Joint Design of Multi-resolution Codes and Intra/Inter-layer Network Coding

by

Tong Wang

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Science in Computer Science and Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2012

Author ....
Department of Electrical Engineering and Computer Science
May 23, 2012

Certified by ....
Muriel Médard
Professor
Thesis Supervisor

Certified by ....
Lizhong Zheng
Associate Professor
Thesis Supervisor

Accepted by ....
Leslie A. Kolodziejski
Chair of the Committee on Graduate Students

Joint Design of Multi-resolution Codes and Intra/Inter-layer Network Coding

by

Tong Wang

Submitted to the Department of Electrical Engineering and Computer Science
on May 23, 2012, in partial fulfillment of the
requirements for the degree of
Master of Science in Computer Science and Engineering

Abstract

In this thesis, we study the joint design of multi-resolution (MR) coding and network coding. The three step coding process consists of MR source coding, layer coding and multi-stream coding. The source coding considers the allocation of codeword length to the two layers and use base layer percentage as the parameter. In the network coding model, we present two coding schemes, intra-layer and inter-layer network coding and use two elements as design parameters: 1) Redundancy at different layer. 2) Whether to code within a layer or across layers. Two metrics, average distortion and total redundancy, are used to characterize quality and efficiency. We define the notion of redundancy plane and provide coding strategies based on this plane. We draw two important conclusions: 1) Inter-layer coding always achieves at least the same performance as intra-layer coding. The mixing of layers never hurts from this perspective. 2) The optimal performance for inter-layer coding is independent of the base layer percentage. Besides the two observations, we also point out that in some region, inter-layer coding can be replaced by changing the parameters without harming the quality or efficiency. Our coding strategies give guidelines for choosing between the two coding schemes in different situations and how to replace inter-layer coding with intra-layer coding while achieving the same performance in terms of quality and efficiency.

Thesis Supervisor: Muriel Médard
Title: Professor

Thesis Supervisor: Lizhong Zheng
Title: Associate Professor

Acknowledgments

First and foremost, I would like to thank my research advisors, Professor Muriel Médard and Professor Lizhong Zheng, for their insightful guidance and continuous support. I'm very grateful to Professor Médard for the invaluable mentorship through the two years' research. I appreciate every advice she offered on not only research, but a long term self development. I sincerely thank Professor Lizhong Zheng, for his patience and tremendous help during the whole process of research. I will always remember the encouragement and kindness I received from him.

I would also like to thank my officemates, Weifei Zeng, Flavio du Pin Calmon, Muriel Rambeloarison and Jason Cloud, who have been a constant source of support and laughter. I have truly enjoyed every interaction with you. I learnt a lot from our conversations on not just research, but a lot of things beyond that. The happy moments in this office are one of the most precious gifts I received at MIT and I will always appreciate them. My thanks also go to all other members in Network Coding and Reliable Communications Group, who made the group like a big family. Thank you all.

To all my other friends in MIT, thank you for making my two year's life at MIT an enjoyable journey. I'm so lucky to getting to know you and experience a new life with you. To my friends in China, thank you for your support and constant care no matter where I am. Specially to Huiying Li, there's no word that can express how lucky I feel to have such a wonderful friend. Thank you for always understanding me, and always being there for me.

Finally, I would like to thank my parents, Yumin Wang, and Aihua Xu, for their unconditional love and support. I could never have come this far without you. You trust me more than I can trust myself. Whenever and wherever I am, your love is my source of strength and courage. This thesis is dedicated to you.

Contents

List of Figures

Chapter 1

Introduction

As both media content (e.g. Youtube videos) over the Internet and wireless devices (e.g. smartphones) become increasingly popular, scalable delivery of rich media content over wireless heterogeneous links is quickly becoming an application with increasing importance. A widely used approach to delivering video is to exploit source-coding techniques such as Multi-resolution Coding (MRC). MRC is a data compression algorithm in which simple, low rate source descriptions are embedded in more complex, high-rate descriptions [5]. Owing to this dependency of the layers, in lossy unreliable networks, even a few bit errors may cause severe degradation in video quality. Therefore, it is necessary for the video encoder to provide adequate error resilience features to protect the video data from channel errors. Redundancy is an important method to deal with network imperfections and provide high reliability. By modifying the transmitted data, i.e, adding redundancy, the probability that the message is successfully transmitted can be increased. Because of the special structure of MRC, two questions arise when redundancy is added: 1) How do we add redundancy to MRC, i.e. what is the structure of redundant codewords? 2) How much redundancy is needed?

One promising answer for the first question is coding for error resilience. One of the examples is forward error correction coding (FEC), where the sender encodes their message in a redundant way by using an error-correcting code. This idea could be pushed further. Instead of adding redundancy at the source, it could also be added at intermediate nodes along the propagation through a network, which fits the idea of a popular technique, network coding [1]. For MRC, the dependence among layers means that special care is needed in determining how they should be handled in the context of network coding. Previous work on network coding with MRC have two approaches, **Intra-layer Network Coding** and **Inter-layer Network Coding**. The difference of them is coding the layers either within one layer or across multiple layers. In this thesis, we're going to study both coding schemes and compare their performance to provide a coding strategy in different situations.

To answer the second question, we know that, while increasing quality, redundancy in transmission also means using more resources, which intrinsically results in the tradeoff between quality and efficiency. Therefore in this thesis, we consider both aspects when deciding the right amount of redundancy. To measure quality, we use rate-distortion characterization of Gaussian sources with squared error distortion. To measure efficiency, we define total redundancy, which combines redundancy at each layer. We want to propose a coding strategy considering both metrics.

In this thesis, we study the joint design of intra-layer and inter-layer network coding with MRC. We build a 3-step system model for the entire coding process: MR source coding, Layer coding and Multi-stream coding. The latter two are done at intermediate nodes of a network, so they together are considered as network coding. In the layer coding block, we introduce intra-layer and inter-layer coding with MRC and use two elements as design parameters: 1) redundancy at different layer; 2) whether to code within a layer or across multiple layers. Then we use the two metrics, average distortion and total redundancy to evaluate the performance and provide coding strategies based on that.

In the following part of the introduction, we first briefly overview the previous literature on 1) MRC with intra-layer and inter-layer coding, 2)Rate-distortion characterization, then present the main contributions and results, and, finally ,outline the organization of this thesis.

1.1 Multi-resolution Codes with Network Coding

Multi-resolution codes (MRC) are data compression algorithms in which simple, low rate source descriptions are embedded in more complex, high-rate descriptions [5]. MRC divides the video into a base layer and one or more enhancement layers. Decoding a higher layer requires all lower layers. The QoE for a user depends on the number of layers he is capable of decoding. Different users may thus experience varying QoS.

Proposed in [1], network coding allows and encourages mixing of data at intermediate nodes of a network, and in the context of MRC, the mixing of layers at intermediate nodes. The combination of MRC and network coding has been considered in literature. Previous work include two different approaches. The earlier work [2, 11, 12, 15] demonstrate substantial improvement in the bandwidth efficiency over traditional methods in the layered multicast setting. These schemes share a main feature that the use of network coding is only inside a layer and not across layers. We refer to this coding as **intra-layer network coding**. The intra-layer constraint does not realize the full potential of network coding. In [4], Dumitrescu first proposed layered multicast that allows network coding of data in different layers, which is referred to as **inter-layer network coding**. Independently, Wu [13] also proposed a cross-layer network coding technique with the main difference of random linear mixing of asymptotical optimality. Reference [10] showed that inter-layer NC indeed helps the delivery of MRC coded media over WiFi, and proposed how to search efficiently for coding strategies online. In [9], the authors proposed a pushback algorithm to generate network codes for single-source multicast of MRC. This algorithm takes advantage of inter-layer as well as intra-layer coding.

While inter-layer network coding is generally believed to have more benefits because it mixes data more aggressively, little has been done to compare rigorously intra-layer network coding and inter-layer network coding with MRC in different settings. Moreover, in the earlier work on MRC with network coding, the MR code is usually given. Here we want to study the joint design of MR and network coding. The design parameters we use affect both the MR source coding and network coding. This joint design allows us more flexibility in choosing the parameters. The investigation of the joint design of MRC and NC is the focus of this thesis.

1.2 Rate-Distortion Characterization

Distortion is a widely used metric for the performance of representation of a source. Information theory sets a fundamental limit on the trade-off between the distortion of a compression algorithm and the number of bits (rate) at the output of the algorithm. Defined in [3], a distortion measure is a mapping:

$$d : \mathcal{X} \times \hat{\mathcal{X}} \rightarrow \mathcal{R}^{+}.$$

from the set of source alphabet-reproduction alphabet pairs into a set of nonnegative real numbers. The distortion $d(x, \hat{x})$ is a measure of the cost of representing the symbol x by the symbol $\hat{x}$. The squared-error distortion,

$$d(x, \hat{x}) = (x - \hat{x})^2,$$

is the most popular distortion measure used for continuous alphabets, such as the complex Gaussian source we use in our model.

In single-resolution source coding, each source symbol is mapped to a binary description with some average per-symbol description length R. The source coding theorem and its converse describe the optimal distortion $D(R)$ theoretically achievable at the given average rate R. Now we consider a multi-resolution source code for describing a source at two resolutions. Instead of having a such rate R, each layer/description corresponds to a certain rate. The description at the first resolution (also called the base layer) is given at rate K_b and achieves an expected distortion of D_b. The description at the second resolution (also called the refinement layer) has rate K_r and achieves expected distortion of D_r. The achievable distortion of the receiver depends on how many layers it is capable of decoding.

In [6], the authors categorize a source as successively refinable if optimal descriptions can be considered as refinements of previous optimal descriptions. The authors lay down necessary and sufficient condition to determine whether a source is successively refinable. A successively refinable code is a special type of multi-resolution code. In [14], the authors derive rate-distortion lower bounds of spatially scalable video coding techniques. The rate-distortion bounds are derived from rate-distortion theory for stationary Gaussian signals where mean squared error is used as the distortion criterion.

In this thesis, we do not go into the details of deriving achievable rate-distortion region for multi-resolution codes. Instead, we use the basic rate-distortion characterization of Gaussian source with squared-error distortion as a metric to evaluate the two coding schemes we present.

1.3 Main Contributions and Results

Our work differs from previous literature in two aspects:

- We consider the joint-design of MRC and network coding. The system model we present considers both MR source coding and network coding. We choose design parameters that are key element in both coding steps.

- We consider intra- and inter-layer network coding from the perspective of rate-distortion. This rate-distortion lies at the heart of MRC.

In this thesis, we develop an end-to-end rate-distortion model for determining the appropriate level of redundancy given the channel estimates in the form of bit erasure rate. We compare the performance of intra-layer coding and inter-layer coding with multi-resolution codes in different situations and run simulations to complement the theoretical analysis. We define the notion of redundancy equivalence line and redundancy plane. Based on the average distortion of the two coding schemes, we divide the redundancy plane into Indifferent Coding Region and Inter-layer Coding

Region, which again consists of Replaceable Inter-layer Coding Region and Irreplaceable Inter-layer Coding Region. We're able to show that:

- Over the whole redundancy plane, inter-layer coding is always better or as good as intra-layer coding. This means 'mixing' never hurts from this perspective. Our work is different from what the previous literature indicated. This is due to the fact that the previous literature only considered fixed codes rather than joint design of MR and NC.

- The optimal performance for inter-layer coding is independent of the base layer percentage. Therefore we can design the source coding and network coding separately. This means, given a fixed MR code, no matter what α is, we can always do inter-layer coding to achieve the optimal performance. But it's not guaranteed for intra-layer coding. Whether the optimal performance is achievable depends on what α is given.

- In indifferent coding region, intra-layer and inter-layer coding are inter-changeable since they achieve the same performance in terms of quality and efficiency. In inter-layer coding region, at any operating point, inter-layer coding can get a smaller average distortion. But in the replaceable inter-layer coding region, performance of inter-layer coding can be achieved by other operating points on the same redundancy equivalence line using intra-layer coding.

1.4 Thesis Organization

The rest of the thesis is organized as follows: In chapter 2, we set up the system model and describe the encoding and decoding processes. Both processes have three steps, MR source coding, layer coding and multi-stream coding. Chapter 3 introduces two performance metrics we use in this thesis: average distortion, to measure the quality of reconstructed signal, and total redundancy, to measure the efficiency of transmission. In this chapter, we also define redundancy equivalence line and redundancy plane. In chapter 4, we calculate the average distortion in both coding schemes and derive the distortion exponent on the entire redundancy plane. Particularly, we find out the optimal operating points to achieve the minimum distortion, and calculate the distortion exponent and total redundancy accordingly. In chapter 5, we divide the redundancy plane into different regions to provide a coding strategy. We discuss the choice of coding schemes in different regions corresponding to different redundancy and network conditions. Chapter 6 shows the simulations results for our work. Chapter 7 summarizes the work and also discusses possible directions and problems about the future work . Mathematical derivations of some results are given in Appendix A. Distortion exponents with different parameters are plotted in Appendix B.

Chapter 2

System Model

In this chapter, we set up the system model which includes the encoding process and decoding process. Both processes have three steps: MR source coding, layer coding, and multi-stream coding, where the latter two are considered as network coding. The layer coding block uses two key elements as design parameters: 1) redundancy at different layers, 2) whether to code within one layer or across multiple layers. Figure 2-1 is an overview of the system.

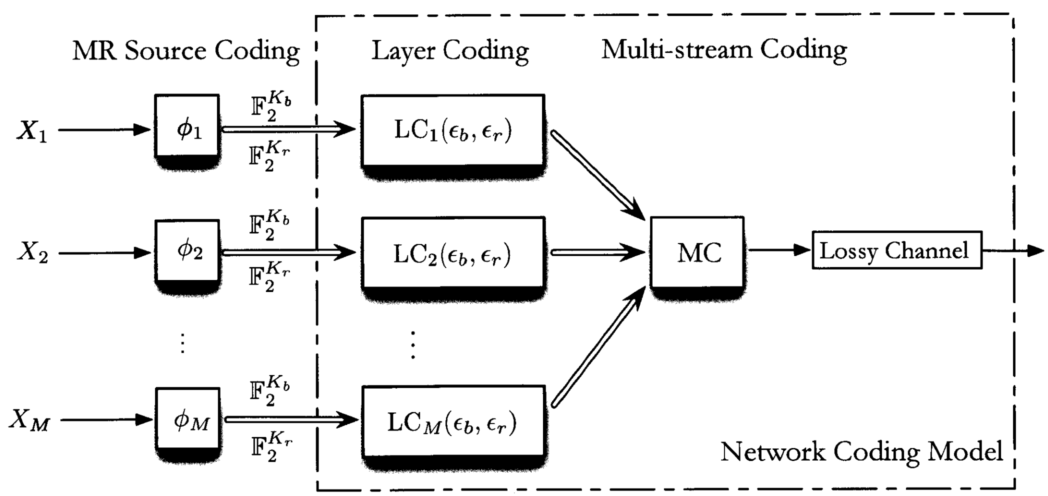

Figure 2-1: System Overview

2.1 Encoding Process

Consider $X_1, X_2, ...X_M$ as zero-mean unit variance complex Gaussian variable, generated by M independent memoryless sources, i.e, $X_i \sim \mathcal{CN}(0,1)$ for $i \in \{1, 2...M\}$. They all undergo a three-step encoding process, where the first one is source coding, the latter two are considered as network coding. In MR source coding and layer coding steps, we assume all coding blocks use the same parameters for all M streams. So, the procedure for each stream is identical before they are coded together in the multi-source coding block. For the sake of simplicity, we look at only one stream for the first two coding steps. Although we do not use index $i \in \{1, 2...M\}$, readers should keep in mind that the same coding steps happen for all M symbols.

2.1.1 MR Source Encoding

The MR source encoder ϕ encodes a source symbol X into a two-layer MRC with a base layer of length K_b and a refinement layer of length K_r. We use vectors $\mathbf{C}_b = \begin{bmatrix} C_b(1) & C_b(2) & ... & C_b(K_b) \end{bmatrix}$ to denote the base layer codeword, and $\mathbf{C}_r = \begin{bmatrix} C_r(1) & C_r(2) & ... & C_r(K_r) \end{bmatrix}$ to denote the refinement layer codeword. For $i \in \{1, 2, ...K_b\}, j \in \{1, 2, ..., K_r\}$, $C_b(i), C_b(j)$ are elements of some finite field GF(q). A vector of length K can then be interpreted as an element in $\mathbb{F}_q^K$, the finite field with q^K elements. Therefore the source encoder $\phi(\cdot)$ is:

$$\phi : \mathcal{X} \rightarrow \mathbb{F}_q^{K_b} \times \mathbb{F}_q^{K_r}.$$

2.1.2 Layer Encoding

Each MRC then undergoes a layer coding block, denoted by LC in figure 2-1. The LC decides whether to code within a layer or across layers, and how much redundancy is needed for each layer. The output is generated by random linear network coding (RLNC) [8]. The layer coding block contains two encoders. The base encoder $\phi_b(\cdot)$ converts $\mathbf{C}_b$ to $\hat{\mathbf{C}}_b$ of length N_b. The refinement encoder $\phi_r(\cdot)$ converts $\mathbf{C}_r$ to $\hat{\mathbf{C}}_r$ of length N_r. ϵ_b denotes the base layer redundancy, and ϵ_r denotes the refinement layer redundancy. The output code lengths for the base layer and refinement layer are defined as

$$N_b \triangleq \frac{K_b}{1-\epsilon_b} \tag{2.1}$$

$$N_r \triangleq \frac{K_r}{1-\epsilon_r}. \tag{2.2}$$

The total length of codewords is

$$N \triangleq N_b + N_r. \tag{2.3}$$

We also define the base layer percentage as

$$\alpha \triangleq \frac{N_b}{N}. \tag{2.4}$$

Now we look at the intra-layer and inter-layer coding separately.

Intra-layer encoding

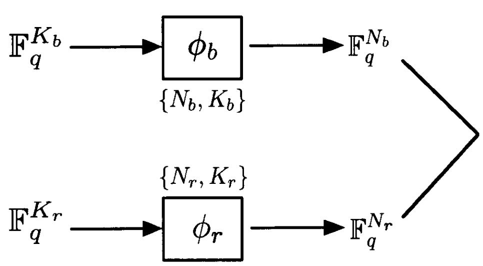

Figure 2-2: Intra-layer coding block

In this scheme, the encoding of each layer is independent. The output codeword consists of elements which are random linear combinations of elements from the same layer. Let a $K_b \times N_b$ matrix $\mathbf{H}_b = \begin{bmatrix} \mathbf{h}_1 & \mathbf{h}_2 & \dots & \mathbf{h}_{N_b} \end{bmatrix}$ represent the base layer encoding function $\phi_b : \mathbb{F}_q^{K_b} \to \mathbb{F}_q^{N_b}$. Elements in $\mathbf{H}_b$ are randomly generated from the same finite field GF(q). The output of the base encoder is therefore a (N_b, K_b) linear code. The jth element in $\hat{C}_b$ is,

$$\hat{C}_b(j) = \sum_{k=1}^{K_b} h_j(k) C_b(k).$$

Similarly for refinement layer, we use a $K_r \times N_r$ matrix $\mathbf{H}_r = \begin{bmatrix} \mathbf{h}_1 & \mathbf{h}_2 & \dots & \mathbf{h}_{N_r} \end{bmatrix}$ to represent the linear function, $\phi_r : \mathbb{F}_q^{K_r} \to \mathbb{F}_q^{N_r}$. The jth element in $\hat{C}_r$ is then

$$\hat{C}_r(j) = \sum_{k=1}^{K_r} h_j(k) C_r(k).$$

Inter-layer encoding

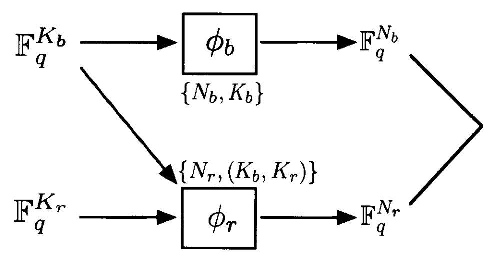

Figure 2-3: Inter-layer coding block

In this scheme, the base layer encoder and generating matrix are the same as in the intra-layer coding case. However, the refinement layer encoder takes in the base layer as well as the refinement layer to generate a random code. The encoding function for the refinement layer is $\phi_r : \mathbb{F}_2^{K_b} \times \mathbb{F}_2^{K_r} \to \mathbb{F}_2^{N_r}$. The size of the encoding matrix is $(K_b + K_r) \times N_r$. The jth element of the refinement code is:

$$\hat{C}_r(j) = \sum_{k=1}^{K_b} h_j(k) C_r(k) + \sum_{l=K_b+1}^{K_b+K_r} h_j(l) C_r^{[i]}(l).$$

2.1.3 Multi-stream Encoding

M codes then undergo a multi-stream coding block. The encoder is a function that maps M codewords of length N to one long codeword of length MN. Denote the multi-stream encoder by ME:

$$\text{ME} : \mathbb{F}_q^N \times \mathbb{F}_q^N \cdots \mathbb{F}_q^N \rightarrow \mathbb{F}_q^{MN}.$$

The layer coding combines different layers and multi-stream coding combines different streams. They together model the network coding process in a network. We then put each element of the codeword into independent channel use and assume each element is erased with probability p. Therefore, each element arrives at the receiver with probability $1 - p$.

2.2 Decoding Process

The decoding process consists of the reversed 3 steps as the encoding process: multi-stream decoding, layer decoding and MR source decoding.

2.2.1 Multi-stream Decoding

The received codeword goes through a multi-stream decoder denoted by MD. It maps the long code of length MN to M MR codes of length N:

$$\text{MD} : \mathbb{F}_q^{MN} \rightarrow \mathbb{F}_q^N \times \mathbb{F}_2^q \cdots \mathbb{F}_q^N.$$

2.2.2 Layer Decoding

Owing to the erasure probability p, the length of received codeword at the receiver is not the original N_b and N_r. We use N_b' and N_r' to denote the actual length of received codeword for base and refinement layer, respectively. $N_b' \leq N_b, N_r' \leq N_r$. Here we relate the erasure of certain elements to eliminating corresponding columns of the encoding matrix. If the ith bit is erased, then at the receiver's side, the equivalent encoding matrix becomes $\mathbf{H}_b' = \begin{bmatrix} \mathbf{h}_1 & \dots & \mathbf{h}_{i-1} & \mathbf{h}_{i+1} & \dots & \mathbf{h}_{N_b'} \end{bmatrix}$. This is the same for the refinement encoding matrix. Then we look at the decoding functions for the two coding schemes individually.

Intra-layer decoding

The base and refinement layer are decoded separately. For the base layer, the equivalent encoding matrix $\mathbf{H}'_b$ at the receiver is a $K_b \times N'_b$ sub matrix of $\mathbf{H}_b$, with $N_b - N'_b$ columns eliminated. The base layer can be decoded if $N'_b \geq K_b$. Similarly for the refinement layer, the equivalent encoding matrix $\mathbf{H}'_r$ at the receiver is a $K_b \times N'_b$ sub matrix of $\mathbf{H}_r$, with $N_r - N'_r$ columns eliminated. Successful decoding requires $N'_r \geq K_r$. The layer decoding block then consists of two independent functions:

$$\begin{aligned}\psi_b :&\mathbb{F}_q^{N_b} \rightarrow \mathbb{F}_q^{K_b}\\ \psi_r :&\mathbb{F}_q^{N_r} \rightarrow \mathbb{F}_q^{K_r}.\end{aligned}$$

The layer decoding block could decode no layers, the base layer, the refinement layer, or both layers.

Inter-layer decoding

In this case, the layer decoding block also has two decoding functions. The base layer decoder is the same as in the previous case. However it is more complicated to decode the refinement layer. Here the equivalent encoding matrix is not just a sub matrix of the original refinement layer encoding matrix. It's constructed from both layers and we can write it as a $(K_b + K_r) \times (N'_b + N'_r)$ matrix $\mathbf{G}_r = \left[\begin{array}{c|c} \mathbf{H}'_b & \\ \hline \mathbf{0} & \mathbf{H}'_r \end{array}\right]$, where $\mathbf{H}'_b$ is a $K_b \times N'_b$ matrix and $\mathbf{H}'_r$ is a $(K_b + K_r) \times N'_r$ matrix. Therefore, $\mathbf{G}_r$ has a right inverse if $N'_r \geq K_r, N'_b + N'_r \geq K_b + K_r$. The two decoding functions in the inter-layer decoding block is:

$$\begin{aligned}\psi_b :&\mathbb{F}_q^{N_b} \rightarrow \mathbb{F}_q^{K_b}\\ \psi_r :&\mathbb{F}_q^{N_b} \times \mathbb{F}_q^{N_r} \rightarrow \mathbb{F}_q^{K_b} \times \mathbb{F}_q^{K_r}.\end{aligned}$$

The layer decoding block could decode none, base layer, or both layers. Notice that the base layer can be decoded alone but the refinement layer cannot be decoded without the base layer. This is because the base encoder encodes only the base layer, while the refinement encoder mixes the base layer's information in the refinement layer. The base decoder can therefore work independently and outputs only the base layer. And the refinement decoder decodes two and the same time or nothing at all.

2.2.3 MR Source Decoding

If the MR source decoder receives two layers from the layer decoder, it uses the refinement decoder to map it back to the alphabet: $\mathbb{F}_2^{K_b} \times \mathbb{F}_2^{K_r} \to \hat{\mathcal{X}}$; if it receives only the base layer, it uses the base layer decoder: $\mathbb{F}_2^{K_b} \to \hat{\mathcal{X}}$; if it receives none of the layers, decoding cannot be done. For intra-layer coding, it is possible that the source decoder gets only the refinement layer from the layer decoder. But it cannot decode anything in this case.

We want to remind the reader that in the LC block, when we say decode one layer, it could refer to the refinement layer if the LC is an intra-layer block. But for MR source decoding block, when we say decode one layer, we mean the mapping at the base decoder, $\mathbb{F}_2^{K_b} \to \hat{\mathcal{X}}$. For later discussions, unless stated otherwise, 'decoding one layer' refers to this mapping, and this one layer is the base layer.

Chapter 3

Performance Metrics

3.1 Quality metric – Average distortion

We consider the squared-error $d(X,\hat{X}) = (X-\hat{X})^2$ as the distortion measure. For i.i.d zero-mean unit-variance complex Gaussian source, the distortion function is $D = e^{-R}$ [3]. When it comes to multi-resolution codes, the quality of representation varies, depending on the number of layers decoded. For a MR code with the base source rate K_b and the refinement source rate K_r, the distortion achieved by the base and refinement decoder is

$$\begin{aligned} D_b &= Ed(X, \psi_b(\phi_b(X))) \\ &= e^{-K_b} \\ &= e^{-(1-\epsilon_b)\alpha N} \end{aligned} \tag{3.1}$$

$$\begin{aligned} D_r &= Ed(X, \psi_r(\phi_r(X))) \\ &= e^{-K_b-K_r} \\ &= e^{-(1-\epsilon_b)\alpha N-(1-\epsilon_r)(1-\alpha)N}. \end{aligned} \tag{3.2}$$

In this model, the receiver could decode no layers, one or two layers, corresponding to probabilities P_0, P_1 and P_2 and distortion $1, D_b$ and D_r. Therefore we use the average distortion considering all these situations. Write the probability P_i in the form of

$P_i = e^{-NE_i(\alpha,\epsilon_b,\epsilon_r)}$, then

$$\begin{aligned} E[D](\alpha,\epsilon_b,\epsilon_r) &= P_0 \cdot 1 + P_1 \cdot D_b + P_2 \cdot D_r \\ &= e^{-NE_0(\alpha,\epsilon_b,\epsilon_r)} + e^{-N[E_1(\alpha,\epsilon_b,\epsilon_r)+\alpha(1-\epsilon_b)]} + e^{-N[E_2(\alpha,\epsilon_b,\epsilon_r)+\alpha(1-\epsilon_b)+(1-\alpha)(1-\epsilon_r)]} \\ &= e^{-NE_0'(\alpha,\epsilon_b)} + e^{-NE_1'(\alpha,\epsilon_b,\epsilon_r)} + e^{-NE_2'(\alpha,\epsilon_b,\epsilon_r)} \end{aligned} \tag{3.3}$$

where

$$E_0'(\alpha,\epsilon_b,\epsilon_r) = E_0(\alpha,\epsilon_b,\epsilon_r) \tag{3.4}$$

$$E_1'(\alpha,\epsilon_b,\epsilon_r) = E_1(\alpha,\epsilon_b,\epsilon_r) + \alpha(1-\epsilon_b) \tag{3.5}$$

$$E_2'(\alpha,\epsilon_b,\epsilon_r) = E_2(\alpha,\epsilon_b,\epsilon_r) + \alpha(1-\epsilon_b) + (1-\alpha)(1-\epsilon_r). \tag{3.6}$$

Often used in the field of information theory, the definition of exponential approximation is introduced here again.

Definition 1: For a sequence $f(1), f(2), ..., f(n), ...$, and a constant $g \in \mathcal{R}$, we say $f(n) \doteq e^{ng}$ if $\lim_{n\to\infty} \frac{\log f(n)}{n} = g$.

Definition 2: We rewrite the average distortion as $E[D](\alpha,\epsilon_b,\epsilon_r) \doteq e^{-NE}$ and call E the **Distortion Exponent**,

$$E \triangleq \min\{E_0', E_1', E_2'\}. \tag{3.7}$$

Then we present the region where the average distortion is defined in. To provide reliable transmission, we assume there is enough redundancy in both layers, i.e, $\epsilon_b, \epsilon_r > p$. Then we introduce the definition of redundancy plane.

Definition 3: Given some p and M, the (ϵ_b, ϵ_r) plane where $\epsilon_b \in (p, 1], \epsilon_b \in (p, 1]$ is called the **Redundancy Plane**.

Our coding strategies are all based on this redundancy plane.

3.2 Efficiency metric – Total redundancy

Another aspect of performance is the efficiency of using the resources. Therefore, besides the quality metric, we also define an efficiency metric, which is the total redundancy of the transmitted codewords. In the model we have presented, we look at the redundancy separately in two layers, ϵ_b and ϵ_r. They show clearly how much redundancy we need to put in the base and refinement layer. Here we introduce another definition which will be used throughout this thesis.

Definition 4: For $\epsilon_b, \epsilon_r \in (p, 1)$, we call the two dimensional (ϵ_b, ϵ_r) plane **Redundancy Plane**.

Instead of evaluating the efficiency layer by layer, we would like to consider the total redundancy. Therefore, we use ϵ to denote the total amount of redundancy of the codewords.

$$\epsilon \triangleq \alpha\epsilon_b + (1-\alpha)\epsilon_r. \tag{3.8}$$

For any give α, we can plot different level of redundancy in figure 3-1. We call these lines **Redundancy Equivalence Lines**. Points on the same line have equal total redundancy. The slope of the lines is decided by α, and so is the direction the total redundancy increases. Intuitively, the higher the line, the greater ϵ, as shown in figure 3-1.

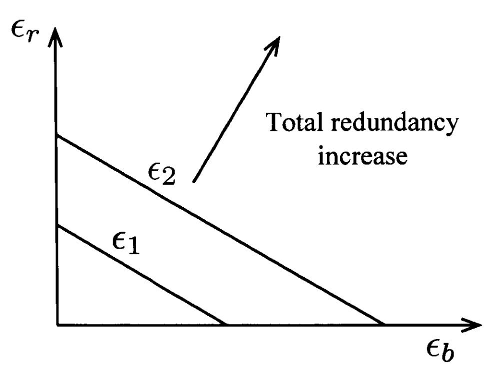

Figure 3-1: Efficiency metric: total redundancy. Points on the same line have the same total redundancy.$\epsilon_2 > \epsilon_1$

Chapter 4

Distortion Analysis on the Redundancy Plane

In this chapter, we analyze the average distortion over the redundancy plane and the total redundancy at the optimal operating points for both intra-layer and inter-layer coding. We first find the probabilities of decoding no layers, base layer and two layers, and then derive the average distortion considering the three situations. Our goal is to find the distortion exponent over the redundancy plane.

4.1 Decoding Probabilities

As previously, we use N_b' and N_r' to denote the length of received base and refinement layer vectors. P_0, P_1 and P_2 are probabilities of decoding no layers, base layer or two layers, respectively. The above notations are the same for intra-layer and inter-layer coding schemes. Since the two cases are discussed separately in different sections, this re-use of rotations will not cause any confusion.

4.1.1 Information Theory Preliminaries

Before we move on to the analysis, we recap some widely used notations and definitions in information theory.

- Entropy of a Bernoulli distribution with parameter p:

$$H(p) = -p\log p - (1-p)\log(1-p). \tag{4.1}$$

- *Kullback-Leibler distance* between two Bernoulli distributions p and q:

$$D(q||p) = q\log\frac{q}{p} + (1-q)\log\frac{1-q}{1-p}. \tag{4.2}$$

- Stirling's Approximation [7]: for $i \in \{0, 1, ...n\}$,

$$\binom{n}{i} = \frac{n!}{i!(n-i)!} \doteq e^{nH(\frac{i}{n})}. \tag{4.3}$$

4.1.2 Intra-layer Coding

The receiver wants to decode all M streams. Here we look at only stream first, and then combine them. First, the receiver cannot decode anything if he fails to decode the base layer. Thus the condition for decoding no layer is $N_b' < K_b$. Second, the receiver decodes only one layer if he can decode the base layer but cannot decode the refinement layer. The condition is $(N_b' \geq K_b) \wedge (N_r' < K_r)$. Third, the receiver can decode two layers if he decodes both the base and refinement layers separately. The condition is $(N_b' \geq K_b) \wedge (N_r' \geq K_r)$. K_b and K_r are the source codeword lengths for the base layer and refinement layer, respectively. We then have

$$P_0 = \Pr\{N_b' < K_b\} \tag{4.4}$$

$$P_1 = \Pr\{(N_b' \geq K_b) \wedge (N_r' < K_r)\} \tag{4.5}$$

$$P_2 = \Pr\{(N_b' \geq K_b) \wedge (N_r' \geq K_r)\}. \tag{4.6}$$

We draw the probability distribution in a 2-D plane in figure 4-1. The horizontal axis is the length of received base layer codeword and the vertical axis is the length of received refinement layer codeword. Then we look at each of the above probabilities.

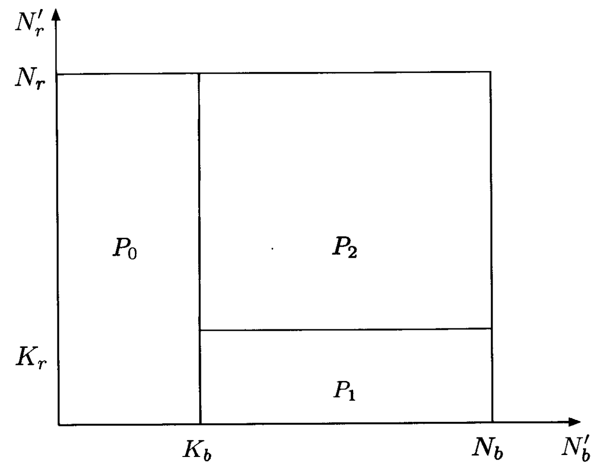

Figure 4-1: Probability of decoding no layers, base layer and two layers in the intra-layer coding scheme.

1. $P_0 = \Pr\{N_b' < K_b\}$

$$P_0 = \Pr\{N_b' < K_b\} = \sum_{i=0}^{K_b-1} \Pr\{N_b' = i\}$$

Let $i' = N_b - i$,

$$P_0 = \sum_{i'=N_b-K_b+1}^{N_b} \Pr\{N_b' = N_b - i'\}$$

Since $N_b = \alpha N, K_b = (1-\epsilon_b)N_b = (1-\epsilon_b)\alpha N$,

$$P_0 = \sum_{i'=\epsilon_b \alpha N+1}^{\alpha N} \binom{\alpha N}{i'} p^{i'}(1-p)^{\alpha N - i'}.$$

Using equation (4.3), we have

$$P_0 = \sum_{i'=\epsilon_b \alpha N+1}^{\alpha N} e^{\alpha N\left[H(\frac{i'}{\alpha N})+\frac{i'}{\alpha N}\log p+(1-\frac{i'}{\alpha N})\log(1-p)\right]}.$$

Let $\gamma_b = \frac{i'}{\alpha N}$, then $\gamma_b \in (\epsilon_b, 1)$,

$$P_0 = \sum_{\epsilon_b}^{1} e^{-\alpha N \cdot D(\gamma_b||p)}$$
$$\doteq e^{-\alpha N \cdot \min_{\gamma_b \in (\epsilon_b,1)} D(\gamma_b||p)}.$$

Since $\epsilon_b > p$, $\min_{\gamma_b \in (\epsilon_b,1)} D(\gamma_b||p) = D(\epsilon_b||p)$, then

$$P_0 \doteq e^{-\alpha N D(\epsilon_b||p)}.$$

Therefore, for the M codewords,

$$E_0(\alpha, \epsilon_b, \epsilon_r) = \alpha M D(\epsilon_b||p). \tag{4.7}$$

2. $P_1 = \Pr\{(N_b' \geq K_b) \wedge (N_r' < K_r)\}$

$$P_1 = \sum_{i=K_b}^{\alpha N} \Pr\{N_b' = i\} \sum_{j=0}^{K_r-1} \Pr\{N_r' = j\}$$

Let $i' = N_b - i, j' = N_r - j$, then

$$P_1 = \sum_{i'=0}^{\epsilon_b \alpha N} \Pr\{N_b' = N_b - i'\} \sum_{j'=\epsilon_r(1-\alpha)N+1}^{(1-\alpha)N} \Pr\{N_r' = N_r - j'\}$$
$$= \sum_{i'=0}^{\epsilon_b \alpha N} \binom{\alpha N}{i'} p^{i'}(1-p)^{\alpha N - i'} \sum_{j'=\epsilon_r(1-\alpha)N+1}^{(1-\alpha)N} \binom{(1-\alpha)N}{j'} p^{j'}(1-p)^{(1-\alpha)N-j'}$$
$$= \sum_{\gamma_b=0}^{\epsilon_b} e^{-\alpha N D(\gamma_b||p)} \sum_{\gamma_r=\epsilon_r}^{1} e^{-(1-\alpha)ND(\gamma_r||p)}$$
$$\doteq e^{-\alpha N \min_{\gamma_b \in [0,\epsilon_b]} D(\gamma_b||p)} \cdot e^{-(1-\alpha)N \min_{\gamma_r \in [\epsilon_r,1]} D(\gamma_r||p)}.$$

Since $\epsilon_b, \epsilon_r > p$, $\min_{\gamma_b \in [0,\epsilon_b]} D(\gamma_b||p) = 0$ and $\min_{\gamma_r \in [\epsilon_r,1]} D(\gamma_r||p) = D(\epsilon_r||p)$, and

$$P_1 \doteq e^{-N(1-\alpha)D(\epsilon_r||p)}.$$

Therefore, for M codewords, we have

$$E_1(\alpha, \epsilon_b, \epsilon_r) = (1-\alpha)MD(\epsilon_r||p). \tag{4.8}$$

3. $P_2 = \Pr\{(N_b' \geq K_b) \wedge (N_r' \geq K_r)\}$

$$P_2 = \sum_{i=K_b}^{\alpha N} \Pr\{N_b' = i\} \sum_{j=K_r}^{(1-\alpha)N} \Pr\{N_r' = j\}$$

Let $i' = N_b - i, j' = N_r - j$, then

$$\begin{aligned}
P_2 &= \sum_{i'=0}^{\epsilon_b \alpha N} \Pr\{N_b' = N_b - i'\} \sum_{j'=0}^{\epsilon_r(1-\alpha)N} \Pr\{N_r' = N_r - j'\} \\
&= \sum_{i'=0}^{\epsilon_b \alpha N} \binom{\alpha N}{i'} p^{i'}(1-p)^{\alpha N - i'} \cdot \sum_{j=0}^{\epsilon_r(1-\alpha)N} \binom{(1-\alpha)N}{j'} p^{j'}(1-p)^{(1-\alpha)N-j'} \\
&= \sum_{\gamma_b=0}^{\epsilon_b} e^{-\alpha N D(\gamma_b||p)} \sum_{\gamma_r=0}^{\epsilon_r} e^{-(1-\alpha)ND(\gamma_r||p)} \\
&\doteq e^{-\alpha N \cdot \min_{\gamma_b \in [0,\epsilon_b]} D(\gamma_b||p)} \cdot e^{-(1-\alpha)N \cdot \min_{\gamma_r \in [0,\epsilon_r]} D(\gamma_r||p)}.
\end{aligned}$$

Since $\epsilon_b, \epsilon_r > p$, $\min_{\gamma_b \in [0,\epsilon_b]} D(\gamma_b||p) = 0$ and $\min_{\gamma_r \in [0,\epsilon_r]} D(\gamma_r||p) = 0$, then

$$P_2 \doteq e^{-N \cdot 0}.$$

Therefore, for M codewords, we have

$$E_2(\alpha, \epsilon_b, \epsilon_r) = 0. \tag{4.9}$$

4.1.3 Inter-layer Coding

The decoding probabilities in this case are similar to the intra-layer coding case, except for P_0. The receiver can decode no layer if he cannot decode the base layer codewords which could come from either the base layer or refinement layer. Therefore, besides the constraint that $N_b' < K_b$, we also need to make sure that refinement layer does not help either, i.e, $N_b' + N_r' < K_b + K_r$. For decoding one layer, the condition is $(N_b' \geq K_b) \wedge (N_r' < K_r)$. For decoding two layers, the condition is $(N_b' \geq K_b) \wedge (N_r' \geq K_r)$. We then have

$$P_0 = P\{(N_b' < K_b) \wedge (N_b' + N_r' < K_b + K_r)\} \tag{4.10}$$

$$P_1 = P\{(N_b' \geq K_b) \wedge (N_r' < K_r)\} \tag{4.11}$$

$$P_2 = P\{(N_r' \geq K_r) \wedge (N_b' + N_r' \geq K_b + K_r)\}. \tag{4.12}$$

Again we plot the probabilities in figure 4-2, and look at each probability.

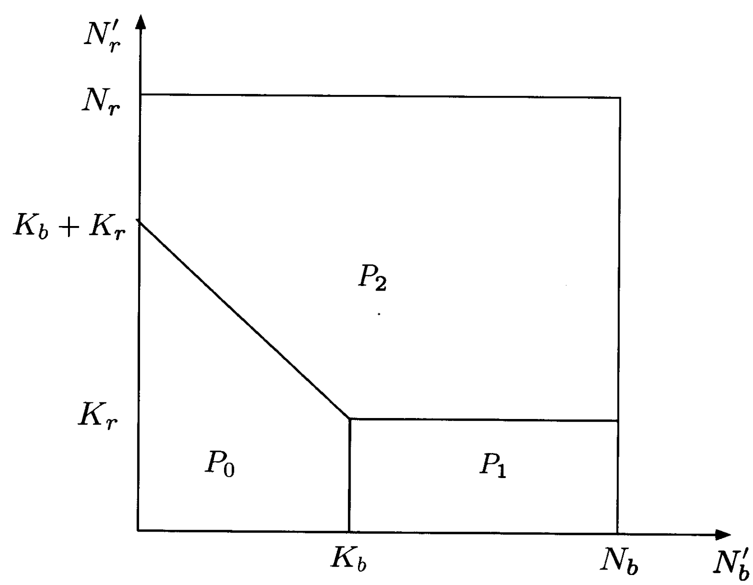

Figure 4-2: Probability of getting no, one and two layers in the inter-layer coding scheme

1. $P_0 = \Pr\{(N_b' < K_b) \wedge (N_b' + N_r' < K_b + K_r)\}$

$$\begin{aligned} P_0 &= \sum_{i=0}^{K_b-1} \Pr\{N_b' = i\}\Pr\{N_b' + N_r' < K_b + K_r | N_b' = i\} \\ &= \sum_{i=0}^{K_b-1} \Pr\{N_b' = i\}\Pr\{N_r' < K_b + K_r - i\} \\ &= \sum_{i=0}^{K_b-1} \Pr\{N_b' = i\} \sum_{j=0}^{K_b+K_r-i-1} \Pr\{N_r' = j\} \end{aligned}$$

Let $i' = N_b - i, j' = N_r - j$, then

$$\begin{aligned} P_0 &= \sum_{i'=\epsilon_b \alpha N+1}^{\alpha N} \Pr\{N_b' = N_b - i'\} \sum_{j'=\epsilon_b \alpha N + \epsilon_r(1-\alpha)N+i+1}^{(1-\alpha)N} \Pr\{N_r' = N_r - j'\} \quad (4.13) \\ &= \sum_{i'=\epsilon_b \alpha N+1}^{\alpha N} \binom{\alpha N}{i'} p^{i'}(1-p)^{\alpha N - i'} \sum_{j'=\epsilon_b \alpha N + \epsilon_r(1-\alpha)N+i'+1}^{(1-\alpha)N} \binom{(1-\alpha)N}{j'} p^{j'}(1-p)^{(1-\alpha)N-j'} \\ &= \sum_{\gamma_b \in [\epsilon_b, 1]} e^{-\alpha N D(\gamma_b || p)} \sum_{\gamma_r \in [(\epsilon_b - \gamma_b)\frac{\alpha}{1-\alpha} + \epsilon_r, 1]} e^{-(1-\alpha)ND(\gamma_r||p)} \\ &\doteq e^{-\alpha N \cdot \min_{\gamma_b \in [\epsilon_b,1]} D(\gamma_b||p) - (1-\alpha)N \cdot \min_{\gamma_r \in [(\epsilon_b - \gamma_b)\frac{\alpha}{1-\alpha}+\epsilon_r, 1]} D(\gamma_r||p)}. \end{aligned}$$

For M codewords, we have

$$\begin{aligned} E_0(\alpha, \epsilon_b, \epsilon_r) &= M \min_{\gamma_b \in [\epsilon_b, 1]} \left[\alpha \cdot D(\gamma_b||p) + (1-\alpha) \cdot \min_{\gamma_r \in [(\epsilon_b - \gamma_b)\frac{\alpha}{1-\alpha}+\epsilon_r, 1]} D(\gamma_r||p)\right] \\ &= \begin{cases} M\alpha D(\epsilon_b||p) + M(1-\alpha)D(\epsilon_r||p) & \epsilon_b \geq \epsilon_r \\ MD\big[\alpha\epsilon_b + (1-\alpha)\epsilon_r||p\big] & \epsilon_b < \epsilon_r \end{cases}. \end{aligned} \quad (4.14)$$

The derivation of equation (4.14) is in Appendix A.

2. $P_1 = \Pr\{(N_b' \geq K_b) \wedge (N_r' < K_r)\}$

This is the same as in the intra-layer case, so we directly use the previous result.

$$E_1(\alpha, \epsilon_b, \epsilon_r) = (1-\alpha)MD(\epsilon_r||p). \quad (4.15)$$

3. $P_2 = \Pr\{(N'_r \geq K_r) \wedge (N'_b + N'_r \geq K_b + K_r)\}$

$$\begin{aligned} P_2 &= \sum_{j=K_r}^{N_r} \Pr\{N'_r = j\}\Pr\{N'_b + N'_r \geq K_b + K_r | N'_r = j\} \\ &= \sum_{j=K_r}^{N_r} \Pr\{N'_r = j\}\Pr\{N'_b \geq K_b + K_r - j\} \\ &= \sum_{j=K_r}^{N_r} \Pr\{N'_r = j\} \sum_{i=K_b+K_r-j}^{N_b} \Pr\{N'_b = i\} \end{aligned}$$

Let $i' = N_b - i, j' = N_r - j$, then

$$\begin{aligned} P_2 &= \sum_{j'=0}^{\epsilon_r(1-\alpha)N} \Pr\{N'_r = N_r - j'\} \sum_{i'=0}^{\alpha\epsilon_b N + (1-\alpha)\epsilon_r N - j'} \Pr\{N'_b = N_b - i'\} \\ &= \sum_{\gamma_r=0}^{\epsilon_r} e^{-(1-\alpha)ND(\gamma_r||p)} \sum_{\gamma_b=0}^{\epsilon_b+(\epsilon_r-\gamma_r)\frac{1-\alpha}{\alpha}} e^{-\alpha ND(\gamma_b||p)} \\ &\doteq e^{-(1-\alpha)N\cdot\min_{\gamma_r\in[0,\epsilon_r]} D(\gamma_r||p) - \alpha N \min_{\gamma_b\in[0,\epsilon_b+(\epsilon_r-\gamma_r)\frac{1-\alpha}{\alpha}]} D(\gamma_b||p)}. \end{aligned}$$

Since $\epsilon_r \in (p, 1]$, $\min_{\gamma_r\in[0,\epsilon_r]} D(\gamma_r||p) = 0$. Besides, because $\gamma_r \in [0, \epsilon_r], \epsilon_b \in (p, 1]$, so $\epsilon_b + (\epsilon_r - \gamma_r)\frac{1-\alpha}{\alpha} > p$, then $\min_{\gamma_b\in[0,\epsilon_b+(\epsilon_r-\gamma_r)\frac{1-\alpha}{\alpha}]} D(\gamma_b||p) = 0$. Thus

$$E_2(\alpha, \epsilon_b, \epsilon_r) = 0. \tag{4.16}$$

4.2 Distortion Exponent

In chapter 3, we show that the average distortion is

$$E[D] = P_0 \cdot 1 + P_1 \cdot D_b + P_2 \cdot D_r. \tag{4.17}$$

In the following part of the chapter we use the probabilities we have derived above to find the distortion exponents for the two coding cases.

4.2.1 Preliminary Function

We introduce an important function which will be used in later analysis. Given p, define $F(x)$ as

$$F(x) = MD(x||p) + x - 1. \tag{4.18}$$

$D(x||p)$ is a K-L distance between two Bernoulli distribution with parameter p and x. As $D(x||p)$ is convex, $F(x)$ is also a convex function. $F(0) > 0, F(1) > 0$ and $F(p) < 0$, so there must exist two solutions to the function $F(x) = 0$, one is in $(0, p]$ and the other in $(p, 1]$. We only need the one in $(p, 1)$, so here we derive for the closed form expression for it. Using a Taylor expansion at p, we can write x as $x = p + \delta$, and $\delta \to 0$. Therefore $D(x||p)$ can be written as:

$$D(x||p) = \frac{\delta^2}{2p(1-p)}. \tag{4.19}$$

Substituting (4.19) it in (4.18), we obtain

$$F(x) = \frac{M}{2p(1-p)}\delta^2 + \delta + p - 1. \tag{4.20}$$

Solving for $F(x) = 0$, we have

$$\delta = \frac{1}{M}\Big(\sqrt{1 + \frac{2M}{p}} - 1\Big)p(1-p).$$

If M is big, then $\frac{2M}{p} >> 1, \sqrt{1 + \frac{2M}{p}} >> 1$, therefore,

$$\delta \approx (1-p)\sqrt{\frac{2p}{M}} \tag{4.21}$$

$$x \approx p + (1-p)\sqrt{\frac{2p}{M}}. \tag{4.22}$$

We use x_0 to denote this solution so that $F(x_0) = 0$. We'll use this solution later in our analysis.

4.2.2 Intra-layer Coding

Based on the previous results, we can then find the distortion exponent in this case. Substitute (4.7),(4.8) and (4.9) in equation (3.4), (3.5) and (3.6) respectively, we get

$$E_0'(\alpha, \epsilon_b, \epsilon_r) = M\alpha D(\epsilon_b||p) \tag{4.23}$$

$$E_1'(\alpha, \epsilon_b, \epsilon_r) = M(1-\alpha)D(\epsilon_r||p) + \alpha(1-\epsilon_b) \tag{4.24}$$

$$E_2'(\alpha, \epsilon_b, \epsilon_r) = \alpha(1-\epsilon_b) + (1-\alpha)(1-\epsilon_r). \tag{4.25}$$

According to the definition of the distortion exponent (3.7), we have

$$E(\alpha, \epsilon_b, \epsilon_r) = \begin{cases} E_0'(\alpha, \epsilon_b, \epsilon_r) & (\epsilon_b, \epsilon_r) \in \mathcal{A}_0 \\ E_1'(\alpha, \epsilon_b, \epsilon_r) & (\epsilon_b, \epsilon_r) \in \mathcal{A}_1 \\ E_2'(\alpha, \epsilon_b, \epsilon_r) & (\epsilon_b, \epsilon_r) \in \mathcal{A}_2 \end{cases} \tag{4.26}$$

We denote the region where $E = E_i'$ as $\mathcal{A}_i$ for $i = 0, 1, 2$. These $\mathcal{A}_i$s will be used later in this thesis. Now we want to find the boundaries for the regions. Define the following functions:

$$\begin{aligned} g_{01}(\alpha, \epsilon_b, \epsilon_r) &= E_0'(\alpha, \epsilon_b, \epsilon_r) - E_1'(\alpha, \epsilon_b, \epsilon_r) \\ &= M\alpha D(\epsilon_b||p) - M(1-\alpha)D(\epsilon_b||p) - \alpha(1-\epsilon_b) \\ &= \alpha F(\epsilon_b) - M(1-\alpha)D(\epsilon_r||p) \\ g_{12}(\alpha, \epsilon_b, \epsilon_r) &= E_1'(\alpha, \epsilon_b, \epsilon_r) - E_2'(\alpha, \epsilon_b, \epsilon_r) \\ &= M(1-\alpha)D(\epsilon_r||p) - (1-\alpha)(1-\epsilon_r) \\ &= (1-\alpha)F(\epsilon_r) \\ g_{20}(\alpha, \epsilon_b, \epsilon_r) &= E_2'(\alpha, \epsilon_b, \epsilon_r) - E_0'(\alpha, \epsilon_b, \epsilon_r) \\ &= \alpha(1-\epsilon_r) + (1-\alpha)(1-\epsilon_r) - M\alpha D(\epsilon_r||p) \\ &= -\alpha F(\epsilon_b) + (1-\alpha)(1-\epsilon_r). \end{aligned}$$

We find the boundaries of $\mathcal{A}_i$s by setting $g_{ij}(\epsilon_b, \epsilon_r)$ to 0, $i, j \in \{0, 1, 2\}$ and calculate the solution.

- Set $g_{01}(\epsilon_b, \epsilon_r) = 0$. We use equation (4.20) to simplify $F(\epsilon_b)$ and abtain an analytical solution.

$$g_{01}(\epsilon_b, \epsilon_r) = \alpha\Big[\frac{M}{2p(1-p)}\delta^2 + \delta + p - 1\Big] - M(1-\alpha)D(\epsilon_r||p).$$

Solving for $g_{01}(\epsilon_b, \epsilon_r) = 0$, we have

$$\begin{aligned}\delta &= \frac{p(1-p)}{M}\left[-1+\sqrt{1+\frac{2M}{p}+\frac{2M^2(1-\alpha)}{\alpha p(1-p)}D(\epsilon_r||p)}\right] \\ &\approx (1-p)\sqrt{\frac{2p}{M}\Big(1+\frac{M(1-\alpha)}{\alpha p(1-p)}D(\epsilon_r||p)\Big)}.\end{aligned}$$

We may write the solution to $g_{01}(\epsilon_b, \epsilon_r) = 0$ as a function of ϵ_r: $\epsilon_{b(01)}(\epsilon_r) = p + \delta = p + (1-p)\sqrt{\frac{2p}{M}\Big(1+\frac{M(1-\alpha)}{\alpha p(1-p)}D(\epsilon_r||p)\Big)}$ for $\epsilon_r \in (p, 1)$. So (ϵ_b, ϵ_r) is in $\mathcal{A}_0$ if $\epsilon_b \in \big(0, \epsilon_{b(01)}(\epsilon_r)\big)$ for $\epsilon_r \in (p, x_0)$, and $\epsilon_b \in \big(0, \epsilon_{b(20)}(\epsilon_r)\big)$ for $\epsilon_r \in (x_0, 1)$. $\epsilon_{b(20)}(\epsilon_r)$ is calculated later.

- Set $g_{12}(\epsilon_b, \epsilon_r) = 0$. Then we get the solution is $\epsilon_r = x_0 = p + (1-p)\sqrt{\frac{2p}{M}}, \epsilon_b \in (p, 1)$. So (ϵ_b, ϵ_r) is in $\mathcal{A}_1$ if $\epsilon_b \in \big(\epsilon_{b(01)}(\epsilon_r), 1\big), \epsilon_r \in (p, x_0)$.

- Set $g_{20}(\epsilon_b, \epsilon_r) = 0$, which is

$$F(\epsilon_b) = \frac{1-\alpha}{\alpha}(1-\epsilon_r).$$

We use the same method, a Taylor expansion, to solve the problem. The above equation is equivalent to

$$\frac{M}{2p(1-p)}\delta^2 + \delta + p - 1 - (1-\epsilon_r)\frac{1-\alpha}{\alpha} = 0.$$

Then we have

$$\delta = \frac{p(1-p)}{M}\left[-1+\sqrt{\frac{2M}{p}[1+\frac{(1-\alpha)(1-\epsilon_r)}{p(1-p)}]}\right]$$
$$\approx (1-p)\sqrt{\frac{2M}{p}\left[1+\frac{(1-\alpha)(1-\epsilon_r)}{p(1-p)}\right]}.$$

Therefore, the solution to $g_{20}(\epsilon_b,\epsilon_r)=0$ is $\epsilon_{b(20}(\epsilon_r)=p+(1-p)\sqrt{\frac{2M}{p}\left[1+\frac{(1-\alpha)(1-\epsilon_r)}{p(1-p)}\right]}, \epsilon_r \in (p,1)$. So (ϵ_b,ϵ_r) is in $\mathcal{A}_2$ if $\epsilon_b \in [\epsilon_{b(20)}(\epsilon_r),1], \epsilon_r \in (x_0,1]$.

For a clear view of the regions, we plot it in figure 4-3 and given an example in figure 4-4.

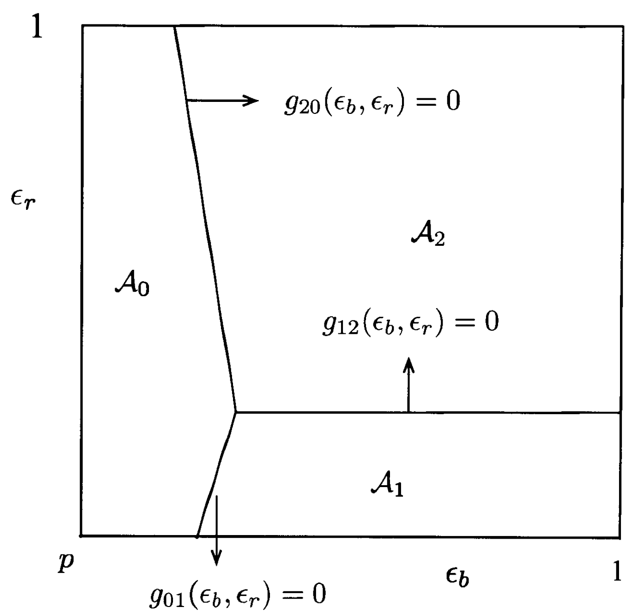

Figure 4-3: E in the intra-layer coding scheme when $p = 0.05, \alpha = 0.5, M = 10$. $\mathcal{A}_0 : \epsilon_b \in \big(p, \epsilon_{b(01)}(\epsilon_r)\big)$ for $\epsilon_r \in (p, x_0)$, and $\epsilon_b \in \big(p, \epsilon_{b(20)}(\epsilon_r)\big)$ for $\epsilon_r \in (x_0, 1), \mathcal{A}_1 : \epsilon_b \in \big(\epsilon_{b(01)}(\epsilon_r), 1\big), \epsilon_r \in (p, x_0), \mathcal{A}_2 : \epsilon_b \in \big(\epsilon_{b(20)}(\epsilon_r), 1\big), \epsilon_r \in (x_0, 1)$.

Next we want to find the maximum distortion exponent. Let $\epsilon^*_{b,\text{intra}}$ and $\epsilon^*_{r,\text{intra}}$ denote the operating points where maximum distortion exponent is achieved, where $E_0' = E_1' = E_2'$, i.e, $g_{01}(\epsilon_b,\epsilon_r)=0, g_{12}(\epsilon_b,\epsilon_r)=0$ and $g_{20}(\epsilon_b,\epsilon_r)=0$. Therefore, we obtain $\epsilon^*_{r,\text{intra}} = x_0 = p+(1-p)\sqrt{\frac{2p}{M}}$, $\epsilon^*_{b,\text{intra}}$ is the solution to $F(\epsilon^*_{b,\text{intra}}) = \frac{1-\alpha}{\alpha}(1-x_0)$.

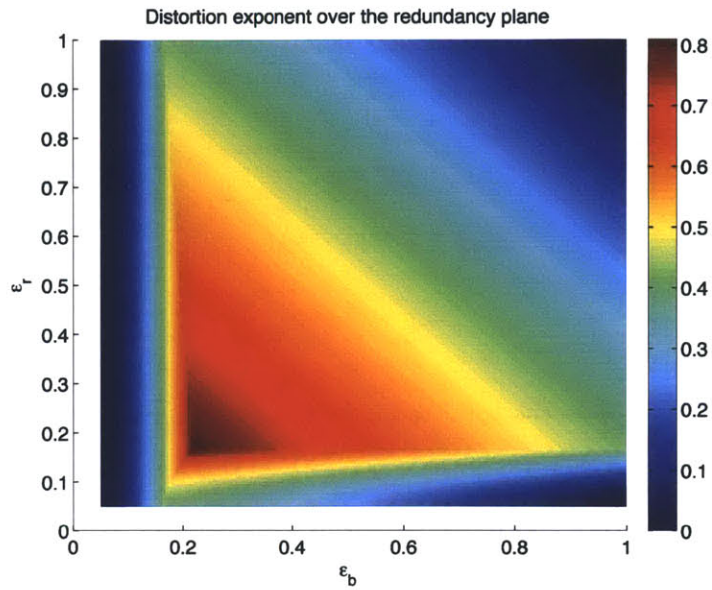

Figure 4-4: E in the intra-layer coding scheme when $p = 0.05, \alpha = 0.5, M = 10$.

Similarly, we use a Taylor expansion to solve the problem. Write $\epsilon^*_{b,\text{intra}}$ as $\epsilon^*_{b,\text{intra}} = p + \delta_b$ and substituting x_0 in this expression, we have

$$\frac{M}{2p(1-p)}\delta_b^2 + \delta_b - (1-p)\Big[1 + \frac{1-\alpha}{\alpha}(1 - \sqrt{\frac{2p}{M}})\Big] = 0 \tag{4.27}$$

Therefore

$$\begin{aligned}
\delta_b &= \frac{p(1-p)}{M}\Big(-1 + \sqrt{1 + \frac{2M}{p}\Big[1 + \frac{1-\alpha}{\alpha}(1 - \sqrt{\frac{2p}{M}})\Big]}\Big) && (4.28)\\
&\approx \frac{p(1-p)}{M}\Big(-1 + \sqrt{1 + \frac{2M}{\alpha p}}\Big) && (4.29)\\
&\approx \frac{p(1-p)}{M}\Big(-1 + \sqrt{\frac{2M}{\alpha p}}\Big) && (4.30)\\
&\approx \frac{p(1-p)}{M}\sqrt{\frac{2M}{\alpha p}} && (4.31)\\
&= (1-p)\sqrt{\frac{2p}{\alpha M}} && (4.32)
\end{aligned}$$

Because $M >> p$, (4.29) follows from $\sqrt{\frac{2p}{M}} << 1$, (4.30) follows from the fact that $\frac{2M}{\alpha p} >> 1$; (4.31) follows from $\sqrt{\frac{2M}{\alpha p}} >> 1$. Then we can obtain an approximate solution for $\epsilon^*_{b,\text{intra}}$ and $\epsilon^*_{r,\text{intra}}$

$$\epsilon^*_{b,\text{intra}} = p + (1-p)\sqrt{\frac{2p}{\alpha M}} \tag{4.33}$$

$$\epsilon^*_{r,\text{intra}} = p + (1-p)\sqrt{\frac{2p}{M}} \tag{4.34}$$

Let $\epsilon^*_{\text{intra}}$ and E^*_{intra} denote the total redundancy and distortion exponent at the optimal operating points, then we have

$$\epsilon^*_{\text{intra}} = p + (1-p)(1+\sqrt{\alpha}-\alpha)\sqrt{\frac{2p}{M}} \tag{4.35}$$

$$E^*_{\text{intra}} = 1 - \epsilon^*_{\text{intra}} = (1-p)\left[1 - (1+\sqrt{\alpha}-\alpha)\sqrt{\frac{2p}{M}}\right] \tag{4.36}$$

Notice that in equation (4.28), (4.29) and (4.30), $\sqrt{\frac{2p}{M}}$ is approximately 0 and therefore neglected because it is compared with 1. But in equation (4.33), (4.34), (4.35) and (4.36), p is also a small value, so $\sqrt{\frac{2p}{M}}$ cannot be neglected.

4.2.3 Inter-layer Coding

In this case, we have

$$E_0' = \begin{cases} M\alpha D(\epsilon_b||p) + M(1-\alpha)D(\epsilon_r||p) & \epsilon_b \geq \epsilon_r \\ MD\left[\alpha\epsilon_b + (1-\alpha)\epsilon_r||p\right] & \epsilon_b < \epsilon_r \end{cases} \tag{4.37}$$

$$E_1' = M(1-\alpha)D(\epsilon_r||p) + \alpha(1-\epsilon_b) \tag{4.38}$$

$$E_2' = \alpha(1-\epsilon_b) + (1-\alpha)(1-\epsilon_r). \tag{4.39}$$

According to the definition of distortion exponent, we have

$$E = \begin{cases} E_0^{'} & (\epsilon_b, \epsilon_r) \in \mathcal{B}_0 \\ E_1^{'} & (\epsilon_b, \epsilon_r) \in \mathcal{B}_1 \\ E_2^{'} & (\epsilon_b, \epsilon_r) \in \mathcal{B}_2 \end{cases} \cdot \tag{4.40}$$

We denote the region where $E = E_i^{'}$ as $\mathcal{B}_i$ where $i = 0, 1, 2$. In order to find the boundaries of $\mathcal{B}_i$s, define $H_{ij} = E_i^{'} - E_j^{'}$. Then

$$\begin{aligned} H_{01} &= E_0^{'} - E_1^{'} \\ &= \begin{cases} \alpha F(\epsilon_b) & \epsilon_b \geq \epsilon_r \\ F\big[\alpha\epsilon_b + (1-\alpha)\epsilon_r\big] - (1-\alpha)F(\epsilon_r) & \epsilon_b < \epsilon_r \end{cases} \\ H_{12} &= E_1^{'} - E_2^{'} \\ &= (1-\alpha)F(\epsilon_r) \\ H_{20} &= E_2^{'} - E_0^{'} \\ &= \begin{cases} -\big[\alpha F(\epsilon_b) + (1-\alpha)F(\epsilon_r)\big] & \epsilon_b \geq \epsilon_r \\ -F(\alpha\epsilon_b + (1-\alpha)\epsilon_r) & \epsilon_b < \epsilon_r. \end{cases} \end{aligned}$$

- Set $H_{01} = 0$, the solution is: $\epsilon_b = x_0$ when $\epsilon_b \geq \epsilon_r$ and $\epsilon_b = h_{01}(\epsilon_r)$ when $\epsilon_b < \epsilon_r$. So (ϵ_b, ϵ_r) is in $\mathcal{B}_0$ if $\epsilon_b \in (p, x_0), \epsilon_r \in (p, x_0)$, and $\epsilon_b \in (p, \frac{x_0-(1-\alpha)\epsilon_r}{\alpha}), \epsilon_r \in (x_0, 1)$.

- Set $H_{12} = 0$, the solution is $\epsilon_r = x_0$. So (ϵ_b, ϵ_r) is in $\mathcal{B}_1$ if $\epsilon_b \in (x_0, 1], \epsilon_r \in (p, x_0)$.

- Set H_{20}=0, the solution is $\epsilon_b = h_{20}(\epsilon_r)$ when $\epsilon_b \geq \epsilon_r$ and $\alpha\epsilon_b + (1-\alpha)\epsilon_r = x_0$ when $\epsilon_b < \epsilon_r$. So (ϵ_b, ϵ_r) is in $\mathcal{B}_2$ if $\epsilon_b \in (\frac{x_0-(1-\alpha)\epsilon_r}{\alpha}, 1), \epsilon_r \in (x_0, \frac{x_0-p\epsilon_b}{1-\alpha})$.

For a clear view of the regions, we plot it in figure 4-5 and given an example in figure 4-6.

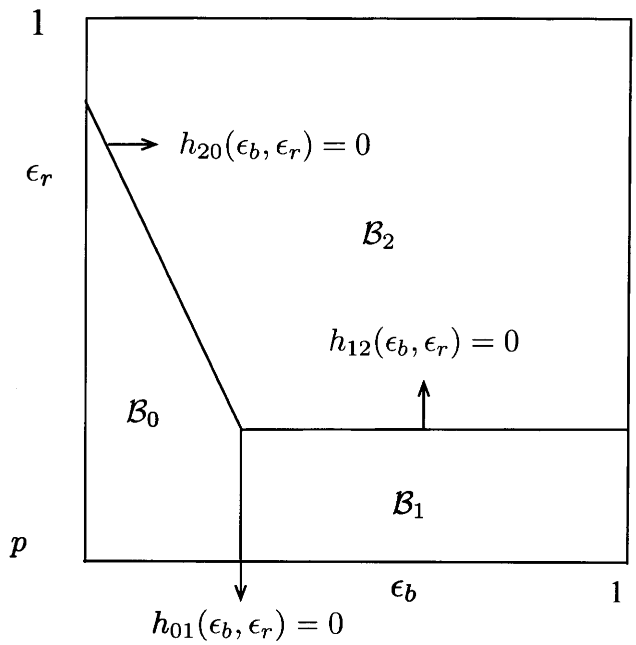

Figure 4-5: Distortion exponent E in a inter-layer coding scheme, $p = 0.05, \alpha = 0.5, M = 10$. $\mathcal{B}_0 : \epsilon_b \in (p, x_0), \epsilon_r \in (p, x_0)$, and $\epsilon_b \in \big(p, \frac{x_0-(1-\alpha)\epsilon_r}{\alpha}\big), \epsilon_r \in (x_0, 1), \mathcal{B}_1 : \epsilon_b \in (x_0, 1), \epsilon_r \in (p, x_0), \mathcal{B}_2 : \epsilon_b \in \big(\frac{x_0-(1-\alpha)\epsilon_r}{\alpha}, 1\big), \epsilon_r \in \big(x_0, \frac{x_0-p\epsilon_b}{1-\alpha}\big)$.

For any α, the minimal distortion is achieved is at the maximal E, where

$$\alpha\epsilon_b + (1-\alpha)\epsilon_r = x_0, \epsilon_b \in (p, x_0). \tag{4.41}$$

Let $\epsilon^*_{\text{inter}}$ and E^*_{inter} denote the total redundancy distortion exponent at the optimal operating points, we have

$$\epsilon^*_{\text{inter}} = x_0 \tag{4.42}$$

$$E^*_{\text{inter}} = 1 - \epsilon^*_{\text{inter}} = (1-p)\Big(1 - \sqrt{\frac{2p}{M}}\Big). \tag{4.43}$$

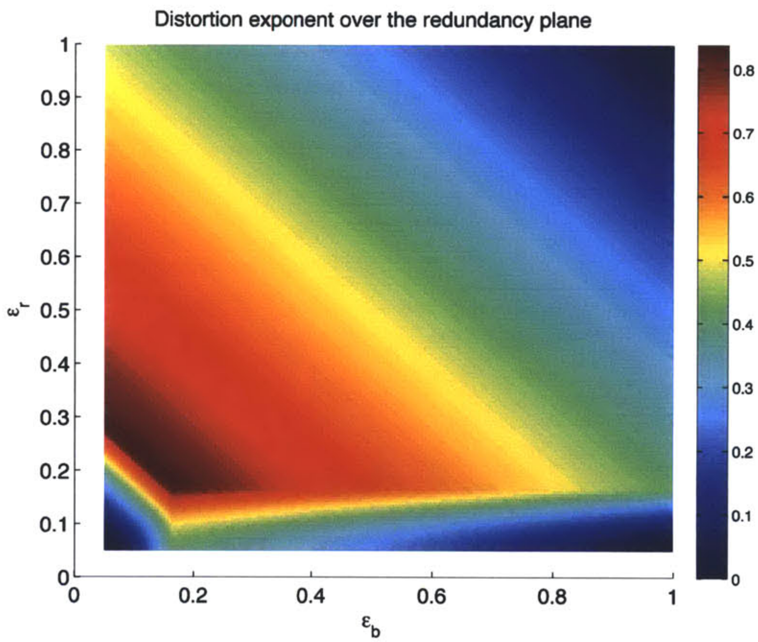

Figure 4-6: Distortion exponent E in a inter-layer coding scheme, $p = 0.05, \alpha = 0.5, M = 10$.

Chapter 5

Coding Strategy

In the previous chapter, we derived the distortion exponents over the redundancy plane for intra-layer coding and inter-layer coding schemes. In this chapter, we first join the redundancy plane for the two coding schemes and then divide the redundancy plane into different regions. Coding strategies are provided considering the two performance metrics, average distortion and total redundancy. Note that the total redundancy of any operating point depends only on its location in the redundancy plane, and is independent of which coding scheme is chosen. Therefore, when we divide the redundancy plane, we first consider which coding scheme is better based on the distortion exponent. After one of the two is chosen, we then use total redundancy as a metric to further divide the regions.

5.1 Dividing the redundancy plane

The redundancy plane can be divided into $\mathcal{A}_i$s for intra-layer coding, and $\mathcal{B}_i$s for inter-layer coding. We've found the boundaries of the regions in the previous chapter and now plot them in the same redundancy plane. These boundaries then divide the redundancy plane into five regions: $\mathcal{A}_1, \mathcal{A}_2, \mathcal{B}_0, \mathcal{B}_1 - \mathcal{A}_1$ and $\mathcal{B}_2 - \mathcal{A}_2$. We use E_{intra} to denote the intra-layer distortion exponent and use E_{inter} to denote the inter-layer

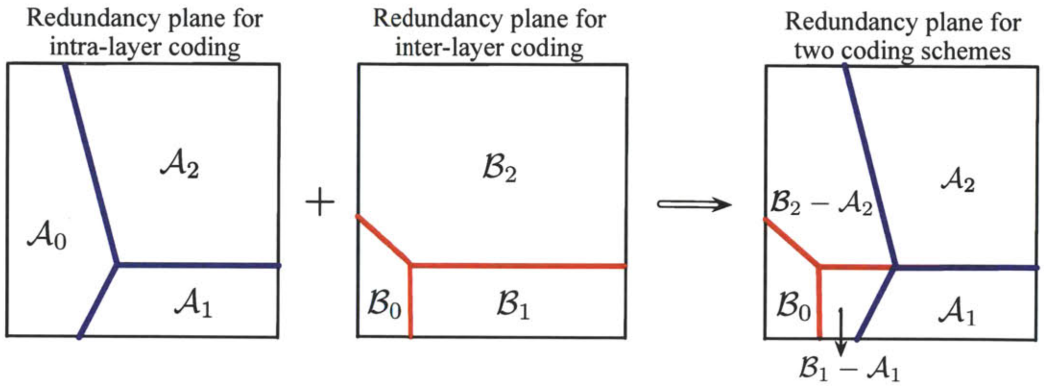

Figure 5-1: The redundancy plane for the two coding schemes, divided into 5 regions

distortion exponent. Compare E_{intra} and E_{inter} in each of the five regions, we have:

$$\mathcal{A}_1 \cup \mathcal{A}_2 : E_{\text{intra}} = E_{\text{inter}}$$
$$\mathcal{A}_0 : E_{\text{intra}} < E_{\text{inter}}.$$

The first coding strategy we provide is to divide the redundancy plane into two regions, according to the distortion exponent.

- **Indifferent coding region**: $\mathcal{A}_1 \cup \mathcal{A}_2$

  At any operating point in this region, both schemes achieve the same performance in terms of average distortion and total redundancy. Therefore, we can choose either of them.

- **Inter-layer coding region**: $\mathcal{A}_0$

  At any operating point in this region, although the total redundancy is the same for both coding schemes, inter-layer coding can achieve a lower distortion. So we choose inter-layer coding over the other.

Thus we can draw our first important conclusion: *Inter-layer coding can always do at least as well as intra-layer coding.* At any point in the redundancy plane, mixing never hurts the performance from this perspective.

In the previous section, we proved that in both coding schemes, the optimal solutions satisfy the following equation: $E = 1 - \epsilon$. Now we can further prove that this also holds as long as $E = E_2'$, i.e, the operating point is in region $\mathcal{A}_2$ in the intra-layer coding scheme and region $\mathcal{B}_2$ in the inter-layer coding scheme. The proof for this is straightforward based on equation (4.25), equation (4.39) and the definition of total redundancy (3.8). Therefore we formulate the statement into a lemma.

Lemma: In region $\mathcal{A}_2$ in the intra-layer coding redundancy plane and $\mathcal{B}_2$ in the intre-layer coding redundancy plane, the distortion exponent E and total redundancy ϵ satisfy: $E = 1 - \epsilon$.

We also have that $E[D] = e^{-NE}$ [3]. So for the points in region $\mathcal{A}_2$ or $\mathcal{B}_2$, the lower the distortion, the less the redundancy. This can be interpreted as there being trade-off between the quality and efficiency. The redundancy lines are drawn in figure 3-1. The direction of decreasing the total redundancy is perpendicular to redundancy lines and moves towards the origin. Therefore, the performance in terms of quality and efficiency improves if moving along this direction, as long as the operating points stay within the region of $\mathcal{A}_2$ for intra-layer coding, and $\mathcal{B}_2$ for inter-layer coding. The optimal operating point is at the intersection point of the three regions for intra-layer coding, and the boundary of $\mathcal{B}_0$ and $\mathcal{B}_2$ for inter-layer coding, as shown in figure 5-2.

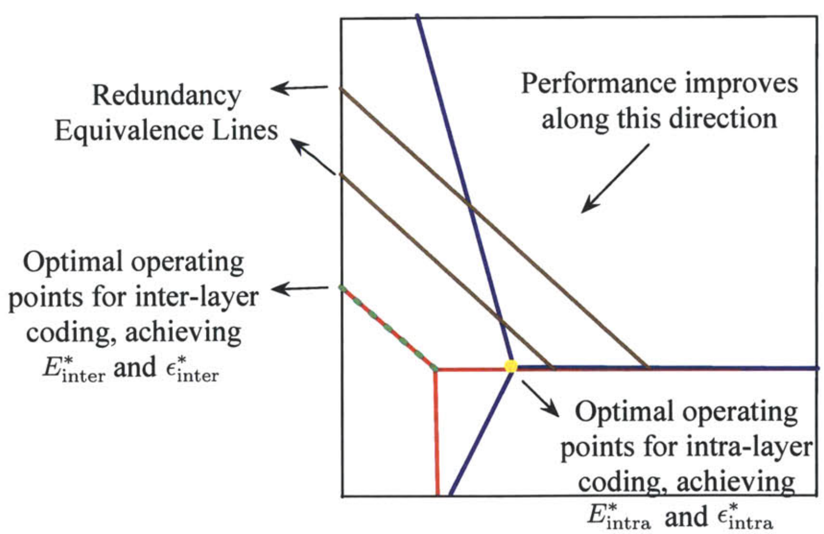

Figure 5-2: Redundancy equivalence lines and optimal operating points for the two coding schemes.

In a redundancy plane, points on the same redundancy equivalence line have equal total redundancy. In region $\mathcal{A}_2$ in the intra-layer coding case and $\mathcal{B}_2$ in the inter-layer coding case, they also have the same distortion level. And we therefore say these points achieve same performance. $\mathcal{B}_2 - \mathcal{A}_2$ belongs to the Inter-layer Coding Region, for a smaller distortion than the other coding scheme. However, for any operating point in this region, we can draw a redundancy equivalence line through that point and into the neighboring region $\mathcal{A}_2$. With either intra-layer or inter-layer coding, points on this line section in $\mathcal{A}_2$ can achieve the same average distortion and total redundancy. So they are equivalent to the original points. Therefore, we call $\mathcal{B}_2 - \mathcal{A}_2$ the **Replaceable inter-layer coding region**, indicated by a yellow shaded area in figure 5-3. The redundancy equivalence line can keep moving down until it hits the yellow line in figure 5-3 where the distortion exponent reaches E^*_{intra}. It is the minimum achievable distortion for intra-layer coding. Below that, there's no equivalent operating points for intra-layer coding to achieve the same performance as the inter-layer coding. So the rest of region $\mathcal{A}_0$ is referred to as **Irreplaceable inter-layer coding region**, .

5.2 Choosing the operating region

To conclude, the coding strategy for a given redundancy plane is:

1. If the given redundancy (ϵ_b, ϵ_r) is in Indifferent coding region, then we could choose either coding scheme, each of which will provide the same performance.

2. If the given redundancy (ϵ_b, ϵ_r) is in inter-layer coding region, then we could always choose the inter-layer coding scheme to get a lower distortion. There are two sub cases:

    - If the operating point is in the Replaceable inter-layer coding region with a fixed total redundancy, but flexible (ϵ_b, ϵ_r), we can move along the redundancy equivalence line into inter-changeable coding region to change for the intra-layer coding scheme.

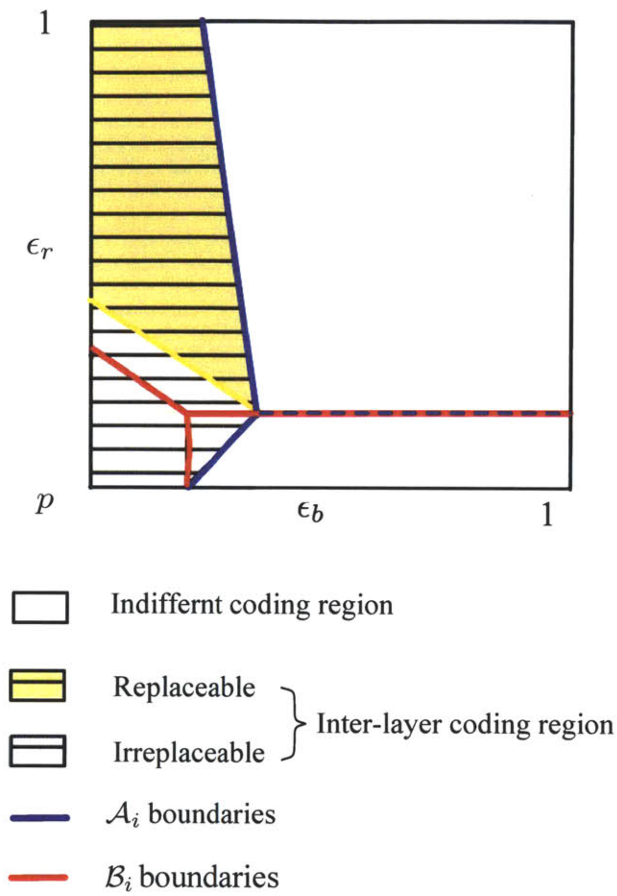

Figure 5-3: Redundancy plane – coding strategy based on different regions

- If the operating point is in the Replaceable inter-layer coding region, then we should always choose inter-layer coding over intra-layer coding.

Chapter 6

Simulations and Analysis

In this chapter, we present simulations to test our analytical results and give an intuitive view of the division of redundancy plane into different regions and how the coding strategy works.

6.1 Contour Maps

The distortion exponents are plotted in figure 4-4 and figure 4-6. Observing the achievable distortions in different regions, we could see that point in $\mathcal{A}_2$ and $\mathcal{B}_2$ can achieve smaller distortions than operating points in other coding regions. They are the recommended coding regions for the two coding schemes. The contour maps of $\mathcal{A}_2$ and $\mathcal{B}_2$ are plotted in figure 6-1 for intra-layer coding, and figure 6-2 for inter-layer coding. The contour lines are evenly spaced, meaning a linear increase of the distortion exponent perpendicular to the lines. This reinforces the lemma we proved before that in $\mathcal{A}_2$ and $\mathcal{B}_2$, the distortion exponent and total redundancy satisfy: $E = 1 - \epsilon$. The two figures also show the operable regions for different requirement of distortion. For example, if the receiver requires a distortion smaller than $e^{-0.8E^*_{\text{intra}}}$ for intra-layer coding, then the operable region is within the orange triangle in figure 6-1.

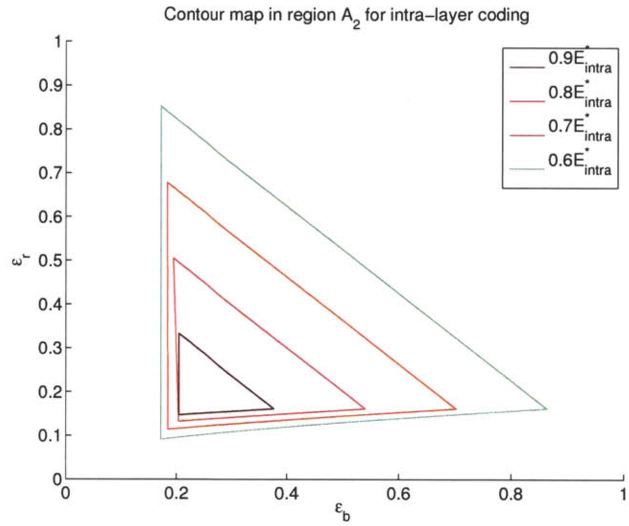

Figure 6-1: Contour map of $\mathcal{A}_2$ for intra-layer coding

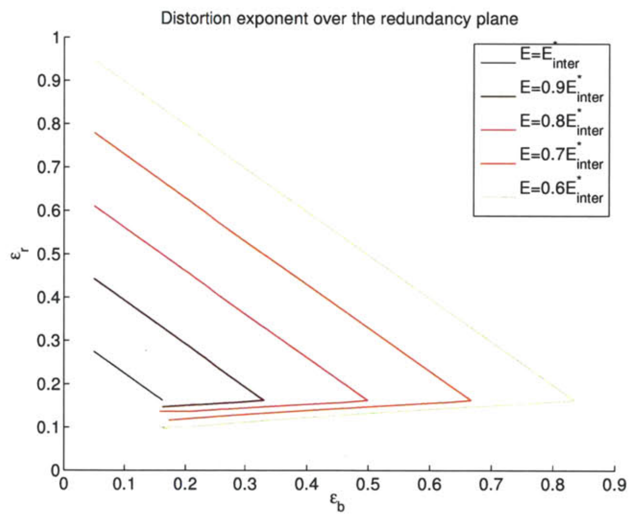

Figure 6-2: Contour map of $\mathcal{B}_2$ for inter-layer coding

6.2 Optimal performance

We consider the optimal achievable performance in the two coding cases. We showed in chapter 4 the minimum distortions achievable for intra-layer and inter-layer coding are as below:

- Intra-layer coding

$$\epsilon^*_{\text{intra}} \approx p + (1-p)(1+\sqrt{\alpha}-\alpha)\sqrt{\frac{2p}{M}} \tag{6.1}$$

$$E^*_{\text{intra}} \approx 1 - \epsilon^*_{\text{intra}} = (1-p)\left[1 - (1+\sqrt{\alpha}-\alpha)\sqrt{\frac{2p}{M}}\right]. \tag{6.2}$$

- Inter-layer coding

$$\epsilon^*_{\text{inter}} \approx p + (1-p)\sqrt{\frac{2p}{M}} \tag{6.3}$$

$$E^*_{\text{inter}} \approx 1 - \epsilon^*_{\text{inter}} = (1-p)\left(1 - \sqrt{\frac{2p}{M}}\right). \tag{6.4}$$

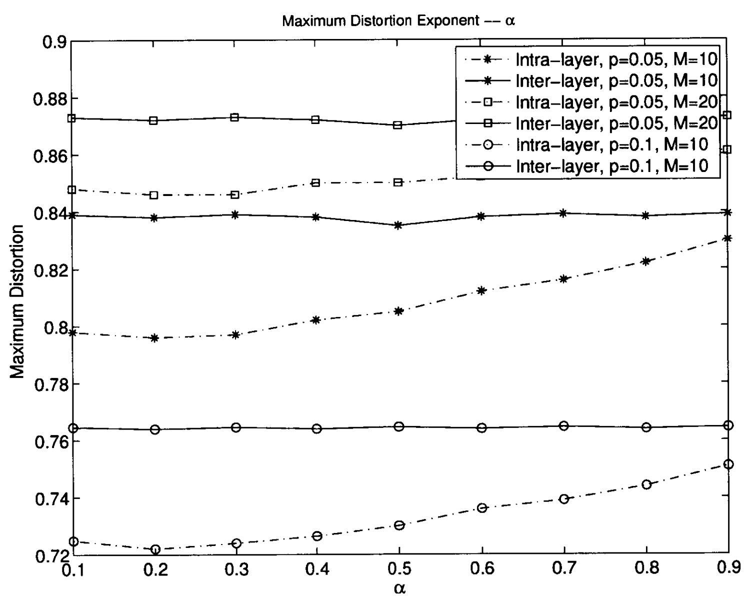

Figure 6-3: Simulation results for different p and M

We look at each coding scheme with different p and M. Figure 6-3 shows the larger p, the greater distortion. Channel erasures hurt the performance by decreasing the decoding probability, and therefore results in a greater distortion. The figure also shows that coding among multi-streams helps performance. The greater M, the smaller the distortion. Thus cooperation among streams are encouraged when the receiver wants to receive multiple streams. No matter what parameter we choose, we may see that inter-layer coding always has a slightly larger distortion exponent than intra-layer coding scheme. As N grows, such difference may translate into considerable benefit in average distortion.

We also observe that in the intra-layer coding case, the maximum E is a function of α as shown in equation (4.36), while the maximum E for inter-layer coding is independent of α, as in equation (4.43). In the latter case, it means no matter what's the α, we can always achieve the same best performance by adjusting the redundancy in the base and refinement layer. Thus we come to our second important conclusion:*For inter-layer coding, the optimal performance is independent of the base-layer percentage.* This could be explained by that inter-layer coding mixes the information of both layers so that the information in the base layer is also carried in the refinement layer and therefore decreases the dependence on the base layer.

6.3 Operable Regions

In the previous part, we compare the performance of two coding schemes with the same parameters, p and M. Now we assume they achieve the same distortion, and see what is the operable regions to achieve the performance.

Assume both coding schemes work in the recommended coding region, $\mathcal{A}_2(\mathcal{B}_2)$. We use some distortion exponent E scaled to different levels to indicate varying requirements for the coding schemes. We plot the operating points for the two coding schemes that achieve the required distortions. Figure 6-4 shows that the operating points lie on redundancy equivalence lines. For the two coding schemes, these line redundancy equivalence line sections overlap in $\mathcal{A}_2$. Inter-layer coding has longer line sections corresponding to a distortion level and therefore more operable points. While ϵ_b can be as small as p for inter-layer coding, intra-layer coding has a tighter constraint on ϵ_b. This is because since the refinement layer does not provide any help in decoding, the base layer has to carry more redundancy to provide the same coding reliability as the inter-layer coding. Therefore region $\mathcal{B}_2 - \mathcal{A}_2$ in figure 6-4 is operable for inter-layer coding but not operable for intra-layer coding due to lack of base layer redundancy. We also point out in chapter 5 that $\mathcal{B}_2 - \mathcal{A}_2$ does not provide more benefit in terms of distortion and total redundancy. And we therefore refer to it as replaceable. The green area in figure 6-4 belongs to the irreplaceable inter-layer coding region. Operating points in this region can achieve smallest distortions, compared to other operating points using either intra-layer or inter-layer coding scheme and can not be replaced by points in other regions.

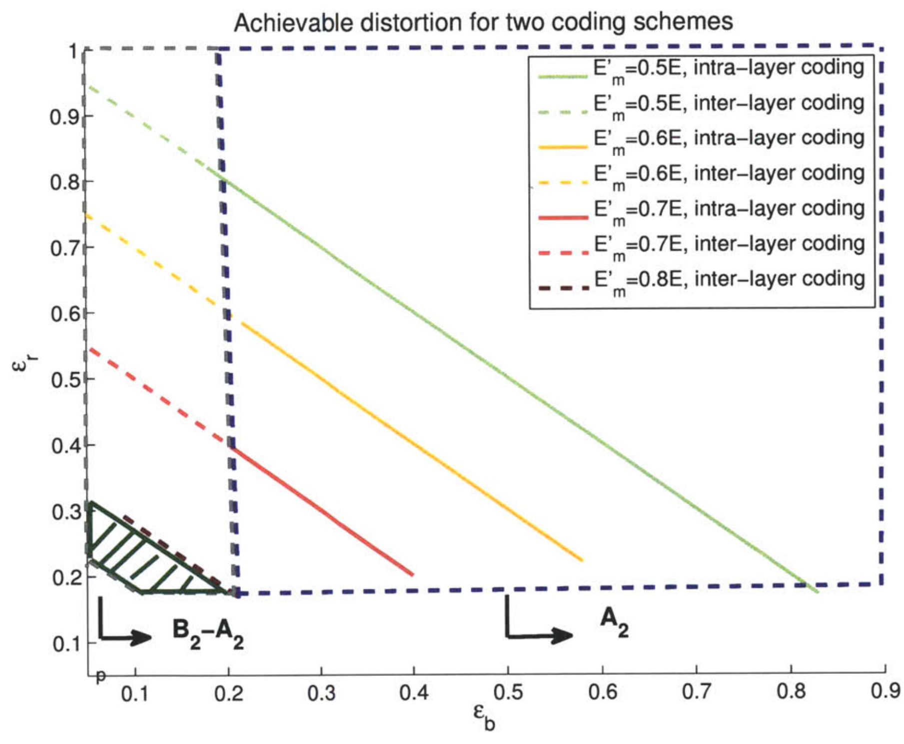

Figure 6-4: Operable regions for intra-layer and inter-layer coding

Chapter 7

Conclusions and Future Work

In this thesis, a joint MR and NC model is developed. We present two coding schemes, intra-layer network coding and inter-layer network coding with MRC. The encoding and decoding process of the two schemes are described in detail, together with the structure of the codes. We provide coding strategies based on the redundancy plane which considers both the quality and efficiency. We show two important conclusions: 1)Inter-layer coding always achieves at least the same performance, in terms of average distortion and total redundancy, as intra-layer coding. The mixing of layers never hurts from this perspective. 2) The optimal performance for inter-layer coding is independent of the base layer percentage. This provides more flexibility in system design. Besides the two observations, we also point out that in some region, inter-layer coding can be replaced by changing the parameters without harming the quality or efficiency. The coding strategies we propose could provide guidelines in designing a joint MR and NC system and answer the questions such as when do we code across the layers or within a layer, and how much redundancy is needed to protect the MRC.

Based on the current work, we can extend the two-layer MRC to multi-layer MRC. There are several aspects we need to consider. First, how does inter-layer network coding mix the layers. In this thesis, we only have two layers, so there is only two mixing options, either mix two layers or not. But for multiple-layer MRC, there are more choices in terms of which layer should mix and how many should be mixed together. Second, in our work, we use the base layer percentage α to represent the allocation of codeword length to base layer in the source coding part. If there are multiple layers in the MRC, we'll need a more complicated joint design of source coding and network coding, considering how the layers are mixed together and how much redundancy put in different layers.

Appendix A

Derivation of E_0' in inter-layer coding

We have

$$E_0 = \min_{\gamma_b \in (\epsilon_b, 1]} \left[\alpha \cdot D(\gamma_b || p) + (1-\alpha) \cdot \min_{\gamma_r \in [(\epsilon_b - \gamma_b)\frac{\alpha}{1-\alpha} + \epsilon_r, 1]} D(\gamma_r || p)\right].$$

We first take a look at $\min_{\gamma_r \in [(\epsilon_b - \gamma_b)\frac{\alpha}{1-\alpha} + \epsilon_r, 1]} D(\gamma_r || p)$. Note that, for a fixed p, $D(\gamma_r || p)$ is a convex function of γ_r and the minimum 0 is achieved at $\gamma_r = p$. Therefore, we discuss the two cases where $(\epsilon_b - \gamma_b)\frac{\alpha}{1-\alpha} + \epsilon_r \leq p$ and $(\epsilon_b - \gamma_b)\frac{\alpha}{1-\alpha} + \epsilon_r > p$.

1. $(\epsilon_b - \gamma_b)\frac{\alpha}{1-\alpha} + \epsilon_r \leq p$
   Then we have $\gamma_b \geq \epsilon_b + (\epsilon_r - p)\frac{1-\alpha}{\alpha} > \epsilon_b$. We also have $\gamma_b \in (\epsilon_b, 1)$, therefore, $\gamma_b \in [\epsilon_b + (\epsilon_r - p)\frac{1-\alpha}{\alpha}, 1]$.

$$\begin{aligned} E_0 &= \min_{\gamma_b \in \epsilon_b + (\epsilon_r - p)\frac{1-\alpha}{\alpha}, 1]} \alpha \cdot D(\gamma_b || p) \\ &= \alpha \cdot D\left[\epsilon_b + (\epsilon_r - p)\frac{1-\alpha}{\alpha} || p\right]. \end{aligned} \tag{A.1}$$

2. $(\epsilon_b - \gamma_b)\frac{\alpha}{1-\alpha} + \epsilon_r > p$
   Then we have $E_0 = \min_{\gamma_b \in [\epsilon_b, 1]} \left[\alpha \cdot D(\gamma_b || p) + (1-\alpha) \cdot D\left[(\epsilon_b - \gamma_b)\frac{\alpha}{1-\alpha} + \epsilon_r || p\right]\right]$ and $\gamma_b < \epsilon_b + (\epsilon_r - p)\frac{1-\alpha}{\alpha}$. We also have $\gamma_b \in (\epsilon_b, 1]$, so we need to compare

$\epsilon_b + (\epsilon_r - p)\frac{1-\alpha}{\alpha}$ with 1 to find the range of γ_b.

(a) $\epsilon_b + (\epsilon_r - p)\frac{1-\alpha}{\alpha} \leq 1$

Then $\gamma_b \in [\epsilon_b, \epsilon_b + (\epsilon_r - p)\frac{1-\alpha}{\alpha}]$

$$E_0 = \min_{\gamma_b \in [\epsilon_b, \epsilon_b + (\epsilon_r - p)\frac{1-\alpha}{\alpha}]} \Big[\alpha \cdot D(\gamma_b||p) + (1-\alpha) \cdot D\big[(\epsilon_b - \gamma_b)\frac{\alpha}{1-\alpha} + \epsilon_r||p\big]\Big]$$

Take the derivative of E_0 with respect of γ_b to find the minimum.

$$\frac{\partial E_0}{\partial \gamma_b} = \alpha\big[\log\frac{\gamma_b}{1-\gamma_b} - \log\frac{(\epsilon_b - \gamma_b)\frac{\alpha}{1-\alpha} + \epsilon_r}{1 - [(\epsilon_b - \gamma_b)\frac{\alpha}{1-\alpha} + \epsilon_r]}\big].$$

Note that the function $\log\frac{x}{1-x}$ increases monotonically with x so $\frac{\partial E_0}{\partial \gamma_b} = 0$ when

$$\gamma_b^* = (\epsilon_b - \gamma_b^*)\frac{\alpha}{1-\alpha} + \epsilon_r$$

So we obtain:

$$\gamma_b^* = \alpha\epsilon_b + (1-\alpha)\epsilon_r.$$

Therefore E_0 decreases when $\gamma_b < \gamma_b^*$ and increases when $\gamma_b > \gamma_b^*$. Incorporating the range, $\gamma_b \in [\epsilon_b, \epsilon_b + (\epsilon_r - p)\frac{1-\alpha}{\alpha}]$, it leads to two following cases:

i. $\epsilon_b \geq \epsilon_r$

In this case, $\gamma_b^* \leq \epsilon_b$. E_0 increases monotonically in the region $[\epsilon_b, \epsilon_b + (\epsilon_r - p)\frac{1-\alpha}{\alpha}]$. So the minimum is at $\gamma_b = \epsilon_b$

$$E_0 = \alpha D(\epsilon_b||p) + (1-\alpha)D(\epsilon_r||p) \tag{A.2}$$

ii. $\epsilon_b < \epsilon_r$

In this case, $\gamma_b^* \in [\epsilon_b, \epsilon_b + (\epsilon_r - p)\frac{1-\alpha}{\alpha}]$. Therefore the minimum is at

$\gamma_b = \gamma_b^*$

$$E_0 = D[\alpha\epsilon_b + (1-\alpha)\epsilon_r||p] \tag{A.3}$$

(b) $\epsilon_b + (\epsilon_r - p)\frac{1-\alpha}{\alpha} > 1$

In this case, $\gamma_b \in [\epsilon_b, 1]$.

$$E_0 = \min_{\gamma_b \in [\epsilon_b, 1]} \left[\alpha \cdot D(\gamma_b||p) + (1-\alpha) \cdot D\left[(\epsilon_b - \gamma_b)\frac{\alpha}{1-\alpha} + \epsilon_r||p\right]\right]$$

We need to compare ϵ_b with $\alpha\epsilon_b + (1-\alpha)\epsilon_r$, which also leads to the following two cases

i. $\epsilon_b \geq \epsilon_r$

Then $\alpha\epsilon_b + (1-\alpha)\epsilon_r \leq \epsilon_b$, so the minimum is achieved at $\gamma_b = \epsilon_b$

$$E_0 = \alpha D(\epsilon_b||p) + (1-\alpha)D(\epsilon_r||p). \tag{A.4}$$

ii. $\epsilon_b < \epsilon_r$

Then $\alpha\epsilon_b + (1-\alpha)\epsilon_r > \epsilon_b$, so the minimum is achieved at $\gamma_b = \gamma_b^*$

$$E_0 = D\left[\alpha\epsilon_b + (1-\alpha)\epsilon_r||p\right]. \tag{A.5}$$

Combining equations (A.1), (A.2), (A.3), (A.4) and (A.1), we have

$$E_0(\alpha) = \begin{cases} \min\left\{\alpha D(\epsilon_b||p) + (1-\alpha)D(\epsilon_r||p), \alpha D\left[\epsilon_b + (\epsilon_r - p)\frac{1-\alpha}{\alpha}||p\right]\right\} & \epsilon_b \geq \epsilon_r \\ \min\left\{D\left[\alpha\epsilon_b + (1-\alpha)\epsilon_r||p\right], \alpha D\left[\epsilon_b + (\epsilon_r - p)\frac{1-\alpha}{\alpha}||p\right]\right\} & \epsilon_b < \epsilon_r \end{cases}$$

Now we want to find the minimum when $\epsilon_b \geq \epsilon_r$ and $\epsilon_b < \epsilon_r$

1. $\epsilon_b \geq \epsilon_r$

Let $F_1(\epsilon_b, \epsilon_r) \triangleq \alpha D\left[\epsilon_b + (\epsilon_r - p)\frac{1-\alpha}{\alpha}||p\right] - \alpha D(\epsilon_b||p) - (1-\alpha)D(\epsilon_r||p)$. Take the derivative of $F_1(\epsilon_b, \epsilon_r)$ with respect to ϵ_b:

$$\begin{aligned}\frac{\partial F_1(\epsilon_b, \epsilon_r)}{\partial \epsilon_b} &= \alpha(\log \frac{h_1}{1-h_1} - \log \frac{p}{1-p}) - \alpha(\log \frac{\epsilon_b}{1-\epsilon_b} - \log \frac{p}{1-p}) \\ &= \alpha(\log \frac{h_1}{1-h_1} - \log \frac{\epsilon_b}{1-\epsilon_b})\end{aligned}$$

where

$$h_1 = \epsilon_b + (\epsilon_r - p)\frac{1-\alpha}{\alpha} > \epsilon_b.$$

So

$$\frac{\partial F_1(\epsilon_b, \epsilon_r)}{\partial \epsilon_b} > 0.$$

$F_1(\epsilon_b, \epsilon_r)$ increases monotonically with ϵ_b. Therefore for $\epsilon_b \geq \epsilon_r$,

$$\begin{aligned}F_1(\epsilon_b, \epsilon_r) &\geq F_1(\epsilon_b = \epsilon_r, \epsilon_r) \\ &= \alpha D\big[\frac{1}{\alpha}\epsilon_r - (\frac{1}{\alpha} - 1)p||p\big] - D(\epsilon_r||p).\end{aligned}$$

We want to compare $F_1(\epsilon_b = \epsilon_r, \epsilon_r)$ with 0. Taking the derivative of $F_1(\epsilon_b = \epsilon_r, \epsilon_r)$ with respect to ϵ_r, we obtain:

$$\frac{\partial F_1(\epsilon_b = \epsilon_r, \epsilon_r)}{\partial \epsilon_r} = \log \frac{g_1}{1-g_1} - \log \frac{\epsilon_r}{1-\epsilon_r},$$

where

$$g_1 = \frac{1}{\alpha}\epsilon_r - (\frac{1}{\alpha} - 1)p > \epsilon_r,$$

so

$$\frac{\partial F_1(\epsilon_b = \epsilon_r, \epsilon_r)}{\partial \epsilon_r} > 0.$$

$F_1(\epsilon_b = \epsilon_r, \epsilon_r)$ increases monotonically with ϵ_r. Therefore, for $\epsilon_r > p$

$$F_1(\epsilon_b = \epsilon_r, \epsilon_r) \geq F_1(\epsilon_b = \epsilon_r, \epsilon_r = p) = 0.$$

Thus, $F_1(\epsilon_b, \epsilon_r) > 0$ for $\epsilon_b \geq \epsilon_r$, which means that

$$\min\Big\{\alpha D(\epsilon_b||p) + (1-\alpha)D(\epsilon_r||p), \alpha D\big[\epsilon_b + (\epsilon_r - p)\frac{1-\alpha}{\alpha}||p\big]\Big\} = \alpha D(\epsilon_b||p) + (1-\alpha)D(\epsilon_r||p).$$

2. $\epsilon_b < \epsilon_r$

Let $F_2(\epsilon_b, \epsilon_r) \triangleq \alpha D\big[\epsilon_b + (\epsilon_r - p)\frac{1-\alpha}{\alpha}||p\big] - D\big[\alpha\epsilon_b + (1-\alpha)\epsilon_r||p\big]$ and take the derivative of F_2 with respect to ϵ_b :

$$\frac{\partial F_2(\epsilon_b, \epsilon_r)}{\partial \epsilon_b} = \alpha(\log\frac{h_{2,1}}{1-h_{2,1}} - \log\frac{h_{2,2}}{1-h_{2,2}})$$

where

$$h_{2,1} \triangleq \epsilon_b + (\epsilon_r - p)\frac{1-\alpha}{\alpha}$$
$$h_{2,2} \triangleq \alpha\epsilon_b + (1-\alpha)\epsilon_r.$$

We compare $h_{2,1}$ and $h_{2,2}$. Let

$$\begin{aligned} h_2 &\triangleq h_{2,1} - h_{2,2} \\ &= \frac{1-\alpha}{\alpha}[\alpha\epsilon_b + (1-\alpha)\epsilon_r - p] \\ &> \frac{1-\alpha}{\alpha}[\alpha\epsilon_b + (1-\alpha)\epsilon_b - p] \\ &= \frac{1-\alpha}{\alpha}(\epsilon_b - p) \\ &> 0. \end{aligned}$$

Therefore $h_{2,1} > h_{2,2}$, and then $\frac{\partial F_2(\epsilon_b,\epsilon_r)}{\partial \epsilon_b} > 0$. So F_2 increases monotonically with $\epsilon_b \in (p, 1)$, so we obtain

$$\begin{aligned} F_2(\epsilon_b, \epsilon_r) &> F_2(\epsilon_b = p, \epsilon_r) \\ &= \alpha D\big[p + (\epsilon_r - p)\frac{1-\alpha}{\alpha}||p\big] - D\big[\alpha p + (1-\alpha)\epsilon_r||p\big]. \end{aligned}$$

Taking the derivative of $F_2(\epsilon_b = p, \epsilon_r)$ with respect to ϵ_r:

$$\frac{\partial F_2(\epsilon_b = p, \epsilon_r)}{\partial \epsilon_r} = (1-\alpha)\big(\log \frac{g_{2,1}}{1-g_{2,1}} - \log \frac{g_{2,2}}{1-g_{2,2}}\big)$$

where

$$\begin{aligned} g_{2,1} &\triangleq p + (\epsilon_r - p)\frac{1-\alpha}{\alpha} \\ g_{2,2} &\triangleq \alpha p + (1-\alpha)\epsilon_r. \end{aligned}$$

To compare $g_{2,1}$ and $g_{2,2}$, define

$$\begin{aligned} g_2 &\triangleq g_{2,1} - g_{2,2} \\ &= (\epsilon_r - p)\frac{(1-\alpha)^2}{\alpha} \\ &> 0. \end{aligned}$$

Then $\frac{\partial F_2(\epsilon_b = p, \epsilon_r)}{\partial \epsilon_r} > 0$, $F_2(\epsilon_b = p, \epsilon_r)$ increases monotonically with $\epsilon_r \in (p, 1)$. Therefore

$$\begin{aligned} F_2(\epsilon_b = p, \epsilon_r) &> F_2(\epsilon_b = p, \epsilon_r = p) \\ &= 0 \end{aligned}$$

So, we have $F_2(\epsilon_b, \epsilon_r) > 0$, i.e

$$\min\Big\{D\big[\alpha\epsilon_b + (1-\alpha)\epsilon_r||p\big], \alpha D\big[\epsilon_b + (\epsilon_r - p)\frac{1-\alpha}{\alpha}||p\big]\Big\} = D\big[\alpha\epsilon_b + (1-\alpha)\epsilon_r||p\big].$$

To conclude, the closed-form expression for E_0 is

$$E_0 = \begin{cases} \alpha D(\epsilon_b || p) + (1-\alpha) D(\epsilon_r || p) & \epsilon_b \geq \epsilon_r \\ D\big[\alpha\epsilon_b + (1-\alpha)\epsilon_r || p\big] & \epsilon_b < \epsilon_r. \end{cases} \tag{A.6}$$

Appendix B

Simulation results

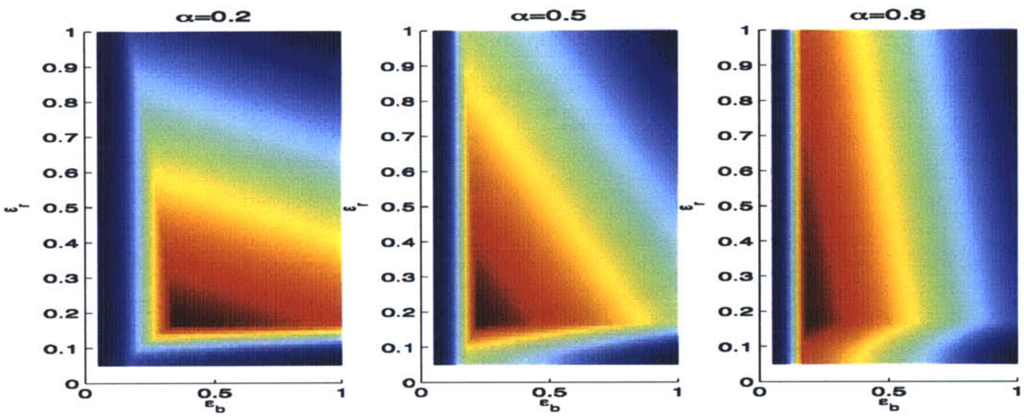

Figure B-1: Intra-layer coding: E with different α when $p = 0.05, M = 10$

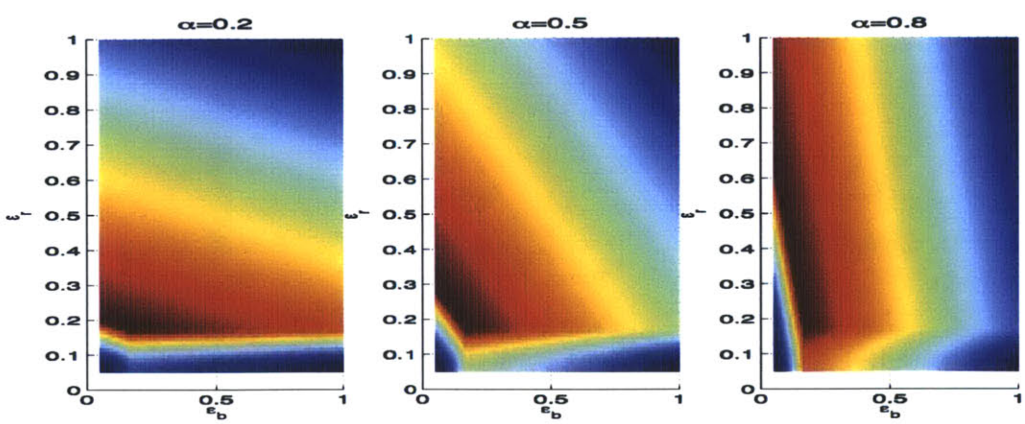

Figure B-2: Inter-layer coding: E with different α when $p = 0.05, M = 10$

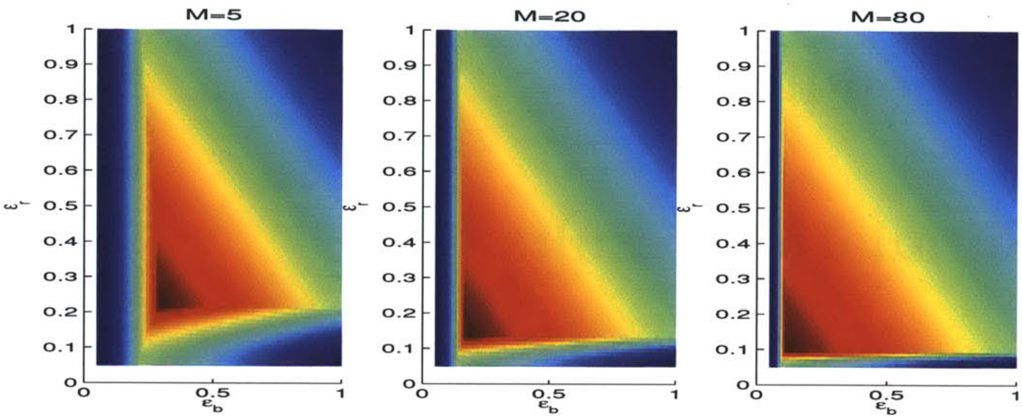

Figure B-3: Intra-layer coding: E with different M when $p = 0.05, \alpha = 0.5$

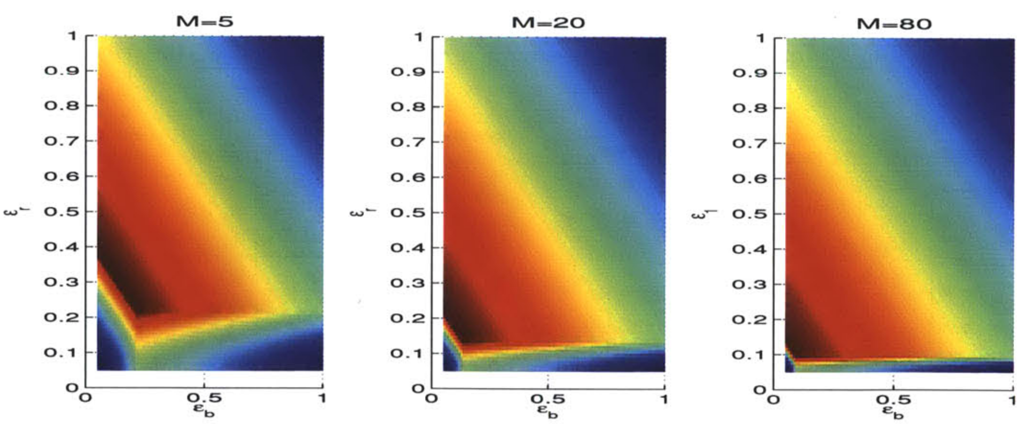

Figure B-4: Inter-layer coding: E with different M when $p = 0.05, \alpha = 0.5$

Bibliography

[1] Rudolf Ahlswede, Ning Cai, Shuo yen Robert Li, and Raymond W. Yeung. Network information flow. *IEEE TRANSACTIONS ON INFORMATION THEORY*, 46(4):1204–1216, 2000.

[2] Xu Chenguang, Xu Yinlong, Zhan Cheng, Wu Ruizhe, and Wang Qingshan. On network coding based multirate video streaming in directed networks. In *Performance, Computing, and Communications Conference, 2007. IPCCC 2007. IEEE Internationa*, pages 332 –339, april 2007.

[3] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley-Interscience, 99th edition, August 1991.

[4] S. Dumitrescu, Mingkai Shao, and Xiaolin Wu. Layered multicast with inter-layer network coding. In *INFOCOM 2009, IEEE*, pages 442 –449, april 2009.

[5] M. Effros. Distortion-rate bounds for fixed- and variable-rate multiresolution source codes. *Information Theory, IEEE Transactions on*, 45(6):1887 –1910, sep 1999.

[6] W.H.R. Equitz and T.M. Cover. Successive refinement of information. *Information Theory, IEEE Transactions on*, 37(2):269 –275, mar 1991.

[7] William Feller. *An Introduction to Probability Theory and Its Applications*. Wiley, New York, 2nd edition, 1957.

[8] T. Ho, M. Médard, R. Koetter, D.R. Karger, M. Effros, Jun Shi, and B. Leong. A random linear network coding approach to multicast. *Information Theory, IEEE Transactions on*, 52(10):4413 –4430, oct. 2006.

[9] MinJi Kim, D. Lucani, Xiaomeng Shi, Fang Zhao, and M. Médard. Network coding for multi-resolution multicast. In *INFOCOM, 2010 Proceedings IEEE*, pages 1 –9, march 2010.

[10] D. Koutsonikolas, Y.C. Hu, Chih-Chun Wang, M. Comer, and A.M.S. Mohamed. Efficient online wifi delivery of layered-coding media using inter-layer network coding. In *Distributed Computing Systems (ICDCS), 2011 31st International Conference on*, pages 237 –247, june 2011.

[11] Niveditha Sundaram, Parameswaran Ramanathan, and Suman Banerjee. Multirate media stream using network coding. In *in Proc. 43rd Annual Allerton Conference on Communication, Control, and Computing*, 2005.

[12] Xiaolin Wu, Bin Ma, and N. Sarshar. Rainbow network flow of multiple description codes. *Information Theory, IEEE Transactions on*, 54(10):4565 –4574, oct. 2008.

[13] Yunnan Wu. Distributing layered content using network coding. In *Sensor, Mesh and Ad Hoc Communications and Networks Workshops, 2008. SECON Workshops '08. 5th IEEE Annual Communications Society Conference on*, pages 1 –4, june 2008.

[14] Rong Zhang and M.L. Comer. Rate distortion analysis for spatially scalable video coding. *Image Processing, IEEE Transactions on*, 19(11):2947 –2957, nov. 2010.

[15] Jin Zhao, Fan Yang, Qian Zhang, Zhensheng Zhang, and Fuyan Zhang. Lion: Layered overlay multicast with network coding. *Multimedia, IEEE Transactions on*, 8(5):1021 –1032, oct. 2006.