathbf{u}$ is expressed as

$$\mathcal{I} = \mathcal{I}_{\text{SZI}} = \{\lfloor K/L \rfloor, 2\lfloor K/L \rfloor, \ldots, L\lfloor K/L \rfloor\}. \qquad (5.7)$$

5.3.2 The ZP-Aware Iterative Decoder with Trellis Pruning

In this work, we consider the use of the standard turbo decoder that consists of the two iterative MAP decoders based on the BCJR algorithm [7].

When the positions of the padding bits are available at the decoder, the corresponding branches in the trellis diagram can be pruned, known as *trellis pruning*. As an example for the RSC encoder with the feedback and feed-forward generating polynomials denoted by $g_0 = (7)_8$ and $g_1 = (5)_8$, the trellis diagram in the MAP decoder with trellis pruning is illustrated in Fig. 5.5(a) for the case of CZP, where the consecutive L trellis branches are pruned for $i \in \mathcal{I}_{\text{CZP}}$. On the other hand, the trellis diagram where the branches are pruned periodically is shown in Fig. 5.5(b), which corresponds to the case of SZI. We note that due to the existence of the pseudo-random interleaver, it is likely that the branches in the MAP decoder for the interleaved sequence are also pruned in a pseudo-random manner in both the cases of CZP and SZI.
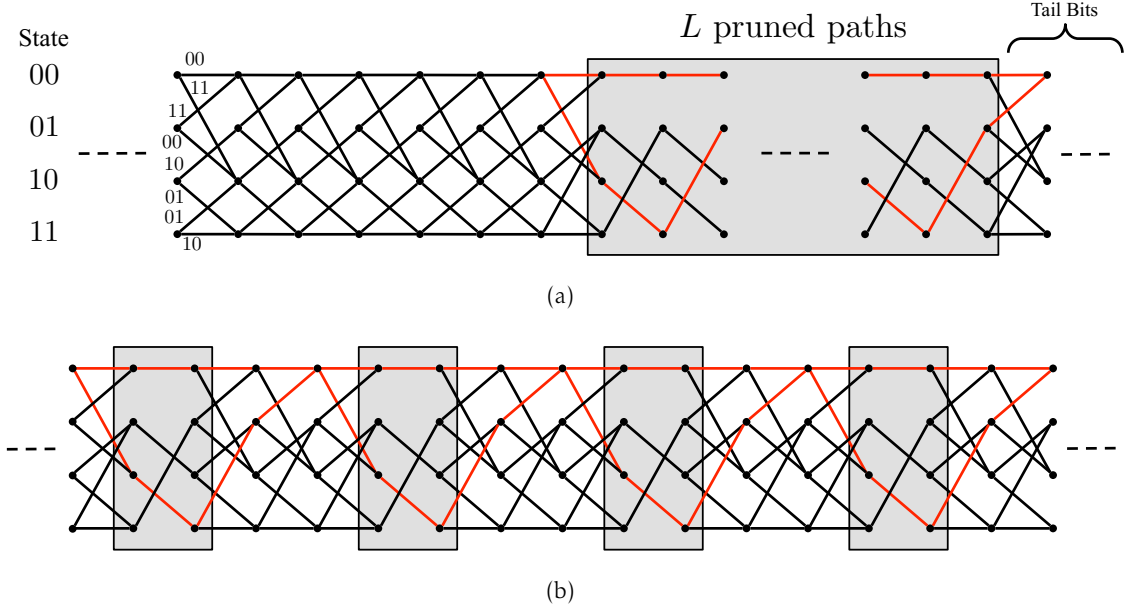
Figure 5.5: Examples of trellis diagrams. (a) CZP where L consecutive branches are pruned. (b) SZI where the branches are periodically pruned.

The trellis pruning increases the free distance in the MAP decoder. When CZP and SZI are compared, CZP can increase the free distance significantly, but only at the end of the trellis section. On the other hand, SZI is able to increase the free distance periodically, and the latter case turns out to be effective for enhancing the error rate performance, which will be highlighted by the subsequent EXIT chart analysis in the next subsection.

For a practice use of PCCC with systematic structure, trellis pruning process can be easily achieved by adjusting the LLR values of the iterative decoders that correspond to the padding bits. Specifically, each *a priori* LLR $\Lambda_{\mathrm{in}}(u_i)$ corresponding to the element u_i is modified such that

$$\Lambda_{\mathrm{in}}(u_i) = \log \frac{p_i(1)}{p_i(0)} = -\infty, \quad \text{for} \quad i \in \mathcal{I}, \tag{5.8}$$

where $p_i(b)$ is the *a priori* probability of the event that the ith bit is $b \in \{0, 1\}$. This is equivalent to the replacement of the *a priori* probability of the corresponding bits as $p_i(0) = 1$ and $p_i(1) = 0$ for $i \in \mathcal{I}$. The block diagram of the proposed iterative decoder is illustrated in Fig. 5.6, where the two MAP decoders correspond to the two RSC encoders and π is the interleaver defined in Fig. 5.4(a), with π^{-1} denoting the corresponding deinterleaver. The block "$-\infty$" serves as an LLR adjustment in (5.8). The final decoded bits are estimated based on the *a posteriori* LLR $\mathbf{\Lambda}^{(2)}$ in the second MAP decoder. As one can see, the proposed LLR adjustment block can be implemented independently of the MAP decoders, and the additional complexity compared to the conventional turbo decoder is negligible as discussed in [73]. When $L = 0$, the proposed decoder is equivalent to the standard iterative decoder without ZP consideration.

5.3.3   EXIT Chart Analysis of CZP and SZI

In order to investigate the effectiveness of the proposed ZP-aware decoding approaches with trellis pruning, we analyze the LLR convergence behavior by the EXIT chart [78].
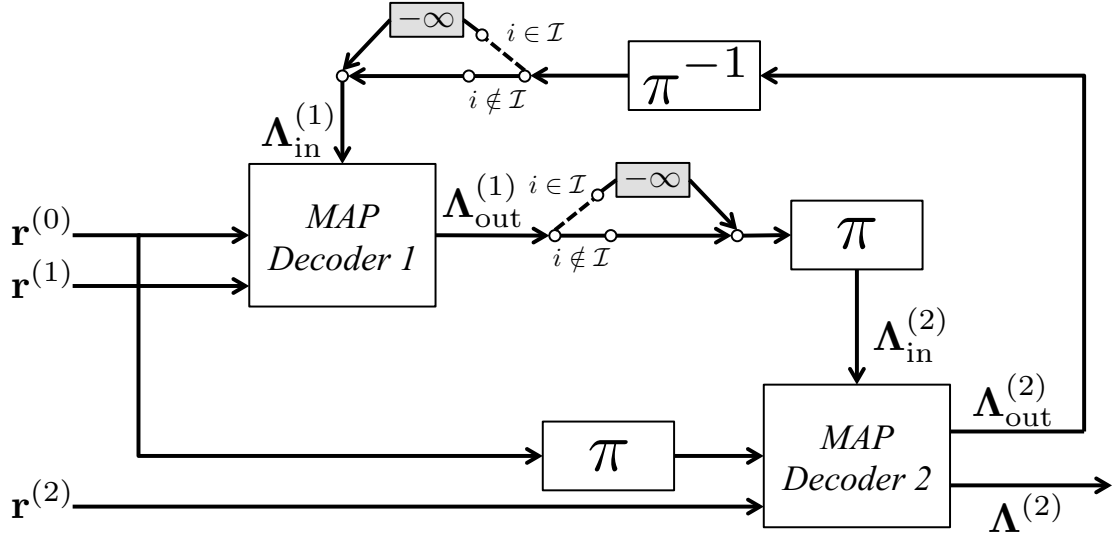
Figure 5.6: A block diagram of the proposed iterative decoding based on MAP decoders.

The EXIT chart is a graph-based method for performance analysis of iterative decoders. For each MAP decoder $k \in \{1,2\}$, let $I_{\text{in}}^{(k)}$ denote the mutual information between the information bits $\{s_i\}$, or equivalently, $\{u_i\}$ where $i \notin \mathcal{I}$, and the provided *a priori* LLR $\Lambda_{\text{in}}^{(k)}$. Likewise, let $I_{\text{out}}^{(k)}$ denote the mutual information between $\{u_i\}$ ($i \notin \mathcal{I}$) and the obtained output extrinsic LLR $\Lambda_{\text{out}}^{(k)}$ after the MAP decoding. The EXIT chart describes the relationship between $I_{\text{in}}^{(k)}$ and $I_{\text{out}}^{(k)}$.

▷ *Analysis of the Proposed Iterative Decoder*

Following [78], we randomly generate the *a priori* LLR $\Lambda_{\text{in}}^{(k)}(u_i)$ that is input to each MAP decoder for each bit $\{u_i\}$, where $i \notin \mathcal{I}$, and obtain the corresponding *a posteriori* LLR $\{\Lambda_{\text{out}}^{(k)}(u_i)\}$.

Let $p_\Lambda(\lambda|u = b)$ denote the conditional probability density function (pdf) of the LLR given that the information bit u is equal to $b \in \{0,1\}$. The mutual information between the information bit and the resulting LLR can be expressed in general as

$$
I = \frac{1}{2} \sum_{b \in \{0,1\}} \int_{-\infty}^{+\infty} p_\Lambda(\lambda|u_i = b)
$$
$$
\cdot \log_2 \left[\frac{2p_\Lambda(\lambda|u_i = b)}{p_\Lambda(\lambda|u_i = 0) + p_\Lambda(\lambda|u_i = 1)} \right] d\lambda, \text{ for } i \notin \mathcal{I}. \tag{5.9}
$$

In the above expression, by the notation u_i with $i \notin \mathcal{I}$ we emphasize the fact that the mutual information in our analysis is derived from the conditional pdf of only the information sequence, i.e., *excluding the zero padding bits* even in the case of ZP cases.

▷ *Results*

As a numerical example, we employ the turbo code with an interleaver size $K = 10^5$, where the S-random interleaver with the spreading parameter $S = \sqrt{K/2} \simeq 223$ is adopted. The two identical RSC encoders with the constraint length of 5 are used, and its generator polynomials are expressed as $g_0 = (37)_8$ for the feed-forward polynomial and $g_1 = (21)_8$ for the feedback polynomial.

Figure 5.7(a) shows the EXIT charts of the proposed iterative decoder with CZP,
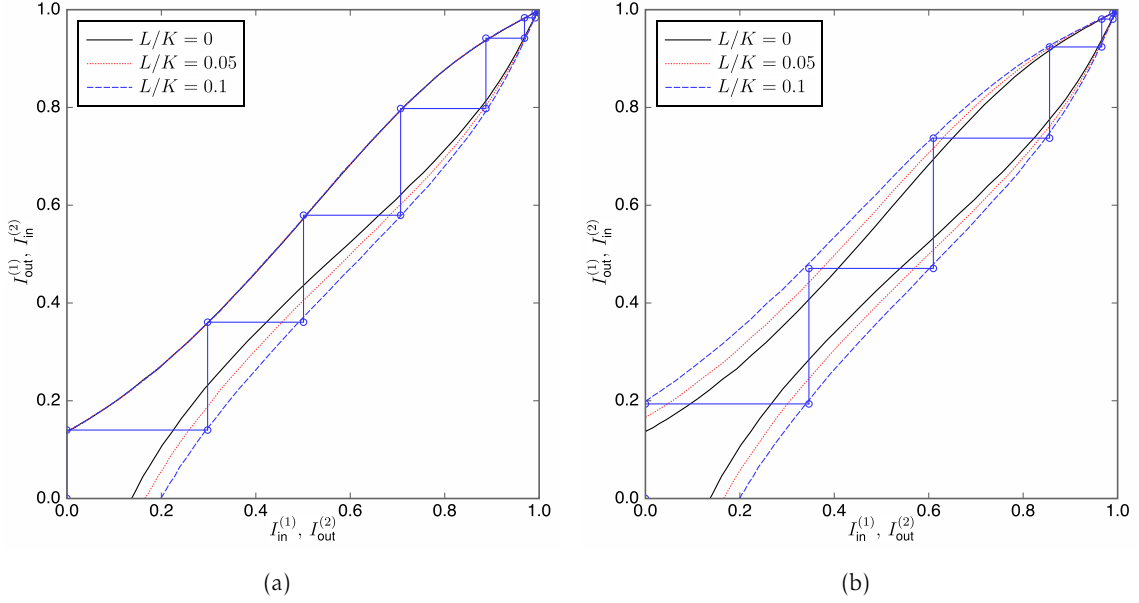
Figure 5.7: The EXIT charts with $L/K = 0$, 0.05, and 0.1 for $E_s/N_0 = -4.6$ dB in the case of (a) CZP and (b) SZI. In the case of CZP, the three curves representing $I_{\text{in}}^{(1)} - I_{\text{out}}^{(1)}$ overlap each other.

where the ratio of the interleaver size and the padding bits, denoted by L/K, is chosen as 0, 0.05, and 0.1, with $L/K = 0$ corresponding to the standard turbo decoder without ZP. The channel SNR is chosen as $E_s/N_0 = -4.6$ dB (0.17 dB in terms of E_b/N_0 for $L/K = 0$, where E_b is the energy per information bit). The mutual information values $I_{\text{in}}^{(k)}$ and $I_{\text{out}}^{(k)}$ correspond to those of the kth MAP decoder, where $k \in \{1, 2\}$. From the result, it is apparent that $I_{\text{out}}^{(1)}$ may not be improved by the proposed decoder, whereas $I_{\text{out}}^{(2)}$ gradually improves as L increases. This suggests that the consecutive insertion of the padding bits does not help improving the overall performance as it only leads to local trellis pruning as observed in Fig. 5.5(a).

On the other hand, Fig. 5.7(b) shows the EXIT charts in the case of SZI with the same parameter setup as those in Fig. 5.7(a). Contrary to the case of CZP, a symmetric expansion of the two curves can be observed, which suggests the effectiveness of the SZI approach. Therefore, the trajectory function in this case leads to convergence with fewer iterations than that in Fig. 5.7(a). Consequently, it is expected that the error performance of SZI should be better than that of CZP, and this performance gap will become significant as the padding bit ratio L/K increases.

5.4   Detection of ZP Sequences

In this section, we first describe the implementation of the padding bit event detector shown in Fig. 5.3. In principle, if the observed number of zero bits in the designated bit positions exceeds some threshold, the detector makes the decision that the zero insertion is applied. An appropriate threshold design is thus critical. An extension of the proposed system to those with variable padding bit lengths is also discussed.

5.4.1 Implementation of Padding Bit Event Detector

As described in Section 5.2.2, it is assumed that the padding bits of length L are generated with probability $p(H_1)$, and the receiver has the perfect knowledge of their positions $\mathcal{I}$. The length of the padding bits is thus represented by a random variable $\ell \in \{0, L\}$.

Let $\mathbf{d} = (d_1, d_2, \ldots, d_L) \in \{0, 1\}^L$ denote the binary vector representing the received and decoded bits corresponding to the zero padding bit positions $\{u_i\}$ with $i \in \mathcal{I}$, and $t = w(\mathbf{d})$ denote the number of zero bits observed in $\mathbf{d}$. Given a threshold U with $0 \le U \le L$ and based on the definitions in (5.2), the proposed detection algorithm is described as

$$t \ge U \Rightarrow \text{Reject } H_0, \text{Accept } H_1,$$
$$t < U \Rightarrow \text{Accept } H_0, \text{Reject } H_1.$$

It follows that when H_1 is accepted in the proposed decoder, we set $\ell = L$, and otherwise we set $\ell = 0$.

5.4.2 Derivation of Threshold U

If the padding bit detector erroneously assumes that the information sequence was zero padded (i.e., the event H_1) while the zero padding was not applied (i.e., the event H_0) at the transmitter, then it may cause severe degradation in terms of error performance by LLR adjusters in the turbo decoder. This incorrect detection is defined as the type I detection error and the probability is denoted as $\Pr\{t \ge U | H_0\}$. On the other hand, if the channel decoder detects the event that the zero padding was not applied (H_0) while it was actually applied (H_1), only the conventional turbo decoder will be invoked and thus may not lead to significant error performance degradation. This error is defined as the type II detection error and that probability is denoted as $\Pr\{t < U | H_1\}$. For the proposed detector to work well, the threshold U should be appropriately determined.

When the event H_0 is chosen at the transmitter, it is reasonable to assume that the bits $\{d_i\}$ in $\mathbf{d}$ are i.i.d. with equal probability for each bit, i.e.,

$$\Pr\{d_i = 0 | H_0\} = \Pr\{d_i = 1 | H_0\} = \frac{1}{2}. \tag{5.10}$$

Consequently, the type I detection error can be expressed as

$$\Pr\{t \ge U | H_0\} = \sum_{k=U}^{L} \binom{L}{k} \frac{1}{2^L}. \tag{5.11}$$

Note that if the length of the actually available padding bits L' is less than L and thus the event H_0 should be chosen at the transmitter, then, rather than randomly choosing 0 or 1 for L' padding bit locations, one can intentionally fill the rest of the null bits by 1 at the transmitter such that we have $\Pr\{d_i = 1 | H_0\} > 1/2$. It is then easy to observe that the resulting type I detection error will become smaller than that of (5.11).

On the other hand, if H_1 is chosen at the transmitter, then we may still assume that the bits $\{d_i\}$ are i.i.d., but with the probability distribution given by

$$\Pr\{d_i = 0 | H_1\} = 1 - \varepsilon, \quad \Pr\{d_i = 1 | H_1\} = \varepsilon, \tag{5.12}$$

where ε is the bit error probability after turbo decoding. It follows that the probability

of the type II detection error is expressed as

$$\Pr\{t < U|H_1\} = 1 - \Pr\{t \geq U|H_1\}$$

$$= 1 - \sum_{k=U}^{L} \binom{L}{k}(1-\varepsilon)^k \varepsilon^{L-k}. \tag{5.13}$$

From (5.11) and (5.13), the probability of the type I detection error $\Pr\{t \geq U|H_0\}$ is a decreasing function with respect to U, while the probability of the type II detection error $\Pr\{t < U|H_1\}$ is a non-decreasing function. Therefore, there is a trade-off between these two detection error probabilities depending on the value of U. Since the type I detection error has more salient effect on the error performance compared to that of type II, we introduce a *significance level* denoted as α, where U is determined such that the condition $\Pr\{t \geq U|H_0\} \leq \alpha$ is satisfied. It follows that a proper U should be determined according to

$$U = \min U' \text{ such that } \Pr\{t \geq U'|H_0\} \leq \alpha. \tag{5.14}$$

The significance level α should be determined to be negligible enough compared to the target frame error rate (FER) performance of the system.

5.4.3   Extension to ZP with Multiple Bit Lengths

We now extend the proposed integrated decoder to the cases where the length of the padding bits can be chosen from a set of multiple patterns. Suppose that the transmitter can choose the length of the zero padding sequences from a finite set $\mathcal{L} = \{L_0, L_1, \ldots, L_M\}$, where without loss of generality we assume that $L_0 = 0$ and they are ordered such that $L_{m-1} < L_m$ for any $m \in \{1, \ldots, M\}$. For each of L_m with $m > 0$, the set of information bit indices, where ZP is applied, is denoted by $\mathcal{I}_m$ and the corresponding decoded bit sequence is denoted by $\mathbf{d}_m = (d_1, \ldots, d_{L_m})$. In this scenario, the condition $\mathcal{I}_{m-1} \subset \mathcal{I}_m$ should be satisfied. We define the following event:

$$H_m : \text{padding bits of length } L_m \text{ are inserted,} \tag{5.15}$$

for $m \geq 1$.

At the receiver, the same detection algorithm as that described in Section 5.2.2 can be applied except that for each m, the detection of H_m should be performed. Under the condition that $\mathcal{I}_{m-1} \subset \mathcal{I}_m$, since the event H_{m-1} is always valid when the event H_m takes place, the number of null bits t can be calculated from only the increment of the decoded bits $\mathbf{d}_m$ relative to $\mathbf{d}_{m-1}$ at the stage of detecting H_m, which we denote by $\mathbf{d}'_m$. Note that the set of the padding bit locations for $\mathbf{d}'_m$ are expressed as $\mathcal{I}'_m = \mathcal{I}_m \cap \bar{\mathcal{I}}_{m-1}$ where $\bar{A}$ denotes a complement of a set $\mathcal{A}$. Therefore, the corresponding set of the threshold values $\mathcal{U} = \{U_1, \ldots, U_M\}$ should be derived based on the number of incremental padding bits $L'_m = L_m - L_{m-1}$ for each m.

In summary, the proposed ZP-aware decoder with unknown ZP sequence for multiple length patterns is described as follows:

Step 1) Set $\ell = 0$ and $m = 0$, and perform I_1 iterations of the proposed decoder.

Step 2) Determine the binary vector $\mathbf{d}'_{m+1}$, and count null bits $t = w(\mathbf{d}'_{m+1})$.

Step 3) If $t \geq U_{m+1}$, increase m by 1 and set $\ell = L_m$. Otherwise, go to Step 5).

Step 4) If $m < M$, go to Step 2).

Step 5) Perform I_2 iterations of the turbo decoding with the number of padding bits $\ell = L_m$.

The additional computational complexity of the above algorithm relative to the conventional turbo decoder is dominated by the *Step 2* above, which corresponds to calculation of the Hamming weight of a vector with at most L_M bits and thus is negligibly small.

5.5 Throughput Analysis

In order to demonstrate the effectiveness of the proposed decoder, we define mathematical expressions of the throughput for the ZP-aware system incorporating the proposed integrated decoder at the receiver and make comparisons with the conventional (ideal ZP-unaware) approaches. For the rest of this chapter, we assume that the number of zero padding bits ℓ is precisely chosen from the candidate set $\mathcal{L}$ such that the reduction of the throughput due to the mismatch of the number of the actual zero padding bits from ℓ is negligible. In this sense, the expressions to be derived may serve as upper bounds for the actual system that has arbitrary zero padding bit length.

5.5.1 Ideal ZP-Unaware System

We start with the derivation of the throughput for the ZP-unaware system where it is assumed that $L' = 0$ with probability $p(H_0) = 1$, i.e., the insertion of dummy bits will never be required. Let ϵ_0 denote the frame error rate (FER) of a given system. Then the maximum throughput of the system can be expressed as

$$T_0 = \frac{(1 - \epsilon_0)K}{N} = (1 - \epsilon_0)R_0, \qquad (5.16)$$

where $R_0 = K/N$ is the effective code rate and is equivalent to the rate of the employed channel code R. This throughput represents the number of the received error-free information bits normalized by the length of the codeword (or a frame length). Since there is no loss associated with the padding bits, we refer to this system as the *ideal* ZP-unaware system.

5.5.2 ZP-Aware Transmitter with a Conventional Decoder

▷ *The Case of* $M = 1$

We now assume that the ZP-aware transmitter with $M = 1$ and the conventional turbo decoder is employed at the receiver, that is, trellis pruning is not implemented and padding bits are just discarded. The throughput of such a ZP-aware transmitter with the conventional decoder can be represented as

$$\tilde{T}_1 = \frac{(1 - \epsilon_0)(K - L)}{N} = (1 - \epsilon_0)R_1, \qquad (5.17)$$

where $R_1 = (K - L)/N$ is an effective code rate when the padding bits are considered as redundant bits and the FER ϵ_0 is the same as that of (5.16).

The average throughput of the probabilistic ZP-aware transmitter with the conventional decoder for the case of $M = 1$ can be expressed as

$$\tilde{T}_{\mathrm{av}} = p(H_0)T_0 + p(H_1)\tilde{T}_1 \tag{5.18}$$

$$= \sum_{m \in \{0,1\}} p(H_m)\tilde{T}_m, \tag{5.19}$$

where $\tilde{T}_0 = T_0$. Since $R_1 < R_0 = R$ when $L > 0$, it follows that $\tilde{T}_1 < \tilde{T}_0$ and thus (5.18) is always lower than (5.16).

▷ *Extension to ZP-Aware Transmitter with Multiple Padding Bit Lengths*

Now we consider the case of multiple padding bit lengths. One can see that the effective code rate of the event H_m for $m \in \{0,\dots,M\}$ is given by

$$R_m = \frac{K - L_m}{N}. \tag{5.20}$$

Similar to (5.18), the average throughput of the ZP-aware transmitter with the conventional decoder for the case of multiple padding bit lengths is given by

$$\tilde{T}_{\mathrm{av}} = \sum_{m=0}^{M} p(H_m)\tilde{T}_m. \tag{5.21}$$

From the above expressions, it is apparent that the insertion of padding bits simply reduces the maximal throughput of the system compared to that of the ZP-unaware system.

5.5.3   ZP-Aware System with the Proposed Decoder

▷ *The Case of $M = 1$*

Next, we examine a possible throughput improvement achieved by the proposed decoder. First, we consider the ZP system with $M = 1$ but the zero padding of length L is always applied, i.e., $p(H_1) = 1$. We further assume that the detection of the padding bit event is always successful. This model serves as our reference when the zero bit insertion is always applied for error correction purpose with an ideal assumption that no detection error occurs. The resulting throughput, achieved by the proposed decoder given that $\ell = L$, can be expressed as

$$T_1 = \frac{(1 - \epsilon_1)(K - L)}{N} = (1 - \epsilon_1)R_1, \tag{5.22}$$

where ϵ_1 is the frame error rate of the system with the proposed decoder and R_1 is the same as that in (5.17).

Let us now consider the proposed adaptive ZP-aware system with $M = 1$. Recall that, in the event that the type II detection error occurs, the standard turbo decoder is employed assuming that $\ell = 0$. In this case, the resulting FER is the same as ϵ_0, while the number of the data bits reduces to $K - L$. It thus follows that the throughput in the event of the type II detection error is the same as $\tilde{T}_1$. On the other hand, the throughput in the event that the type I detection error occurs can be considered as zero since the decoding process is likely to fail with high probability.

Combining the above observations, the average throughput of the proposed ZP-aware system with $M = 1$ incorporating the integrated decoder can be expressed as

$$
\begin{aligned}
T_{\text{ZP}} =\, & p(H_0)(1 - \Pr\{t \geq U|H_0\})T_0 \\
& + p(H_1)(1 - \Pr\{t < U|H_1\})T_1 \\
& + p(H_1)\Pr\{t < U|H_1\}\tilde{T}_1.
\end{aligned}
\tag{5.23}
$$

When the detection is performed perfectly, we have $\Pr\{t \geq U|H_0\} = \Pr\{t < U|H_1\} = 0$ and thus (5.23) reduces to

$$
\begin{aligned}
T_{\text{ZP}}^* &= p(H_0)T_0 + p(H_1)T_1 \\
&= T_0 + p(H_1)(T_1 - T_0),
\end{aligned}
\tag{5.24}
$$

which suggests that the improvement of the throughput by $p(H_1)(T_1 - T_0)$ can be achieved from the conventional approach without ZP if the condition $(T_1 - T_0) > 0$ is satisfied. The simulation results in the next section reveal that this condition is satisfied with practical turbo codes in the wide SNR region of interest and thus the proposed system may outperform the conventional system.

Note that from (5.24), we notice that when the detector is successful, we have

$$
T_{\text{ZP}}^* = \sum_{m \in \{0,1\}} p(H_m)T_m,
\tag{5.25}
$$

which is equivalent to the expectation of the throughputs over the systems with and without padding bits.

▷ *The Case of Multiple Padding Bit Lengths*

Let ϵ_m denote the achievable FER of the system with the event H_m, where the detection of padding bits is assumed to be always successful as $\ell = L_m$. The throughput for the case of H_m given that $\ell = L_m$ is then expressed as

$$
T_m = (1 - \epsilon_m)R_m,
\tag{5.26}
$$

where R_m is the same as (5.20). Similar to (5.25), the expectation of the throughputs when the detection errors are negligible is given by

$$
T_{\text{ZP}}^* = \sum_{m=0}^{M} p(H_m)T_m.
\tag{5.27}
$$

5.5.4 Expression Based on Finite Blocklength Theory

In this subsection, the expression for the achievable throughput is derived based on the finite blocklength theory [79, 80]. According to [79], the rate achieved by the code with blocklength N and the target FER ϵ is well approximated by

$$
R = C - \sqrt{\frac{V}{N}}Q^{-1}(\epsilon) + \frac{\log N}{2N},
\tag{5.28}
$$

where Q^{-1} is the inverse Q-function. The capacity C for an AWGN channel is given by

$$
C = \frac{1}{2}\log(1 + P),
\tag{5.29}
$$

where P is the channel SNR. Furthermore,

$$V = \frac{P}{2} \frac{P+2}{(P+1)^2} (\log e)^2 \tag{5.30}$$

is called as a channel dispersion.

From (5.28), the FER can be expressed as a function of the rate R as

$$\hat{\epsilon}(R) = Q\left(\sqrt{\frac{N}{V}} \left(C - R + \frac{\log N}{2N} \right) \right). \tag{5.31}$$

For each effective code rate R_m, the achievable FER can be given by (5.31) with $R = R_m$. Consequently, when the detection error is negligible, we have the following expression for the theoretical throughput upper bound:

$$T^*_{\mathrm{UB},M} = \sum_{m=0}^{M} p(H_m)(1 - \hat{\epsilon}(R_m)) R_m. \tag{5.32}$$

Note that when $M = 0$ and with $p(H_0) = 1$, the above expression $T^*_{\mathrm{UB},0}$ corresponds to the upper bound of throughput of the ZP-unaware system given in (5.16).

5.6   Numerical Results

The effectiveness of the proposed ZP approaches, together with the integrated ZP detection, is evaluated for several scenarios through extensive computer simulations. The throughput numerically calculated based on the finite blocklength theory is also shown as a reference.

5.6.1   Comparisons of CZP, SZI, and No ZP Cases

We first compare the FER performance of CZP and SZI. In this scenario, the number of the information bits is $K = 4000$ and the single ZP case with $L = 200$ bits is evaluated. We initially consider the case with the event probability $p(H_1) = 1$, i.e., the ZP is always applied. We set the significance level as $\alpha = 10^{-5}$, and the threshold estimated by (5.14) is $U = 131$. The iteration numbers in the MAP decoders are set as $I_1 = 8$ and $I_2 = 12$, which are empirically chosen large enough such that the performance improvement achieved by the proposed detection is clearly observed.

Figure 5.8 compares the FER performance of CZP and SZI. As a reference, the cases without any detection error at the proposed receiver are also plotted, which may serve as the lower bounds that can be achieved by actual turbo codes with trellis pruning and ideal ZP event detection. In the figure, the FER curves calculated by (5.31) are also plotted, where $\hat{\epsilon}(R_0)$ corresponds to that of the ZP-unaware system without any padding bit (i.e., $L = 0$), and $\hat{\epsilon}(R_1)$ corresponds to the ZP-aware system as an ideal reference for SZI or CZP. As we have seen in Section 5.3.3, SZI outperforms CZP by sparse trellis pruning, and the coding gain of SZI relative to the ZP-unaware system is twice as large as that of CZP. The SNR improvement for a given targeted FER between the ZP-unaware system and that of the SZI with perfect detection case is slightly lower than that between $\hat{\epsilon}(R_0)$ and $\hat{\epsilon}(R_1)$ suggested by the finite blocklength theory, and this shows that the coding gain is not fully achieved by the simple trellis pruning. Nevertheless, considering its low cost implementation based on simple insertion of known bits without any further
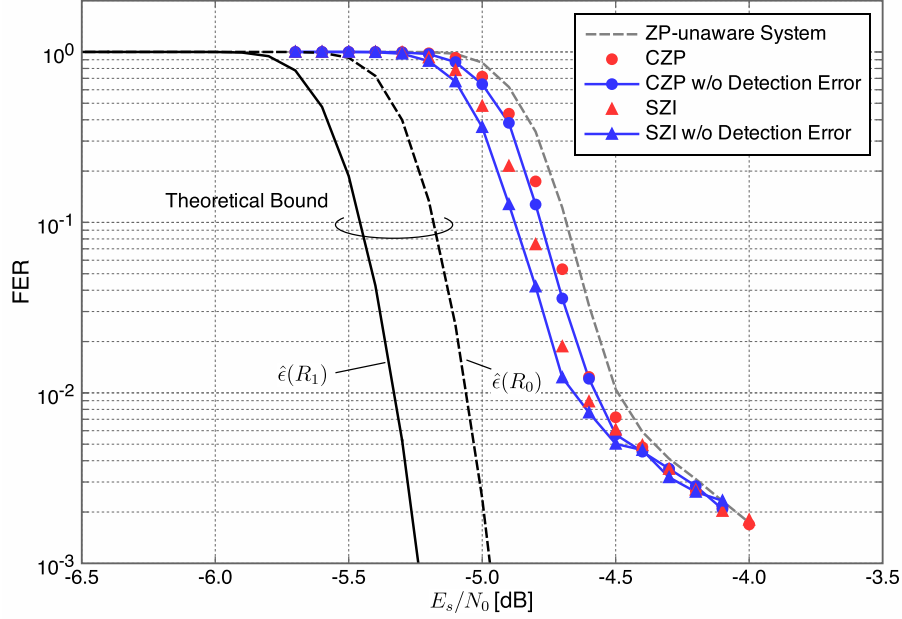
Figure 5.8: Comparison of FER performance for CZP and SZI with $L = 200$, $K = 4000$ and $p(H_1) = 1$, where the cases without any detection error at the proposed decoder and the theoretical bounds of FER for the systems with and without padding bits are also plotted.

modification of the original code structure at the transmitter, the use of trellis pruning may be a practically attractive approach for enhancing system reliability. Furthermore, we observe that the ZP-aware system with the proposed detector achieves FER close to that with perfect detection, which confirms that our design criterion for the event detection is appropriate.

5.6.2 Evaluation of The Throughput Performance for Single Padding Pattern

Next, we investigate the throughput performance of the integrated decoder for the system with the binary detection case, i.e., $\ell \in \{0, L\}$ where $L = 256$ and $K = 4096$. Since the error correcting capability of SZI is always superior to that of CZP, only SZI will be considered in what follows. Setting the significance level as $\alpha = 10^{-5}$, the threshold is estimated as $U = 163$. The iteration numbers in the MAP decoders are set as $I_1 = 8$ and $I_2 = 12$. With reference to the analysis given in Section 5.5, the following three cases are evaluated:

1. the ideal ZP-unaware system, where the throughput is given by T_0 of (5.16),

2. SZI and the proposed decoder incorporating the padding bit detector with $p(H_1) = 0.2$, where the throughput is given by T_{ZP} of (5.23), and

3. the expected throughput with $p(H_1) = 0.2$ ignoring detection error, which is given by T_{ZP}^* of (5.24).

In addition, the corresponding upper bounds based on the finite blocklength analysis (5.32) are also shown as a reference. Note that the above throughput values are obtained by the *hybrid simulation* where only the values for ϵ_0 and ϵ_1 are replaced by those
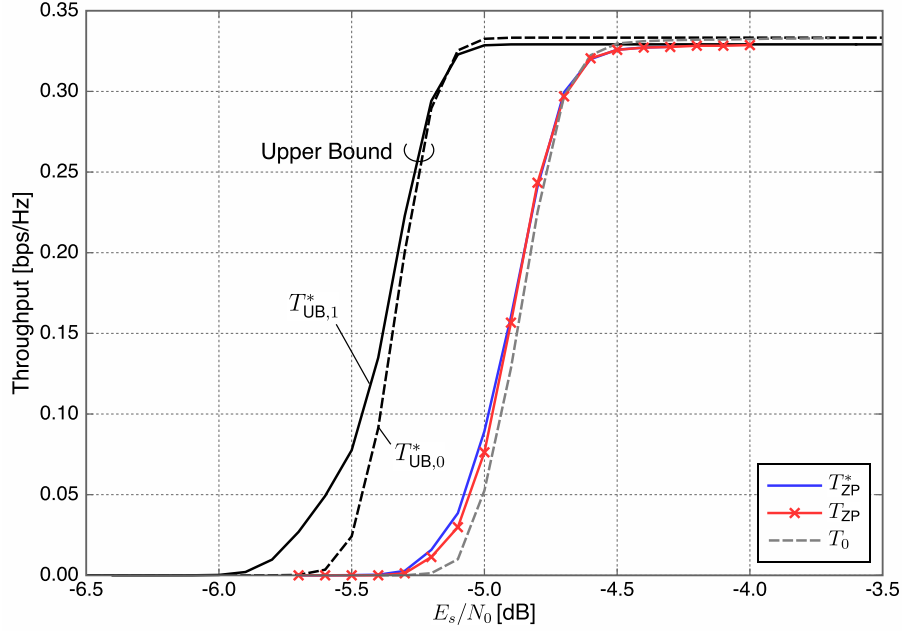
Figure 5.9: Throughput comparison for SZI with $L = 256$, $K = 4096$ and $p(H_1) = 0.2$ with and without detection error and the ZP-unaware system. The upper bounds for ZP-aware and -unaware systems are also plotted.

obtained through simulations and the resulting throughputs are calculated using the expressions in (5.16)–(5.24). We employ E_s/N_0 as a parameter of the channel SNR, rather than E_b/N_0, recalling that our purpose is based on the assumption that the proposed decoding technique is valid only when the padding bits are inserted, and since they do not carry any useful information, if unexploited they are simply discarded at the receiver.

Figure 5.9 illustrates the comparison of throughput. First of all, we observe that T_{ZP}^* outperforms T_0 in the low SNR region, even though padding bits reduce the number of information bits carried by one transmission frame. Therefore, the probabilistic trellis pruning improves the total throughput by the fact that $(T_1 - T_0) > 0$ in (5.24). By comparing T_{ZP} and T_{ZP}^*, slight difference is observed in the low SNR region because of the detection error of padding bits. Nevertheless, the results suggest that the precise detection of padding bits can be achieved in the middle SNR range. It is thus apparent that the proposed padding bit event detector effectively works such that the reliability of communication is improved, thereby improving the throughput.

By comparing the throughput of the ZP-unaware system T_0 and those of the other ZP-aware systems, we observe some degradation of the maximum throughput for the latter case in high channel SNR region, and this is caused by the loss of the throughput due to the padding bit insertion, whereas such loss is not observed in the ZP-unaware system by the ideal assumption that padding bits need not be inserted. This characteristic is also observed in the analytical throughput upper bounds $T_{UB,0}^*$ and $T_{UB,1}^*$. Nevertheless, due to the improvement in error correction performance, the benefit of the proposed approach even over the ideal ZP-unaware system is significant when they are compared in the low SNR region.
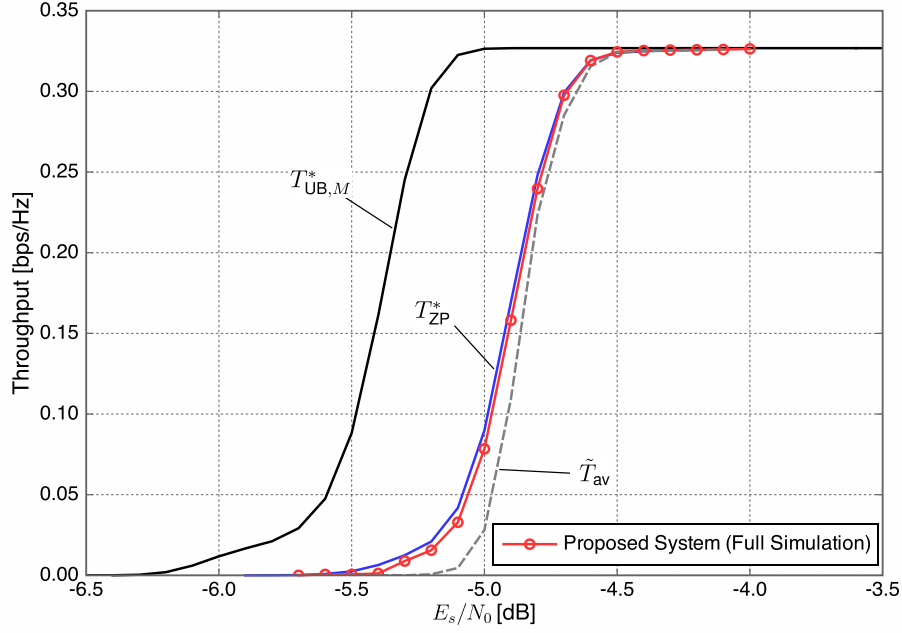
Figure 5.10: The throughput comparisons of the ZP-aware system with multiple padding bit patterns with $M = 4$, where $\mathcal{L} = \{0, 64, 128, 256, 512\}$ for the conventional and proposed decoders. The throughput of the proposed decoder system is obtained by full simulations. The theoretical upper bound is also plotted.

5.6.3   Multiple Padding Bit Patterns

Finally, aiming at more practical transmission scenarios, the throughput of the proposed system with $K = 4096$ is evaluated, where the multiple padding bit patterns defined by $\mathcal{L} = \{L_0, L_1, L_2, L_3, L_4\} = \{0, 64, 128, 256, 512\}$ are employed. In this example, we choose the *a priori* probabilities of the padding event as $p(H_0) = 0.5$, $p(H_1) = 0.25$, $p(H_2) = 0.125$, $p(H_3) = 0.0625$, and $p(H_4) = 0.0625$. The corresponding set of the thresholds is found as $\mathcal{U} = \{50, 50, 89, 163\}$ in the case of significance level $\alpha = 10^{-5}$. For comparison, the ZP-aware system with the conventional decoder is also evaluated where padding bits are inserted with the same probabilities at the transmitter but the standard turbo decoding is always employed at the receiver. The corresponding throughput is given by $\tilde{T}_{\mathrm{av}}$ of (5.21). In this scenario, the throughput of the proposed system is also evaluated by *full simulation*; we actually calculate the ratio of the total number of effective information bits (i.e., excluding padding bits) in correctly decoded frames to the total number of codewords in the entire frame.

Figure 5.10 illustrates the throughput performance of the proposed system by full simulation, the average throughput of the proposed system with ideal padding event detection given by T_{ZP}^*, and the average throughput of the ZP-aware system with the conventional decoder $\tilde{T}_{\mathrm{av}}$. The theoretical upper bound for the throughput based on the finite blocklength analysis i.e., $T_{\mathrm{UB},M}^*$, is also plotted as a reference, which exhibits similar tendency to the actual performance achieved by the practical systems compared here.

The simulation results demonstrate that compared to the conventional ZP-aware transmission approach, the throughput can be improved for any channel SNR by the proposed decoding method. In fact, we observe that the proposed decoder achieves performance close to that achieved by the ideal error-free event detector. From these

results, it can be expected that the proposed integrated decoder can operate effectively with improved throughput in practical transmission systems where the length of the ZP is not known *a priori* at the receiver.

5.7   Conclusion

In this chapter, we have proposed a novel approach for improving the throughput of the system by exploiting the padding bits inherent in the transmitted sequence as a consequence of the mismatch between the information length and the length of the data received from the upper layer. This improvement is achieved by trellis pruning in the process of the iterative decoder. Upon the allocation of ZP bits, it has been shown that SZI is more effective than CZP, which is supported by the EXIT chart analysis and FER performance. Furthermore, we have proposed an integrated decoder that detects the length of the ZP without side information at the receiver. The results have shown that the proposed integrated decoder improves the throughput of the overall system compared to the conventional system, where padding bits are simply discarded at the receiver.

C H A P T E R 6
Conclusion

In this thesis, we have proposed several frameworks exploiting statistical properties observed in any modern source codec. Our approach is based on the cross-layer design of the OSI reference model, where the source coding process is equivalent to the upper layer and the channel coding is implemented in the physical layer.

In Chapter 1, we have given a general overview of the system combining source and channel coding and reviewed several conventional studies on JSCC. The problems inherent in them were identified, which motivated the subsequent studies in the thesis.

In Chapter 2, we have described the principles of source and channel coding especially aiming at multimedia transmission. We have also demonstrated the image transmission system as an example.

Chapter 3 reveals the achievable R-D function by JSCC system using a subband coded audio signal and RCPC codes. It has been shown that JSCC scheme can achieve lower R-D function than SSCC by taking into account the CSI.

In Chapter 4, we have proposed a new approach for designing the UEP system with graceful degradation for multimedia broadcasting service of existing source codec. To this end, a new measure that quantifies the average quality of the videos received by uniformly distributed users has been introduced. It has been shown that the MLC with UEP property can achieve the best performance among other comparable systems with and without UEP in terms of average received quality within a coverage area. The results for individual video sequences have revealed that the system design based on video database is applicable to most scenes in digital broadcasting.

In Chapter 5, we have proposed a novel approach for improving the throughput of the system by exploiting the padding bits inherent in the transmitted sequence as a consequence of mismatch caused by entropy coding in the upper layer. Furthermore, we have proposed an integrated decoder that detects the length of the ZP without side information at the receiver. The results have shown that the proposed decoding technique can improve throughput performance compared to the conventional zero-padded system where padding bits are discarded at the receiver.

Remaining issues that should be considered in future are as follows:

- In Chapter 4, only a quasi-static channel state is considered, but in practice, time-variant components such as multipath fading should be considered. In this case, it is hard to numerically estimate $D(R)$ since the distortion of a specific P-picture depends on all decoding results of the previous pictures in the corresponding group of pictures (GOP) in MPEG-4, and this dependence is even stronger in H.264/AVC or other modern codecs with B-pictures. Another possible approach is to directly estimate an R-A function by experimentally transmitting videos at every distance from the transmitter, which is also a laborious task. Therefore, the extension of this issue is left as future work.

- The extension of the MLC with other source codec should be considered in Chapter 4. Specifically for H.264/SVC [46], the quality scalability is suitable for our

assumption of employing graceful degradation. For *coarse-grain quality scalability (CGS)*, where the bit-rate allocation for each layer is inflexible, the cumulative rate R in Fig. 4.3 can be considered as the number of layers and it takes positive integer values. As for *medium-grain quality scalability (MGS)*, which has the flexibility of adjusting bit-rate for each layer, the bit-rate can replace R as positive integer values.

- In Chapter 4, the bit-rates of sources can be further taken into consideration for analyzing robustness against bit errors in terms of JSCC.

- In Chapter 5, we have assumed that the padding bit lengths are chosen only from a few set. However, in practical systems, the number of padding bits could be arbitrary. It should be revealed whether the detection of padding bits is possible in such a framework. If not, the limitation of the detection capability should be discussed.

Bibliography

[1] I.-T. Rec., *H.264 and ISO/IEC 14496-10 (MPEG-4 AVC), Advanced video coding for generic audiovisual services*, 2003.

[2] T. Wiegand, "Overview of the H. 264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560 –576, 2003.

[3] "YouTube." [Online]. Available: https://www.youtube.com

[4] R. I.-T. H.265, *ISO/IEC 23008-2: high efficiency video coding*, 2013.

[5] G. J. Sullivan, J. R. Ohm, W. J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, 2012.

[6] C. E. Shannon, "Communication in the presence of noise," *Proc. IRE*, vol. 37, no. 1, pp. 10–21, 1949.

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

[8] R. G. Gallager, "Low-density parity-check codes," *IEEE Trans. Inf. Theory*, vol. 8, pp. 21–28, 1962.

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

[10] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, Jul. 1948.

[11] S. Vembu, S. Verdu, and Y. Steinberg, "The source-channel separation theorem revisited," *IEEE Trans. Inf. Theory*, vol. 41, no. 1, pp. 44–54, 1995.

[12] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2005.

[13] M. Jeanne, J. C. Carlach, and P. Siohan, "Joint source-channel decoding of variable-length codes for convolutional codes and turbo codes," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 10–15, 2005.

[14] M. Fresia, F. Pérez-Cruz, H. V. Poor, and S. Verdú, "Joint source and channel coding," *IEEE Signal Process. Mag.*, vol. 27, no. 6, pp. 104–113, 2010.

[15] Y. Hu, J. Garcia-Frias, and M. Lamarca, "Analog joint source-channel coding using non-linear curves and MMSE decoding," *IEEE Trans. Commun.*, vol. 59, no. 11, pp. 3016–3026, 2011.

[16] O. Fresnedo, F. J. Vazquez-Araujo, J. Garcia-Frias, M. Gonzalez-Lopez, and L. Castedo, "Comparison between analog joint source-channel coded and digital BICM systems," in *Proc. IEEE Int. Conf. Commun.*, 2011, pp. 1–5.

[17] M. J. Ruf and J. W. Modestino, "Operational rate-distortion performance for joint source and channel coding of images," *IEEE Trans. Image Process.*, vol. 8, no. 3, pp. 305–320, Jan. 1999.

[18] V. Stankovic, R. Hamzaoui, and Z. Xiong, "Efficient channel code rate selection algorithms for forward error correction of packetized multimedia bitstreams in varying channels," *IEEE Trans. Multimed.*, vol. 6, no. 2, pp. 240–248, 2004.

[19] H. Kim, R. Annavajjala, P. C. Cosman, and L. B. Milstein, "Source-channel rate optimization for progressive image transmission over block fading relay channels," *IEEE Trans. Commun.*, vol. 58, no. 6, pp. 1631–1642, Jun. 2010.

[20] F. Zhai, Y. Eisenberg, T. N. Pappas, R. Berry, and A. K. Katsaggelos, "Rate-distortion optimized hybrid error control for real-time packetized video transmission," *IEEE Trans. Image Process.*, vol. 15, no. 1, pp. 40–53, 2006.

[21] H. Shi, P. K. M. Ho, and V. Cuperman, "Combined speech and channel coding for mobile radio communications," *IEEE Trans. Veh. Technol.*, vol. 43, no. 4, pp. 1078–1087, 1994.

[22] A. Bernard, X. Liu, R. D. Wesel, and A. Alwan, "Speech transmission using rate-compatible trellis codes and embedded source coding," *IEEE Trans. Commun.*, vol. 50, no. 2, pp. 309–320, 2002.

[23] H. Houas, I. Fijalkow, and C. Baras, "Robust and extreme unequal error protection scheme for the transmission of scalable data over OFDM systems," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2008, pp. 2973–2976.

[24] A. Khalifeh and H. Yousefi'zadeh, "Optimal Audio Transmission Over Error-Prone Wireless Links," *IEEE Trans. Multimed.*, vol. 12, no. 3, pp. 204–214, 2010.

[25] H. Imai and S. Hirakawa, "A new multilevel coding method using error-correcting codes," *IEEE Trans. Info. Theory*, vol. 23, pp. 371–377, 1977.

[26] U. Wachsmann, R. F. H. Fischer, and J. B. Huber, "Multilevel codes: Theoretical concepts and practical design rules," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1361–1391, Jul. 1999.

[27] K. Fazel and M. J. Ruf, "Combined multilevel coding and multiresolution modulation," in *Proc. IEEE Int. Conf. Commun.*, vol. 2, 1993, pp. 1081–1085.

[28] R. H. Morelos-Zaragoza, M. P. C. Fossorier, S. Lin, and H. Imai, "Multilevel coded modulation for unequal error protection and multistage decoding. I. Symmetric constellations," *IEEE Trans. Commun.*, vol. 48, no. 2, pp. 204–213, 2000.

[29] M. Isaka, M. P. C. Fossorier, R. H. Morelos-Zaragoza, S. Lin, and H. Imai, "Multilevel coded modulation for unequal error protection and multistage decoding. II. Asymmetric constellations," *IEEE Trans. Commun.*, vol. 48, no. 5, pp. 774–786, 2000.

[30] S.-H. Chang, M. Rim, P. C. Cosman, and L. B. Milstein, "Optimized unequal error protection using multiplexed hierarchical modulation," *IEEE Trans. Inf. Theory*, vol. 58, no. 9, pp. 5816–5840, Sep. 2012.

[31] M. O. Zamkotsian, K. P. Peppas, F. Lazarakis, and P. G. Cottis, "Multilevel spatial hierarchical modulation: An efficient scheme for unequal error protection under Rician fading," *IEEE Trans. Veh. Technol.*, vol. 64, no. 11, pp. 5177–5186, 2015.

[32] H. Meric and J. M. Piquer, "DVB-S2 spectrum efficiency improvement with hierarchical modulation," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2014, pp. 4331–4336.

[33] A. C. Bovik, *Handbook of Image and Video Processing*, 2nd ed.   Elsevier Academic Press, 2005.

[34] J. Huang, K. Peng, S. Member, and C. Pan, "Scalable video broadcasting using bit division multiplexing," *IEEE Trans. Broadcast.*, vol. 60, no. 4, pp. 701–706, 2014.

[35] Yu Wang, Lap-Pui Chau, and K.-H. Yap, "Bit-rate allocation for broadcasting of scalable video over wireless networks," *IEEE Trans. Broadcast.*, vol. 56, no. 3, pp. 288–295, 2010.

[36] W. Yao, L. P. Chau, and S. Rahardja, "Joint rate allocation for statistical multiplexing in video broadcast applications," *IEEE Trans. Broadcast.*, vol. 58, no. 3, pp. 725–728, 2012.

[37] S. Nazir, D. Vukobratovic, V. Stankovic, I. Andonovic, K. Nybom, and S. Gronroos, "Unequal error protection for data partitioned H.264/AVC video broadcasting," *Multimed. Tools Appl.*, vol. 74, no. 15, pp. 5787–5809, 2015.

[38] I.-T. R. T.81, *ISO/IEC 10918-1: Information technology — Digital compression and coding of continuous-tone still images — requirements and guidelines*, 1994.

[39] I. W. International Standard, *ISO/IEC 13818-3*, *Information technology – generic coding of moving pictures and associated audio information – Part 3: Audio*.   Geneva, Switzerland: International Organization for Standardization, 1998.

[40] ——, *ISO/IEC 14496-3: Information technology — Coding of audio-visual objects — Part 3: Audio*, 1999.

[41] A. Said and W. A. W. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 3, pp. 243–250, 1996.

[42] A. McCree, K. Truong, E. B. George, T. P. Barnwell, and V. Viswanathan, "A 2.4 kbit/s MELP coder candidate for the new US Federal Standard," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, vol. 1, 1996, pp. 200–203.

[43] P. Noll, "MPEG digital audio coding," pp. 59–81, 1997.

[44] C. Christopoulos, A. Skodras, and T. Ebrahimi, "The JPEG2000 still image coding system: An overview," *IEEE Trans. Consum. Electron.*, vol. 46, no. 4, pp. 1103–1127, 2000.

[45] Mozilla,   "Mozilla   JPEG   Encoder   Project."   [Online].   Available: https://github.com/mozilla/mozjpeg

[46] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the Scalable Video Coding Extension of the H . 264 / AVC Standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1103–1120, 2007.

[47] W. Li, "Overview of fine granularity scalability in MPEG-4 video standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 3, pp. 301–317, 2001.

[48] M. Domański, A. Łuczak, and S. Maćkowiak, "Spatio-temporal scalability for MPEG video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 7, pp. 1088–1093, 2000.

[49] T. Wiegand, L. Noblet, and F. Rovati, "Scalable Video Coding for IPTV Services," *IEEE Trans. Broadcast.*, vol. 55, no. 2, pp. 527–538, 2009.

[50] G. J. Conklin and S. S. Hemami, "A comparison of temporal scalability techniques," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 6, pp. 909–919, 1999.

[51] V. K. Goyal, "Multiple description coding: Compression meets the network," *IEEE Signal Process. Mag.*, vol. 18, no. 5, pp. 74–93, 2001.

[52] D. Huffman, "A method for the construction of minimum-redundancy codes," *Proc. IRE*, vol. 40, no. 9, pp. 1098–1001, 1952.

[53] Y. Takishima, M. Wada, and H. Murakami, "Reversible variable length codes," *IEEE Trans. Commun.*, vol. 43, no. 2/3/4, pp. 158–162, 1995.

[54] S. Kumar and L. Xu, "RVLC decoding scheme for improved data recovery in MPEG-4 video coding standard," *Real-Time Imaging*, vol. 10, no. 5, pp. 315–323, 2004.

[55] J. Hagenauer, "Rate-compatible punctured convolutional codes (RCPC codes) and their applications," *IEEE Trans. Commun.*, vol. 36, no. 4, pp. 389–400, 1988.

[56] J. Hagenauer, N. Seshadri, and C. E. W. Sundberg, "The performance of rate-compatible punctured convolutional codes for digital mobile radio," *IEEE Trans. Commun.*, vol. 38, no. 7, pp. 966–980, Jul. 1990.

[57] R. Gallager, "Low-density parity-check codes," *IRE Trans. Inf. Theory*, 1962.

[58] S. Benedetto and G. Montorsi, "Design of parallel concatenated convolutional codes," *IEEE Trans. Commun.*, vol. 44, no. 5, pp. 591–600, 1996.

[59] G. Ungerboeck, "Channel coding with multilevel / phase signals," *IEEE Trans. Inf. Theory*, vol. IT-28, pp. 55—-67, 1982.

[60] Y. Watanabe and H. Ochiai, "An optimal code rate design of multilevel coded modulation for multimedia data transmission employing unequal error protection," in *Proc. IEEE Glob. Commun. Conf.*, Dec. 2014, pp. 1423–1428.

[61] P. Westerink, J. Biemond, and D. Boekee, "An optimal bit allocation algorithm for sub-band coding," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 1988, pp. 757–760.

[62] Y. Watanabe and H. Ochiai, "A novel design and modeling of UEP-based compressed video broadcasting with multilevel coded modulation," *IEEE Trans. Broadcast.*, vol. 62, no. 3, pp. 598–609, Sep. 2016.

[63] A. Goldsmith, *Wireless Communications*. Cambridge Univ. Press, 2005.

[64] T. S. Rappaport, *Wireless Communications: Principles and Practice*.   Prentice Hall, Upper Saddle River, NJ, 1996.

[65] M. F. Sabir, R. W. Heath, and A. C. Bovik, "Joint source-channel distortion modeling for MPEG-4 video." *IEEE Trans. Image Process.*, vol. 18, no. 1, pp. 90–105, Jan. 2009.

[66] International Standard, *ISO/IEC 14496-2: Information technology — Coding of audio-visual objects — Part 2: Visual.*   ISO/IEC/JTC1/SC29 WG11, 2001.

[67] "Xiph.org   ::   Derf's   Test   Media   Collection."   [Online].   Available: https://media.xiph.org/video/derf/

[68] "FFmpeg." [Online]. Available: https://www.ffmpeg.org/

[69] A. S. Tanenbaum and D. Wetherall, *Computer Networks*, 5th ed.   Pearson Prentice Hall, 2011.

[70] D. Skordoulis, Q. Ni, H. H. Chen, A. P. Stephens, C. Liu, and A. Jamalipour, "IEEE 802.11n MAC frame aggregation mechanisms for next-generation high-throughput WLANs," *IEEE Wirel. Commun.*, vol. 15, no. 1, pp. 40–47, 2008.

[71] K. Ishibashi, H. Ochiai, and R. Kohno, "Embedded forward error control technique (EFECT) for low-rate but low latency communications," *IEEE Trans. Wirel. Commun.*, vol. 7, no. 5, pp. 1456–1460, May 2008.

[72] S. ten Brink, "Code doping for triggering iterative decoding convergence," in *Proc. IEEE Int. Symp. Inf. Theory*, vol. 54, no. 3, 2001, p. 235.

[73] S. Pfletschinger and F. Sanzi, "Error floor removal for bit-interleaved coded modulation with iterative detection," *IEEE Trans. Wirel. Commun.*, vol. 5, no. 11, pp. 3174–3181, 2006.

[74] S. V. Maiya and T. E. Fuja, "Cooperation via trellis pruning," *IEEE Trans. Commun.*, vol. 59, no. 6, pp. 1563–1569, Jun. 2011.

[75] O. Collins and M. Hizlan, "Determinate state convolutional codes," *IEEE Trans. Commun.*, vol. 41, no. 12, pp. 1785–1794, 1993.

[76] C. Wang and C. Chao, "Path-compatible pruned convolutional (PCPC) codes," *IEEE Trans. Commun.*, vol. 50, no. 2, pp. 213–224, 2002.

[77] C.-h. Wang and Y. Lin, "Combined puncturing and path pruning for convolutional codes and the application to unequal error protection," *IEEE Trans. Commun.*, vol. 56, no. 9, pp. 1385–1389, 2008.

[78] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. Commun.*, vol. 49, no. 10, pp. 1727–1737, 2001.

[79] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, 2010.

[80] M. Hayashi, "Information spectrum approach to second-order coding rate in channel coding," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 4947–4966, 2009.

[81] Y. Watanabe and H. Ochiai, "Exploiting padding bits for improvement of channel decoder performance," in *Proc. IEEE Glob. Commun. Conf.*, 2015.

This page intentionally left blank

Publications

Journal Papers

[J-1] Y. Watanabe and H. Ochiai, "A novel design and modeling of UEP-based compressed video broadcasting with multilevel coded modulation," *IEEE Trans. Broadcast.*, vol. 62, no. 3, pp. 598–608, Sep. 2016.

[J-2] Y. Watanabe and H. Ochiai, "A novel channel decoder exploiting inherent padding bits for fixed-length frame transmission," *IEEE Trans. Commun.*, (in preparation).

[J-3] Y. Watanabe and H. Ochiai, "Global motion model using accelerometer and gyroscope," *EURASIP Journal on Image and Video Process.*, (in preparation).

International Conference Papers

[I-1] Y. Watanabe and H. Ochiai, "A general rate-distortion model and its optimization for multimedia joint source-channel coding," in *Proc. IEEE International Workshop on Multimedia Signal Processing (MMSP)*, Banff, Alberta, Canada, Sep. 2012 (On-going work poster).

[I-2] Y. Watanabe and H. Ochiai, "An optimal code rate design of multilevel coded modulation for multimedia data transmission employing unequal error protection," in *Proc. IEEE Global Commun. Conf. (GLOBECOM'14)*, Austin, TX, Dec. 2014.

[I-3] Y. Watanabe and H. Ochiai, "Exploiting padding bits for improvement of channel decoder performance," in *Proc. IEEE Global Commun. Conf. (GLOBECOM'15)*, San Diego, CA, Dec. 2015.

Domestic Conference Papers

[D-1] 渡辺, 落合, "サブバンド符号化を用いたオーディオ情報源に対する不均一誤り訂正符号化方式," 電子情報通信学会ソサイエティ大会, A-4-13, 北海道大, 2011 年 9 月.

[D-2] 渡辺, 落合, "雑音のある通信路に対するサブバンド符号化を用いたオーディオ伝送のための 不均一誤り訂正符号化方式," 情報理論とその応用シンポジウム (SITA'11), 岩手県鴬宿, 2011 年 12 月.

[D-3] 渡辺, 落合, "情報源・通信路結合符号のための Rate-Distortion モデルの最適化," 電子情報通信学会ソサイエティ大会, A-6-10, 富山大, 2012 年 9 月.

[D-4] 渡辺, 落合, "リアルタイムマルチメディア伝送のための不均一誤り訂正符号化方式," 電子情報通信学会総合大会, A-6-13, 岐阜大, 2013 年 3 月.

[D-5] 渡辺, 落合, "マルチレベル符号化を用いた階層画像伝送システム," 電子情報通信学会ソサイエティ大会, A-4-15, 福岡工大, 2013 年 9 月.

[D-6] 渡辺, 落合, "マルチレベル符号化を用いたスケーラブル画像伝送システムの設計," 電子情報通信学会総合大会, A-6-6, 新潟大, 2014 年 3 月.

[D-7] 渡辺, 落合, "パディングビットを利用した誤り訂正符号の復号特性の改善に関する検討,"
電子情報通信学会総合大会, A-2-8, 九州大, 2016 年 3 月.