APPROVAL

Name: Upul Samarawickrama

Degree: Doctor of Philosophy

Title of Thesis: Design and Application of New Multiple Description Coding Schemes

Examining Committee: Dr. Carlo Menon
Chair

Dr. Jie Liang, Senior Supervisor

Dr. Ivan Bajić, Supervisor

Dr. Daniel Lee, Supervisor

Dr. Sami Muhaidat, Internal Examiner

Dr. Jun Chen, External Examiner
Assistant Professor of Electrical and Computer Engineering, McMaster University

Date Approved: May 09, 2011
Declaration of Partial Copyright Licence

The author, whose copyright is declared on the title page of this work, has granted to Simon Fraser University the right to lend this thesis, project or extended essay to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users.

The author has further granted permission to Simon Fraser University to keep or make a digital copy for use in its circulating collection (currently available to the public at the “Institutional Repository” link of the SFU Library website <www.lib.sfu.ca> at: <http://ir.lib.sfu.ca/handle/1892/112>) and, without changing the content, to translate the thesis/project or extended essays, if technically possible, to any medium or format for the purpose of preservation of the digital work.

The author has further agreed that permission for multiple copying of this work for scholarly purposes may be granted by either the author or the Dean of Graduate Studies.

It is understood that copying or publication of this work for financial gain shall not be allowed without the author's written permission.

Permission for public performance, or limited permission for private scholarly use, of any multimedia materials forming part of this work, may have been granted by the author. This information may be found on the separately catalogued multimedia material and in the signed Partial Copyright Licence.

While licensing SFU to permit the above uses, the author retains copyright in the thesis, project or extended essays, including the right to change the work for subsequent purposes, including editing and publishing the work in whole or in part, and licensing other parties, as the author may desire.

The original Partial Copyright Licence attesting to these terms, and signed by this author, may be found in the original bound copy of this work, retained in the Simon Fraser University Archive.

Simon Fraser University Library
Burnaby, BC, Canada
Abstract

Multiple description coding is a source coding based solution for packet loss problem in communication networks. In this thesis we first present a two-stage algorithm for two-channel multiple description predictive coding. In this algorithm, a predictive encoder is used in the first stage of each description. The second stage is designed to refine the joint reconstruction from the first stage. Theoretical analysis and simulation results show that this method is very efficient in the high rate multiple description coding of strongly correlated sources. We then present a low complexity $M$-channel multiple description coding scheme which is designed based on two-rate coding and staggered quantization. For correlated sources, a two-rate predictive coding is used in each description. The application of the proposed scheme in lapped transform based image coding is also investigated, and the optimal transform is obtained. Experimental results using both 1-D memoryless sources and 2-D images demonstrate the superior performance of the proposed scheme. Next, a previously developed prediction compensation based two-channel algorithm is extended to generate more than two descriptions. Application of this algorithm in lapped transform based image coding shows a performance competitive to the state-of-the-art algorithms. Finally, this algorithm is further improved using a three layer design and sequential prediction. This improved algorithm is applied to lapped transform based image coding, where the corresponding optimal lapped transform is formulated and obtained. Image coding results show that this method outperforms other latest schemes.
To my beloved teachers, Mr. Premadasa Jayaweera and the late Mr. Harishchandra

Gamachchige
Acknowledgments

First and foremost, I would like to express my sincere gratitude to my senior supervisor, Professor Jie Liang, for his continuous advice, guidance and encouragement throughout the course of this work. His insightful vision and broad knowledge in signal compression and coding has been a great resource in completing this thesis. His kind and caring support, optimism and patience lead me through the challenges encountered during this work. I consider myself privileged to have had the opportunity to work under his guidance.

I would like to thank Professor Ivan Bajić and Professor Daniel Lee for serving as supervisors and Professor Sami Muhaidat for serving as the internal examiner on the thesis examination committee. I would also like to thank the external examiner, Professor Jun Chen from McMaster University, and the defence chair, Professor Carlo Menon for their valuable time.

I am very grateful to all my friends in the multimedia communications laboratory for their continuous encouragement and for making my years at Simon Fraser University a pleasant experience. Further, I would like to thank the faculty and the staff of the School of Engineering Science for their support during this work.

This acknowledgement would not be complete without thanking my family. I extend my heartfelt gratitude to my wife and to my parents. Without their continuous support and sacrifices this thesis would have not been possible.
Contents

Approval ii
Abstract iii
Dedication iv
Acknowledgments v
Contents vi
List of Tables ix
List of Figures x

1 Introduction 1
  1.1 Multiple Description Coding 3
  1.1.1 Rate Distortion Results for MD coding 6
  1.2 Motivation 7
  1.3 Objectives and Contributions 11
  1.4 Acronyms 13

2 A Two-Stage Algorithm for MD Predictive Coding 14
  2.1 Introduction 14
  2.2 The Modified MDSQ (MMDSQ) 15
  2.3 Two-Stage MD predictive coding 15
### Two-Rate Coding with Staggered Quantization

3.1 Introduction .................................................. 25

3.2 System Design and Performance Analysis .......................... 26
3.2.1 System Design for Memoryless Sources .......................... 26
3.2.2 System Design for Correlated Sources ............................ 29
3.2.3 Effect of Prediction Quantization ................................. 30
3.2.4 Performance Analysis ......................................... 32
3.2.5 Distortion expressions for two-rate coding without staggered quantization .............................................. 38

3.3 Optimal Design for MD Image Coding ............................. 39
3.3.1 Overview of the Time-Domain Lapped Transform .......... 40
3.3.2 Optimal Design of the Transform ............................... 42

3.4 Experimental Results ............................................ 45
3.4.1 Experimental Results for 1-D Sources ........................... 45
3.4.2 Experimental Results for MD Image Coding .................. 47

3.5 Summary .......................................................... 49

### M-Channel MDC with Prediction-Compensation

4.1 Introduction ...................................................... 64

4.2 Prediction Compensated MDC with Many Descriptions .......... 65
4.2.1 Bit allocation ............................................... 67

4.3 MD Coding of a First Order Gauss-Markov Source ............... 70

4.4 Optimal Design for Lapped Transform based MD Image Coding 71
4.4.1 Overview of the System ...................................... 71
4.4.2 The Wiener Filter ........................................... 73
4.4.3 Objective Function and Optimal Rate Allocation ............ 74

4.5 Application in Image Coding ..................................... 78
List of Tables

1.1 List of acronyms. .............................................. 13
List of Figures

1.1 A digital communication system. .............................................. 2
1.2 Two channel multiple description coding. ................................. 3
1.3 DPCM encoder. ........................................................................ 8
1.4 DPCM decoder. ........................................................................ 8

2.1 The proposed MD predictive encoder. ........................................ 16
2.2 The quantizer partitions for $x(n)$ resulting from the first-stage predictive quantizers. ......................................................... 19
2.3 The central SNR vs side SNR of the proposed method and the reference method in [10] for MD coding of a GM(1) source with $\rho = 0.9$ and unit variance at $R = 3$ and $R = 5$ bits/sample/channel. ...................... 23
2.4 End-to-end average distortion of the proposed method and the reference method in [10] for a speech signal coded at $R = 3$ and $R = 4$ bits/sample/channel. ............................................................... 24

3.1 Three-channel MDC of a memoryless source with the proposed method. The entropy coding is not shown. Darker arrows represent higher-rate ($R_0$) coded samples, and lighter arrows represent lower-rate ($R_1$) coded samples. (a) Encoding; (b) Decoding when only two descriptions are available. ......................................................... 27
3.2 The staggered lower-rate quantizers in the $M-1$ descriptions and the joint quantizer. .............................................................. 28
3.3 All possible configurations of the received finer cells for $M = 5$ and $k = 3$. 35
<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.4</td>
<td>All possible configurations of the received descriptions for the finer cell pattern in Fig. 3.3 (a). Dashed lines denote lost descriptions.</td>
</tr>
<tr>
<td>3.5</td>
<td>Performance gain of the staggered quantizers: $\alpha_{M,k}$ for different $M$ and $k$.</td>
</tr>
<tr>
<td>3.6</td>
<td>Forward and inverse time-domain lapped transforms.</td>
</tr>
<tr>
<td>3.7</td>
<td>Side distortions $D_1, D_2, D_3$ vs central distortion $D_4$ of the proposed method for four-description coding of a unit variance GM(1) source with correlation coefficient 0.9 and $R = 5$ bits/sample/description.</td>
</tr>
<tr>
<td>3.8</td>
<td>Side distortion vs. central distortion of the proposed method and the MDLVQ in [24] for four description coding of a Gaussian source of unit variance.</td>
</tr>
<tr>
<td>3.9</td>
<td>Expected distortion vs loss probabilities of the proposed method and the MDLVQ in [24] for four description coding of a Gaussian source of unit variance.</td>
</tr>
<tr>
<td>3.10</td>
<td>The side PSNR vs central PSNR for three-description coding of Barbara at total bit rate of 1 bit/pixel.</td>
</tr>
<tr>
<td>3.11</td>
<td>The side PSNR vs central PSNR for three-description coding of Lena at total bit rate of 1 bit/pixel.</td>
</tr>
<tr>
<td>3.12</td>
<td>The side PSNR vs central PSNR for three-description coding of Boat at total bit rate of 1 bit/pixel.</td>
</tr>
<tr>
<td>3.13</td>
<td>The side PSNR vs central PSNR for four-description coding of Barb at total bit rate of 1 bit/pixel.</td>
</tr>
<tr>
<td>3.14</td>
<td>The side PSNR vs central PSNR for four-description coding of Lena at total bit rate of 1 bit/pixel.</td>
</tr>
<tr>
<td>3.15</td>
<td>The side PSNR vs central PSNR for four-description coding of Boat at total bit rate of 1 bit/pixel.</td>
</tr>
<tr>
<td>3.16</td>
<td>The side PSNR vs central PSNR for four-description coding of image Barbara with TRPCSQ and TDLT-based RD-MDC with total rate of 1 bit/pixel.</td>
</tr>
</tbody>
</table>
3.17 The side PSNR vs central PSNR for four-description coding of image Lena with TRPCSQ and TDLT-based RD-MDC with total rate of 1 bit/pixel. ................................................................. 60
3.18 The expected PSNR for images Barbara and Lena for four-description coding with TRPCSQ and TDLT-based RD-MDC with total rate of 1 bit/pixel .......................... 61
3.19 Reconstructed image Barbara with two descriptions for $M = 4$ and central PSNR of 35.13 dB. (a) JPEG 2000-based RD-MDC (26.11 dB). (b) TDLT-based RD-MDC (28.07 dB). (c) TRPCSQ with single description optimized filters (28.68 dB). (d) TRPCSQ with optimized filters (28.93 dB). ................................................................. 62
3.20 Reconstructed image Barbara with three descriptions for $M = 4$ and central PSNR of 35.13 dB. (a) JPEG 2000-based RD-MDC (28.89 dB). (b) TDLT-based RD-MDC (30.23 dB). (c) TRPCSQ with single description optimized filters (31.54 dB). (d) TRPCSQ with optimized filters (31.74 dB). ................................................................. 63

4.1 Examples of prediction and bit allocation with five (top) and six (bottom) descriptions, respectively. Each circle represents one sample. The number in each circle is the polyphase it belongs to. Each sample in polyphase $j$ is predicted by two samples in polyphase 0, with rate $R_{\Delta(j)}$ for the residual encoding. ................................................................. 67
4.2 Average SNR $10 \log_{10}(D_M)$ vs loss probability $p$ for PC-MDC and TRPCSQ in [35] at $M = 3$, $M = 4$ and $M = 6$. The source is first order Gauss-Markov with unit variance and $\rho = 0.95$. .................................................. 81
4.3 Side distortion vs central distortion for 3 description coding of image Barbara at 1 bits per pixel total rate. ................................................................. 82
4.4 Side distortion vs central distortion for 3 description coding of image Lena at 1 bits per pixel total rate. ................................................................. 83
4.5 Side distortion vs central distortion for 3 description coding of image Boat at 1 bits per pixel total rate. ................................................................. 84
5.7 Three description coding results for image Lena at total bit rate of 1 bits per pixel. .......................................................... 114
5.8 Three description coding results for image Barbara at total bit rate of 1 bits per pixel. .......................................................... 115
5.9 Three description coding results for image Boat at total bit rate of 1 bits per pixel. ............................................................. 116
5.10 Four description coding results for image Lena at total bit rate of 1 bits per pixel. ............................................................ 117
5.11 Four description coding results for image Barbara at total bit rate of 1 bits per pixel. ......................................................... 118
5.12 Four description coding results for image Boat at total bit rate of 1 bits per pixel. ............................................................. 119
5.13 Nine description coding results for image Lena at total bit rate of 1.25 bits per pixel. .......................................................... 120
5.14 Nine description coding results for image Barbara at total bit rate of 1.25 bits per pixel. ...................................................... 121
5.15 Nine description coding results for image Boat at total bit rate of 1.25 bits per pixel. .......................................................... 122
5.16 Four description coding results for the proposed method and TRPCSQ in [35] for image Boat at total bit rate of 1 bits per pixel. The PSNR values $D_4$ and $D_1$ of the proposed method are tuned to be equal to TRPCSQ ($D_4 = 38.79$ dB and $D_1 = 24.65$ dB). (a) Proposed, two descriptions (27.34 dB) (b) TRPCSQ, two description (27.57 dB) . 123
5.17 Four description coding results for the proposed method and TRPCSQ in [35] for image Boat at total bit rate of 1 bits per pixel. The PSNR values $D_4$ and $D_1$ of the proposed method are tuned to be equal to TRPCSQ ($D_4 = 38.79$ dB and $D_1 = 24.65$ dB). (a) Proposed, three descriptions (31.57 dB) (b) TRPCSQ, three descriptions (30.80 dB) . 124
Chapter 1

Introduction

Transmission of information over a digital communication system in general involves three basic operations, namely, source coding, channel coding, and transmission as depicted in Figure 1.1. The input to the source encoder can be either an analog or digital signal. The source encoder generates an efficient digital representation of the input signal usually in binary format. The main objective of source encoding is to obtain a representation with as few symbols as possible (compression), which is done by removing the redundancy present in the input signal. Source coding can be mainly categorized into two types as lossy and lossless coding. The lossless coding preserves all the information in the input signal while, as its name implies, lossy coding involves a loss of some information. The function of the channel encoder is to add redundancy to its input in a controlled manner to achieve robustness against the errors caused in the transmission channel. The added redundancy helps the channel decoder to recover reliably the transmitted data. The function of channel decoder and the source decoder is to estimate the digital signal at the input of the respective encoder at the sending end.

In some digital communication systems, data is transmitted in blocks called packets or frames. Data communication networks, such as the Internet suffer from the problem of packet losses caused by congestion in the network. In some situations, this problem can be solved by retransmission of lost packets. However, this requires a feedback path to request retransmission and therefore cannot be used in broadcast
type communications. Retransmission may also not be possible in some real time communication systems (such as in audio and video communication), where the allowable delay is limited. In case of transmission of an analog source over a packet loss network, an alternative source coding based solution for the above problem is multiple description coding (MDC) where, several versions referred to as “descriptions” of the same message are generated and transmitted such that, each describes the message with a certain acceptable quality. When more than one description is available at the receiver they can be combined to reconstruct the message at a higher quality. The term “message” may refer to one source sample or a group of source samples. Different descriptions of the same message are transmitted over separate communication paths and hence it is less likely to lose all of them simultaneously. Therefore, with this method a certain minimum quality of service can be guaranteed at the receiver. Another application of MDC suggested in literature is in wireless mobile networks. In wireless mobile communication, the transmitted signal arrives at the receiver over multiple paths. These multipath signals may, from time to time, interfere destructively with each other, causing a weak signal level at the receiver. This phenomenon is known as multipath fading. A frame under deep fading can be considered as lost and therefore multiple description coding can be used in this situation as well. Recently, a significant amount of research has been done in MDC and many methods have been developed to generate multiple descriptions from a source.
1.1 Multiple Description Coding

MDC was invented in the 1970s as a speech coding method for achieving reliability in telephone networks. When a link fails in a telephone network the calls have to be diverted to a standby link. These standby links add extra cost to the system. On the other hand they are not used in the normal operation and therefore can be considered as a waste of resources. The number of standby links can be greatly reduced if the information in a call can be split into two streams and sent over two channels such that each independently gives an acceptable quality. An odd-even sample separation based speech coding techniques was developed in [11] where the odd numbered voice samples and even numbered voice samples are send over two different channels. When both the channels work the speech was reconstructed at a higher quality. On the other hand, when one fails, samples on the other channel are used to reconstruct the voice at a lower quality. A good review on MDC can be found in [9].

Fig 1.2 shows two channel MD coding. Note that channel coding and modulation are not shown in the figure. The MD encoder generates two descriptions and sends them over two channels. If only one description is received corresponding side decoder decodes it and the resulting distortion is called side distortion. If both descriptions are received, the central decoder reconstructs the source at a higher quality and the resulting distortion is called central distortion.
MDC was studied as an information theoretical problem in late 1970s and early 1980s by several authors. A rate distortion analysis of the two channel MDC was provided in [8] for a memoryless source and the achievable rate region was constructed. It was shown in [25] that this indeed is the rate distortion region in the case of a Gaussian source with MSE distortion measure. Shannon-type bounds for two channel MD coding of memoryless sources were obtained by [58], and the results were extended to sources with memory in [57]. Information-theoretic analyses of MDC with more than two channels can be found in, e.g., [54, 27, 41]. Among them, an achievable rate-distortion (R-D) region for the MDC of memoryless sources is given in [54], by generalizing the two-description result in [8]. An improved achievable region is obtained in [27] by encoding a source in $M$ stages, where the $k$-th stage refines the previous stages and can only be decoded when at least $k$ descriptions are received. The scheme is based on the theory of source coding with side information and distributed source coding (DSC) [6]. Recently, another improved scheme is developed in [41], which can achieve points outside the achievable region in [27].

Not many practical MDC algorithms were developed until Vaishampayan [51] in 1993 provided the framework for the implementation of multiple description scalar quantizers (MDSQ) for two channels. His method of MD generation consists of scalar quantization followed by an index assignment. Since Vaishampayan’s initial work on MDSQ, several methods of MD coding have been developed. A correlating transform based two channel MDC technique was proposed in [55] where a linear transform is applied on source vectors (an ordered tuple of source samples) to generate correlated transform coefficients. In this method, when some of the coefficients are lost, they can be estimated from the available coefficients by exploiting the correlation introduced by the linear transform. A modified MDSQ (MMDSQ) using two-stage coding is developed in [43], which avoids the index assignment problem of the MDSQ and still achieves the high rate asymptotic performance of the entropy constrained MDSQ of [52].

While most of the MDC algorithms available in the literature are limited to two channels, there are several potential application for MDC that would require more than two channels. We refer to MDC with more than two channels as $M$-channel MDC.
with implicit assumption that $M > 2$. In [21, 28] multiple descriptions are generated using erasure correcting codes to provide unequal loss protections (ULP) to different layers of a scalable source code. However, the method only has good performance when at least $n$ descriptions are received, for some pre-specified $n$, because it is usually optimized to minimize the expected distortion for a given channel loss probability. In addition, the optimization algorithm is not straightforward.

In [4], the MDSQ in [51] is extended to more than two channels via combinatorial optimization. However, the scheme only has one degree of freedom and only assigns the index symmetrically around the main diagonal of the index matrix. A general symmetric $M$-channel MDC should have $M - 1$ degrees of freedom, which allow more flexibility in tuning the redundancy. Another extension of the MDSQ with $M - 1$ degrees of freedom is developed in [44], which shares some similarities to the method in [27], i.e., the coding consists of multiple stages such that each stage refines the preceding stages. However, both algorithms become quite complicated as the number of channels increases.

A lattice vector quantization based MDC method is presented in [24], where $M$ descriptions are generated by uniquely assigning each point in a finer central lattice to $M$ points in a sublattice. The method also involves an index assignment problem, which increases the complexity in design and implementation.

In another class of MDC methods, the source splitting method is used to generate different descriptions. This approach was pioneered by Jayant in [11]. In [14], the transform coefficients are split into two parts. Each part is quantized into one description. Each description also includes the coarsely quantized result of the other part, which improves the reconstruction when the other description is lost. A similar approach is developed in [20] for $M$ descriptions using the SPIHT framework. Recently the method in [14] is generalized to JPEG 2000 in [46] for two-description coding, where each JPEG 2000 code-block is coded at two rates, one in each description. The rate allocation is determined by Lagrangian optimization. The method is called RD-MDC in [46]. To get balanced descriptions, the RD-MDC needs to partition all code-blocks into two subsets with similar R-D curves, and exhaustive search is needed for this operation. A classification-based method is proposed in it with better tradeoff
between complexity and performance.

In [1], the method in [46] is extended to $M$-channel case, where each JPEG 2000 code-block is still encoded at two rates. The higher-rate coded code-blocks are divided into $M$ subsets and are assigned to $M$ descriptions. Each description also carries the lower-rate codings of the remaining code-blocks. One benefit of [46, 1] is that they maintain a good compatibility with JPEG 2000, e.g., each description can be decoded by a standard JPEG2000 decoder.

In [45], a multi-rate method with $M - 1$ degrees of freedom is developed, which generalizes the two-rate method in [1]. In this method, each subset of the source is coded at $M$ different rates, one for each description. However, as in [1], only the finest version of each subset from all received descriptions is used for the reconstruction. Other available representations are simply discarded. In addition, finding the optimal bit allocation is more complicated than the two-rate method.

1.1.1 Rate Distortion Results for MD coding

Consider two channel MD coding of a source. Let $R_1$ and $R_2$ be rates on channel 1 and channel 2 respectively and $D_0$, $D_1$, and $D_2$ be central and side distortions respectively. Assume that MD coding is done in vectors of $n$ source samples. A quintuple $(R_1, R_2, D_0, D_1, D_2)$ is said to be achievable if, for some positive integer $n$, there exist a MD coder with rates $R_1$ and $R_2$ and distortions $D_0$, $D_1$, and $D_2$. The rate distortion region for this MD coding is the closure of the set of achievable quintuples $(R_1, R_2, D_0, D_1, D_2)$. In [25] Ozarow found the rate distortion region for a memoryless Gaussian source of variance $\sigma^2$. In this case, the rate distortion region is given by

$$D_1 \geq \sigma^2 2^{-2R_1} \tag{1.1}$$
$$D_2 \geq \sigma^2 2^{-2R_2} \tag{1.2}$$
and

\[
D_0 \geq \begin{cases} 
\frac{\sigma^2 2^{-2(R_1+R_2)}}{1-(\sqrt{(1-D_1/\sigma^2) D_2/\sigma^2}-\sqrt{D_1 D_2/\sigma^4})^2} & \text{if } D_1 + D_2 > \sigma^2 + D_0 \\
\sigma^2 2^{-2(R_1+R_2)} & \text{otherwise.}
\end{cases}
\] (1.3)

With the assumption that \( R_1 = R_2 = R \gg 1 \) and \( D_1 = D_2 \approx 2^{-2(1-\nu)R} \) where \( 0 < \nu \leq 1 \), from (1.3) it can be shown that

\[
D_0 D_1 \geq \frac{1}{4} \sigma^4 2^{-4R}.
\] (1.4)

This is called the distortion product bound, and frequently used in the literature for evaluating the efficiency of a MDC scheme. In [57] shannon-type bound were derived for a general stationary source. The bounds found in [57] are similar to (1.1) - (1.3) except that the variance \( \sigma^2 \) in (1.1) - (1.3) is replaced with the entropy rate power \( P_x \) of the source. Specifically, the distortion product bound for a general stationary source is given by

\[
D_0 D_1 \geq \frac{1}{4} P_x^2 2^{-4R}.
\] (1.5)

### 1.2 Motivation

Most of the MD coding algorithms developed so far are based on traditional source coding techniques such as prediction, transform and quantization. However, the adaptation of these techniques for MD coding is not trivial. For example, given the source probability density function, we can find the optimal single descriptions scalar quantizer when the bit rate is sufficiently large. However, the design of optimal MD scalar quantizer under the same conditions is still an open problem. This is true regarding the design of predictive and transforms based MD coders too. There is still ongoing work to improve the adaptation of these techniques for MD coding.

Most of the real life signal sources are correlated. In traditional source coding the objective is to minimize the bit rate by removing the correlation in the source using prediction, transforms or filter banks. While multiple description coding is a source
CHAPTER 1. INTRODUCTION

Figure 1.3: DPCM encoder.

Figure 1.4: DPCM decoder.
coding method, it’s different from traditional source coding in that it has to guard against channel losses. This channel loss protection needs some extra redundancy which results in a correlation among the descriptions. Some MDC methods such as source splitting leaves some of the inherent source correlation or redundancy so that in the case of description losses, missing descriptions can be estimated from the received descriptions using this residual correlation. However, most of the practical MDC methods are designed with the assumption that the source is first decorrelated using a transform or filter banks, and the MDC is applied to already decorrelated signal. For example multiple descriptions are generated in [3] by applying MDSQ after performing the decorrelating block transform. More specifically this is equivalent to replacing the quantizer of a single description block transform coded system with MDSQ. Furthermore, it is shown in [3] that the optimal transform and bit allocation in this case are identical to those in single description coding. However, application of this methodology, i.e. separating the decorrelating stage from the MDC stage is not straightforward in case of predictive coding.

To understand this, consider the differential pulse-coded modulation (DPCM) coding shown in Fig 1.3 and Fig 1.4. The source sample $x_n$ is predicted by $\bar{x}_n$ and the resulting prediction error $e_n$ is quantized and sent to the decoder. Note that the DPCM encoder also contains the DPCM decoder to reconstruct the source samples. These reconstructed source samples are used in predicting the future samples. This way, it avoids the predictor mismatch in the encoder and the decoder. To generate multiple descriptions let’s assume we replace the quantization in this DPCM coding with MDSQ, as it is done in [3] in a transform coded system. The problem in this approach is that the encoder does not have the knowledge of the descriptions received by the decoder. Therefore, it cannot guarantee that the predictors are equal in the encoder and the decoder, which leads to drift errors in the reconstructed samples. This example shows that the application of predictive coding along with multiple description coding is not as straightforward as for transform coding. On the other hand, predictive coding is an important technique is source coding. While correlation can be removed using transforms there are some situations predictive coding is essential. For
example, the temporal correlation in video coding is removed using motion compensation which is a form of predictive coding. Another example for predictive coding is in JPEG coding standard where the DC coefficients of a discrete cosine transformed block is predicted from the DC coefficients of the already encoded blocks. Therefore it is important to design multiple description predictive coders.

In [53], the MD scalar quantizer (MDSQ) developed in [51] is applied to the two channel MD predictive coding of video, where each description uses a separate predictive encoder with a quantizer equivalent to a side quantizer in the MDSQ. To maintain the necessary offset between the side quantizers, the predictor output is quantized by an additional quantizer, which renders the prediction sub-optimal. A similar MD predictive coding method is presented in [22] without using the quantizer in the prediction loop. This compromises the optimal offset between the side quantizers, leading to a higher central distortion. In [10], a sub-optimal second-order predictor is used for the predictive encoding of speech signals, which leaves some correlations in the prediction residual. The two descriptions are obtained by splitting the prediction residual into the odd-indexed and even-indexed parts. When only one description is available, the remaining correlation in the prediction residual is utilized to obtain a sufficiently accurate estimate of the source.

Another area that needs further research is $M$-channel MDC. Most of the studies carried out so far on MDC have concentrated only on the two channel MD coding problem. MDSQ for example is designed only for the two channel case and how to extend it for more than two channels with a sufficient degree of freedom is not known. On the other hand, there is a number of potential applications for MD coding that require more than two descriptions. For example, a typical internet image would require about 10 packets of size 576 bytes in a network using Internet Protocol Version 6 (IPV6) [9]. An IP packet can lose independently and therefore it has to be considered as a description when MD coding is used. Therefore a MD image coding algorithm should be capable of generating more than two descriptions when used in such an IP network. Therefore, it is important to design efficient $M$-channel MDC coding algorithms.
CHAPTER 1. INTRODUCTION

Images are widely transmitted on today’s data communication and wireless networks in which MDC has potential applications due to packet loss problem and multipath fading. Therefore, it is important to design efficient MD image coding methods. However, design of good MD image coding is challenging, because there is no good model for images. Sometimes an algorithm with nice theoretical performance may not work well in practice. Hence it is important to design improved MD image coding schemes, by taking full advantage of the properties of natural images and successful techniques in single description image coding as well as previous MD image coding.

1.3 Objectives and Contributions

This work has four main objectives

1. Design, analysis and evaluation of new methods for MD predictive coding

2. Improving existing MD coding methods

3. Design, analysis and evaluation of new $M$-channel MD coding algorithms

4. Design of efficient MD image coding methods

Four MD coding algorithms are developed in this work with these objectives. In Chapter 2 a two-stage algorithm is developed for two channel multiple description predictive coding. In this algorithm, a predictive encoder is used in the first stage of each description. The two encoders are designed to create staggered quantizers for the input, which allow further refinement of the joint reconstruction in the second stage. Theoretical analysis and simulation results show that this method is very efficient in the high rate multiple description coding of strongly correlated sources.

Chapter 3 presents a low complexity $M$-channel multiple description coding scheme in which each description carries one subset of the input with a higher bit rate and the rest with a lower bit rate. The lower-rate codings in different descriptions are designed to be mutually refinable using staggered scalar quantizers. For correlated sources, a two-rate predictive coding is used in each description. Closed-form expressions of
the distortions are derived when different numbers of descriptions are received. The application of the proposed scheme in lapped transform based image coding is also investigated, and the optimal transform is obtained. Experimental results using both 1-D memoryless sources and 2-D images demonstrate the superior performance of the proposed scheme.

When applied to MD image coding, the two-channel prediction compensated MDC (PC-MDC) scheme in [38] achieves better results than those in [43, 46], and represents the state of the art in two-description image coding. In Chapter 4, motivated by the superior performance of the two-channel PC-MDC in image coding, the prediction-compensated approach is generalized to $M$-channel case. The expressions for various distortions and the optimal bit allocation are derived. The optimal lapped transform is designed for the proposed algorithm using a three descriptions coding example. Image coding experiments show that the proposed algorithm competes well with the state-of-the-art techniques. However, an analysis using a Gauss-Markov source reveals some deficiency in the proposed method.

In Chapter 5, the prediction compensation based MDC scheme in Chapter 4 is further improved using a three-layer design and sequential prediction. In each description, a subset of the source samples is encoded in the first layer. In the second layer, the remaining subsets are encoded sequentially by predicting from the already encoded subsets. As a result, each description can produce a coarse reconstruction of the source. When multiple descriptions are received, a refined reconstruction is obtained by fusing all coarse reconstructions. In addition, a third layer is included to refine the reconstruction when only one description is lost, which dominates when the probability of channel error is low. The closed-form expressions of the expected distortion are derived for 1-D sources when different numbers of descriptions are received. The proposed scheme is then applied to lapped transform based image coding, where the corresponding optimal lapped transform is formulated and obtained. Image coding results show that the proposed method outperforms other latest schemes.

Results and findings of this research has been published in two journal papers and five conference papers, and another journal paper has been accepted for publication. The two-stage algorithm for two channel multiple description predictive coding is
presented in [30]. The low complexity $M$-channel multiple description coding scheme with two-rate coding and staggered quantization is presented in [35], [31] and [32]. The two-channel prediction compensated MDC scheme is presented in [38] and [39]. Finally, the $M$-channel prediction compensation based MDC scheme with three layers is presented in [33] and [34].

### 1.4 Acronyms

Table 1.1 defines the acronyms used in this thesis.

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCT</td>
<td>Discrete cosine transform</td>
</tr>
<tr>
<td>DPCM</td>
<td>Differential pulse-coded modulation</td>
</tr>
<tr>
<td>DSC</td>
<td>Distributed source coding</td>
</tr>
<tr>
<td>IDCT</td>
<td>Inverse discrete cosine transform</td>
</tr>
<tr>
<td>KLT</td>
<td>Karhunen-Loeve Transform</td>
</tr>
<tr>
<td>LMMSE</td>
<td>Linear minimum mean squared error</td>
</tr>
<tr>
<td>MDC</td>
<td>Multiple description coding</td>
</tr>
<tr>
<td>MDLVQ</td>
<td>Multiple description lattice vector quantization</td>
</tr>
<tr>
<td>MDSQ</td>
<td>Multiple description scalar quantizer</td>
</tr>
<tr>
<td>MMDSQ</td>
<td>Modified multiple description scalar quantizer</td>
</tr>
<tr>
<td>MSE</td>
<td>Mean square error</td>
</tr>
<tr>
<td>PC-MDC</td>
<td>Prediction compensated multiple description coding</td>
</tr>
<tr>
<td>PSD</td>
<td>Power spectral density</td>
</tr>
<tr>
<td>PSNR</td>
<td>Peak signal to noise ratio</td>
</tr>
<tr>
<td>R-D</td>
<td>Rate-distortion</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal to noise ratio</td>
</tr>
<tr>
<td>TDLT</td>
<td>Time-domain lapped transform</td>
</tr>
<tr>
<td>TLMDC</td>
<td>Three-layer multiple description coding</td>
</tr>
<tr>
<td>TRPCSQ</td>
<td>Two-rate predictive coding and staggered quantization</td>
</tr>
<tr>
<td>ULP</td>
<td>Unequal loss protections</td>
</tr>
<tr>
<td>WSS</td>
<td>Wide sense stationary</td>
</tr>
</tbody>
</table>

Table 1.1: List of acronyms.
Chapter 2

A Two-Stage Algorithm for MD Predictive Coding

2.1 Introduction

Linear prediction is an important technique in source coding for removing the redundancy in correlated signals. However, as pointed out in Section 1.2, integrating prediction with MDC is not as straightforward as integrating other decorrelating techniques such as transforms. In this chapter we present a new multiple description predictive coder design based on modified MDSQ (MMDSQ) developed in [43]. MMDSQ avoids the index assignment problem of the MDSQ in [51] and still achieves the high rate asymptotic performance of the entropy constrained MDSQ of [52]. Motivated by the simplicity and the good performance of MMDSQ, we extend this technique to MD predictive coding. Our algorithm preserves the simplicity of MMDSQ with little sacrifice of the predictor optimality for highly correlated sources. We first give a brief introduction into MMDSQ followed by the system design of our algorithm. Expressions for the central and side distortions are derived and the distortion product is compared with the lower bound in [57]. The performance of the proposed algorithm is experimentally evaluated and compared with the MD predictive coding algorithm in [10]. Theoretical analysis and simulation results show that the proposed method is a good complement to [10] for the high rate MD coding of strongly correlated sources.
CHAPTER 2. A TWO-STAGE ALGORITHM FOR MD PREDICTIVE CODING

2.2 The Modified MDSQ (MMDSQ)

In the MMDSQ [43], a source sample is quantized by two uniform scalar quantizers whose cells are staggered by an offset of $q/2$, where $q$ is the quantization step size. The combination of the two quantizers yields a joint quantizer with step size $q/2$. The outputs of the two quantizers form the base layers of the two descriptions, respectively. Their joint quantization error is further quantized by a finer second-stage uniform scalar quantizer, whose output is split into two halves to form the second layer of each description. When only one description is available, the decoder ignores the second-stage bits and reconstructs the source with the first-stage data. The corresponding distortion, the side distortion, is given by the first-stage quantizer. When both descriptions are available, the decoder reconstructs the source using data from both layers. The resulting distortion, called the central distortion, is governed by the second-stage quantizer.

As pointed out in Section 1.1.1, the performance of a MD coder can be evaluated by the distortion product, which is the product of its side distortion and central distortion. In [43], it is shown that despite the simplicity of the MMDSQ, its distortion product is identical to that of the entropy constrained MDSQ in [52].

In [43], the MMDSQ is applied to image coding, after decorrelating the input by wavelet transform. In this chapter, we apply this technique to prediction based multiple description coding and investigate its application in speech and audio coding.

2.3 Two-Stage MD predictive coding

Consider a source of the following first-order Gauss-Markov (GM(1)) model

$$x(n) = \rho x(n-1) + \omega(n), \quad (2.1)$$

where $x(n)$ is the $n$-th source sample, $\omega(n)$ is a white Gaussian noise sample that is uncorrelated with $x(n)$, and $\rho$ is the correlation coefficient. In a DPCM encoder $x(n)$ is predicted by $\hat{x}(n-1)$, where $\hat{x}(n-1)$ is the reconstruction of $x(n-1)$. The
Figure 2.1: The proposed MD predictive encoder.
prediction error $e(n) = x(n) - \rho \hat{x}(n-1)$ is quantized and transmitted to the decoder. The reconstruction of $x(n)$ is obtained by $\hat{x}(n) = \rho \hat{x}(n-1) + \hat{e}(n)$, where $\hat{e}(n)$ is the reconstruction of $e(n)$.

Our objective is to design two DPCM encoders to create the effect of two staggered quantizer partitions on the input so as to apply the MMDSQ coding scheme. The proposed encoder is depicted in Fig. 2.1, where each description uses a separate scalar-quantizer-based DPCM encoder. The average of reconstructions $\hat{x}_1(n)$ and $\hat{x}_2(n)$ is used as a refined reconstruction of the source. The corresponding error $e(n)$ is quantized by another scalar quantizer and the result is evenly split into the two descriptions, as in the MMDSQ. $Q_1, Q_2$ and $Q_3$ are all uniform scalar quantizers and their outputs are entropy-coded.

Note that since the quantization in the DPCM encoders are applied to $e_1(n)$ and $e_2(n)$ rather than to $x(n)$, simply employing staggered quantizers $Q_1$ and $Q_2$ will not generate the desired quantizer partitions for $x(n)$, because by $x(n) = e_i(n) + \rho \hat{x}_i(n-1)$ ($i = 1, 2$), the quantizer partitions for $x(n)$ and $e_i(n)$ are shifted by $\rho \hat{x}_i(n-1)$. Therefore the predictor needs to be designed jointly with the quantizers.

To achieve MMDSQ-like side quantizer partitions on $x(n)$, in this chapter we use the sub-optimal predictors $\bar{x}_i(n) = \hat{x}_i(n-1)$ instead of the optimal $\rho \hat{x}_i(n-1)$, as shown in Fig. 2.1. To understand this solution, consider the example in Fig. 2.2, which shows the resulting quantizer partitions for $x(n)$ with given $\bar{x}_1(n)$ and $\bar{x}_2(n)$, where $q$ is the step size of both $Q_1$ and $Q_2$, and $\Delta$ is the offset between the partitions for $x(n)$. The figure also shows a particular $x(n)$ and the corresponding reconstructions $\hat{x}_1(n)$ and $\hat{x}_2(n)$. We assume that the rates in the first-stage encoders are sufficiently high such that the source is approximately uniformly distributed within a cell. Therefore, the optimal reconstruction using the outputs of the two first-stage encoders is $\frac{1}{2}(\hat{x}_1(n) + \hat{x}_2(n))$, which is the midpoint of the intersection of the two staggered cells that $x(n)$ belongs to. The width of this intersection can take two possible values, i.e., $\Delta$ with probability $\frac{\Delta}{q}$ and $q - \Delta$ with probability $\frac{q - \Delta}{q}$. Under the high rate quantization assumption the associated MSE $E(e^2(n))$ for a given $\Delta$ can be
expressed as

\[ E(e^2(n)) = \frac{\Delta \Delta^2}{q} \cdot \frac{1}{12} + \frac{q - \Delta (q - \Delta)^2}{q} \cdot \frac{1}{12} = \frac{1}{12q} (\Delta^3 + (q - \Delta)^3). \]  

(2.2)

This MSE is minimized when \( \Delta = q/2 \), i.e., when the reconstructions of the two quantizers are offset by \( q/2 \), or in other words, \( \hat{x}_1(n) - \hat{x}_2(n) = \pm q/2 \). One way to guarantee this is to have \( \bar{x}_1(n) - \bar{x}_2(n) = \pm q/2 \) which in turn requires \( \hat{x}_1(n-1) - \hat{x}_2(n-1) = \pm q/2 \) since \( \bar{x}_i(n) = \hat{x}_i(n-1) \) in our coder.

As a result, if the offset is optimal at time \((n-1)\) it will be optimal at time \(n\) too. Therefore, to achieve the optimum offset for all samples we only need to ensure this for the first sample. In our encoder, this is obtained by using optimally staggered \( Q_1 \) and \( Q_2 \) (shifted by \( q/2 \)) for the first sample \( x(1) \), with \( \bar{x}_1(0) = \bar{x}_2(0) = 0 \), and thereafter using non-staggered identical midtread quantizers for \( Q_1 \) and \( Q_2 \).

If we choose the optimal predictors \( \rho \hat{x}_i(n-1) \) and still want to maintain \( \hat{x}_1(n) - \hat{x}_2(n) = \pm q/2 \) for all \( n \), an offset of \( \pm (1 - \rho)q/2 \) between \( Q_1 \) and \( Q_2 \) would be needed, depending on whether \( \hat{x}_1(n-1) - \hat{x}_2(n-1) \) is \( q/2 \) or \(-q/2\). This can be achieved by, for example, fixing the partition of \( Q_1 \) and shifting that of \( Q_2 \) for each \( n \). However, when only the second description is available, \( \hat{x}_1(n-1) \) is unknown and an extra bit about the sign of \( \hat{x}_1(n-1) - \hat{x}_2(n-1) \) has to be included in the second description to reconstruct \( \hat{e}_2(n) \) and \( \hat{x}_2(n) \). This will significantly reduce the coding efficiency of the method, especially at low rates, and will also require the estimation of the correlation coefficient \( \rho \), as in [10]. Both problems are avoided in our method by using \( \bar{x}_i(n) = \hat{x}_i(n-1) \).

As will be proved in Sec. 2.4, the loss of performance due to the use of the sub-optimal prediction is small when the source is highly correlated such as in densely sampled speech or audio signals.
2.4 Performance analysis

We now derive the expressions for the central distortion and the side distortion of the proposed MD predictive coding for the GM(1) source in (2.1). Let the rate in the first and second stage quantizers be $R_1$ and $R_2$, respectively. Therefore, the overall rate in each description is $R = R_1 + R_2/2$. We assume $R_1$ and $R_2$ are high enough so that the relevant quantization errors are approximately uniformly distributed.

When only one description is available, the reconstruction error is given by the first-stage encoder. Since $R_1$ is large, the variance, $\sigma^2_{e_{i(n)}}$, of the prediction error at the first-stage encoder $i$ is given by

$$\sigma^2_{e_{i(n)}} = E(x(n) - \hat{x}_i(n-1))^2 \approx E(x(n) - x(n-1))^2 = \left(\frac{2}{1 + \rho}\right) \sigma^2_{\omega},$$

(2.3)

where $\sigma^2_{\omega}$ is the variance of $\omega(n)$. Note that the prediction error $e_i(n)$ is not in general Gaussian since it is a function of the quantized version of the source sample $x_i(n-1)$. However, when the quantization rate is high the distribution of prediction error can be approximated by a Gaussian, thus when entropy coding is used, the side distortion
is given by

\[ D_1 = \left( \frac{2}{1 + \rho} \right) \left( \frac{2\pi e}{12} \right) \sigma_\omega^2 2^{-2R_1}. \tag{2.4} \]

If the optimal prediction \( \hat{\rho} \hat{x}_i(n - 1) \) is used, the term \( 2/(1 + \rho) \) will disappear in (2.3) and (2.4). Hence using the sub-optimal prediction loses only 0.22 dB and 0.11 dB in \( D_1 \) for \( \rho = 0.9 \) and \( \rho = 0.95 \), respectively.

When both descriptions are available at the decoder, the reconstruction error is determined by the second-stage quantizer. At high rates the error in the first-stage quantizer is uniform with variance \( D_1 \), and the input \( e(n) \) to the second-stage quantizer is also uniform with variance \( D_1/4 \). Therefore, the distortion given by the second-stage quantizer is

\[
D_0 = \frac{1}{4} D_1 2^{-2R_2} = \frac{1}{4} \left( \frac{2}{1 + \rho} \right) \left( \frac{2\pi e}{12} \right) \sigma_\omega^2 2^{-2(R_1 + R_2)}. \tag{2.5}
\]

Hence, the distortion product of the proposed algorithm is given by

\[
D_0 D_1 = \frac{1}{4} \left( \frac{2}{1 + \rho} \right)^2 \left( \frac{2\pi e}{12} \right)^2 \sigma_\omega^4 2^{-4R}. \tag{2.6}
\]

From (1.5) the following lower bound of the distortion product for MD coding of a GM(1) source can be obtained:

\[
D_0 D_1 \geq \frac{1}{4} \sigma_\omega^4 2^{-4R}. \tag{2.7}
\]

Comparing (2.6) and (2.7), our algorithm is away from the lower bound by 9.09 – 20 \( \log_{10}(1 + \rho) \) dB, which is 3.51 dB and 3.29 dB for \( \rho = 0.90 \) and \( \rho = 0.95 \) respectively. Out of this loss, 3.07 dB is due to the use of entropy-coded scalar quantizers.

We now derive the expected distortion of the system. Consider the transmission of two descriptions over two lossy channels with a loss probability of \( p \). In multiple description coding when both the descriptions are lost some other solution such as error concealment or retransmission must be used. In this derivation we assume retransmission in such a situation and therefore the end-to-end average distortion
$D_{av}$ can be written as

$$D_{av} = \frac{(1-p)^2}{k} D_0 + \frac{2p(1-p)}{k} D_1,$$

(2.8)

where $k = 1 - p^2$. To minimize $D_{av}$ with the constraint of $2R_1 + R_2 \leq 2R$, the following Lagrangian cost function can be defined

$$J_p = \frac{(1-p)^2}{k} D_0 + \frac{2p(1-p)}{k} D_1 + \lambda(2R_1 + R_2 - 2R),$$

(2.9)

and the optimal bit allocation is given by $R_2 = \frac{1}{2} \log_2 \left( \frac{1-p}{8p} \right)$ and $R_1 = R - \frac{R_2}{2}$. The minimal $D_{av}$ can thus be obtained. This shows that as the increase of $p$, more bits should be allocated to $R_1$ in order to reduce $D_1$. In practical implementations, we need to impose the constraints that $R_1 \in [0, R]$ and $R_2 \in [0, 2R]$.

### 2.5 Experimental results

In this section, the performance of the proposed algorithm is experimentally evaluated and compared with that of the algorithm in [10]. We assume that ideal entropy coding is employed and hence the entropy is used as the rate in all the simulations.

Fig. 2.3 shows the tradeoff between the central and the side distortions for a unit variance GM(1) source with $\rho = 0.9$ and $R = 3$ and 5 bits/sample/channel, respectively. Although the central distortion of [10] improves with the rate $R$, it can only achieve a very limited range of side distortion. Its minimal $D_1$ is determined by the inter-description prediction residual and is independent of the bit rate. In this example, the corresponding maximum side signal to noise ratio (SNR) is only about 12.8 dB, which may not meet the minimal requirement of some applications. Compared with this, the side SNR of our method can be improved significantly by varying the bit allocation. However, our algorithm is worse than [10] when the rate in the first stage is very low (i.e., when the side SNR is low). This is in fact an inherent problem of the MMDSQ. It becomes more pronounced in our method because the predictive encoders are less efficient in removing the source correlation at low rates.
As a summary, our method can be used as an alternative to [10] when its side SNR is not satisfactory.

Figure 2.4 (b) shows the end-to-end average distortion of the MD coding of a CD format (44.1kHz and 16 bits/sample) male speech signal with $R = 3$ and $R = 4$, respectively. The correlation coefficient of the signal is 0.99, which is typical for CD format speech and audio signals. It can be seen that our method outperforms [10] when the packet loss probability is greater than $10^{-3}$ and $10^{-4}$, respectively. This verifies the suitability of our method for high rate and high packet loss probability scenarios. The fact that our method does not need to estimate the correlation coefficient is also attractive to some applications.

### 2.6 Summary

This chapter presents a simple two-stage multiple description predictive coding scheme, based on the modified MDSQ in [43]. Theoretical and experimental results show that this method can serve as a good complement to the method in [10] for the high rate MD coding of strongly correlated sources.
Figure 2.3: The central SNR vs side SNR of the proposed method and the reference method in [10] for MD coding of a GM(1) source with $\rho = 0.9$ and unit variance at $R = 3$ and $R = 5$ bits/sample/channel.
Figure 2.4: End-to-end average distortion of the proposed method and the reference method in [10] for a speech signal coded at $R = 3$ and $R = 4$ bits/sample/channel.
Chapter 3

Two-Rate Coding with Staggered Quantization

3.1 Introduction

In this chapter, a scalar quantizer-based low-complexity $M$-channel MDC scheme is designed with good asymptotic and practical performances, especially in image coding. In the proposed algorithm, the source is first divided into $M$ subsets. Similar to [14, 20, 46, 1], each description in our method carries higher-rate information about one subset and lower-rate information about other subsets of the source. However, our method introduces two improvements. First, the lower-rate codings in different descriptions are designed to be mutually refinable by using staggered quantization, which generalizes the staggered quantizer in [43] for two-description coding. Secondly, for correlated sources, our method removes the correlation in the source by using two-rate predictive coding.

We derive closed-form expressions of the expected distortions of the proposed scheme for different numbers of received descriptions. Simulations with 1-D sources show that our method outperforms the lattice quantizer method in [24]. We also apply our scheme in lapped transform-based MD image coding, and formulate the optimization of the lapped transform for our scheme. For fair comparison, we implement the RD-MDC of [1] in the same lapped transform framework. Image coding
results demonstrate the superior performance of our method over [1].

### 3.2 System Design and Performance Analysis

In this section, we present the proposed MDC scheme with two-rate predictive coding and staggered quantization (TRPCSQ). We start with the system design for memoryless sources, followed by the design and analysis of the more general case with correlated sources.

#### 3.2.1 System Design for Memoryless Sources

Consider $M$-channel multiple description coding of a memoryless source $\{x(n)\}$ using uniform scalar quantizer and entropy coding. Fig. 3.1 illustrates the encoder of the proposed method for the three-channel case. We fictitiously partition the source samples into blocks of size $M$, with the $k$-th block containing samples $x(kM)$ to $x((k+1)M-1)$. By this partition, the $i$-th samples from all blocks, i.e., $\{x(kM+i)\}$ (for all $k$), form the $i$-th polyphase of the source.

To get the $i$-th description, we quantize the $i$-th samples of all blocks, i.e., the $i$-th polyphase, with a finer quantization step $q_0$. The resulting bit rate after entropy coding is $R_0$. Other samples are quantized with a coarser quantization step $q_1$, which results in a rate of $R_1$. $R_1$ is usually much smaller than $R_0$. As a result, each sample $x(n)$ is quantized into $R_0$ bits in exactly one description and $R_1$ bits in the remaining $M-1$ descriptions. The average bits per sample in each description is thus $R = (R_0 + (M-1)R_1)/M$. This two-rate coding concept is similar to the two-rate coding schemes in [14, 20, 46, 1].

A key difference between our method and [14, 20, 46, 1] is that the lower-rate quantizers in different descriptions of our method are designed to be mutually refinable, in order to reduce the distortion when some descriptions are lost. This is achieved by maintaining a shift or staggering among their quantization cells. Although the concept of staggered quantization is not new, its application and detailed analysis for $M$-channel MDC have not been reported.
Figure 3.1: Three-channel MDC of a memoryless source with the proposed method. The entropy coding is not shown. Darker arrows represent higher-rate ($R_0$) coded samples, and lighter arrows represent lower-rate ($R_1$) coded samples. (a) Encoding; (b) Decoding when only two descriptions are available.

Assume that sample $x(n)$ is coded into $R_0$ bits in Description 0 and $R_1$ bits in other descriptions. Assume also that the quantization step size is $q_1$ in the lower-rate quantizers. Fig. 3.2 shows how the lower-rate quantizers for $x(n)$ in Descriptions 1 to $M - 1$ are staggered. The uniform staggering is used due to its simplicity, where the shifts between neighboring quantizers are identical. To facilitate the performance analysis, we also define a joint quantizer of the $M - 1$ lower-rate quantizers. Since they are uniformly staggered, their combination yields a refined quantization partition with step size of

$$\Delta = q_1 / (M - 1). \quad (3.1)$$

It is also possible to introduce staggering between the partitions of the higher-rate quantizer and the lower-rate ones. However, the improvement is negligible, because in practice $R_0$ is usually much higher than $R_1$. Therefore, this method is not used in order to reduce the complexity of the system.
CHAPTER 3. TWO-RATE CODING WITH STAGGERED QUANTIZATION

Figure 3.2: The staggered lower-rate quantizers in the $M - 1$ descriptions and the joint quantizer.

At the receiver, when all the descriptions are available, the reconstruction for each sample is simply obtained from the description in which it is coded at the higher rate $R_0$. This is similar to [1]. Therefore the two methods have the same redundant bit rate, i.e., $(M - 1)R_1$.

When only a subset of the descriptions is received, as in Fig. 3.1(b), we first directly reconstruct those samples that are coded at the higher rate $R_0$ in any of the received descriptions. Each of the remaining samples is jointly reconstructed from all received descriptions. The joint decoding first finds the intersection of all the quantization cells that the sample belongs to in all received descriptions. The final reconstruction is then computed using the optimal reconstruction value of the joint bin. For example, at high rates, the midpoint can be used. Since all received lower-rate bits are used to refine the reconstruction, our method can achieve better performance than [1].
3.2.2 System Design for Correlated Sources

We next present the proposed $M$-channel MDC for a stationary correlated source \( \{x(n)\} \). In this case, each description still encodes all source samples sequentially, with one polyphase coded at a higher rate and others at a lower rate. In addition, to remove the source correlation, a modified DPCM system is used in each description, \textit{i.e.}, each sample \( x(n) \) is predicted from the previously reconstructed samples in the same description, and the prediction residual is encoded.

There are two differences between our modified DPCM and the conventional one. The first being that two-rate coding is used in each DPCM, \textit{i.e.}, in the \( i \)-th description, the prediction errors corresponding to the \( i \)-th polyphase samples are coded with the higher rate \( R_0 \), whereas the errors of other samples are coded with the lower rate \( R_1 \). The second difference is that the predictions for lower-rate coded samples are also quantized in order to achieve uniformly staggered quantizers, as explained later.

Note that the DPCM is applied within each description in our method, and we assume each description is either perfectly received or completely lost, as generally assumed in MDC framework. Therefore there is no drifting problem in our scheme.

At the decoder side, each description uses a separate DPCM decoder. When not all descriptions are received, we first reconstruct all the higher-rate coded samples from each received description and directly use them as the final reconstructions. Other samples need to be jointly reconstructed from all the available lower-rate coded outputs of the DPCM decoders. As in the case of memoryless case, the lower-rate coding is designed to be mutually refinable using staggered quantization. However, due to the use of predictive coding, the design and analysis of mutually refinable quantizers in the lower-rate coding are not straightforward.

Let \( \bar{x}_i(n) \) and \( e_i(n) \) be the prediction and prediction error for sample \( x(n) \) in the \( i \)-th description, respectively, \textit{i.e.},

\[
x(n) = \bar{x}_i(n) + e_i(n).
\]

In DPCM, the prediction error \( e_i(n) \) is quantized. This induces a quantization partition on \( x(n) \) with the same step size. However, the partition of \( x(n) \) is shifted from
that of $e_i(n)$ by $\bar{x}_i(n)$.

In the following discussion, without loss of generality, we assume that at time $n$ the prediction error $e_0(n)$ of $x(n)$ in Description 0 is coded at the higher rate $R_0$, and $e_i(n), i = 1, ..., M - 1$, are coded at the lower rate $R_1$ with a step size of $q_1$. When only lower-rate coded versions of $x(n)$ are received, we can reconstruct $x(n)$ from the intersection of all the induced partitions for $x(n)$, leading to a better reconstruction quality than the individual decoders.

Our objective is to design the DPCM coders such that the induced lower-rate quantization partitions for $x(n)$ from Description 1 to Description $M - 1$ are uniformly staggered, thereby simplifying the decoder implementation. This kind of staggering can be achieved by imposing the following two conditions: (1) The $M - 1$ lower-rate quantizers for $e_i(n), i = 1, ..., M - 1$, are uniformly staggered, as shown in Fig. 3.2. (2) The corresponding predictions $\bar{x}_i(n), i = 1, .., M - 1$, are also quantized by a uniform quantizer with the same quantization step $q_1$ before being used to predict the source sample $x(n)$. This is the second difference of our DPCM from the conventional one. Note that the predictions for higher-rate coded samples are not quantized.

To see this, notice from Fig. 3.2 that the uniform staggering is periodic with period $q_1$, i.e., the partition in any description can be shifted by an integer multiple of $q_1$ without changing the uniform staggering. The second condition above ensures that the partition for $x(n)$ in description $i$ is shifted from that of $e_i(n)$ by a multiple of $q_1$. Therefore the uniform staggering across different descriptions is still preserved.

The quantization of the prediction was previously employed in [53] to maintain the required offset between the side quantizers of a DPCM-based two-description MDC. The scheme developed here can be viewed as the generalization of [53] to more than two descriptions. We also give the detailed performance analysis of such a generalization in this chapter.

### 3.2.3 Effect of Prediction Quantization

In this part, we show that although the quantization of the prediction $\bar{x}_i(n)$ makes the prediction sub-optimal, the degradation is negligible when the quantization step
CHAPTER 3. TWO-RATE CODING WITH STAGGERED QUANTIZATION

$q_1$ is sufficiently small. Moreover, the variance of the prediction error $e_i(n)$ will be approximately the same for all $i$ and $n$, as in the standard DPCM. The analysis is based on high rate assumption, but as in many source coding problems, the optimal design based on the high rate model usually still has good performance at low rates [23].

To illustrate these results, consider the MD coding of a first-order Gauss-Markov source, i.e.,

$$x(n) = \rho x(n-1) + \omega(n), \quad (3.3)$$

where $\rho$ is the correlation coefficient, and $\omega(n)$ is a white noise sample with variance $\sigma^2_\omega$. In this case, the optimal prediction for $x(n)$ in Description $i$ at high rates is

$$\bar{x}_i(n) = \rho \hat{x}_i(n-1), \quad (3.4)$$

where $\hat{x}_i(n-1)$ is the reconstruction of $x(n-1)$ in Description $i$.

Let $q_{p,i}(n)$ be the error introduced by the quantization of the prediction $\rho \hat{x}_i(n-1)$ with a step size of $q_1$. The final prediction error of $x(n)$ in Description $i$ becomes

$$e_i(n) = x(n) - (\rho \hat{x}_i(n-1) + q_{p,i}(n)). \quad (3.5)$$

The quantization of $e_i(n)$ with the same step size $q_1$ introduces a quantization error $q_{e,i}(n)$, i.e.,

$$\hat{e}_i(n) = e_i(n) + q_{e,i}(n). \quad (3.6)$$

The reconstruction of $x(n)$ in Description $i$ is

$$\hat{x}_i(n) = \hat{e}_i(n) + (\rho \hat{x}_i(n-1) + q_{p,i}(n)). \quad (3.7)$$

From (3.5), (3.6), and (3.7), we get

$$\hat{x}_i(n) - x(n) = \hat{e}_i(n) - e_i(n) = q_{e,i}(n). \quad (3.8)$$

Therefore with the quantization of the prediction, the final reconstruction error still
equals to the quantization error of the prediction residual, as in the standard DPCM system [40]. Thus the prediction residual $e_i(n)$ in (3.5) can be written as

$$e_i(n) = x(n) - (\rho x(n-1) + \rho q_{e,i}(n-1) + q_{p,i}(n))$$

$$= \omega(n) - (\rho q_{e,i}(n-1) + q_{p,i}(n)).$$

(3.9)

Strictly speaking, $e_i(n)$ above is non-stationary, because the quantization steps for the prediction and prediction residual vary with time, and $q_{p,i}(n)$ only appears when the sample is coded at the lower rate. However, note that $q_{e,i}(n-1)$ and $q_{p,i}(n)$ are in the same order since they are quantization errors of two quantizers with the same step $q_1$. Therefore when the rates are sufficiently high, the quantization step $q_1$ is much smaller than $\sigma_\omega$, the standard deviation of $\omega(n)$. Thus the two quantization noises in (3.9) are negligible, and $e_i(n)$ can be considered as approximately stationary with a variance of $\sigma_\omega^2$. This is the same as in the standard DPCM, where the quantization noise $q_{e,i}(n-1)$ of $x(n-1)$ is also ignored in the predictor design and performance analysis (see page 114 of [40]).

3.2.4 Performance Analysis

We now derive closed-form expressions of the expected distortion when different numbers of descriptions are received. The solution will be verified in Sec. 3.4. We use $D_k$ ($k = 0, ..., M$) to represent the expected distortion when $k$ descriptions are received. The solution for correlated sources is derived first, from which the memoryless case can be easily deduced. For comparison purpose, the distortions are also derived in Sec. 3.2.5 for memoryless sources when the staggering is not used in the lower-rate quantizers.

We first derive $D_M$ when all descriptions are available. In this case, each sample is reconstructed using the description in which it is coded at the higher rate $R_0$. As shown in (3.8) and (3.9), the reconstruction error in the two-rate DPCM system is still given by the quantization error of the prediction residual, and the residual is approximately stationary at high rates. Therefore the R-D formula of the standard entropy coded scalar quantizer can still be applied, i.e., the distortion $D_M$ with $M$
descriptions received is
\[ D_M = \frac{1}{12} \epsilon^2 2^{-2R_0}. \] (3.10)
where \( \epsilon^2 \) is the entropy power of the prediction error [40].

When \( k < M \) descriptions are available, only \( k/M \) portion of the source samples can be reconstructed from the higher-rate coding. The rest \( (M-k)/M \) portion of the samples need to be jointly reconstructed from the lower-rate coding using all received descriptions. Therefore the expected distortion \( D_k \) is
\[ D_k = \frac{1}{M} \left( k \frac{2\pi e}{12} \epsilon^2 2^{-2R_0} + (M-k)D'_k \right), \] (3.11)
where \( D'_k \) is the expected distortion of the jointly reconstructed samples from the \( k \) received descriptions, which will be derived next.

When only one description is received, \( i.e., k = 1 \), no joint decoding can be applied. Similar to (3.10), the distortion is simply
\[ D'_1 = \frac{2\pi e}{12} \epsilon^2 2^{-2R_1}. \] (3.12)

We now analyze the case with \( 1 < k < M \). Assume that \( x(n) \) is coded at the lower rate \( R_1 \) in all received descriptions. In the following, we denote a quantization cell of the lower-rate quantizer as a **macrocell**, and denote the macrocells that \( x(n) \) belongs to in the received descriptions as the **received macrocells**. In addition, a cell in the joint quantization partition in Fig. 3.2 is called a **finer cell**.

In our scheme, when \( k \) of the \( M \) descriptions are received, the reconstruction of the lower-rate coded sample \( x(n) \) is chosen to be the midpoint of the intersection of all received macrocells, which is nearly optimal at high rates. In this case, the source is approximately uniformly distributed within the intersection. Therefore if the intersection width is \( l\Delta \), the expected distortion for \( x(n) \) would be \( \frac{1}{12} (l\Delta)^2 \), where the intersection width is determined by the leftmost and rightmost received macrocells.

The minimal width of the intersection is clearly \( \Delta \). To find the maximal possible width of the intersection, note that any two received macrocells have a shift of at least \( \Delta \). Therefore the shift between the leftmost and rightmost received macrocells is at
least \((k - 1)\Delta\). The maximal intersection of \(k\) received macrocells is thus \((M - k)\Delta\).

Hence the expected distortion \(D'_k\) for joint decoding is

\[
D'_k = \sum_{l=1}^{M-k} p_{M,k}(l) \frac{(l\Delta)^2}{12},
\]

where \(p_{M,k}(l)\) is the probability of having an intersection of \(l\Delta\) among the \(k\) received macrocells. In this chapter, we assume different descriptions are transmitted independently with the same probability of loss. Therefore \(p_{M,k}(l)\) can be obtained by the following combinatorial method.

To calculate \(p_{M,k}(l)\), we index the finer cells in each macrocell from 1 to \(M - 1\). Since each sample is coded at lower rate in \(M - 1\) descriptions, and the quantizers in these descriptions are uniformly staggered, the finer cell indices of the sample in the \(M - 1\) descriptions must be mutually distinctive. Therefore, if only \(k\) descriptions are received, the number of possible configurations of the received finer cell indices is \(\binom{M-1}{k}\). Since we assume the description loss probabilities are independent and identically distributed, each configuration will have a probability of \(1/\binom{M-1}{k}\). Moreover, the received finer cell index configurations can be classified into \(M - k\) categories, according to their intersection widths.

It should be pointed out that each decoder only knows the macrocell index of a sample, but not its finer cell index. The finer cell indices are only used here to facilitate the calculation of \(p_{M,k}(l)\).

If we sort the received finer cell indices in increasing order, and let \(I_i\) represent the \(i\)-th index after the sort, the leftmost and the rightmost received macrocells will be determined by \(I_k\) and \(I_1\). Clearly, the leftmost and the rightmost received macrocells are shifted by \(I_k - I_1\) finer cells, and the number of finer cells in their intersection is

\[
l = M - 1 - (I_k - I_1).
\]

For example, when \(M = 5\), the four sorted finer cell indices for \(x(n)\) are \((1, 2, 3, 4)\). If \(k = 3\), there are 4 possible configurations of the received finer cell indices, i.e., \((1, 2, 3), (2, 3, 4), (1, 2, 4)\), and \((1, 3, 4)\), as shown in Fig. 3.3, together with their
corresponding received macrocells. The possible macrocell intersection widths are $\Delta$ and $2\Delta$.

Note that each set of possible received finer cell indices can appear in different orders and in different received descriptions, depending on the location of the sample. For example, Fig. 3.4 shows the possible scenarios for the pattern in Fig. 3.3 (a). It can be seen that all of them have the same intersection width and reconstruction error, as all of them can be converted to the format in Fig. 3.3 (a) by reordering the descriptions. Comparing Fig. 3.3 and Fig. 3.4, it is apparent that the sorted finer cell indices provide a description-invariant and location-invariant way of analyzing the performance of the staggered quantizers.
To find the expression of $p_{M,k}(l)$, we use the following two facts. First, when the intersection width is $l\Delta$, there are $l$ possible pairs of $(I_1, I_k)$ for the finer cell indices of the rightmost and the leftmost received macrocells, because the sample can be in any of the $l$ finer cells in the intersection. In Fig. 3.3, the possible $(I_1, I_k)$ pairs for $l = 2$ are (1, 3) and (2, 4), whereas the only pair for $l = 1$ is (1, 4).

Second, for a given pair of $(I_1, I_k)$, the finer cell indices of the remaining $k - 2$ received macrocells can take any $k - 2$ values between them without affecting the intersection width. Since the maximal number of finer cell indices between them is $I_k - I_1 - 1$, there are $\binom{I_k - I_1 - 1}{k - 2}$ different combinations of the received finer cell indices. This can be seen from Fig. 3.3 (c-d), where $(I_1, I_k) = (1, 4)$, therefore $\binom{I_k - I_1 - 1}{k - 2} = 2$, i.e., the third finer cell index can take the value of 2 or 3, leading to the combinations of $(1, 2, 4)$ and $(1, 3, 4)$.
The analysis above shows that $p_{M,k}(l)$ is given by

$$p_{M,k}(l) = \frac{l^{(M-2-k-l)}_{k-2}}{l^{(M-1)}_{k-2}} = \frac{(M-1)}{(M-k)}$$,

where we use the relationship $I_k - I_1 = M - 1 - l$ from (3.14).

Plugging into (3.13), we get

$$D'_k = \frac{\Delta^2}{12(M-k)} \sum_{l=1}^{M-k} \left( \frac{M-2-l}{k-2} \right)^3.$$

Using $\Delta = q_1/(M-1)$ and the result of entropy coded scalar quantizer, this can be rewritten as

$$D'_k = \frac{1}{(M-1)^2(M-k)} \left( \sum_{l=1}^{M-k} \left( \frac{M-2-l}{k-2} \right)^3 \right) \frac{2\pi e^2}{12} 2^{2-2R_1}.$$

Define a new variable $\alpha_{M,k} = 1$ for $k = 1$, and

$$\alpha_{M,k} = \frac{1}{(M-1)^2(M-k)} \left( \sum_{l=1}^{M-k} \left( \frac{M-2-l}{k-2} \right)^3 \right)$$

for $1 < k \leq M - 1$, (3.12) and (3.17) can be unified into

$$D'_k = \frac{2\pi e}{12} \alpha_{M,k} \epsilon^2 2^{-2R_1}, \quad 1 \leq k \leq M - 1.$$

The distortion $D_k$ in (3.11) thus becomes

$$D_k = \frac{1}{M} \left( k^{2\pi e / 12} \epsilon^2 2^{-2R_0} + (M-k)\frac{2\pi e}{12} \alpha_{M,k} \epsilon^2 2^{-2R_1} \right).$$

This will be verified in Sec. 3.4 and compared with the MDLVQ method in [24].

Given the distortion formulas for correlated sources, the distortion for memoryless sources can be easily obtained. The only difference in the proposed algorithm for correlated sources and memoryless sources is that predictive coding is not used in the
latter. Instead of quantizing the prediction errors, for memoryless sources the two-rate quantization is applied directly to the source samples. The staggered quantizers are still used in the lower-rate quantizers, as in the correlated source case. Therefore, if we denote $\epsilon^2$ to be the entropy power of the source rather than the entropy power of the prediction error, equations (3.10) to (3.20) still hold for memoryless sources.

3.2.5 Distortion expressions for two-rate coding without staggered quantization

In this part, we determine the amount of improvement provided by the staggered quantization. For this purpose we derive the distortions of two-rate coding of a memoryless source without using the staggered quantizers. In this case, the coding and the description formation is similar to the proposed method, with the only difference that the quantizers in the lower-rate coding are not staggered. We call this simplified method the two-rate coding scheme (TRC).

At the receiver, if a sample is coded at the higher rate in any of the received description, its reconstruction is obtained from that description. Otherwise, its reconstruction is obtained from any of the available lower-rate coded samples, which are identical. The distortions of higher-rate coded samples and lower-rate coded samples are simply $\frac{2\pi e}{12} \epsilon^2 2^{-2R_0}$ and $\frac{2\pi e}{12} \epsilon^2 2^{-2R_1}$, respectively, where $\epsilon^2$ is the entropy power of the source. Therefore, the expected distortion $D_{k,\text{TRC}}$ of the simplified TRC scheme with $k \in [1, M]$ received descriptions is

$$D_{k,\text{TRC}} = \frac{1}{M} \left( k \frac{2\pi e}{12} \epsilon^2 2^{-2R_0} + (M - k) \frac{2\pi e}{12} \epsilon^2 2^{-2R_1} \right).$$  

(3.21)

It can be seen from (3.20) and (3.21) that the only difference of the two methods lies in the term $\alpha_{M,k}$. Therefore, for the same $R_0$ and $R_1$, the two methods have the same performance when all descriptions are available ($k = M$). Also, since $\alpha_{M,1} = 1$, the two methods are also equivalent when only one description is received ($k = 1$). In all other cases, the proposed method with the staggered quantizers can have better performance due to the joint decoding.
To evaluate the performance gain of the proposed method for other values of $k$, we consider the case where $R_0$ is sufficiently larger than $R_1$ such that the term $k \frac{2 \pi e}{12} e^{2-2R_0}$ is negligible in (3.20) and (3.21) compared to the second term. In this case, $D_k$ in (3.20) is approximately less than $D_{k, \text{TRC}}$ in (3.21) by a factor of $\alpha_{M,k}$. Fig. 3.5 shows the value of $\alpha_{M,k}$ for different values of $k$ and $M$. For example, when $M = 6$ and $k = 5$, $\alpha_{M,k}$ is approximately -14 dB, i.e., $D_5$ is less than $D_{5, \text{TRC}}$ by 14 dB in six-description MDC. Therefore, the staggered quantizer-based refinable coding provides significant improvements over the non-staggered case. It can also be seen that for a given $M$, the improvement increases almost linearly with $k$ (in log scale). As $M$ increases, the maximal improvement is also increased.

### 3.3 Optimal Design for MD Image Coding

In this section we apply our scheme to image coding, where block transform is used instead of the DPCM. The time-domain lapped transform (TDLT) developed in [48] is adopted, which generalizes the DCT by applying time-domain pre/postfilters. The TDLT has been selected by the forthcoming JPEG XR standard [36], which is a low-cost alternative to JPEG 2000 with competitive performance.

To get MDC, the transformed blocks are partitioned into $M$ subsets after the DCT. In each description, one subset is coded with a higher bit rate $R_0$ and other subsets are coded with a lower rate $R_1$. The lower-rate coding is made mutually refinable using staggered quantization. The DC coefficients are encoded with two-rate predictive coding as explained in section 3.2.

A nice property of the TDLT is that the pre/postfilters can be optimized for different applications. Since our MDC framework described above differs from the traditional single description coding, the optimal filters found in [48] may not be optimal in our scheme. In this section, we formulate the optimization of the pre/postfilters and the bit allocation for our framework.
CHAPTER 3. TWO-RATE CODING WITH STAGGERED QUANTIZATION

Figure 3.5: Performance gain of the staggered quantizers: $\alpha_{M,k}$ for different $M$ and $k$.

3.3.1 Overview of the Time-Domain Lapped Transform

The original lapped transform reported in [19] is applied as a post filter after DCT at block boundaries. On the other hand, in TDLT, a prefilter is applied before DCT at block boundaries, and a post filter is applied at the same locations after inverse DCT. This makes TDLT more compatible and easy to integrate with the existing DCT based software or hardware codec implementations. Lapped transform can be designed to achieve different requirement. In single description coding, the usual design goal is to achieve better compression. However, lapped transform can also be designed to spread out information of a block into the neighboring blocks which can be used in
recovering a block in an error scenario.

Fig. 3.6 shows the block diagrams of the forward and inverse TDLT for 1-D data. An $L \times L$ prefilter $P$ is applied at the boundary of two blocks ($L$ is the block size). The $L$-point DCT $C$ is then applied to each block. As a result, the basis functions of the TDLT cover two blocks. At the decoder, the inverse DCT and the postfilter $T$ at block boundaries are applied. $P$ and $T$ have the following structures to yield linear-phase filters [48]:

$$P = W \text{diag}\{I, V\} \ W$$ (3.22)

and

$$T = P^{-1} = W \text{diag}\{I, V^{-1}\} \ W,$$ (3.23)

where

$$W = \frac{1}{\sqrt{2}} \begin{bmatrix} I & J \\ J & -I \end{bmatrix},$$ (3.24)

$I$ and $J$ are $\frac{L}{2} \times \frac{L}{2}$ identity matrix and counter-identity matrix, respectively. The
matrix $V$ is an $\frac{L}{2} \times \frac{L}{2}$ invertible matrix that can be optimized for different purposes. diag\{$A, B$\} denotes a block diagonal matrix with matrices $A$ and $B$ on the diagonal, and zeros elsewhere.

Let $P = \begin{bmatrix} P_0^T & P_1^T \end{bmatrix}^T$ where $P_0$ and $P_1$ contain the first and the last $L/2$ rows of the prefilter $P$, respectively. Define $P_{12} = \text{diag}\{P_1, P_0\}$. The $L \times 2L$ forward transform is given by $F = CP_{12}$.

To obtain the inverse transform, let $T = \begin{bmatrix} T_0 & T_1 \end{bmatrix}$, where $T_0$ and $T_1$ are the first and the last $L/2$ columns of $T$, respectively. Define $T_{21} = \text{diag}\{T_1, T_0\}$. The $2L \times L$ inverse transform is thus $G = T_{21}C^T$.

### 3.3.2 Optimal Design of the Transform

Given the target bit rate $R$ and the probability $p$ of losing each description, our objective is to find the optimal prefilter and postfilter in the TDLT and the optimal bit allocation $R_0$ and $R_1$ that minimize the following expected distortion

$$D = \sum_{k=0}^{M} p_k D_k,$$

(3.25)

where $p_k$ is the probability of receiving $k$ descriptions, and $D_k$ is the corresponding expected reconstruction error. When $k = 0$, $D_k$ is chosen as the variance of the input.

In the proposed MDC scheme, when $k$ descriptions are available, $k$ out of $M$ DCT blocks will be reconstructed from the higher-rate coding and the rest will be jointly reconstructed from lower-rate coding. After postfiltering at block boundaries, each reconstructed image block contains the contributions from the quantization errors of two DCT blocks. When $R_0$ and $R_1$ are sufficiently high, the quantization errors in different DCT blocks are approximately uncorrelated, and their contributions to the reconstruction error are additive. Therefore $D_k$ can still be written as

$$D_k = \frac{1}{M} \left( kD_M + (M - k)D'_k \right),$$

(3.26)

where $D_M$ is the average reconstruction error caused by the DCT blocks coded with
rate $R_0$, which is also the reconstruction error when all descriptions are available. $D'_k$ is the reconstruction error caused by the joint decoding of lower-rate coded DCT blocks.

Plugging (3.26) into (3.25), we get

$$D = \frac{1}{M} \sum_{k=1}^{M} p_k k D_M + \frac{1}{M} \sum_{k=1}^{M} (M - k) p_k D'_k + p_0 D_0$$  \hspace{1cm} (3.27)

$$= \frac{\mu_k}{M} D_M + \frac{1}{M} D' + p_0 D_0,$$

where $\mu_k$ is the expected value of $k$ (which can be obtained when the loss probability of each description is independent and identical), and

$$D' = \sum_{k=1}^{M} (M - k) p_k D'_k.$$  \hspace{1cm} (3.28)

To find $D_M$, assume that $y(n)$ is a higher-rate coded DCT block with quantization error $q_y(n)$. After the inverse TDLT, the reconstruction error becomes $Gq_y(n)$. As usual, we assume the quantization noises of different subbands are uncorrelated. Therefore the average reconstruction error per sample is

$$D_M = \frac{1}{L} \sum_{i=0}^{L-1} ||g_i||^2 \sigma_{q_y(i)}^2,$$  \hspace{1cm} (3.29)

where $\sigma_{q_y(i)}^2$ is the variance of the quantization noise of the $i$-th entry of $y(n)$, and $g_i$ is the $i$-th column of $G$. As in Sec. 3.2, at high rates $\sigma_{q_y(i)}^2$ can be written as

$$\sigma_{q_y(i)}^2 = \frac{2\pi e}{12} \epsilon^2(i) 2^{-2R_0i},$$  \hspace{1cm} (3.30)

where, for $i = 0$, $\epsilon^2(i)$ is the entropy power of the prediction error of the DC coefficient, since we use predictive coding for the DC coefficient. For other $i$, $\epsilon^2(i)$ is the entropy power of the respective DCT coefficient. $R_0i$ is the bits allocated to the $i$-th coefficient of a higher-rate coded DCT block, such that $\frac{1}{L} \sum_{i=0}^{L-1} R_{0i} = R_0$. 
Upon the optimal bit allocation [13] for rates $R_0$, the minimal value for (3.29) is given by

$$D_M = \frac{2\pi e}{12} \left( \prod_{i=0}^{L-1} \|g_i\|^2 \epsilon^2(i) \right)^{\frac{1}{L}} 2^{-2R_0} \triangleq \frac{2\pi e}{12} \epsilon^2 2^{-2R_0}. \quad (3.31)$$

To find $D'$ in (3.28) we need to find the distortion $D'_k$. In this case, let $y(n)$ be a lower-rate coded DCT block. Similar to (3.29), $D'_k$ can be written as

$$D'_k = \frac{1}{L} \sum_{i=0}^{L-1} \|g_i\|^2 \sigma^2_{q_y(k,i)}, \quad (3.32)$$

where $\sigma^2_{q_y(k,i)}$ is the quantization noise variance of the $i$-th entry of $y(n)$ after joint de-quantization from $k$ received descriptions. Therefore, from (3.19), $\sigma^2_{q_y(k,i)}$ is given by

$$\sigma^2_{q_y(k,i)} = \frac{2\pi e}{12} \alpha_{M,k} \epsilon^2(i) 2^{-2R_1}, \quad (3.33)$$

where $\epsilon^2(i)$ is the same as in (3.30), whereas $\alpha_{M,k}$ is independent of $i$ and is given by (3.18). Substituting (3.32) and (3.33) in (3.28) we have

$$D' = \sum_{k=1}^{M} (M - k) p_k \frac{1}{L} \sum_{i=0}^{L-1} \|g_i\|^2 \frac{2\pi e}{12} \alpha_{M,k} \epsilon^2(i) 2^{-2R_1}$$

$$= \frac{2\pi e}{12} \sum_{k=1}^{M} (M - k) p_k \alpha_{M,k} \left( \frac{1}{L} \sum_{i=0}^{L-1} \|g_i\|^2 \epsilon^2(i) 2^{-2R_1} \right) \quad (3.34)$$

$$= \frac{2\pi e}{12} \beta_M \frac{1}{L} \sum_{i=0}^{L-1} \|g_i\|^2 \epsilon^2(i) 2^{-2R_1},$$

where $\beta_M = \sum_{k=1}^{M} (M - k) p_k \alpha_{M,k}$ is a constant independent of $i$. Minimizing $D'$ under the constraint $\frac{1}{L} \sum_{i=0}^{L-1} R_{1i} = R_1$ yields

$$D' = \frac{2\pi e}{12} \beta_M \left( \prod_{i=0}^{L-1} \|g_i\|^2 \epsilon^2(i) \right)^{\frac{1}{L}} 2^{-2R_1} = \frac{2\pi e}{12} \beta_M \epsilon^2 2^{-2R_1}. \quad (3.35)$$

Our objective is to minimize the average distortion $D$ by optimizing the TDLT,
in particular, the matrix $V$ in the prefilter. From (3.27), (3.31) and (3.35) it is clear that a TDLT is optimal if and only if it minimizes

$$
\left( \prod_{i=0}^{L-1} ||g_i||^2 \epsilon^2(i) \right)^{\frac{1}{L}}.
$$

There is no closed-form solution to this optimization, but a numerical optimization can be used to find the optimal TDLT.

Under the constraint of $R_0 + (M-1)R_1 = MR$, the optimal bit allocation that minimizes $D$ is

$$
R_0 = R + \frac{M-1}{2M} \log_2 \frac{\mu_k(M-1)}{\beta_M},
$$

$$
R_1 = R - \frac{1}{2M} \log_2 \frac{\mu_k(M-1)}{\beta_M}.
$$

With the assumption that the input is a first order Gauss-Markov (GM(1)) source with correlation coefficient 0.95. The optimized matrix $V$ in the TDLT prefilter $P$ is found to be

$$
V = \begin{bmatrix}
0.9490 & 0.7577 & 0.2961 & 0.1526 \\
-0.5526 & 0.8846 & 0.5673 & 0.1524 \\
0.1093 & -0.3700 & 1.0589 & 0.3169 \\
-0.0309 & 0.0035 & -0.1304 & 1.1391
\end{bmatrix}.
$$

### 3.4 Experimental Results

In this section we compare the performances of the proposed two-rate predictive coding with staggered quantization algorithm (TRPCSQ) and two state-of-the-art MDC methods in [24, 1].

#### 3.4.1 Experimental Results for 1-D Sources

We first verify the theoretical distortions derived in Sec. 3.2.4. Fig. 3.7 compares the theoretical curves and the simulation results for four description coding of a unit
variance first order Gauss-Markov (GM(1)) source with correlation coefficient 0.9. The rate is fixed at $R = 5$ bits per sample per description. The curves are obtained by choosing different $R_0$ and $R_1$ such that $R_0 + (M - 1)R_1 = MR$. In the simulation, the entropy of the encoder output is used as the measure of the bit rate. It can be seen that the theoretical results agree well with the experimental results.

Fig. 3.8 compares the proposed method and the $M$-channel lattice vector quantizer based MDC (MDLVQ) in [24] with $A_2$ lattice. The four-description coding of a unit variance Gaussian memoryless source with $R = 5$ bits/sample/description is tested. Only the theoretical results of the MDLVQ are included. According to [24], the differences between the theoretical and simulation results of the MDLVQ are negligible at this bit rate.

It can be seen that the distortion $D_3$ of the proposed TRPCSQ method is about 3 dB better than the MDLVQ in most cases. $D_2$ of the MDLVQ is comparable to the proposed method. Only $D_1$ of the MDLVQ is better than our method by about 2 dB. However, when the packet loss probability is small, the probability of receiving only one description is very low. Therefore the proposed method should outperform the MDLVQ in most cases when the expected distortion is considered. This is verified in Fig. 3.9, where the optimal expected distortion is plotted for TRPCSQ and MDLVQ for different loss probabilities. The redundancy is tuned in both methods to achieve the minimum expected distortion for each loss probability. It can be seen the expected distortion of our method is 2 dB better than the MDLVQ when the packet loss probability is below $10^{-2}$. Although the improvement diminishes as the packet loss probability increases, the MDLVQ can only outperform when the packet loss probability is greater than 0.5, which is unlikely in practice.

Note that the complexity of the lattice quantizer-based MDLVQ is significantly higher than our scalar quantizer based method. In addition, the MDLVQ involves an index assignment problem, which does not have a unique solution and may cause unbalanced descriptions. Another advantage of our method is that it can be applied to multiple description image coding with good performance, as shown next.
3.4.2 Experimental Results for MD Image Coding

In this part, the proposed method is applied to the TDLT-based image coding [48]. The transformed blocks are uniformly partitioned into $M$ subsets. For example, if $M = 4$, a checkerboard pattern is used. In each description, one subset is coded with a higher bit rate and other subsets are coded with a lower rate. The prediction for each DC coefficient is the average of the DC coefficients of the top and left neighboring blocks. The entropy coding method in [49] is then used to encode the two-rate quantized DC prediction residuals and other coefficients.

Since there are very few $M$-channel MD image coding software, we compare our method with the two-rate RD-MDC method in [1], whose code is available at [29]. Note that [1] is based on wavelet, and one of its goals is to maintain compatibility with JPEG 2000. In addition, the algorithm in [1] is not fully optimized to yield more than two descriptions. Therefore this comparison should not be used to judge against the method in [1]. To get fair comparison, we also implement the RD-MDC method in the TDLT framework, and the results are reported in this part.

In the image coding experiments, we use peak signal to noise ratio (PSNR) as the performance measure, which is defined as below.

$$PSNR = 10 \log_{10} \left( \frac{\text{MAXVAL}^2}{\text{MSE}} \right)$$  \hspace{1cm} (3.39)

where, MAXVAL is the maximum pixel value, which is equal to 255 in all the image coding experiments in this thesis, since 8 bits are used to represent a gray scale pixel of the original images being coded.

Fig. 3.10 - 3.15 show the side PSNR vs central PSNR of our method and RD-MDC in [1] for three-description and four-description coding of the $512 \times 512$ images Barbara, Boat and Lena at the total bit rate of $MR = 1$ bits/pixel. The optimized prefilter in (3.38) is used in our method, and the header bytes of the JPEG 2000 are excluded in bit rate calculation. Given the same central PSNR, the side PSNR of the proposed method outperforms the RD-MDC in almost all the cases. Note that part of the gain is due to the different codecs used in the two methods.
CHAPTER 3. TWO-RATE CODING WITH STAGGERED QUANTIZATION 48

In order to get a fair comparison between our method and the two-rate RD-MDC in [1], we also implement the RD-MDC in the TDLT framework. The transformed blocks are partitioned into $M$ subsets in the same way as in our method. Each description carries one subset with a higher rate and others with a lower rate, as in [1]. To ensure fairness, we use the single-description optimized pre/postfilters in both algorithms. This leads to some degradations of the performance of our method. We also apply DC prediction in the RD-MDC before two-rate coding, as in our method. This improves the performance of the RD-MDC. In a summary, the only difference between our TRPCSQ and the TDLT-based RD-MDC is that staggered quantizers are used in TRPCSQ for the lower-rate coding, whereas only simple two-rate coding is used in the RD-MDC.

The two TDLT-based methods are compared in Fig. 3.16 - 3.18 for four-description coding. Fig. 3.16 and 3.17 show the tradeoff between the side and central PSNRs. Compared to JPEG 2000-based RD-MDC in Fig. 3.13 and 3.14, the TDLT-based RD-MDC has better performance for Barbara and is similar for Lena. Our method has better performance than TDLT-based RD-MDC when two or three descriptions are received, but is slightly worse with one description because of the staggered quantizers, but this case has lower probability in practice. In Fig. 3.18 the optimal expected PSNR is plotted for TRPCSQ and RD-MDC at different loss probabilities, where the redundancy is tuned to achieve the maximum expected PSNR for each loss probability. The superior performance of our method can be clearly observed.

Fig. 3.19 and Fig. 3.20 show portions of the image Barbara reconstructed from two and three descriptions respectively when the image is coded into four descriptions. The central distortion is fixed at 35.13 dB, with $R_0 \approx 0.66$ and $R_1 \approx 0.11$ bits in the three TDLT-based methods. The TDLT-based RD-MDC is 1.96 dB and 1.34 dB better than the JPEG 2000-based RD-MDC in the two cases, due to different codecs and the DC prediction. Our method with the single description optimized filters further improves the result of TDLT-based RD-MDC by 0.61 and 1.31 dB. This is a fair comparison. The optimized filters for our scheme generates an additional gain of 0.25 and 0.20 dB, respectively.

It should be noted that the idea of staggered quantization can also be used in both
our method and the RD-MDC method to design the higher-rate quantizer and the lower-rate quantizers, such that the former can be refined by the latter when more than one description is received. This improvement has been used in [47] for two-description coding. However, it only has notable gains in the high redundancy region, where the quantization step sizes are similar in the higher-rate and lower-rate quantizers. In addition, the generalization to the $M$-description case is not straightforward, and it is not immediately clear how large the gain can be. A systematic investigation of this topic is beyond the scope of this work.

3.5 Summary

This chapter presents an $M$-channel MDC method using two-rate coding and staggered quantizers. Closed-form expressions of the expected distortions of the system are derived for different number of received descriptions. The method is also applied to lapped transform-based multiple description image coding. Experimental results with 1-D and 2-D data show that this method achieves better performance than other state-of-the-art schemes. The image coding results can be further improved. For example, more advanced prediction can be used instead of simple DC prediction, and the entropy coding can be fine-tuned based on the characteristics of the prediction residuals and the staggered quantizers.

Finally, although the constraint of network packet size is not considered in this chapter, our method can be modified in a number of ways to be used in packet networks. For example, the image can be divided into slices, and each slice is coded by our method, such that each description can be fit into one packet. Alternatively, after encoding the image into a number of descriptions, each description can be divided into multiple packets, by modifying the approach in [5].

In the next two chapters, we generalize the prediction compensation based MDC method in [38] to $M$-channels and propose some improvements.
Figure 3.7: Side distortions $D_1, D_2, D_3$ vs central distortion $D_4$ of the proposed method for four-description coding of a unit variance GM(1) source with correlation coefficient 0.9 and $R = 5$ bits/sample/description.
Figure 3.8: Side distortion vs. central distortion of the proposed method and the MDLVQ in [24] for four description coding of a Gaussian source of unit variance.
Figure 3.9: Expected distortion vs loss probabilities of the proposed method and the MDLVQ in [24] for four description coding of a Gaussian source of unit variance.
Figure 3.10: The side PSNR vs central PSNR for three-description coding of Barbara at total bit rate of 1 bit/pixel.
Figure 3.11: The side PSNR vs central PSNR for three-description coding of Lena at total bit rate of 1 bit/pixel.
Figure 3.12: The side PSNR vs central PSNR for three-description coding of Boat at total bit rate of 1 bit/pixel.
Figure 3.13: The side PSNR vs central PSNR for four-description coding of Barb at total bit rate of 1 bit/pixel.
Figure 3.14: The side PSNR vs central PSNR for four-description coding of Lena at total bit rate of 1 bit/pixel.
Figure 3.15: The side PSNR vs central PSNR for four-description coding of Boat at total bit rate of 1 bit/pixel.
Figure 3.16: The side PSNR vs central PSNR for four-description coding of image Barbara with TRPCSQ and TDLT-based RD-MDC with total rate of 1 bit/pixel.
TABLE 3.3: The central and side PSNR for four-description coding of image Lena with TRPCSQ and TDLT-based RD-MDC with total rate of 1 bit/pixel.

<table>
<thead>
<tr>
<th>Description</th>
<th>Central PSNR $D_4$ (dB)</th>
<th>Side PSNR $D_i$ (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1 RD-MDC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D1 TRPCSQ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D2 RD-MDC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D2 TRPCSQ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D3 RD-MDC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D3 TRPCSQ</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 3.17: The side PSNR vs central PSNR for four-description coding of image Lena with TRPCSQ and TDLT-based RD-MDC with total rate of 1 bit/pixel.
Figure 3.18: The expected PSNR for images Barbara and Lena for four-description coding with TRPCSQ and TDLT-based RD-MDC with total rate of 1 bit/pixel.
Figure 3.19: Reconstructed image Barbara with two descriptions for $M = 4$ and central PSNR of 35.13 dB. (a) JPEG 2000-based RD-MDC (26.11 dB). (b) TDLT-based RD-MDC (28.07 dB). (c) TRPCSQ with single description optimized filters (28.68 dB). (d) TRPCSQ with optimized filters (28.93 dB).
Figure 3.20: Reconstructed image Barbara with three descriptions for $M = 4$ and central PSNR of 35.13 dB. (a) JPEG 2000-based RD-MDC (28.89 dB). (b) TDLT-based RD-MDC (30.23 dB). (c) TRPCSQ with single description optimized filters (31.54 dB). (d) TRPCSQ with optimized filters (31.74 dB).
Chapter 4

$M$-Channel MDC with Prediction-Compensation

4.1 Introduction

In [38], a prediction compensated MDC scheme (PC-MDC) is developed for the two-channel case, where the source is partitioned into two subsets, and each subset is encoded as the base layer of one description. Each description also encodes the prediction residual of the other subset, using the reconstruction of the base layer as the prediction reference. This is more efficient than the two-rate coding in [14, 46]. When applied to the time-domain lapped transform-based image coding [48], PC-MDC achieves better results than those in [43, 46], and represents the state of the art in two-description image coding.

In this chapter, motivated by the superior performance of the two-channel PC-MDC in image coding, we generalize the prediction-compensated approach to $M$-channel case. We derive expressions for various distortions and the optimal bit allocation. We also design the optimal lapped transform for the proposed algorithm using a three descriptions coding example. This MDC scheme is further improved in the next chapter using new design methodologies.
4.2 Prediction Compensated MDC with Many Descriptions

In this section we describe the proposed $M$-channel prediction compensated MDC (PC-MDC) algorithm and derive the formulas for various distortions and the optimal bit allocation for $M$ description coding of a 1-D wide sense stationary (WSS) Gaussian source.

We fictitiously partition the source samples into blocks of size $M$, with the $k$-th block containing samples $x(kM)$ to $x((k+1)M-1)$. By this partition, the $i$-th samples from all blocks, i.e., $\{x(kM + i)\}$ (for all $k$), form the $i$-th polyphase of the source. Let’s assume that the descriptions are indexed from 0 to $M-1$. In description $i$, the $i$-th polyphase $x_i$ is coded independent of other polyphases, which we call as intra-coding analogous to [38]. The reconstruction of this polyphase is used to predict all other polyphase components, and the prediction residuals are encoded in the same description. Similar to [38], we refer to this predictive coding as inter-coding. Note that each description carries information about all the source samples. Given each description, a reconstruction of the entire input can be obtained. In the receiver side, whenever an intra-coded version is received for a polyphase, it is used to obtain the reconstruction. The reconstruction of any other polyphase is obtained by averaging all available inter-coded reconstructions of that polyphase.

Let’s consider a wide sense stationary (WSS) Gaussian source with power spectral density (psd) $S_x(\omega)$. After downsampling by $M$, each polyphase has a psd

$$S_{x_i}(\omega) = \frac{1}{M} \sum_{k=0}^{M-1} S_x(\frac{\omega - 2\pi k}{M}). \quad (4.1)$$

Suppose the bit rate for intra-coding is $R_0$ (bits/sample/description), the MSE of the downsampled input is

$$d_I = \epsilon 2^{-2R_0} \exp \left( \frac{1}{2\pi} \int_{-\pi}^{\pi} \log S_{x_i}(\omega) d\omega \right) \triangleq \epsilon 2^{-2R_0} w_0^2, \quad (4.2)$$
where $\epsilon$ depends on the quantization and entropy coding scheme. For entropy coded scalar quantization, $\epsilon = \pi e/6$.

In each description, we use one polyphase to predict each of the remaining polyphases, and encode the prediction residual. Let’s assume that the polyphases are indexed from 0 to $M - 1$. For wide sense stationary signals, when polyphase $i$ is used to predict polyphases $k = i + j \mod M$ and $l = i - j \mod M$ the variances of the prediction residual are identical.

Therefore the number of distinct residual variances is

$$N = \left\lfloor \frac{M}{2} \right\rfloor. \quad (4.3)$$

The prediction residual between polyphase $i$ and polyphase $k = i \pm j \mod M$ is encoded with bit rate $R_{\Delta(j)}$, where $\Delta(j)$ is the index of the residual variance. It is also the distance of the two polyphases. More precisely,

$$\Delta(j) = \begin{cases} j & \text{if } j \in [1, N], \\ M - j & \text{if } j \in [N + 1, M - 1]. \end{cases} \quad (4.4)$$

Two examples with even $M$ and odd $M$ are shown in Fig. 4.1.

The linear minimum mean squared error (LMMSE) or Wiener filter $H_k(\omega)$ for predicting polyphase $k = i \pm j \mod M$ using polyphase $i$ is [15]

$$H_k(\omega) = \frac{S_{x_ki}(\omega)}{S_{x_i}(\omega)}, \quad (4.5)$$

and the psd of the prediction error is

$$S_{e_{\Delta(j)}}(\omega) = S_{x_k}(\omega) - \frac{|S_{x_ki}(\omega)|^2}{S_{x_i}(\omega)}. \quad (4.6)$$

In our method, we encode the prediction error at a bit rate lower than the intra bit rate $R_0$. 
Figure 4.1: Examples of prediction and bit allocation with five (top) and six (bottom) descriptions, respectively. Each circle represents one sample. The number in each circle is the polyphase it belongs to. Each sample in polyphase $j$ is predicted by two samples in polyphase 0, with rate $R_{\Delta(j)}$ for the residual encoding.

The MSE of the residual part is

$$d_{p,\Delta(j)} = \epsilon 2^{-2R_{\Delta(j)}} \exp \left( \frac{1}{2\pi} \int_{-\pi}^{\pi} \log S_{\epsilon\Delta(j)}(\omega) d\omega \right)$$

$$\triangleq \epsilon 2^{-2R_{\Delta(j)}} w_{\Delta(j)}^2. \quad (4.7)$$

4.2.1 Bit allocation

Assuming the probability of losing each description is $p$, the expected distortion is given by

$$D_M = \sum_{i=0}^{M} p_i D_{M,i}, \quad (4.8)$$

where $p_i = \binom{M}{i} p^{M-i} (1-p)^i$ is the probability of receiving $i$ descriptions, and $D_{M,i}$ is the corresponding MSE.

The MSE $D_{M,i}$ is related to the bit allocation among the intra part and the inter parts. When $i$ descriptions are received, each polyphase is reconstructed by intra coding and inter coding with a probability of $i/M$, and $(M-i)/M$, respectively. The average intra-coded distortion for each polyphase is simply $\epsilon w_0^2 \sigma_x^2 2^{-2R_0}$. To find the average distortion of an inter-coded polyphase, we assume description 0 is lost without losing generality. In this case, only the inter codings of polyphase 0 are
available since it is intra-coded only in description 0. Let’s assume that, out of the remaining \( M - 1 \) descriptions, \( i \) are received. The number of possible combinations for this \( i \) descriptions is \( \binom{M-1}{i} \). The reconstruction error of polyphase 0 in the \( k \)-th such combination is

\[
\sigma_{i,k}^2 = \epsilon \frac{1}{i^2} \sum_{j \in [1,M-1], I_{i,k,j} = 1} w_{\Delta(j)}^2 2^{-2R_{\Delta(j)}}. \tag{4.9}
\]

where \( I_{i,k,j} = 1 \) if the \( j \)-th description is received in the \( k \)-th combination and 0 otherwise. Note that \( \epsilon w_{\Delta(j)}^2 2^{-2R_{\Delta(j)}} \) is the distortion of the inter coding of the polyphase 0 in description \( j \).

The average distortion for polyphase 0 is the average of \( \sigma_{i,k}^2 \) over all \( \binom{M-1}{i} \) combinations:

\[
\sigma_i^2 = \frac{1}{\binom{M-1}{i}} \sum_{k=1}^{\binom{M-1}{i}} \sigma_{i,k}^2. \tag{4.10}
\]

Combining (4.9) and (4.10), it is easy to see that each polyphase \( j \) \( (j \in [1, M - 1]) \) contributes \( \frac{\binom{M-1}{i} i}{M-1} \) times to the summation. Therefore

\[
\sigma_i^2 = \epsilon \frac{1}{\binom{M-1}{i}} \frac{1}{i^2} \sum_{j=1}^{M-1} \frac{\binom{M-1}{i} i}{M-1} w_{\Delta(j)}^2 2^{-2R_{\Delta(j)}}
\]

\[
= \epsilon \frac{1}{i(M-1)} \sum_{j=1}^{M-1} w_{\Delta(j)}^2 2^{-2R_{\Delta(j)}} \tag{4.11}
\]

By symmetry of the system, this is the average distortion of every inter-coded polyphase when \( i \) descriptions are received.

As a summary, the average distortion of the entire input with \( i \) \( (i > 0) \) received descriptions is

\[
D_{M,i} = \frac{\epsilon \sigma_i^2}{M} \left( i w_{\Delta(j)}^0 2^{-2R_0} + \frac{M-i}{i(M-1)} \sum_{j=1}^{M-1} w_{\Delta(j)}^2 2^{-2R_{\Delta(j)}} \right). \tag{4.12}
\]

Plugging this into (4.8), we can then find the optimal bit allocation to minimize the
expected distortion $D$, subject to the bit rate constraint

$$\frac{1}{M} \left( R_0 + \sum_{j=1}^{M-1} R_{\Delta(j)} \right) = R. \quad (4.13)$$

Using Lagrangian multiplier, we can find the following condition for $k = 1, \ldots, N$:

$$\left( \frac{1}{M} \sum_{i=1}^{M} i p_i \right) w_0^{2^2-2R_0} = \left( \frac{1}{M} \sum_{i=1}^{M} \frac{p_i(M - i)}{i(M - 1)} \right) w_k^{2^2-2R_k}. \quad (4.14)$$

Define

$$u_0 = \frac{1}{M} \sum_{i=1}^{M} i p_i, \quad u_1 = \frac{1}{M} \sum_{i=1}^{M} \frac{p_i(M - i)}{i(M - 1)}, \quad (4.15)$$

and

$$c_{M,0} = u_0 w_0^2, \quad c_{M,k} = u_1 w_k^2. \quad (4.16)$$

The relationship between $R_k$ and $R_0$ can be found to be

$$R_k = R_0 - \frac{1}{2} \log_2 \frac{c_{M,0}}{c_{M,k}}. \quad (4.17)$$

Substituting into (4.13), $R_0$ is obtained as

$$R_0 = R + \frac{1}{2M} \log_2 \left( \prod_{j=1}^{M-1} \frac{c_{M,0}}{c_{M,\Delta(j)}} \right). \quad (4.18)$$
4.3 MD Coding of a First Order Gauss-Markov Source

If the input is a first order Gauss-Markov signal, its samples satisfy the following relationship
\[ x(n) = \rho x(n - 1) + e(n), \]
(4.19)
where \( e(n) \) is zero-mean Gaussian noise with variance \( (1 - \rho^2)\sigma_x^2 \), and \( \sigma_x^2 \) is the variance of \( x \). \( \rho \) is called the correlation coefficient.

When we apply our MDC algorithms for this source, each polyphase is still a first order Gauss-Markov signal with correlation coefficient \( \rho^M \). If DPCM is used to intra-code each polyphase, the MSE of each polyphase is \( \epsilon (1 - \rho^{2M})\sigma_x^2 - 2R_0 \), so
\[ w_0^2 = (1 - \rho^{2M})\sigma_x^2. \]
(4.20)

The linear MMSE (Wiener) filter between polyphases \( i \) and \( k = i \pm j \pmod{M} \) is a two-tap filter given by
\[
\mathbf{h}_j = \begin{bmatrix} h_j(0) & h_j(1) \end{bmatrix} = \begin{bmatrix} \rho^j & \rho^{M-j} \end{bmatrix} \begin{bmatrix} 1 & \rho^M \\ \rho^M & 1 \end{bmatrix}^{-1}
= \frac{1}{1 - \rho^{2M}} \begin{bmatrix} \rho^j - \rho^{2M-j} & \rho^{M-j} - \rho^{M+j} \end{bmatrix}.
\]
(4.21)
The variance of the prediction residual is
\[
w_j^2 = (1 + h_j^2(0) + h_j^2(1))
- 2h_j(0)\rho^j - 2h_j(1)\rho^{M-j} + 2h_j(0)h_j(1)\rho^M \sigma_x^2.
\]
(4.22)

Substituting Eq. (4.20) - (4.22) in the results obtained in Section 4.2.1 we can find the average distortion \( D_M \) for a given description loss probability \( p \). Fig 4.2 shows average SNR \( 10 \log_{10}(1/D_M) \) vs loss probability \( p \) when a first order Gauss-Markov source with \( \rho = 0.95 \) and unit variance is MD coded with 3, 4 and 6 descriptions respectively. The bit rate used is 3 bits per sample per description. The figure compares
CHAPTER 4. M-CHANNEL MDC WITH PREDICTION-COMPENSATION

the proposed prediction compensated MDC algorithm with two rate predictive coding and staggered quantization (TRPCSQ) in [35] which is also presented in Chapter 3. It can be seen that TRPCSQ outperforms PC-MDC in all the cases and the performance gap increases with $M$. In particular, the gap at $M = 6$ and $p = 0.01$ is 7.4 dB. The large performance gap between TRPCSQ and PC-MDC suggests a deficiency in PC-MDC design. It also suggests that there is room for further improvement. This will be studied in the next chapter. On the other, as will be shown in Section 4.5, practical image coding performance of PC-MDC is pretty good, and can be better than TRPCSQ in some cases.

4.4 Optimal Design for Lapped Transform based MD Image Coding

In this section we apply the proposed MDC scheme to image coding, for which transform coding needs to be used. The time-domain lapped transform (TDLT) framework developed in [48] is adopted for this purpose, which improves the performance of the DCT-based system by applying time-domain pre/postfilters. An overview of the TDLT is given in Section 3.3.1.

4.4.1 Overview of the System

In applying our algorithm to MD image coding, instead of partitioning the source samples directly, we partition them into $M$ polyphases at the block level as in [38, 35] to improve the coding efficiency. As a result, all predictions and refinements will be performed at the block level. In addition, practical FIR Wiener filters will be used.

To understand how our algorithm is applied in this block based framework, let’s first consider 3 description coding. Fig 3.6 shows the forward and inverse block transformation of the source with TDLT, where $P$ and $T$ are the prefilter and the postfilter respectively, and $L$ is the block size. To generate three descriptions, the prefiltered blocks $\{s(k)\}$ are split into three subsets $\{s_0(k) = s(3k)\}$, $\{s_1(k) = s(3k + 1)\}$, and $\{s_2(k) = s(3k + 2)\}$. After DCT, quantization, and entropy coding, each
subset of blocks form the base layers of one description. We call these blocks the *intra blocks* of each description. Analogous to the prediction described in Section 4.2, in each description we predict the remaining subsets from the intra blocks and the prediction residual is included as another layer, usually at much lower bit rate than the base layer. The prediction is obtained by using a Wiener filter. We refer to these prediction residual blocks as the *inter blocks*.

At the decoder side, if all descriptions are received, the decoded intra blocks from all descriptions are combined. The postfilter is then applied to obtain the reconstructed signal. The inter block bits in each description are simply discarded. These data are therefore the redundant data of the MDC system. If some descriptions are lost, the missing blocks are first estimated from the received intra blocks by Wiener filter. The decoded inter blocks (residuals) are then added to the estimation before postfiltering.

In three description coding, as described above, one third of the image blocks is intra-coded and the remaining two third is inter-coded in each description. As a result, if only two descriptions are received, two subsets have intra-coded blocks while the remaining subset has only inter-coded blocks. In this case, for the subset that there is no intra coding available, we take the average of the two reconstructions from the inter codings. This can reduce the MSE by one half.

The design of the system includes determining the optimal prefilter for the lapped transform, the optimal prediction filter for lost blocks, the optimal bit allocation between the intra blocks and inter blocks, as well as the bit allocation within a block. The objective is to minimize the expected distortion at the receiver.

Before we delve into the details, we first define the various bit rates used. We use $R$ (in bits/pixel/description)) to represent the bit rate of each description. $R_0$ and $R_1$ denote the average bits for each intra block pixel and each inter block pixel, respectively. For example, if there are three descriptions, the bit rate for each description is $R = \frac{1}{3}(R_0 + 2R_1)$ bpp/description, or $3R = R_0 + 2R_1$. Note that, for simplicity, we assume all inter-coded blocks to have the same average per pixel bit rate. In contrast, in Section 4.2, bit rate of an inter-coded polyphase in a description varied with the polyphase index.
In what follows, we use $x(i)$, $s(i)$, $y(i)$ and $q_y(i)$ to denote the $i$-th block of prefilter input, DCT input, DCT output and quantization noise of intra blocks, respectively. Note that $x(i)$ is aligned with the prefilter, whereas $s(n)$ is aligned with the DCT as shown in Fig. 3.6. We also use $d(i)$, $u(i)$ and $q_u(i)$ to denote the $i$-th inter block prediction residual, its DCT output and quantization noise, respectively. The reconstruction of a variable is denoted by the hat operator.

### 4.4.2 The Wiener Filter

We use Wiener filter to predict the inter blocks from intra blocks in each description. Here we only give the formulas for the first description of three description coding case, where $\{s(3k)\}$ are intra-coded. The formulas for other cases can be derived similarly. Let’s define the two neighboring intra blocks as

$$s_{I2} = \begin{bmatrix} s_T^{(3k)} & s_T^{(3k+3)} \end{bmatrix}^T,$$

(4.23)

The Wiener filters for estimating $s^{(3k+1)}$ and $s^{(3k+2)}$ from $s_{I2}$ are

$$H_1 = R_{s^{(3k+1)s_{I2}}}R_{s_{I2}s_{I2}}^{-1},$$

$$H_2 = R_{s^{(3k+2)s_{I2}}}R_{s_{I2}s_{I2}}^{-1},$$

(4.24)

As usual, the quantization noise is ignored in the Wiener filter. All matrices in the Wiener filters can be obtained by assuming an input signal model (AR(1) model is used in this chapter) and considering the effect of the prefilter in the TDLT. The details can be found in [38].

As in [18], we also normalize the Wiener filter so that all row sums are equal to 1. In addition, a special Wiener filter is used at the boundary to predict a missing block from only one neighboring block.

To apply the Wiener filter in (4.24) to 2-D images, we first estimate each row of a lost block using the horizontal neighbors, and then estimate each column of the block using vertical neighbors. The average of the horizontal and vertical estimations is used as the final prediction.
4.4.3 Objective Function and Optimal Rate Allocation

Given the target bit rate $R$ and the probability $p$ of losing one description, our objective is to find the optimal prefilter $P$ and the optimal bit allocation $R_0$ and $R_1$ for intra and inter blocks that minimize the expected distortion. We demonstrate how to determine the optimal prefilter and the bit allocation using 3 description coding as an example. We use $D_i$ to denote the MSE when $i$ descriptions are received. We also assume that all descriptions are balanced, i.e., they have the same rate and distortion. Since we are considering the 3 description coding case for demonstration purpose, the expected distortion $D$ is given by

$$D = (1 - p)^3 D_3 + 3p(1 - p)^2 D_2 + 3p^2(1 - p)D_1 + p^3 \sigma^2_x. \quad (4.25)$$

where $\sigma^2_x$ is the source variance. Note that the scenario of losing all descriptions is also considered in this expression.

Let $D_I$ and $D_P$ be the MSE contributed by intra-coded blocks and inter-coded blocks respectively.

When all three descriptions are received, only intra-coded data are used in reconstruction. Thus $D_3 = D_I$. When only one description is received, one third of the blocks is intra-coded, whereas two thirds of the blocks are inter-coded. Therefore the MSE is $D_1 = 1/3 (D_I + 2D_P)$. If two descriptions are received, each of them can generate a reconstruction of the missing intra-coded blocks from the lost description. By averaging the two reconstructions, the MSE can be halved. Therefore the overall MSE is $D_2 = 1/3 (2D_I + 1/2D_P)$.

To find $D_I$, let the quantization noise of $y(n)$ be $q_y(n)$. After the inverse TDLT, the reconstruction error becomes $Gq_y(n)$, where $G$ is the $2L \times L$ inverse transform that combines IDCT and the postfilter $T$ as described in Section 3.3.1. As usual, we assume the quantization noises of different subbands are uncorrelated. Therefore the MSE of the reconstruction is

$$D_I = \frac{1}{L} \sum_{i=0}^{L-1} ||g_i||^2 \sigma^2_{q_y(i)}, \quad (4.26)$$
where $\sigma^2_{q_y(i)}$ is the variance of the quantization noise of the $i$-th entry of $y(n)$, and $g_i$ is the $i$-th column of $G$. At high rates, $\sigma^2_{q_y(i)}$ can be written as

$$\sigma^2_{q_y(i)} = \epsilon \sigma^2_{y(i)} 2^{-2R_0},$$

where $\epsilon$ is a constant that depends on the input statistics and the quantization scheme. For entropy constrained scalar quantization of Gaussian sources, we have $\epsilon = \pi e/6$. $R_0$ is the bits allocated to the $i$-th entry of an intra block, and the average intra block bit rate is $\frac{1}{L} \sum_{i=1}^{L} R_{0i} = R_0$. $\sigma^2_{y(i)}$ is the variance of the $i$-th entry.

Upon optimal bit allocation [13], the minimal value for (4.26) is given by

$$D_I = \epsilon \left( \prod_{i=0}^{L-1} ||g_i||^2 \sigma^2_{y(i)} \right)^{\frac{1}{L}} 2^{-2R_0} \triangleq \epsilon \sigma^2_{L,I} 2^{-2R_0}.$$ (4.28)

This is in fact the objective function of the single description coding (SDC). For block transforms, the minimum value of (4.28) is achieved by the KLT, of which the DCT is a close approximation, if the input follows an AR(1) model with strong correlation. For lapped transforms and longer filter banks, there is no closed form expression for the optimal solution, but numerical optimization method can be used to find the solution that minimizes (4.28).

In single description coding, a performance measure called the coding gain is defined based on $\sigma^2_{L,I}$ in Eq. (4.28),

$$\text{coding gain} = 1/ \left( \prod_{i=0}^{L-1} ||g_i||^2 \sigma^2_{y(i)} \right)^{\frac{1}{L}} = 1/\sigma^2_{L,I}. $$ (4.29)

When $L = 8$, the coding gain of the optimized TDLT in [48] is 9.62 dB for AR(1) inputs with correlation $\rho = 0.95$. This is substantially higher than the 8.83 dB of the DCT.

If only one description is received, the error contributed by intra-coded blocks is the same as above. To find $D_P$, the MSE caused by inter-coded blocks, let's consider description 0, where the blocks $\{s(3k)\}$ are intra-coded and blocks $\{s(3k + 1)\}$ and
\{s(3k+2)\} are inter-coded. We first write the reconstruction of \(s(3k+j) \ (j = 1, 2)\) as

\[\hat{s}(3k+j) = \hat{d}(3k+j) + s_{H_1}(3k+j), \tag{4.30}\]

where \(\hat{d}(3k+j)\) and \(s_{H_1}(3k+j)\) are the reconstruction of the prediction residual \(d(3k+j)\) and the prediction of \(s(3k+j)\) from \(\hat{s}(3k)\) and \(\hat{s}(3k+3)\), respectively. Since the prediction residual \(d(3k+j)\) at the encoder is given by

\[d(3k+j) = s(3k+j) - s_{H_1}(3k+j), \tag{4.31}\]

the following relationship can be obtained from (4.30) and (4.31):

\[\hat{s}(3k+j) - s(3k+j) = \hat{d}(3k+j) - d(3k+j). \tag{4.32}\]

In other words, the reconstruction error of \(s(3k+j)\) is equal to that of the prediction residual \(d(3k+j)\). This is indeed a property of any differential coding system [40].

Since \(d(3k+j)\) and \(u(3k+j)\) are related by a DCT transform, we have

\[\hat{d}(3k+j) - d(3k+j) = C^T q_u(3k+j), \tag{4.33}\]

where \(C\) is the DCT matrix. Therefore, after postfiltering, the reconstruction error becomes \(Gq_u(3k+j)\), and its MSE is

\[D_P(3k+j) = \frac{1}{L} \sum_{i=0}^{L-1} \epsilon ||\hat{g}_i||^2 \sigma^2_{u_i}(3k+j)2^{-2R_1(3k+j)}. \tag{4.34}\]

where \(\sigma^2_{u_i}(3k+j)\) is the \(i\)-th diagonal element of the autocorrelation matrix of \(u(3k+j)\) as given by

\[R_{uu}(3k+j) = C\{R_{ss}(3k+j) - H_j R_{s_2s}(3k+j)\}C^T. \tag{4.35}\]

Since Eq. (4.34) has the same format as the reconstruction error caused by \(q_v(n)\) in (4.26), the derivation from (4.26) to (4.28) can be applied here, and the minimal
value of $D_P(3k + j)$ after optimal bit allocation is therefore

$$D_P(3k + j) = \epsilon \left( \prod_{i=0}^{L-1} ||g_i||^2 \sigma_{u_i}^2(3k + j) \right)^{\frac{1}{L}} 2^{-2R_1}. \tag{4.36}$$

Due to the symmetry in predicting $s(3k + 1)$ and $s(3k + 2)$ from $[s(3k), s(3k + 3)]$, the prediction error variances $\sigma_{u_i}^2(3k + 1)$ and $\sigma_{u_i}^2(3k + 2)$ should be equal. Therefore,

$$D_P = \epsilon \left( \prod_{i=0}^{L-1} ||g_i||^2 \sigma_{u_i}^2 \right)^{\frac{1}{L}} 2^{-2R_0} \triangleq \epsilon \sigma_{L,P}^2 2^{-2R_1}. \tag{4.37}$$

The expected distortion is thus

$$D = (1 - p)^2 D_I + 3p(1 - p)^2 \frac{1}{3} \left( 2D_I + \frac{1}{2} D_P \right) + 3p^2 (1 - p) \frac{1}{3} (D_I + 2D_P) + p^3 \sigma_x^2. \tag{4.38}$$

After simplification, we get

$$D_3 = (1 - p) \epsilon \sigma_{L,t}^2 2^{-2R_0} + \frac{1}{2} p(1 - p)(1 + 3p) \epsilon \sigma_{L,P}^2 2^{-2R_1} + p^3 \sigma_x^2. \tag{4.39}$$

The remaining bit allocation issue is to determine the optimal $R_0$ and $R_1$ that minimize the expected distortion $D$, subject to $R_0 + 2R_1 = 3R$.

$$\min (1 - p) \epsilon \sigma_{L,t}^2 2^{-2R_0} + \frac{1}{2} p(1 - p)(1 + 3p) \epsilon \sigma_{L,P}^2 2^{-2R_1}, \tag{4.40}$$

s.t. $R_0 + 2R_1 = 3R$.

The problem can be solved by Lagrangian method. After considering the constraints
of $R_0 \leq R$ and $R_1 \geq 0$, the optimal bit allocation can be found to be

$$R_0 = \min \left( R, R + \frac{1}{3} \log_2 \frac{4\sigma_{L,I}^2}{p(1 + 3p) \sigma_{L,P}^2} \right),$$

$$R_1 = \max \left( 0, R - \frac{1}{6} \log_2 \frac{4\sigma_{L,I}^2}{p(1 + 3p) \sigma_{L,P}^2} \right). \tag{4.41}$$

Eq. (4.41) shows that more bits should be allocated to encode the prediction residual when the description loss probability $p$ is higher or when $\sigma_{L,P}^2$ is larger (i.e., when the data are more difficult to predict). Notice that $R_1 = 0$ when $R < \frac{1}{6} \log_2 \left( \frac{4\sigma_{L,I}^2}{p(1 + 3p) \sigma_{L,P}^2} \right)$. This is the threshold below which no prediction residual is sent. In this case the method reduces to the approach in [18]. It should be emphasized that this threshold is derived based on the first-order Gauss-Markov model. For non-stationary signals like natural images, sending prediction residual is beneficial even at very low bit rates, because these bits are spent at regions with strong edges, and can thus significantly improve the reconstruction quality.

### 4.5 Application in Image Coding

The system designed in Section 4.4 is for 1-D signals. To apply it to 2-D images, we follow the conventional separable approach, i.e., for each block, the transforms and filters are applied row by row, then column and column. In particular, to apply the 1-D FIR Wiener filter in (4.4.2) to 2-D images, we first estimate each row of a target block using the co-located rows from the two horizontal neighbors, and then estimate each column of the block using the co-located columns from the two vertical neighbors. The average of the horizontal and vertical estimations is used as the final prediction.

Another design issue is how to partition the image blocks into different subsets. To improve the efficiency of the prediction, the blocks should be partitioned such that more neighboring blocks of a missing block can be available for prediction. The optimal partition pattern under this criterion for any value of $M$ is found in [2].
particular, when $M = K^2$, the optimal partition pattern is simply a $K \times K$ square matrix. As another example, when $M = 3$, the polyphase index of the $(i, j)$-th image block is $(i - j) \mod 3$.

Given the number of desired descriptions $M$ and the corresponding block partition, the next step is to design the corresponding optimal TDLT transform and Wiener filters. Since separable approach is used, and the system is designed based on the 1-D signal model, where we consider the horizontal and vertical partition patterns, and use them to design the corresponding filters as described in Section 4.4.

As a result, the filters in $K^2$-description image coding are actually the same as those in $K$-description coding of 1-D signals. A special case is $M = 4$, where the block partition pattern is a $2 \times 2$ matrix. Therefore the minimal distance of blocks in the same polyphase is 2, and the design of the transform and Wiener filters reduce to the two-description method in [38].

### 4.5.1 Image Coding Results

Fig 4.3 - 4.8 show MD coding results for $M = 3$ and $M = 4$ for images Barbara, Lena and Boat, respectively. The proposed algorithm is compared with two rate MD predictive coding with staggered quantization (TRPCSQ) in [35] and rate distortion MDC (RD-MDC) in [1]. RD-MDC results are obtained using the codec available at [29]. From these results it can be seen that the proposed algorithm consistently outperforms RD-MDC, in most cases with a significant margin. Except for image Barbara, the proposed algorithm has a comparable performance to TRPCSQ when central PSNR is smaller and when more than one description is received. On the other hand, the proposed algorithm tend to be better than TRPCSQ when the central PSNR is large. It can also be seen that, except in three description coding of image Barbara, the proposed algorithm performs better than TRPCSQ when only one description is received.
4.6 Summary

In this chapter, the prediction-compensated multiple description coding in [38] is extended to generate more than two descriptions. Expressions for various distortions are derived, and the optimal bit allocation is determined. Comparison with TRPCSQ in [35] for MD coding of a first order Gauss-Markov source shows that the proposed method has a significant deficiency. However, application in MD image coding shows that its performance is comparable to that of the TRPCSQ.

In the next chapter, we propose further modifications to PC-MDC to improve both its theoretical and practical performances.
Figure 4.2: Average SNR $10 \log_{10}(D_M)$ vs loss probability $p$ for PC-MDC and TR-PCSQ in [35] at $M = 3$, $M = 4$ and $M = 6$. The source is first order Gauss-Markov with unit variance and $\rho = 0.95$. 
Figure 4.3: Side distortion vs central distortion for 3 description coding of image Barbara at 1 bits per pixel total rate.
Figure 4.4: Side distortion vs central distortion for 3 description coding of image Lena at 1 bits per pixel total rate.
Figure 4.5: Side distortion vs central distortion for 3 description coding of image Boat at 1 bits per pixel total rate.
CHAPTER 4.  M-CHANNEL MDC WITH PREDICTION-COMPENSATION

Figure 4.6: Side distortion vs central distortion for 4 description coding of image Barbara at 1 bits per pixel total rate.
Figure 4.7: Side distortion vs central distortion for 4 description coding of image Lena at 1 bits per pixel total rate.
Figure 4.8: Side distortion vs central distortion for 4 description coding of image Boat at 1 bits per pixel total rate.
Figure 4.9: Three description coding results for the proposed method and TRPCSQ in [35] for image Lena at total bit rate of 1 bits per pixel. The PSNR values $D_3$ of the proposed method and TRPCSQ are tuned to be equal to TRPCSQ 39.06 dB. (a) Proposed, single description (31.99 dB) (b) TRPCSQ, single description (31.91 dB)
Figure 4.10: Three description coding results for the proposed method and TRPCSQ in [35] for image Lena at total bit rate of 1 bits per pixel. The PSNR values $D_3$ of the proposed method and TRPCSQ are tuned to be equal to TRPCSQ 39.06dB. (a) Proposed, two descriptions (34.99 dB) (b) TRPCSQ, two descriptions (35.07 dB)
Chapter 5

A Three-Layer MDC Scheme

5.1 Introduction

In Chapter 4, two-channel prediction compensated MDC (PC-MDC) scheme in [38] is generalized for the $M$-channel case. Analysis in Chapter 4 reveals that, despite reasonable performance in MD image coding, there is a significant deficiency in the performance of this $M$-channel PC-MDC algorithm when compared with the two rate coding algorithm in [35] using a first order Gauss-Markov source. In this chapter, we present an improved $M$-channel MD coding algorithm based on prediction compensation. The improvement of the new algorithm over PC-MDC in Chapter 4 comes from two new design techniques, namely, sequential prediction and three layer design.

The algorithm proposed in this chapter consists of three layers. In the first layer of each description, a subset of the source samples is encoded similar to intra-coding of PC-MDC in Chapter 4. In the second layer, the remaining subsets are encoded sequentially by predicting from all of the already encoded subsets. In contrast, the prediction in PC-MDC in Chapter 4 used only the base layer subset as the prediction reference. Having more samples in the prediction reference improves the prediction in the proposed algorithm. The third layer encoding in the proposed algorithm is designed to improve the performance when there is only one description lost, which also occurs the most frequently among all error scenarios when the loss probability of each description is very small.
CHAPTER 5. A THREE-LAYER MDC SCHEME

The organization of the chapter is as follows. In Sec. 5.2, we describe the framework of the proposed scheme, and analyze the performance of the proposed scheme for 1-D Gaussian sources. In Sec. 5.3, we modify our scheme for lapped transform-based MD coding, and formulate the optimization of the corresponding lapped transform. In Sec. 5.4, the performance of the proposed method in MD image coding is demonstrated and compared with other methods.

5.2 System Description and Performance Analysis

In this section, we describe the proposed three-layer multiple description coding (TLMDC), and analyze its R-D performance for a 1-D wide sense stationary (WSS) Gaussian source. In Sec. 5.3, the scheme will be modified for block transform-based image coding.

5.2.1 System Description

In our scheme, the source samples \( \{x(n)\} \) are partitioned into \( M \) polyphases \( x_i \), \( i = 0, \ldots, M - 1 \), with samples in the \( i \)-th polyphase given by \( x_i(n) = x(nM + i) \). As discussed in Sec. 5.1, each description in our method contains three layers. In the first layer of description \( i \), the \( i \)-th polyphase \( x_i \) is coded at bit rate \( R_0 \) bits/sample. In the second layer, the remaining polyphases are encoded sequentially by predicting from the already encoded polyphases. As a result, the first two layers of each description carry information about all the source samples. Given each description, a reconstruction of the entire input can be obtained. The second layer bit rate, \( R_1 \), is lower than \( R_0 \), and the predictively coded polyphases usually have lower quality than the directly coded polyphase in Layer 1.

When all descriptions are received, only the first layers of all descriptions are used to reconstructed the source. If some descriptions are lost, we first generate a coarse reconstruction of the source from the first two layers of each received description. All coarse reconstructions are then fused to obtain a refined reconstruction as follows. If a polyphase is coded in Layer 1 of any received description, its corresponding
reconstruction will be used directly. Otherwise its second layer reconstructions from all the received descriptions are averaged to get the refined reconstruction for this polyphase.

When the loss probability of each description is very low, as is the case in most network conditions, the dominating error scenario is where there is only one lost description. Therefore, a third layer is designed in our scheme to improve the performance in this case. As higher quality reconstructions of \( M - 1 \) polyphases are already available from the first layers of the received \( M - 1 \) descriptions, the goal of the third layer is to refine the quality of the remaining polyphase, which is usually lower than that of other polyphases, even after averaging all the coarse reconstructions.

Since there are only \( M \) possibilities of losing one description, we can afford to consider each case in the decoder. To get balanced descriptions, \( i.e., \) all descriptions are approximately of the same size, the refinement bits of the target polyphase in each case are evenly split into the third layers of the \( M - 1 \) received descriptions, which is a generalization of the second layer in [43] for two-description coding. As a result, the third layer of description \( i \) contains \( \frac{1}{M-1} \) of the refinement bits for all other polyphases \( j \neq i \). The splitting can be skipped if the requirement of balanced descriptions is not critical. In that case, all the third layer bits for the \( i \)-th polyphase can be included in, \( e.g., \) the \( ((i + 1) \mod M) \)-th descriptions.

The bitstream structures of all descriptions are illustrated in Fig. 5.1 for \( M = 4 \), where each row represents the bits for one description. Each row is divided into three layers, and each layer contains bits for different polyphases, as denoted by \( x_i \) in each box. The predictive coding in Layer 2 is represented by horizontally connected boxes, whereas the splitting in Layer 3 is represented by vertically connected boxes.

Note that when there are only two descriptions, \( i.e., \) \( M = 2 \), Layer 2 and Layer 3 in our scheme can be merged, and the scheme reduces to the two-layer prediction-compensated MDC in [38].
5.2.2 Sequential Prediction in Layer 2

As shown in [38], predictive coding-based MDC achieves better R-D performance than direct two-rate coding. However, two options exist when generalizing the two-description predictive coding scheme in [38] to more than two descriptions. If each description simply uses one polyphase or subset to predictively encode all other subsets as in Chapter 4, the coding efficiency will deteriorate as the increase of $M$, due to the diminished correlation among different subsets. In the proposed algorithm, this problem is resolved by using sequential prediction, where all previously encoded subsets are used to predict the next subset. The orthogonality principle is also used to reduce the complexity of the sequential prediction.

To illustrate the second layer coding, consider description 0, where the 0-th polyphase $x_0$ is encoded in the first layer. In the second layer, we first predict polyphase $x_1$ from the decoded 0-th polyphase $\hat{x}_0$.

Let $S_x(\omega)$ be the power spectral density (psd) of the source. Each polyphase $x_i$
thus has the same psd given by [50]

$$S_0(\omega) = \frac{1}{M} \sum_{k=0}^{M-1} S_x(\omega - \frac{2\pi k}{M}).$$  \hfill (5.1)

Assuming high rate coding, the linear minimum mean squared error (LMMSE) or Wiener filter $H_1(\omega)$ for predicting $x_1$ from $\hat{x}_0$ is given by [15]

$$H_1(\omega) = \frac{S_{x_{10}}(\omega)}{S_0(\omega)},$$  \hfill (5.2)

where $S_{x_{ij}}(\omega)$ is the cross spectral density between $x_i$ and $x_j$. The resulting prediction error $e_1$ is orthogonal to $x_0$ and has a psd of

$$S_1(\omega) = S_0(\omega) - \frac{|S_{x_{10}}(\omega)|^2}{S_0(\omega)}. \hfill (5.3)$$

The prediction errors $\{e_1(n)\}$ are coded at bit rate $R_1$ in the second layer.

The next step is to use both $\hat{x}_0$ and $\hat{x}_1$ to predictively encode polyphase $x_2$. This is more efficient than using $\hat{x}_0$ alone, as the correlation between $x_1$ and $x_2$ is generally stronger than that between $x_0$ and $x_2$.

Since $x_1$ can be decomposed into the projection on $x_0$ (via the Wiener filter) and the projection error $e_1$, which is orthogonal to $x_0$, the LMMSE prediction of $x_2$ from $x_0$ and $x_1$ is equivalent to the summation of separate LMMSE predictions from $x_0$ and $e_1$. It can be shown that the psd of the resulting prediction error $e_2$ is given by

$$S_2(\omega) = S_0(\omega) - \frac{|S_{x_{20}}(\omega)|^2}{S_0(\omega)} - \frac{|S_{x_{2e_1}}(\omega)|^2}{S_1(\omega)}, \hfill (5.4)$$

where $S_{x_{ie_j}}(\omega)$ is the cross spectral density between $x_i$ and $e_j$.

The prediction of other polyphases can be obtained in a similar manner, i.e., the psd of the prediction error of the $j$-th polyphase is given by

$$S_j(\omega) = S_0(\omega) - \frac{|S_{x_{j0}}(\omega)|^2}{S_0(\omega)} - \sum_{i=1}^{j-1} \frac{|S_{x_{je_i}}(\omega)|^2}{S_i(\omega)}. \hfill (5.5)$$
CHAPTER 5. A THREE-LAYER MDC SCHEME

This prediction error is coded at bit rate $R_j$ in the second layer. By the property of predictive coding, the MSE of the source equals to that of the prediction residual ([40], pp. 114). Since the residual of linear prediction is also Gaussian, the MSE of the residual (and the source) can be written as [40]

$$d_j = \delta \exp \left( \frac{1}{2\pi} \int_{-\pi}^{\pi} \log S_j(\omega) d\omega \right) 2^{-2R_j} \triangleq \delta \gamma_j^{2} 2^{-2R_j}, \quad (5.6)$$

where $\delta$ is a constant that depends on the input statistics and the quantization scheme. For Gaussian sources and entropy-constrained scalar quantizer, $\delta = \pi e/6$ [40]. Eq. (5.6) is also applicable to Layer 1 when $j = 0$.

5.2.3 The Third Layer Coding

When $M-1$ descriptions are available, high-quality reconstructions of $M-1$ polyphases can be obtained from the first layers of these descriptions. The goal of the third layer is to refine the other polyphase. Two approaches can be employed in this layer.

Method 1

In the first method, the target polyphase is predicted from all other $M-1$ polyphases that are reconstructed from the first layer coding. The prediction residual is encoded at rate $R_M$ and split among the $M-1$ descriptions.

This prediction is similar to the prediction in the second layer in Section 5.2.2 when $j = M-1$. The difference is that the prediction in Sec. 5.2.2 involves one first-layer-coded polyphase and $M-2$ second-layer-coded polyphases, whereas here the prediction uses $M-1$ first-layer-coded polyphases. However, at high rate, this difference can be neglected. This simplification is commonly used in the performance analysis of predictive coding ([40], pp. 114).

In this case, the psd of the prediction error is given by (5.5) when $j = M-1$. This error is coded at rate $R_M$. The corresponding MSE is given by

$$d_{M} = \delta \gamma_{M-1}^{2} 2^{-2R_M}. \quad (5.7)$$
Method 2

In the second method, the third layer is used to refine the average of the reconstructions of the target polyphase from the second layer decoding of all received descriptions. In these \( M - 1 \) descriptions, the target polyphase is predictively coded at rate \( R_i \) to \( R_{M-1} \), respectively. Assuming the reconstruction errors from the second layer codings of different descriptions are uncorrelated, the average of the \( M - 1 \) second layer reconstructions has the following prediction residual variance

\[
\sigma^2_e = \frac{1}{(M-1)^2} \sum_{j=1}^{M-1} \delta \gamma_j^2 2^{-2R_j}.
\]  

(5.8)

In the third layer, we encode this residual with a bit rate of \( R_M \). The MSE due to this coding is

\[
d_M = \delta \sigma^2_e 2^{-2R_M}.
\]  

(5.9)

5.2.4 R-D Performance Analysis

We now investigate the R-D performance of the proposed MDC scheme. Let \( D_k \) denote the expected distortion when there are \( k \) received descriptions. When \( k = M \), i.e., all descriptions are received, all source samples are reconstructed from the first layer. Hence, the distortion is given by (5.6) with \( j = 0 \):

\[
D_M = \delta \gamma_0^2 2^{-2R_0}.
\]  

(5.10)

When \( k < M \), we first reconstruct the \( k \) polyphases coded in the first layer. The reconstructions of the other polyphases are obtained from either the second layer or third layer, depending on the value of \( k \). Therefore,

\[
D_k = \frac{1}{M}(k D_M + (M - k) D'_k),
\]  

(5.11)

where \( D'_k \) is the expected distortion for the other \( M - k \) polyphases.

When \( k = M - 1 \), \( D'_k \) is obtained from the third layer coding and is given by \( d_M \).
in (5.7) or (5.9), depending on which method is used in the third layer.

When \( k < M - 1 \), \( D'_k \) is given by the second layer coding. In this case, assuming, without loss of generality, that Description 0 is lost and \( x_0 \) needs to be reconstructed from the second layers of the \( k \) received descriptions. Therefore there are \( \binom{M-1}{k} \) possible combinations of the indices of the received descriptions. Let \( \mathcal{I}_l \) be the \( l \)-th index combination. If Description \((M-j)\) is received, \( x_0 \) will be coded at rate \( R_j \) in its second layer. We use the average of all the reconstructions of \( x_0 \) from the \( k \) received descriptions as the final reconstruction. Assuming that the reconstruction errors are uncorrelated in different descriptions, the reconstruction error after the average can be written as

\[
D'_{kl} = \frac{1}{k^2} \sum_{M-j \in \mathcal{I}_l} \delta \gamma_j^2 2^{-2R_j}.
\] (5.12)

The distortion \( D'_k \) is the average of all \( D'_{kl} \). Hence,

\[
D'_k = \frac{1}{\binom{M-1}{k}} \sum_{l} D'_{kl}.
\] (5.13)

Eq. (5.13) is symmetric for each \( j \in \{1, 2, \ldots, M-1\} \). Therefore each \( \gamma_j^2 2^{-2R_j} \) appears exactly \( k \binom{M-1}{k} / (M-1) \) times in (5.13). Thus

\[
D'_k = \frac{1}{\binom{M-1}{k}} k^2 \sum_{j=1}^{M-1} k \binom{M-1}{k} / (M-1) \delta \gamma_j^2 2^{-2R_j}
\] (5.14)

\[
= \frac{1}{k(M-1)} \sum_{j=1}^{M-1} \delta \gamma_j^2 2^{-2R_j}.
\]

Given the expressions of all \( D_k \), and assuming that the probability of losing each description is \( p \), the overall expected distortion is

\[
D = \sum_{k=0}^{M} p_k D_k,
\] (5.15)

where \( p_k = \binom{M}{k} p^{M-k} (1-p)^k \) is the probability of receiving \( k \) descriptions.

Our objective is to minimize the overall expected distortion subject to the rate
constraint of \( R \) bits/sample/description,

\[
\frac{1}{M} \sum_{j=0}^{M} R_j = R. \tag{5.16}
\]

Using Lagrangian multiplier method, we can find the following solutions for the two methods of the third layer coding.

**Method 1**

When Method 1 is used in the third layer, the optimal bit allocation can be found to be

\[
R_j = R_0 - \frac{1}{2} \log_2 \frac{c_0}{c_j}, \quad j \in \{1, 2, \ldots, M\}, \tag{5.17}
\]

where

\[
c_0 = \frac{1}{M} \delta \gamma_0^2 \sum_{i=1}^{M} i p_i,
\]

\[
c_j = \frac{1}{M} \delta \gamma_j^2 \sum_{i=1}^{M-2} p_i (M-i) \frac{1}{i(M-1)}, \quad j \in \{1, 2, \ldots, M-1\}
\]

\[
c_M = \frac{\rho M^{-1}}{M} \delta \gamma_{M-1}.
\]

From (5.16) and (5.17), \( R_0 \) can be obtained as

\[
R_0 = \frac{M}{M+1} R + \frac{1}{2(M+1)} \log_2 \left( \prod_{j=1}^{M} \frac{c_0}{c_j} \right). \tag{5.19}
\]

Under this bit allocation, the minimal overall expected distortion is

\[
D = p_0 \sigma_x^2 + (M + 1) \left( \prod_{j=0}^{M} c_j \right)^{\frac{1}{M+1}} \left( \prod_{j=0}^{M} c_j \right)^{-2} R, \tag{5.20}
\]

where \( \sigma_x^2 \) is the source variance.
CHAPTER 5. A THREE-LAYER MDC SCHEME

Method 2

When Method 2 is used in the third layer, the optimal bit allocation becomes

\[ R_j = R_0 - \frac{1}{2} \log_2 \left( \frac{(M-2)c_0}{(M-1)c_j} \right), \quad j \in \{1, \ldots, M-1\} \]  \hspace{1cm} (5.21)

\[ R_M = \frac{1}{2} \log_2 \left( \frac{c_M \delta(M-2)}{c_{M-1}(M-1)^2} \right). \]  \hspace{1cm} (5.22)

From (5.21), (5.22) and (5.16), \( R_0 \) can be found as

\[ R_0 = R + \frac{1}{2M} \log_2 \left( \frac{c_{M-1}(M-2)^M-2}{c_M \delta(M-1)^M-3} \prod_{j=1}^{M-1} \frac{c_0}{c_j} \right). \]  \hspace{1cm} (5.23)

The corresponding minimal expected distortion is

\[ D = p_0 \sigma_x^2 + M \left( \frac{c_M \delta(M-1)^{M-3}}{c_{M-1}(M-2)^{M-2}} \prod_{j=0}^{M-1} c_j \right)^{\frac{1}{M}} 2^{-2R}. \]  \hspace{1cm} (5.24)

Comparing (5.20) and (5.24), it can be seen that the distortion in Method 2 decays faster with the bit rate \( R \) than in Method 1. In other words, there is a relative rate loss in Method 1 compared to Method 2.

5.2.5 Example with GM(1) Sources

In this section, we compare the performance of the proposed three-layer MDC (TLMDC) scheme with other methods for a first-order Gaussian-Markov (GM(1)) source, which can be modeled as

\[ x(n) = \rho x(n-1) + v(n), \]  \hspace{1cm} (5.25)

where \( x(n) \) is the \( n \)-th sample, \( v(n) \) is a Gaussian white noise sample independent of \( x(n) \), and \( \rho \) is the correlation coefficient. After downsampling, each polyphase itself
CHAPTER 5. A THREE-LAYER MDC SCHEME

is a GM(1) source with a correlation coefficient of $\rho^M$ and power spectral density

$$S_0(\omega) = \frac{\sigma_x^2 (1 - \rho^{2M})}{1 - 2\rho^M \cos \omega + \rho^{2M}}. \quad (5.26)$$

From (5.5) it can be seen that evaluation of $S_j(\omega)$ requires the knowledge of $S_{x_je_i}(\omega)$ and $S_i(\omega)$ for $i < j$. It can be shown that $S_{x_je_i}(\omega)$ is given by

$$S_{x_je_i}(\omega) = S_{x_ji}(\omega) - \frac{S_{x_i0}(\omega) S_{x_j0}(\omega)}{S_0(\omega)} - \sum_{t=1}^{i-1} \frac{S_{x_ie_i}(\omega) S_{x_je_i}(\omega)}{S_t(\omega)}. \quad (5.27)$$

When $l > k$, we can express $S_{x_{lk}}$ as

$$S_{x_{lk}}(\omega) = \frac{\sigma_x^2 (\rho^{l-k} - \rho^{2M-l+k} + \rho^M e^{j\omega} (\rho^{k-l} - \rho^{l-k}))}{1 - 2\rho^M \cos \omega + \rho^{2M}}. \quad (5.28)$$

These results can be used to recursively evaluate (5.5) and (5.27).

Fig. 5.2 shows $10 \log_{10}(1/D)$ at different packet loss probabilities for six-description coding of a unit-variance GM(1) source with $\rho = 0.95$ and $R = 3$ bits/sample/description. Three versions of our scheme are reported, i.e., Method 1, Method 2, and no third layer at all. For comparison purpose, the results of the TRPCSQ in [35] and PC-MDC in Chapter 4 are also included. This figure reveals several important facts about the proposed MD coding method. The only difference between the proposed method without third layer and PC-MDC is that the former uses sequential prediction. In Fig. 5.2, the proposed method without layer 3 is always better than PC-MDC, which suggests that the sequential prediction is more efficient than the prediction method in PC-MDC. Comparing the curves for different versions of the proposed scheme we can also see that the third layer coding improves the performance significantly at lower loss probabilities, as expected. It can be seen from the figure that TLMDC has a better performance than TRPCSQ when loss probability is low and vice-versa. For the configuration in Fig. 5.2, Method 2 is better than Method 1. However, the performance gap decreases as $p$ decreases.
Fig. 5.3 shows how the expected SNR \(10 \log_{10}(1/D)\) varies with the correlation coefficient \(\rho\) for six-description coding of a unit-variance GM(1) source with description loss probability \(p = 0.01\) and \(R = 3\) bits/sample/description. Higher source correlation results in improved SNR as expected. Note that the performance difference between the proposed method without layer 3 and PC-MDC decreases at lower correlation. This is because, the improved sequential prediction in the proposed method without layer 3 becomes less effective at lower source correlation.

As pointed out earlier, Method 1 has a rate deficiency compared to Method 2. Fig. 5.4 shows a plot of \(10 \log(1/D)\) vs. bit rate at different loss probabilities for four-description coding of the unit-variance GM(1) source with \(r = 0.95\), which shows that the rate deficiency of Method 1 increases as \(R\). Note that when \(p = 0.1\), the distortion has an asymptote around \(-40\) dB, because the distortion at high loss probabilities and high bit rates is dominated by \(p_0 \sigma^2_x\).

### 5.3 Optimal Design for Block Transform and Application in Image Coding

In this section we apply the proposed MDC scheme to image coding. As in Chapter 4 the time-domain lapped transform (TDLT) framework developed in [48] is adopted, which improves the performance of a DCT-based system by applying time-domain pre/postfilters.

Similar to Chapter 4, instead of partitioning the source samples directly, we partition them into \(M\) polyphases at the block level to improve the coding efficiency. As a result, all predictions and refinements will be performed at the block level. In addition, practical FIR Wiener filters will be used. Due to the three-layer structure of the proposed method, the predictions and refinements are different from Chapter 4.

A nice property of the TDLT is that the pre/postfilters can be optimized for different applications. Since our MDC framework described above differs from the traditional single description coding, the optimal filters found in [48] may not be
optimal in our scheme. An overview of the TDLT is given in Section 3.3.1. In this section, we formulate the optimization of the pre/postfilters and the corresponding Wiener filters for our framework.

5.3.1 The FIR Wiener Filter

In the following discussion, we use $x(i)$, $s(i)$ and $y(i)$ to denote the $i$-th block of prefilter input, DCT input and DCT output, respectively. Note that $x(i)$ is aligned with the prefilter, whereas $s(n)$ is aligned with the DCT, as shown in Fig. 3.6. The
reconstruction of a variable is denoted by the hat operator.

In the second layer of the proposed MDC scheme, the blocks that are not coded in the first layer are coded with sequential prediction using Wiener filters. To reduce the complexity, instead of using all previously coded data in the prediction (as in Sec. 5.2), in the image coding, we only use the two nearest neighboring blocks to interpolate or predict the target block.

Consider the $i$-th description in $M$ description coding, where the DCT blocks $\{y(Ml + i)\}, l = 0, 1, 2, \ldots$, are coded in the first layer with bit rate $R_0$. In the second layer, we first predict the prefilter output blocks $\{s(Ml + i + 1)\}$ using the
Figure 5.4: Expected SNR $10 \log_{10}(1/D)$ vs total bit rate/sample for MD coding of a GM(1) source with $M = 4$, $\rho = 0.95$.

reconstructed \{\hat{s}(Ml+i)\} from the first layer. Let $s_{il,2}$ be the two nearest neighboring blocks

$$s_{il,2} = \left[ s^T(Ml+i) \ s^T(M(l+1)+i) \right]^T, \quad (5.29)$$

The Wiener filters for estimating $s(Ml+i+1)$ from $s_{il,2}$ is

$$H_1 = R_{s(Ml+i+1)s_{il,2}} R_{s_{il,2}s_{il,2}}^{-1}. \quad (5.30)$$
As in [38], the quantization noise is ignored in the Wiener filter. The resulting prediction error is DCT coded at bit rate $R_1$.

Next, we interpolate $\{s(Ml+i+2)\}$ from $\{\hat{s}(Ml+i+1)\}$ and $\{\hat{s}(M(l+1)+i)\}$, which are the nearest available blocks to the left and right sides of $\{s(Ml+i+2)\}$. In general, the block $s(Ml+i+j), j \in \{1, 2, \ldots, M-1\}$ in description $i$ is interpolated from $\hat{s}(Ml+i+j-1)$ and $\hat{s}(M(l+1)+i)$, and the prediction residual is coded at rate $R_j$ after the DCT. The resulting interpolation error is coded at rate $R_j$ after DCT.

All matrices in the FIR Wiener filters can be obtained if the input statistics and the prefilter in the TDLT are known. Here we assume the input has the GM(1) model with correlation coefficient 0.95. The details can be found in [38]. As in [18, 38], we also normalize the Wiener filter to have unit row sum. In addition, a special Wiener filter is used at the boundary to predict a missing block from only one neighboring block.

### 5.3.2 Optimal Design of the Lapped Transform

Given the target bit rate $R$ and the probability $p$ of losing each description, our objective is to find the optimal prefilter and postfilter in the TDLT that minimize the expected distortion in (5.15), i.e., $D = \sum_{k=0}^{M} p_k D_k$.

In the proposed MDC scheme, when $k$ descriptions are available, $k$ out of $M$ DCT blocks are reconstructed from the first layer coding, and the rest are reconstructed from second layer (if $k < M - 1$) or third layer coding (if $k = M - 1$). After postfiltering at block boundaries, each block contains the contributions from two DCT blocks. When the bit rates are sufficiently high, the quantization noises in different DCT blocks are approximately uncorrelated, and their contributions to the reconstruction error are additive. Therefore $D_k$ can still be written as (5.11). The only difference from Sec. 5.2 is that $D_M$ and $D'_k$ here are obtained from block transform and quantization rather than direct quantization.
Plugging (5.11) into (5.15), we get

\[ D = \frac{1}{M} \sum_{k=1}^{M} p_k k D_M + \frac{1}{M} \sum_{k=1}^{M} (M - k) p_k D'_k + p_0 \sigma_x^2 \]

\[ = \frac{\mu_k}{M} D_M + \sum_{k=1}^{M-2} \frac{M - k}{M} p_k D'_k + \frac{1}{M} p_{M-1} D'_{M-1} + p_0 \sigma_x^2, \]

(5.31)

where \( \mu_k \) is the expected value of \( k \) (which can be obtained when the loss probability of each description is independent and identical), and \( \sigma_x^2 \) is the variance of the source.

We next find the expressions of different terms in (5.31).

To find \( D_M \), assume that \( y_0(n) \) is a first-layer-coded DCT block with quantization noise \( q_{y_0}(n) \). After the inverse TDLT, the reconstruction error becomes \( G q_{y_0}(n) \), where \( G \) is the \( 2L \times L \) inverse transform that combines IDCT and the postfilter \( T \) as described in Section 3.3.1. Assuming the quantization noises of different subbands are uncorrelated, the average reconstruction error per sample is

\[ D_M = \frac{1}{L} \sum_{i=0}^{L-1} ||g_i||^2 \sigma_{q_{y_0}}^2 (i), \]

(5.32)

where \( \sigma_{q_{y_0}}^2 (i) \) is the variance of the \( i \)-th entry of \( q_{y_0}(n) \), and \( g_i \) is the \( i \)-th column of \( G \). As in Sec. 5.2, at high rates, \( \sigma_{q_{y_0}}^2 (i) \) can be written as

\[ \sigma_{q_{y_0}}^2 (i) = \delta \sigma_{y_0}^2 (i) 2^{-2R_{0i}}, \]

(5.33)

where \( \sigma_{y_0}^2 (i) \) and \( R_{0i} \) are the variance and the allocated bits of the \( i \)-th entry of \( y_0(n) \), respectively, and the bit allocation satisfies \( \frac{1}{L} \sum_{i=0}^{L-1} R_{0i} = R_0 \).

The distortion \( D'_k \) in (5.31) with \( k < M - 1 \) is the reconstruction error of a second-layer-coded block, which is obtained after averaging the \( k \) second-layer-decoded blocks from \( k \) descriptions, each having a different rate.

As in (5.14), when averaged over all the possible combinations, the variance of the
quantization noise of the $i$-th DCT coefficient is given by

$$D'_{k,i} = \frac{1}{k(M-1)} \sum_{j=1}^{M-1} \delta \sigma^2_{y_j}(i) 2^{-2R_{ji}},$$  \hspace{1cm} (5.34)$$

where $\sigma^2_{y_j}(i)$ is the variance of the $i$-th entry of a prediction residual block $y_j(n)$, and the bit allocation for this block satisfies $\frac{1}{L} \sum_{i=0}^{L-1} R_{ji} = R_j$.

Similar to (5.32), after the inverse transform, the average reconstruction error per sample is

$$D'_k = \frac{1}{L} \sum_{i=0}^{L-1} \|g_i\|^2 \frac{1}{k(M-1)} \sum_{j=1}^{M-1} \delta \sigma^2_{y_j}(i) 2^{-2R_{ji}},$$  \hspace{1cm} (5.35)$$

Thus, the second term in (5.31) can be expressed as

$$\sum_{k=1}^{M-2} \frac{M-k}{M} p_k D'_k = \alpha_{M,p} \sum_{j=1}^{M-1} \frac{1}{L} \sum_{i=0}^{L-1} \|g_i\|^2 \delta \sigma^2_{y_j}(i) 2^{-2R_{ji}},$$  \hspace{1cm} (5.36)$$

where,

$$\alpha_{M,p} = \frac{1}{M} \sum_{k=1}^{M-2} \frac{M-k}{k(M-1)}.$$  \hspace{1cm} (5.37)$$

Finally, the distortion $D'_{M-1}$ in (5.31) is the reconstruction error of a third-layer-coded block. In Sec. 5.2.3, two methods are proposed for the third layer coding. However, it is observed from image coding experiments that the second method performs consistently better than the first method. Hence only the second method is considered from now on. In this method, a third-layer-coded block is first reconstructed by averaging all $M-1$ second-layer decodings of the block. The residual is further coded in the third layer with rate $R_M$.

When averaged over all $M-1$ second-layer codings, the MSE of the quantization noise is obtained by letting $k = M-1$ in (5.34). This error is further reduced by the third-layer coding with rate $R_{Mi}$ such that $\frac{1}{L} \sum_{i=0}^{L-1} R_{Mi} = R_M$. The final MSE of the
i-th DCT coefficient is thus

\[
\sigma^2_{\delta y_M}(i) = \left( \frac{\delta}{(M-1)^2} \sum_{j=1}^{M-1} \delta \sigma^2_{y_j}(i) 2^{-2R_{ji}} \right) 2^{-2R_{Mi}}. \tag{5.38}
\]

After inverse transform, the MSE \(D'_{M-1}\) for a third-layer coded sample is

\[
D'_{M-1} = \frac{\delta}{L(M-1)^2} \sum_{i=0}^{L-1} \|g_i\|^2 \sum_{j=1}^{M-1} \delta \sigma^2_{y_j}(i) 2^{-2(R_{ji}+R_{Mi})}. \tag{5.39}
\]

Substituting (5.32), (5.36) and (5.39) in (5.31) we obtain

\[
D = p_0 \sigma^2_x + u_0 \sum_{i=0}^{L-1} \|g_i\|^2 \delta \sigma^2_{y_0}(i) 2^{-2R_{0i}} + u_1 \sum_{j=1}^{M-1} \sum_{i=0}^{L-1} \|g_i\|^2 \delta \sigma^2_{y_j}(i) 2^{-2R_{ji}} + u_2 \sum_{j=1}^{M-1} \sum_{i=0}^{L-1} \|g_i\|^2 \delta \sigma^2_{y_j}(i) 2^{-2(R_{ji}+R_{Mi})}, \tag{5.40}
\]

where

\[
u_0 = \frac{1}{LM} \sum_{k=1}^{M} p_k k, \\
u_1 = \frac{\alpha_{M,p}}{L}, \\
u_2 = \frac{\delta}{LM(M-1)^2} p_{M-1}. \tag{5.41}
\]

The Lagrangian method can be used to minimize the expected distortion, subject to the constraint

\[
R = \frac{1}{LM} \sum_{j=0}^{M} \sum_{i=0}^{L-1} R_{ji}. \tag{5.42}
\]
It can be shown that at the optimal bit allocation, the expected distortion is

\[ D = p_0 \sigma_x^2 + \]
\[ \delta LM(M - 1) \frac{M-1}{M} (u_0 u_2) \frac{1}{M} \left( \frac{u_1}{M-2} \right)^{\frac{M-2}{M}} \beta(P) 2^{-2R}, \]  

(5.43)

where

\[ \beta(P) = \left( \prod_{j=0}^{M-1} \prod_{i=0}^{L-1} \sigma_y^2(i) \right)^{\frac{1}{LM}} \left( \prod_{i=0}^{L-1} ||g_i||^2 \right)^{\frac{1}{L}}. \]  

(5.44)

We can then design the lapped transform prefilter \( P \) (more precisely, \( V \) in (3.22)) to minimize \( \beta(P) \), which is the only part in (5.43) that depends on the TDLT filters. Since there is no closed-form solution for this step, we use Matlab to find the optimized solution. In order to easily control the tradeoff between the optimized transforms for single description coding and multiple description coding, we use a weighted objective function

\[ J = wD + (1 - w)J_0, \]  

(5.45)

where \( 0 \leq w \leq 1 \), and \( J_0 \) is the single description coding gain of the transform, which is defined as

\[ J_0 = 1/ \left( \prod_{i=0}^{L-1} ||g_i||^2 \sigma_y^2(i) \right)^{\frac{1}{L}}, \]  

(5.46)

where \( \sigma_y^2(i) \) is the variance of the \( i \)-th entry of the DCT coefficient \( y(n) \).

5.3.3 Application in image coding

The system designed above is for 1-D signals. To apply it to 2-D images, we follow the conventional separable approach as in Chapter 4, i.e., for each block, the transforms and filters are applied row by row, then column by column. To apply the 1-D FIR Wiener filter in (5.30) to 2-D images, we first estimate each row of a target block using the co-located rows from the two horizontal neighbors, and then estimate each column of the block using the co-located columns from the two vertical neighbors. The average of the horizontal and vertical estimations is used as the final prediction.
CHAPTER 5. A THREE-LAYER MDC SCHEME

The same partitioning pattern as in Chapter 4 is used to partition the image blocks into different polyphases or groups. In this partitioning, to improve the efficiency of the prediction, the blocks are partitioned such that more neighboring blocks of a missing block can be available for prediction. The optimal partition pattern for any value of $M$ is found in [2]. In particular, when $M = K^2$, the optimal partition pattern is simply a $K \times K$ square matrix. As another example, when $M = 3$, the polyphase index of the $(i, j)$-th image block is $(i - j) \mod 3$.

Since separable approach is used, and the system is designed based on the 1-D signal model, we find the horizontal and vertical distance between blocks in the same polyphase, and use them to design the optimal TDLT transform and Wiener filters. As a result, the filters in $K^2$-description image coding are actually the same as those in $K$-description coding of 1-D signals. It can also be seen that the filters for 9-description image coding are identical to 3-description image coding. Another special case is $M = 4$, where the block partition pattern is a $2 \times 2$ matrix. Therefore the minimal distance of blocks in the same polyphase is 2, and the design of the transform and Wiener filters reduce to the two-description method in [38].

5.4 Experimental Results

In this section, the proposed three-layer MDC (TLMDC) method is applied to the TDLT-based image coding [48]. The entropy coding method in [49] is used to encode the quantized DCT transform coefficients.

We compare the proposed method with the rate-distortion-based multiple description coding (RD-MDC) in [1], the MDC scheme with two-rate predictive coding and staggered quantization (TRPCSQ) in [35] and PC-MDC in Chapter 4. We use the RD-MDC codec available at [29], which does not provide a way to generate descriptions optimized for a particular probability. For a given bit rate, the codec has a parameter which decides the central and side distortions at the same time. RD-MDC has only one degree of freedom. Therefore, when the central distortion is decided, all the other distortions are determined automatically. In contrast, our method has two degrees of freedom which enables us to decide up to two distortion values at the same time.
### Figure 5.5: Side distortion vs central distortion for 3 description coding of image Lena at 1 bits per pixel total rate.

<table>
<thead>
<tr>
<th>Central PSNR $D_3$ (dB)</th>
<th>Side PSNR $D_i$ (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>D2 TLMDC</td>
</tr>
<tr>
<td></td>
<td>D2 RD–MDC</td>
</tr>
<tr>
<td></td>
<td>D1 both methods</td>
</tr>
</tbody>
</table>

- Central PSNR $D_3$: 37.8, 38, 38.2, 38.4, 38.6, 38.8, 39, 39.2, 39.4
- Side PSNR $D_i$: 29, 30, 31, 32, 33, 34, 35, 36, 37
CHAPTER 5. A THREE-LAYER MDC SCHEME

Figure 5.6: Side distortion vs central distortion for 4 description coding of image Lena at 1 bits per pixel total rate.
Fig. 5.5 and Fig. 5.6 compare the proposed method with RD-MDC for three and four description coding of the Lena image, respectively. The side PSNRs are plotted against the central PSNR at total rate of 1 bits per pixel. The RD-MDC results are generated using the codec at [29]. To get a fair comparison, for each PSNR combination generated by the RD-MDC codec, we tune our codec to obtain the same central PSNR $D_M$ and the side PSNR $D_1$ as the RD-MDC. It can be seen from both figures that our method outperforms RD-MDC.

Note that the RD-MDC in [1] is based on wavelet, and one of its goals is to maintain compatibility with JPEG 2000. To get fair comparison, we also implement the RD-MDC method in the TDLT framework. That is, the transformed blocks are partitioned into $M$ subsets in the same way as in our method. Each description carries one subset with a higher rate and others with a lower rate, as in [1]. It is shown in [35] that TDLT-based RD-MDC is slightly better than wavelet-based RD-MDC.

In Fig. 5.7 - 5.15, the optimal expected PSNR is plotted at different loss probabilities for three description, four description and nine description coding of images Lena, Barbara and Boat, respectively. Note that the RD-MDC used in this experiment is based on TDLT framework. The redundancy is tuned in each case to achieve the maximum expected PSNR for each loss probability. In three description coding, the proposed method performs slightly better than TRPCSQ, RD-MDC and PC-MDC when the loss probability is low. In four description coding, it can be clearly seen that the proposed method has superior performance compared to the rest of the methods compared. The improvement of the proposed method over TRPCSQ and RD-MDC is even greater in 9 description coding case.

Fig. 5.17 shows four description coding results for the proposed method and TRPCSQ in [35] for image Boat at total bit rate of 1 bits per pixel. The PSNR values for reconstruction from all four descriptions ($D_4$) and from only one description ($D_1$) for the proposed method are tuned to be equal to TRPCSQ. This is possible since the proposed method has two degrees-of-freedom. The corresponding PSNR values are $D_4 = 38.79$ dB and $D_1 = 24.65$ dB for both methods. Under this setting the PSNR value $D_2$ when there are two descriptions is 27.34 dB for the proposed method and 27.57 dB for TRPCSQ. From the images we can see that the proposed method has
achieved better visual quality than TRPCSQ even in this case where TRPCSQ has a slightly better PSNR value. When reconstructed from three descriptions, both visual quality and the PSNR value are better in the proposed method than in TRPCSQ. The visual quality improvement in the proposed method in three descriptions case can be clearly seen in the sky in the image.
Figure 5.8: Three description coding results for image Barbara at total bit rate of 1 bits per pixel.

5.5 Summary

This chapter presents an $M$-channel MDC method using three-layer coding. The closed-form expressions of the expected distortions of the system are derived for different number of received descriptions. The method is also applied to lapped transform-based multiple description image coding. Experimental results show that this method achieves better performance than other state-of-the-art schemes. The image coding results can be further improved. For example, more advanced 2-D filters can be used instead of 1-D filters, and the entropy coding can be fine-tuned based on the
Figure 5.9: Three description coding results for image Boat at total bit rate of 1 bits per pixel.

characteristics of the prediction residuals.
Figure 5.10: Four description coding results for image Lena at total bit rate of 1 bits per pixel.
Figure 5.11: Four description coding results for image Barbara at total bit rate of 1 bits per pixel.
Figure 5.12: Four description coding results for image Boat at total bit rate of 1 bits per pixel.
Figure 5.13: Nine description coding results for image Lena at total bit rate of 1.25 bits per pixel.
Figure 5.14: Nine description coding results for image Barbara at total bit rate of 1.25 bits per pixel.
Figure 5.15: Nine description coding results for image Boat at total bit rate of 1.25 bits per pixel.
Figure 5.16: Four description coding results for the proposed method and TRPCSQ in [35] for image Boat at total bit rate of 1 bits per pixel. The PSNR values $D_4$ and $D_1$ of the proposed method are tuned to be equal to TRPCSQ ($D_4 = 38.79$ dB and $D_1 = 24.65$ dB). (a) Proposed, two descriptions (27.34 dB) (b) TRPCSQ, two description (27.57 dB)
Figure 5.17: Four description coding results for the proposed method and TRPCSQ in [35] for image Boat at total bit rate of 1 bits per pixel. The PSNR values $D_4$ and $D_1$ of the proposed method are tuned to be equal to TRPCSQ ($D_4 = 38.79$ dB and $D_1 = 24.65$ dB). (a) Proposed, three descriptions (31.57 dB) (b) TRPCSQ, three descriptions (30.80 dB)
Chapter 6

Conclusions and Future Directions

6.1 Conclusions

In this thesis our main objective is developing new multiple description coding methods and improving existing schemes, focusing particularly on MD predictive coding, $M$-channel MD coding and their application in image coding. Several new algorithms were proposed, analyzed and experimentally evaluated. All the proposed algorithms consist of predictive coding as a main component while two-rate coding in Chapter 3 can be simplified for memoryless sources as well.

A two-stage predictive coding algorithm is presented for two-channel MD coding in Chapter 2. This algorithm is based on modified multiple description scalar quantization in [43], and, as a result, inherits its simplified structure. Theoretical analysis and simulation results show that this method is very efficient in the high rate multiple description coding of strongly correlated sources. It is also observed that this method can serve as a complement to the method in [10].

Chapter 3 presents an $M$-channel MDC method using two-rate coding and staggered quantization. This method can be applied for memoryless sources as well as correlated source. For correlated sources it employs predictive coding. Application of this method in lapped transform-based multiple description image coding is demonstrated. Experimental results with 1-D and 2-D data show that this method achieves better performance than other state-of-the-art schemes.
CHAPTER 6. CONCLUSIONS AND FUTURE DIRECTIONS

The two-channel prediction-compensated multiple description coding in [38] is extended to generate more than two descriptions in Chapter 4. Design of lapped transform for MD image coding is presented using a 3 description coding example. Image coding experiments show that this method is competitive to the other state of the art techniques compared. However, comparison with two-rate coding in Chapter 3 for MD coding of a first order Gauss-Markov source shows that this method has a significant deficiency. Subsequently, an improved prediction compensation based \(M\)-channel MDC method is presented in Chapter 5. The improvement of this method over the method in Chapter 4 came from the sequential prediction and three-layer design which handles the low loss probability scenario more efficiently. Analysis using a first order Gauss-Markov source and experimental evaluation using MD image coding further validate this improvement.

6.2 Suggestions for Further Study

As explained in Chapter 1, designing MD predictive coders is not straight-forward due to the predictor mismatch problem. On the other hand predictive coding is essential in video coding to deal with the temporal and spacial correlation, and the predictor mismatch problem has been an issue in designing MD video coders [16]. In this work we developed several MD predictive coding algorithms that do not have the predictor mismatch problem. It would be interesting to investigate how these algorithms can be used in MD video coding.

Another issue that needs further study is designing efficient predictors, specially in MD image coding. For example, in applying prediction compensation based MDC to image coding, we used the simple 1-D filters. However, the image data has two-dimensional correlation structure. Consequently, more advanced 2-D filters would result in better prediction than 1-D filters[17]. This also raises the question of designing the corresponding optimal transform using 2D signal model, which has not been well studied.
Bibliography


and improvements due to an odd–even sample-interpolation procedure. IEEE


and selective quantization. Proc. SPIE Conf. on Visual Communications and


[16] Yen-Chi Lee, Y. Altunbasak, and R.M. Mersereau. An enhanced two-stage mul-
tiple description video coder with drift reduction. IEEE Trans. Circ. Syst. for

resilient time domain lapped transform. Proc. IEEE Int. Conf. Acoust., Speech,

441, Feb. 2007.


degradation over packet erasure channels through forward error correction. IEEE


IEEE/DIMACS Workshop on Coding and Quantization, Piscataway, NJ, Oct.
1993.


