Modeling of Scalable Video Content for Multi-user Wireless Transmission

by

Hassan Bader Mansour
B.E., American University of Beirut, 2003
M.A.Sc., The University of British Columbia, 2005

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

in

The Faculty of Graduate Studies
(Electrical and Computer Engineering)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

August, 2009

© Hassan Bader Mansour 2009
Abstract

This thesis addresses different aspects of wireless video transmission of scalable video content to multiple users over lossy and under-provisioned channels. Modern wireless video transmission systems, such as the Third Generation Partnership Project (3GPP)’s high speed packet access (HSPA) networks and IEEE 802.11-based wireless local area networks (WLANs) allow sharing common bandwidth resources among multiple video users. However, the unreliable nature of the wireless link results in packet losses and fluctuations in the available channel capacity. This calls for flexible encoding, error protection, and rate control strategies implemented at the video encoder or base station.

The scalable video coding (SVC) extension of the H.264/AVC video standard delivers quality scalable video bitstreams that help define and provide quality of service (QoS) guarantees for wireless video transmission applications. We develop real-time rate and distortion estimation models for the coarse/medium granular scalability (CGS/MGS) features in SVC. These models allow mobile video encoders to predict the packet size and corresponding distortion of a video frame using only the residual mean absolute difference (MAD) and the quantization parameter (QP).

This thesis employs different cross layer resource allocation techniques that jointly optimize the video bit-rate, error protection, and latency control algorithms in pre-encoded and real-time streaming scenarios. In the first scenario, real-time multi-user streaming with dynamic channel throughput and packet losses is solved by controlling the base and enhancement layer quality as well as unequal erasure protection (UXP) overhead to minimize the frame-level distortion. The second scenario considers pre-encoded scalable video streaming in capacity limited wireless channels suffering from latency problems and packet losses. We develop a loss distortion model for hierarchical predictive coders and employ dynamic UXP allocation with a delay-aware non-stationary rate-allocation streaming policy. The third scenario addresses the problem of efficiently allocating multi-rate IEEE 802.11-based network resources among multiple scalable video streams using temporal fairness constraints. We present a joint link-adaptation at the physical (PHY) layer and a dynamic packet dropping mechanism in the network or medium access control (MAC) layer for multi-rate wireless networks. We demonstrate that these methods result in significant performance gains over existing schemes.
# Table of Contents

Abstract ................................................................. ii

Table of Contents ..................................................... iii

List of Tables .......................................................... vii

List of Figures ........................................................ viii

Acronyms ................................................................. xii

Acknowledgements ...................................................... xiv

Dedication ................................................................. xv

Co-authorship Statement ............................................... xvi

1 Introduction and Overview ........................................ 1

1.1 Introduction ......................................................... 1

1.2 Thesis Statement and Objectives ................................. 3

1.2.1 Thesis Statement ................................................ 3

1.2.2 Thesis Objectives ................................................ 4

1.3 Overview of H.264/AVC and the Scalable Video Coding Extension ................................. 4

1.3.1 The H.264/AVC Video Coding Standard ....................... 4

1.3.2 The Scalable Video Coding Extension ......................... 9

1.4 Video Rate and Distortion Modeling ............................... 11

1.4.1 Generalized Rate-Distortion Model ......................... 12

1.4.2 Real-time Rate-Distortion Models ............................ 12

1.5 Overview of Wireless Video Transmission Applications .......... 16

1.5.1 Streaming Services ............................................ 16

1.5.2 Conversational Services ...................................... 17

1.5.3 Application Layer Forward Error Correction for Video Transmission .................. 18
# Table of Contents

1.6 Examples of Wireless Transmission Networks ................................................. 20  
  1.6.1 High Speed Packet Access ................................................................. 20  
  1.6.2 IEEE 802.11-based WLANs ................................................................. 21  
  1.6.3 Optimal Bit-rate Allocation for Wireless Video Transmission ................... 22  
1.7 Thesis Contributions and Organization ......................................................... 24  
  1.7.1 Summary of Contributions ................................................................. 24  
  1.7.2 Thesis Organization ............................................................................. 25  
1.8 References ................................................................................................. 27  

2 Rate and Distortion Modeling of CGS/MGS Coded Scalable Video Content ........ 32  
  2.1 Introduction .............................................................................................. 32  
   2.1.1 Main Contributions ............................................................................ 33  
  2.2 Analytical Derivation of the Rate-Distortion Behavior ............................... 34  
   2.2.1 Distribution of Residual Samples ...................................................... 34  
   2.2.2 Base-layer Models ............................................................................ 35  
   2.2.3 Enhancement-layer Models ............................................................. 43  
  2.3 Empirical Modeling of the Rate-Distortion Behavior ................................... 46  
   2.3.1 Prediction Error Estimation Model .................................................... 47  
   2.3.2 Distortion Model .............................................................................. 48  
   2.3.3 Rate Model ....................................................................................... 49  
  2.4 Real-Time Joint Rate and Protection Allocation for Multi-User Scalable Video  
    Streaming ................................................................................................. 50  
   2.4.1 System Overview .............................................................................. 51  
   2.4.2 Joint Rate and Protection Allocation for Multiple Scalable Video Streams 53  
   2.4.3 Simulation Results ............................................................................ 56  
  2.5 Conclusion ............................................................................................... 58  
  2.6 References ............................................................................................... 59  

3 Channel Aware Multi-User Scalable Video Streaming over Lossy Under- 
  provisioned Channels ....................................................................................... 61  
  3.1 Introduction .............................................................................................. 61  
   3.1.1 Summary of Contributions ................................................................ 62  
  3.2 Literature Review ...................................................................................... 63  
  3.3 Video Streaming System Overview ............................................................. 65  
  3.4 Application Layer Error Protection ............................................................ 66  
   3.4.1 Rate-Minimized Adaptive UXP Allocation ........................................ 66  
  3.5 Packet Delay Restrictions in Video Streaming ............................................ 67
Table of Contents

3.5.1 Packet Delay Analysis ................................................. 68
3.5.2 Play-out Deadline and the Maximum NALU Size .................. 69
3.5.3 UXP Delay Analysis ................................................... 70
3.6 Loss-Distortion Modeling ................................................. 72
  3.6.1 Hierarchical Prediction Structure and the Picture Copy (PC) Conceal-
        ment in SVC ............................................................ 72
  3.6.2 Proposed Loss-Distortion Model for SVC ......................... 74
3.7 Multi-User Bit-rate Allocation .......................................... 76
  3.7.1 Stationary Rate-Distortion Behavior ............................ 78
  3.7.2 Non-Stationary Rate-Distortion Behavior ....................... 78
  3.7.3 Delay-Aware Non-Stationary Behavior .......................... 79
3.8 On the Non-convexity of the Rate Allocation Problem .............. 79
3.9 Simulations and Results ................................................ 82
  3.9.1 Scheme Definition .................................................. 82
  3.9.2 Performance Testing ............................................... 84
3.10 Conclusion ........................................................................ 88
3.11 References ...................................................................... 89

4 Dynamic Resource Allocation for MGS H.264/AVC Video Transmission over
  Link-Adaptive Networks ..................................................... 92
4.1 Introduction ..................................................................... 92
  4.1.1 Main Contributions ................................................... 93
  4.1.2 Overview of Related Work ......................................... 94
4.2 General Overview of the Resource Allocation Problem ............ 95
  4.2.1 System Overview ....................................................... 96
  4.2.2 Problem Description ................................................ 96
  4.2.3 Overview of Existing Resource Allocation Solutions .......... 97
  4.2.4 Brief Overview of the Proposed Solution ...................... 98
4.3 Proposed Resource Allocation Scheme .................................. 100
  4.3.1 Joint Link Adaptation and MAC/Network Traffic Control .. 100
  4.3.2 Solution of the Global Optimization Problem ............... 102
4.4 Performance Evaluation .................................................. 108
4.5 Conclusion ...................................................................... 112
4.6 References ...................................................................... 113
## Table of Contents

5 Discussion and Conclusions ......................................................... 115
   5.1 Significance of the Research ................................................. 115
   5.2 Potential Applications ........................................................ 116
   5.3 Contributions ......................................................................... 117
   5.4 Future Research ...................................................................... 118
      5.4.1 Enhanced Rate-Distortion Modeling ................................. 118
      5.4.2 Increasing Video Streaming Complexity ............................ 119
   5.5 References .............................................................................. 120
List of Tables

1.1 List of Modulation and Coding Schemes for an HSDPA System ............. 21
1.2 Valid MCS Indices for 2x2 802.11n WLAN ........................................ 23

2.1 Comparison of Rate (Size in Bits) and Distortion (PSNR) Estimation in terms
   of RMSE ................................................................................................. 43
2.2 Extended List of Modulation and Coding Schemes ................................. 52

3.1 HSDPA Modulation and Coding Schemes ............................................. 85

4.1 Valid MCS Indices for 2x2 802.11n WLAN ........................................ 104
## List of Figures

1.1 General structure of the H.264/AVC hybrid video encoder ........................................... 5
1.2 Example of the types of scalability that can be provided by an SVC bitstream .................. 9
1.3 Comparison between the acyclic dependency graphs of non-scalable H.264/AVC (top) and scalable SVC (bottom). The SVC dependency graph shows the hierarchical predictive structure of an SVC coder .................................................. 10
1.4 Illustration of the quantization refinement coding performed in a CGS enhancement layer on a 4x4 transform block .......................................................... 11
1.5 General structure of the proposed video streaming system over 3G wireless networks. The link layer feeds back the estimated channel capacity $C$, RLC loss rate $p$, and the packet delay $d$ to the streaming server ........................................ 17
1.6 Simplified illustration of a video conferencing system over 3G HSDPA/HSUPA network. Every device has the capability of receiving and transmitting a video stream over a downlink / uplink channel, respectively. The channel quality is reported to a central controller located at the base station which in turn dictates the transmission rates to the wireless users. .................................................. 18
1.7 Standard UXP scheme with FGS layer-based classes. Each RTP packet is also fragmented into $m$ RLC frames. The value of $m$ depends on the link-layer specifications ........................................ 19
1.8 BER performance of STBC modes of an 802.11n MIMO PHY in quasi-static fading channel, perfect channel estimation assumed [48] ........................................... 22
2.1 Simplified encoder block diagram showing the possible extraction points for rate-distortion model parameters .......................................................... 33
2.2 Illustration of the pixel domain residual histogram and the Laplacian distribution fit for selected Intra- and Inter-coded frames from the Foreman sequence. .... 35
2.3 Performance of the derived distortion model for six reference video sequences encoded at QPs of 38, 32, and 26 .................................................. 39
2.4 Performance of the derived rate model for six reference video sequences encoded at QPs of 38, 32, and 26 .................................................. 41
2.5 Performance of the derived scalable distortion model for the enhancement layer distortion of six reference video sequences with a base layer encoded at a QP = 38 and CGS enhancement layer encoded at a QP = 32. .......................... 45

2.6 Performance of the derived scalable rate model for six reference video sequences with the base layer encoded at QP = 38 and CGS enhancement layer encoded at QP = 32. .......................... 46

2.7 Relationship of the derived MSE with respect to the parameter $\lambda$ and the quantization parameter QP. The figures show a linear relationship of the MSE with $\lambda$ and an exponential relationship with the QP. .......................... 48

2.8 Comparison between the modeled distortion and rate estimates and the actual data. .......................................................... 50

2.9 Schematic diagram of the live multi-user scalable video streaming environment. The mobile users share the same downlink channel. The channel SNR is reported to the streaming server which accordingly adjusts the encoding quality of the scalable video sources to satisfy the corresponding throughput constraint. ....... 51

2.10 Plot of the UXP failure rate function $P_F(.)$ showing the non-convex shape of the function over the full range of $k$ and the convex region over the reduced feasible interval. .................................................. 55

2.11 Simulated SNR values in a correlated slowly fading channel (a) and the corresponding achievable peak data rates (b) for 10% packet loss rate. ............ 57

2.12 Comparison between the four defined streaming schemes in terms of average PSNR. 58

3.1 The plot shows that unequal error protection (UEP) results naturally in a non-convex optimization problem as witnessed by the concavity of the UEP function compared to linearity of equal error protection (EEP). The protected bit-rate region is bounded by the available channel capacity. ......................... 62

3.2 General structure of the proposed video streaming system over 3G wireless networks. The link layer feeds back the estimated channel capacity $C$, RLC loss rate $p$, and the packet delay $d$ to the streaming server. The rate and distortion functions, $R_u$ and $D_u$, are defined in (3.21). ......................... 65

3.3 Timing diagram for the transmission of a NALU indexed by $l$. The NALU is fragmented into $m$ RLC frames each of which can have a different transmission delay $d$. .......................................................... 69

3.4 Comparison between the acyclic dependency graphs of non-scalable H.264/AVC (top) and scalable SVC (bottom). The SVC dependency graph shows the hierarchical predictive structure of an SVC coder. This prediction structure plays an important role in loss-distortion modeling as discussed in Section 3.6. ....... 72
3.5 (a) Dyadic hierarchical structure of two GOPs with 4 temporal scalability levels. The frame loss dependency among B-frames can be seen in the groups of empty frames where the lost frames are those that have an X underneath. The remaining empty frames are affected during decoding due to their dependency on the lost frames. (b) Example of an SVC base layer coding sequence with a fixed GOP size G and an intra-refresh period $T_{\text{refresh}}$. The $\Delta$s indicate the mean square error between the specified frames. ................................. 73

3.6 The figures show the performance of the loss-distortion model, derived in Section 3.6, in estimating the MSE of the sequences Foreman and Bus with packet losses of 3%, 5%, and 10% as a function of the Intra-refresh period. Error bounds are shown for a 95% confidence interval. ................................. 76

3.7 Example of convex relaxation using the linear envelope of the UEP constraint function. The optimal solution $x^*$ to the convex-relaxed problem is non-feasible, therefore, we achieve a feasible solution by projecting onto the feasible non-convex set using the projection operator $P(x^*)$ defined by $P(u) = \arg \min_x \|x - u\|^2_2$ s.t. $x \in \mathcal{X}$, where $\mathcal{X}$ is the feasible set. ................................. 81

3.8 Comparison of the Y-PSNR performance between the four streaming schemes FUFR, OUFR, OUOR, and MOUOR without considering the effect of packet delay as described in Section 3.7.2. ................................. 83

3.9 Comparison of the Y-PSNR performance between the four streaming schemes FUFR, OUFR, OUOR, and DMOUOR with the effect of packet delay as described in Section 3.7.3. ................................. 84

3.10 Comparison of the Y-PSNR performance between the four schemes DMOUOR, OUOR, OUFR, and FUFR for individual video sequences under 3.6 Mbps, 5.3 Mbps, and 7.2 Mbps capacity and 10% PLR. ................................. 87

3.11 Comparison of the Y-PSNR performance between the three schemes DMOUOR, OUOR, and OUFR for different play-out deadlines over a channel with 5.3 Mbps capacity and 3% PLR. ................................. 87

4.1 General diagram of the wireless video streaming system in which a transmitter (base-station or access-point) serves multiple video users/clients. The video streaming service is allocated a fixed share of the bandwidth that the transmitter can support, such that all video users compete for a portion of the allocated bandwidth share. ................................. 94
The Joint operation of the Link-Adaptation and Traffic-Control modules in our proposed framework. The algorithm combines the resources \((\vec{R}, N, \vec{\tau})\) assigned to all users allowing dynamic allocation of these resources on a short term basis. This results in more efficient use of the limited capacity \(\eta\) and higher overall received video quality \(\vec{D}\) of all transmitted streams that share the same wireless channel.

BER performance of STBC modes of an 802.11n MIMO PHY in quasi-static fading channel, perfect channel estimation assumed \cite{14}.

Example of the achievable packet error rate (PER) curves and their continuous approximation as a function of the PHY transmission rates for different SNRs.

Plot of the function \(\rho(.)\) showing the non-convex shape of the function over the full range of \(W\) and the convex region over the reduced feasible interval.

Example of the mapping process involved in discretizing the continuous valued optimal rate points \(w^*_l\) to the achievable discrete PHY rates \(C'_l\) or \(C''_l\), where \(l \in \{0, 1, 2\}\).

Comparison in average Y-PSNR performance between the three resource allocation schemes: JLATC (proposed), ILA, and CLD. The performance of each scheme is plotted versus different channel SNRs and the following service temporal share values: (a) \(\eta = 0.4\), (b) \(\eta = 0.5\), (c) \(\eta = 0.65\).

Individual stream performance in terms of Y-PSNR using the JLATC, ILA, and CLD resource allocation schemes for different channel SNR values and a service temporal share \(\eta = 0.4\).

Comparison of bandwidth utilization and corresponding video Y-PSNR between the three resource allocation schemes: JLATC (proposed), ILA, and CLD. The plots correspond to a streaming temporal share budget \(\eta = 0.5\) and channel SNR = 12.5 dB for (a) and (b), and SNR = 15dB for (c) and (d).
Acronyms

3GPP  3rd Generation Partnership Project
AMC  Adaptive Modulation and Coding
ARQ  Automatic Repeat Request
BER  Bit Error Rate
CABAC  Context Adaptive Binary Arithmetic Coding
CAVLC  Context Adaptive Variable Length Coding
CGS  Coarse Grain Scalability
CQI  Channel Quality Indicator
DCT  Discrete Cosine Transform
DVB-H  Digital Video Broadcast - Handheld
FEC  Forward Error Correction
FER  Frame Error Rate
FGS  Fine Grain Scalability
GOP  Group Of Pictures
HD  High Definition
HSDPA  High Speed Downlink Packet Access
HSUPA  High Speed Uplink Packet Access
IP  Internet Protocol
ITU  International Telecommunication Union
LTE  Long Term Evolution
MAD  Mean Absolute Difference
MAC  Media Access Control
MANET  Mobile Adhoc Network
MB  Macroblock
Acronyms

MBMS  Multimedia Broadcast Multicast
MCS  Modulation and Coding Scheme
MGS  Medium Grain Scalability
MPEG  Motion Picture Experts Group
MSE  Mean Square Error
NALU  Network Abstraction Layer Unit
PHY  Physical Layer
PSNR  Peak Signal to Noise Ratio
PER  Packet Error Rate
QoS  Quality of Service
QP  Quantization Parameter
RDO  Rate-Distortion Optimization
RLC  Radio Link Control
RS  Reed Solomon
RTP  Real-time Transport Protocol
SATD  Sum of Absolute Transform Differences
SD  Standard Definition
SNR  Signal to Noise Ratio
SVC  Scalable Video Coding
TB  Transmission Block
TTI  Transmission Time Interval
UBC  The University of British Columbia
UE  User Equipment
UEP  Unequal Error Protection
UMTS  Universal Mobile Telecommunications System
UXP  Unequal Erasure Protection
VCL  Video Coding Layer
WLAN  Wireless Local Area Network
Acknowledgements

To begin with, I would like to thank my supervisors, Dr. Panos Nasiopoulos and Dr. Vikram Krishnamurthy, for their time, patience, and support throughout my degree. I have known Dr. Nasiopoulos for over five years. During these years, his kind, attentive, and cheerful approach has helped me strive through the rough stretches that accompany graduate student life. In effect, he was not just a graduate supervisor, he has become more of a friend and life mentor. Dr. Krishnamurthy is an exceptional academic whose dedication has been inspirational. His supervision has helped me develop my analytical skills and taught me that “life is a constrained optimization problem.” I would also like to express my gratitude to the committee members: Dr. Jane Wang, Dr. Victor Leung, and Dr. Tim Salcudean, for their constructive feedback.

I would like to thank my research group colleagues: Lino for his humour, Qiang for the discussions on life and divinity, Victor for his carefree attitude, Sergio for his nonchalance and the rides to Whistler, Matthias for all the times he agreed, Jack for bringing out the teacher in me, Colin for his knowledge and objectivity, Di for her eclectic name, Rupin for his positivity, and an un-named someone for testing the limits of my patience.

To Dr. Ozgur Yilmaz, thank you for all the thoughtful discussions and your contagious appreciation for the beauty of knowledge.

To the staff of “The Boulevard,” thank you for your hospitality, friendliness, and that human aspect which is scarce in engineering life. Without you this thesis would probably still be a work in progress.

To my parents, it would suffice to say that this would never have been possible without you. Keep up the good work! To my brother, Wissam, thank you for your support and encouragement. Look who’s a doctor too!

Last but not least, I would like to thank my best friends Rayan Saab, Lina Darwiche, and Ilker Hacihaliloglu for always being there. Rayan, you’re welcome. Lina, thank you for listening. Haci, I’ll get your name right one day.
“As people become more intelligent they care less for preachers and more for teachers.”

Robert G. Ingersoll
Co-authorship Statement

This research presents work conducted by Hassan Mansour, in collaboration with Dr. Panos Nasiopoulos, Dr. Vikram Krishnamurthy, and Dr. Yaser P. Fallah.

Note, this is a Manuscript based thesis constructed around the three manuscripts described below:

**Manuscript 1:** *Rate and Distortion Modeling of CGS/MGS coded Scalable Video Content.* This manuscript is the work of Hassan Mansour who received suggestions and feedback from his supervisors, Dr. Nasiopoulos and Dr. Krishnamurthy.

**Manuscript 2:** *Channel Aware Multi-User Scalable Video Streaming over Lossy Under- provisioned Channels: Modeling and Analysis.* This manuscript is the work of Hassan Mansour who received suggestions and feedback from his supervisors, Dr. Nasiopoulos and Dr. Krishnamurthy.

**Manuscript 3:** *Dynamic Resource Allocation for MGS H.264/AVC Video Transmission over Link-Adaptive Networks.* This manuscript is the joint work of Hassan Mansour and Dr. Fallah. Hassan Mansour was responsible for developing the video rate and distortion models, running performance evaluations, formulation and solution of the optimization problem, interpretation of results and writing the manuscript in collaboration with Dr. Fallah. Dr. Nasiopoulos and Dr. Krishnamurthy provided guidance and editorial input into the creation of the manuscript.

The first and last chapters of the thesis were written by Hassan Mansour, with editing assistance and consultation from Dr. Panos Nasiopoulos and Dr. Vikram Krishnamurthy.

Manuscripts 2 and 3 have been published in the IEEE Transactions on Multimedia. This material is posted here with permission of the IEEE. Such permission of the IEEE does not in any way imply IEEE endorsement of any of the University of British Columbia’s products or services. Internal or personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution must be obtained from the IEEE by writing to pubs-permissions@ieee.org. By choosing to view this material, you agree to all provisions of the copyright laws protecting it.
Chapter 1

Introduction and Overview

1.1 Introduction

The emergence of digital video coding during the last two decades has revolutionized the manner in which different sectors of human society interact. Coupled with a leap in wireless data transmission technology, mobile digital video content is allowing us to experience the dawn of an era of video mobility in which the moving picture can accompany us everywhere in our day to day activities. The impact of these services has opened a global market for new multimedia services stretching from the entertainment industry to the business sector to creating a virtual arena for political debate and news reporting.

Digital video hit its first great success with the advent of the MPEG-2 digital video coding standard in 1994 [1]. Following its standardization, MPEG-2 video became the technology of choice for the worldwide delivery of standard definition (SD) and high definition (HD) digital television systems. The content has been delivered using traditional systems based on MPEG-2 Systems (ITU-T H.222.0) [2] for transmission over satellite, cable, and terrestrial channels, and for storage onto DVDs [3]. Moreover, MPEG-2 coded content has also been used in conversational video conferencing services over H.320 systems [4].

Today, digital video applications include multimedia messaging, video streaming, and video conferencing over a wide range of modern transmission and storage systems. These include mobile TV, internet video streaming over wired and wireless networks, and SD and HD TV broadcast to DVD and Blu-ray disc players [4]. The diversity of these new digital video services and network infrastructures call for higher coding efficiency and flexibility on the part of the coded video bitstream. Different transmission media, for instance, can have varying bandwidth capabilities. Meanwhile, higher coding efficiency allows for a better utilization of the available data-rate by either increasing the number of transmission channels or delivering higher quality video to the consumer for the same number of channels. To address these issues, the H.264 video coding standard or MPEG-4/AVC [5] and its scalable extension SVC [6] were developed and have already been adopted as part of several media transmission standards, such as, 3GPP’s multimedia broadcast/multicast services (MBMS) [7], digital video broadcast for handheld devices (DVB-H) [8,9], and Qualcomm’s MediaFlo solution [10] to name a few.
Chapter 1. Introduction and Overview

Through several advanced coding tools, the H.264/AVC video coding standard achieves double the coding efficiency of MPEG-2 Video while providing a network friendly architecture that allows for facilitated fragmentation and encapsulation of the coded video elements. The H.264/AVC design incorporates a highly efficient video coding layer (VCL), and a network abstraction layer (NAL) to encapsulate the VCL in an appropriate manner suitable for transport stream delivery or for storage media [3]. The scalable extension of H.264/AVC, known as SVC, extends previous attempts at achieving bitstream scalability with MPEG-2 Video, H.263 [11], and MPEG-4 Visual [12] without the significant loss in coding efficiency and increased decoder complexity that have hampered such attempts. Therefore, the objective of SVC was to allow the delivery of a video bitstream containing sub-bitstreams that are independently decodable with “a complexity and reconstruction quality similar to that achieved using the existing H.264/AVC design” [4].

Recently, broadband ad hoc wireless networks and high-throughput mobile 3G/3.5G networks, such as, IEEE 802.11-based wireless local area networks (WLAN) [13,14], high-speed packet access (HSPA) [15–17] networks, multimedia broadcast and multicast services (MBMS) [7], and the next next generation of HSPA long term evolution (LTE) [18] networks, have been deployed at a pace that will ensure that Internet-based video services can now be taken by consumers “on-the-go”. As a result, the number of subscribers to video streaming and conversational services is at a steady rise, which calls for new streaming and scheduling policies that manage the large volume of video payload being transmitted over shared wireless links [19,20]. These modern video transmission systems using mobile networks are based on the real-time transport protocol (RTP) over an Internet protocol (IP) architecture (RTP/IP) [4,21]. However, the channel dynamics resulting from the nature of the wireless medium, inherently a broadcast medium, results in a tremendous set of challenges facing the support of a demanding applications such as video transmission [22].

Wireless channel link quality varies in time and space as a result of interference, multipath fading, and the adaptive resource sharing mechanisms used to manage the dynamic data rate requirements from a varying number of users joining and leaving the network in an unpredictable manner. For video transmission applications, additional challenges arise due to the bandwidth-intensity, loss-tolerance, and delay-sensitivity that characterize the service [20,22]. Moreover, a variety of devices with different throughput and display capabilities, from smart phones with small screens and limited power to high-end PCs with HD displays, add to the complexity of the task at hand. To address these challenges, cross layer optimized resource allocation algorithms coupled with scalable video coding have become the topic of significant research in the wireless video transmission community.

Cross layer optimization techniques attempt to satisfy the end-to-end requirements of the
video application by increasing the information exchange between the protocol layers. Such
information exchange allows a streaming server or other media aware network elements to
optimally adapt the transmission to the underlying channel conditions while satisfying the
video application requirements [22]. Consequently, the dynamic and complex components in the
transmission system, namely, the video payload and channel behavior, need to be analytically
modeled to maintain a tractable and solvable optimization problem. Since channel behavior
modeling has been extensively studied in the research literature [23][24], we will focus our
attention on deriving bit-rate and distortion models that characterize the coded video bitstream
and incorporate these models within different scenarios of multi-user video transmission.

In the remainder of this chapter, we present our thesis statement and list the research objectives. Next, we give an overview of the H.264/AVC video coding standard and its scalable extension, discuss previous attempts at modeling video sources. We then describe the requirements for wireless video applications and give a background overview of the methodologies used to ensure that these requirements are met. Finally, we list the thesis contributions and present the organization of the following chapters.

1.2 Thesis Statement and Objectives

1.2.1 Thesis Statement

In this thesis, we address different aspects of wireless video transmission of scalable video content to multiple users over under-provisioned links. We approach the problem from a media-aware resource allocation perspective, in which the video payload is characterized in terms of its bit-rate and expected distortion (a measure of the decoded video quality) at the receiver. Our research has shown that modeling scalable video content provided by the CGS/MGS feature of SVC using the Laplacian distribution as a source distribution model, q-domain distortion, and entropy and conditional entropy results in a more accurate approximation of the video rate and distortion behavior. Moreover, we demonstrate that by optimizing the allocation of application layer forward error correction (AL-FEC) and encoding bit-rates, considerable improvement in decoded video PSNR can be achieved under capacity and latency constrained channels. Furthermore, we show that jointly optimizing the link-adaptation and traffic control modules in multi-rate PHY networks with a temporal share fairness scenario results in a better utilization of the allocated service time as well as an improvement in received video quality.
1.2.2 Thesis Objectives

Based on the description in the previous section, it is evident that the transmission of digital video content is an important and challenging topic. Its applications involve significant research abridging the video compression and data communication fields. Moreover, the involved difficulties vary with the underlying network infrastructure that will support the transmitted multimedia content. Therefore, the goals of this research include the following:

1. Develop accurate scalable video rate and distortion models that capture the dynamic behavior of scalable video content.

2. Study the different requirements of modern video transmission networks, namely, 3G’s HSPA networks and IEEE 802.11-based WLANs.

3. Customize application layer forward error correction schemes that could overcome the limitations of the above mentioned networks.

4. Develop resource allocation and traffic control policies that ensure seamless transition and graceful degradation in displayed video quality when faced with transmission errors.

To address these objectives, we develop scalable video rate and distortion models that capture the dynamic behavior of scalable video content encoded using the CGS/MGS feature in SVC. Moreover, we implement cross-layer optimization schemes that utilize the derived models to deliver the best possible video quality while addressing the limitations and constraints of the wireless network infrastructure.

1.3 Overview of H.264/AVC and the Scalable Video Coding Extension

The evolution of digital video coding standards started with the conception of MPEG-1 (ITU-T H.261) as early as 1992 and was succeeded by continued efforts at maximizing the coding efficiency with MPEG-2 Video (ITU-T H.262), MPEG-4 Visual (ITU-T H.263 and it’s enhancements), and finally the latest international video coding standard MPEG-4/AVC (ITU-T H.264) [3].

1.3.1 The H.264/AVC Video Coding Standard

General Description

A video sequence is composed of a succession of still images/frames that are captured at a rate that varies between 15 frames per second (fps) and an unspecified upper limit depending on
Chapter 1. Introduction and Overview

the target video application. The basic design principles of video compression employed in the above mentioned standards can be summarized by a series of processes. In what follows, we will focus on the processes employed in the H.264/AVC video coding standard:

- An uncompressed image $I_o$ undergoes a predictive coding process that captures the spatial redundancies (Intra-prediction) within the same frame or temporal redundancies (Inter-prediction) across consecutive frames. The predictive process replaces these redundancies with control information (Intra-prediction modes, block sizes, motion vectors, reference indices,...) that allows a decoder to reconstruct an approximation (prediction image) of the original uncompressed image.

- The difference between the original image $I_o$ and its prediction $I_p$ is called the residual image $I_r$. A transform coding process is then applied to the residual image in which every block is transformed using an integer approximation of the 2-dimensional discrete cosine transform (DCT). The H.264/AVC standard supports two transform block sizes of 4x4 and 8x8 pixels.

- The transformed residual samples next undergo a quantization process followed by an entropy coding process of the quantized transform coefficients. Quantization is the stage which introduces loss in the encoding process for better compression. H.264/AVC defines two entropy coding processes, the context-adaptive variable length coding (CAVLC) and the context adaptive binary arithmetic coding (CABAC). The latter process is highly efficient and approaches the true data entropy at the cost of higher computational complexity.

Figure 1.1: General structure of the H.264/AVC hybrid video encoder.

Fig. 1.1 shows a diagram of the different components of an H.264/AVC video encoder. An uncompressed input picture is divided into macroblocks of 16x16 pixels each before it goes
through the different encoding processes described above. After quantization, the quantized samples go through inverse-quantization, inverse-transform, motion-compensation, and finally deblocking filtering. The last stage is used to remove any blocking artifacts that arise from the quantization of the integer cosine transform coefficients. The video encoder retains a copy of the decoded frames in a decoded picture buffer. These buffered frames are used as references for Inter-prediction of other frames. Three types of coded video frames are supported in H.264/AVC: I-frames, P-frames, and B-frames. An I-frame is coded using Intra-frame prediction as an independent self-contained compressed picture. P and B pictures, on the other hand, are Inter-frame predicted from previous pictures only, in the case of P frames, or from both previous and future pictures in B frames [3].

### Transform and Quantization

The transform and quantization stages in H.264/AVC are of extreme importance in the derivation of rate and distortion models. We will therefore give a more detailed description of these two processes as shown in [25].

**Transform:** The purpose of introducing a transform is to reduce the spatial correlation to improve compression. Since the smallest prediction block employed in H.264/AVC is of size 4x4 pixels, the standard uses a 4x4 transform block size which also leads to a significant reduction in ringing artifacts compared to the 8x8 transform size used in previous standards.

The discrete cosine transform has been the transform of choice in previous video coding standards. This is due to its close approximation of the statistically optimal Karhunen-Loève transform for image signals [25]. Moreover, given an $N \times N$ signal matrix $x$, the DCT coefficients $\hat{x}$ can be computed as the linear transformation $\hat{x} = HxH^T$, where $H$ is the forward transform matrix with the $k$th row, $n$th column element $H_{kn}$ given by:

$$H_{kn} = c_k \sqrt{\frac{2}{N}} \cos \left( n + \frac{1}{2} \right) \frac{k\pi}{N}, \quad (1.1)$$

where $c_0 = \sqrt{2}$, and $c_k = 1$ for $k > 0$. However, the DCT entries are irrational numbers that require a floating point representation which is platform dependent. Therefore, the transform used in H.264/AVC is an integer approximation of the 4x4 DCT achieved by rounding scaled entries of the DCT matrix to the nearest integer as shown below:

$$H = \text{round}(2.5 \times H_{DCT}). \quad (1.2)$$

---

**Chapter 1. Introduction and Overview**
This results in the following integer transform matrix

\[ H = \begin{bmatrix}
1 & 1 & 1 & 1 \\
2 & 1 & -1 & -2 \\
1 & -1 & -1 & 1 \\
1 & -2 & 2 & -1
\end{bmatrix}. \]  

(1.3)

The standard then defines a scaled inverse transform matrix \( H_i \) given by

\[ H_i = \begin{bmatrix}
1 & 1 & 1 & 1/2 \\
1 & 1/2 & -1 & -1 \\
1 & -1/2 & -1 & 1 \\
1 & -1 & 1 & -1/2
\end{bmatrix}, \]  

(1.4)

which results in the relationship

\[ H_i D^{-1} H = I, \]  

(1.5)

where \( D \) is a diagonal matrix given by \( D = \text{diag}\{4, 5, 4, 5\} \).

Since the integer transform and its scaled inverse are non-orthonormal matrices, a 4x4 block \( x \) can only be reconstructed by first scaling the transformed coefficients \( \hat{x} \) with a matrix \( V \) before applying the inverse transform. This can be illustrated as follows:

\[
\hat{x} = H x H^T, \\
x = H_i (\hat{x} \oslash V) H_i^T,
\]

(1.6)

where \( \oslash \) indicates element by element division, and \( V \) is given by

\[
V = D E D^T = \begin{bmatrix}
16 & 20 & 16 & 20 \\
20 & 25 & 20 & 25 \\
16 & 20 & 16 & 20 \\
20 & 25 & 20 & 25
\end{bmatrix}, \]  

(1.7)

where \( E \) is a 4x4 matrix of 1s.

**Quantization:** Let \( X \) be a 4x4 block from the residual frame \( I_r \), and let \( \hat{X} \) be the corresponding transformed block. The H.264/AVC encoder performs a scalar quantization of the coefficients \( \hat{X} \) using only integer multiplication and right bit-shifts to avoid any division \[25\].

Let \( \hat{X}_q \) and \( \hat{X}_r \) be the quantized and reconstructed block samples of \( X \), respectively. A
typical scalar quantization and reconstruction process can be performed by

\[ \hat{X}_q = \text{sign}(\hat{X}) \left\lfloor \frac{\hat{X}}{Q_s} + f \right\rfloor, \quad \text{and} \quad \hat{X}_r = \hat{X}_q Q_s, \]  

(1.8)

where \( Q_s \) is the quantization step, and \( f \in [0, 0.5] \) is a rounding factor that controls the quantization near zero. Note that in the above equation and in the remainder of this thesis, the addition of a scalar to a matrix denotes adding the scalar to each element of the matrix. Also note that the sign function is defined elementwise on the matrix \( X \), such that,

\[
\text{sign}(x_{uv}) = \begin{cases} 
1, & \text{if } x_{uv} > 0 \\
0, & \text{if } x_{uv} = 0 \\
-1, & \text{if } x_{uv} < 0
\end{cases}
\]  

(1.9)

where \( u \) and \( v \) are the row and column indices.

In order to reduce complexity and avoid division operations, the H.264/AVC encoder implements the quantization process by multiplying the transformed signal by a scalar multiplier extracted from periodic quantization followed by right bit-shifts to reduce the dynamic range of the quantized coefficients. The standard defines a quantization parameter (QP) which is used to select the multiplier from the quantization tables. Let \( Q \in \{0, 1, \cdots, 51\} \) be the encoding QP, the quantization and reconstruction operations can be summarized as follows

\[
\hat{X}_q = \text{sign}(\hat{X}) \left\lfloor \left( \left\lfloor \frac{\hat{X}}{Q_m} \right\rfloor \cdot M(Q_m) + f 2^{15+Q_e} \right) 2^{-(15+Q_e)} \right\rfloor, \\
\hat{X}_r = \hat{X}_q \cdot S(Q_m) 2^{Q_e},
\]

(1.10)

where \( \cdot \) is the Hadamard product which denotes elementwise multiplication, \( Q_m = Q \mod 6 \), \( Q_e = \lfloor Q/6 \rfloor \), \( M(Q_m) \) is a 4x4 forward scaling matrix, and \( S(Q_m) \) is the reconstruction matrix. The scaling matrices \( M(Q_m) \) and \( S(Q_m) \) are derived from standard defined quantization tables [5], and have the following relationship

\[
M(Q_m) \cdot S(Q_m) \cdot V \cong 2^{21},
\]

(1.11)

where \( V \) is the matrix defined in (1.7).

Finally, the reconstructed residual signal \( X_r \) is calculated by performing the inverse transform and rounding operations shown below

\[
X_r = \left\lfloor \left( H_i \hat{X}_r H_i^T + 2^5 \right) 2^{-6} \right\rfloor.
\]

(1.12)
1.3.2 The Scalable Video Coding Extension

Overview of SVC features

The scalable video coding (SVC) extension of H.264/AVC provides spatial, temporal, and SNR (quality) scalability which ensures a graceful degradation in video quality when faced with channel loss [4].

Spatial scalability is a term used to describe cases in which a different subset of the video bitstream can be pruned to result in a decoded content with a reduced spatial resolution. Similarly, temporal scalability results in the presentation of the source with a reduced frame rate. On the other hand, quality scalability describes the case where a sub-bitstream provides the same spatio-temporal resolution as the full bitstream except for a lower quality or signal-to-noise ratio (SNR) [4]. Fig. 1.2 illustrates the different kinds of scalability with a base layer, an SNR first enhancement layer, and a spatial second enhancement layer.

![Fig. 1.2: Example of the types of scalability that can be provided by an SVC bitstream.](image)

The SVC design aims at providing the different kinds of scalability with the minimum loss in coding efficiency and as little increase as possible in decoding complexity compared with the single layer H.264/AVC design. To achieve these requirements, spatial scalability is provided using the concept of inter-layer prediction in which a considerable amount of enhancement layer coding information, such as, motion vectors, intra-modes, and texture data, are derived from lower layers. This reduces the amount of information that needs to be separately encoded in the enhancement layer bitstream.

The concept of temporal scalability is achieved by means of a hierarchical coding structure illustrated in Fig. 1.3. Temporal scalability is based on a Group of Picture (GOP) basis with one key-picture per GOP. The hierarchical prediction structure imposes a coding dependency between the B-frames of a GOP. The benefits of this dependency are translated into increasing
the coding efficiency compared to non-hierarchical predictive AVC, and enabling temporal scalability, where B-frames at the lower hierarchical level can be discarded during transmission to produce a bitstream with a lower temporal resolution.

![Non-hierarchical H.264/AVC Structure]

### Figure 1.3: Comparison between the acyclic dependency graphs of non-scalable H.264/AVC (top) and scalable SVC (bottom). The SVC dependency graph shows the hierarchical predictive structure of an SVC coder.

A coded video frame in SVC can be encapsulated into one or several Network Abstraction Layer Units (NALUs), such that the base layer and each CGS enhancement layer is embedded in a separate NALU. Rate control is also achieved using the quality scalable progressive refinement NALUs which can be truncated arbitrarily. This helps in supporting flexible bit rate adaptation of the coded video as a response to channel quality information that is fed back from the link level feedback mechanism such as the automatic repeat request (ARQ).

#### SNR Scalability in SVC

In SVC, quality scalability is achieved using medium-grained or coarse-grained scalability (MGS/CGS) where scalable enhancement packets deliver quality refinements to a preceding layer representation by re-quantizing the residual signal using a smaller quantization step size and encoding only the quantization refinements in the enhancement layer packets. Fig. 1.4 illustrates the quantization refinement coding performed in a CGS enhancement layer. The original transform coefficients are first quantized by the base-layer. The enhancement layer quantizes the base-layer quantization error using a finer quantization step size.

Moreover, MGS/CGS enhancement layer packets are coded using the key-picture concept where the coding loop is closed at the highest scalability layer for key-pictures only, while the
loop is closed at the base-layer quality for the remaining pictures. This approach achieves a tradeoff between the coding efficiency of enhancement layers and the drift at the decoder [4]. These new features in SVC result in a divergence from existing rate-distortion models, calling for the development of improved models that can accurately capture the runtime rate-distortion behavior.

During the development of the SVC standard, another form of scalability called fine granular scalability (FGS) was proposed and integrated into the SVC coding engine. FGS was inherited from MPEG-4 PFGS [27] in which the refinement in quantization coefficients are encoded using bit-plane coding. A more advanced form of FGS was then designed specifically for SVC which uses cyclic block coding [28, 29] in an attempt to reduce the decoder complexity. However, despite the efforts to simplify the decoding process of FGS, it was removed from the Phase I standardization of SVC. FGS remains an ongoing research project as part of Phase II of the SVC standardization with efforts being focused on achieving a more simplified decoding structure.

1.4 Video Rate and Distortion Modeling

The advent of video applications into the wireless scene has called for highly accurate media-aware coder control and streaming algorithms to manage the difficulties of wireless transmission.

Video rate and distortion modeling lies at the core of model-based coder control algorithms,
whereby a coded video packet is abstracted in terms of its rate $R_s$ and the distortion $D_s$ of the resulting coded picture [30]. The rate is defined as the size in bytes of the video packet and the distortion is often measured in terms of the mean-square-error (MSE) or peak-signal-to-noise-ratio (PSNR). Rate-distortion (R-D) models are used in real-time coder and transmission control algorithms to predict the rate and distortion of a video packet prior to the completion of the encoding process. However, scalability and encoder complexity add a variety of parameters that render traditional rate and distortion models inaccurate. For instance, the interdependency between rate-distortion optimized (RDO) motion estimation and rate-control in H.264/AVC is described as a “chicken and egg” dilemma [31]. Therefore, it is important that the encoder parameters used in rate-distortion modeling be accessible as early as possible in the encoding process.

1.4.1 Generalized Rate-Distortion Model

One of the earliest and more accurate video rate-distortion models [32] devises an empirical formula that relates the video packet bit-rate $R_s$ and its distortion $D_s$. This relationship is expressed as follows:

$$D_s(R_s) = D_o + \frac{\theta_o}{R_s - R_o}$$ (1.13)

where $D_o$, $\theta_o$, and $R_o$ are model parameters that depend on the video content and encoder. A major advantage of the model shown in (1.13) is that it expresses the video distortion as a convex function of the bit-rate. However, this model can only express the time-average behavior of a coded video stream and does not reflect the temporal variation that is normally exhibited during video encoding.

Another rate distortion model presented in [33] expresses the rate-distortion relationship as an exponential function scaled by the variance of the transform coefficients. This model can be expressed as follows:

$$D_s(R_s) = \sigma^2 e^{-a R_s},$$ (1.14)

where $a$ is a sequence dependent parameter.

1.4.2 Real-time Rate-Distortion Models

Several models have been proposed to predict the rate and distortion of a video packets prior to the completion of the encoding process [30, 34].
Classic MPEG-2 models

In [34], simplified MPEG-2 rate control models are used to estimate the rate and distortion of coded video frames. These models are given by the mean absolute difference (MAD) between every two successive frames $\zeta_M$ and the quantization step size $Q_s$ as shown below:

$$D(Q) = aQ_s, \quad R(\zeta_M, Q_s) = b\frac{\zeta_M}{Q_s},$$

(1.15)

where $a$ and $b$ are model parameters derived empirically for each sequence. This model has several limitations. The most significant of which are the lack of distinction between Intra and Inter coded frames, and the independence of the distortion model from the sequence complexity measure $\zeta_M$. Moreover, this model does not result in an accurate prediction of the rate and distortion of the H.264/AVC coded video packets due to the new coding techniques adopted in H.264/AVC.

Distribution-based Rate-Distortion models

For more accurate representations of the video rate-distortion behavior, several models have been proposed based on the $q$-domain and the $\rho$-domain of the residual frame $I_r$ [30,35,37]. The $q$-domain refers to modeling the source rate and distortion as a function of the quantization step size and the complexity of the residual signal based on a statistical fit to the distribution of residual samples. In the $\rho$-domain the rate and distortion are modeled as a function of $\rho$, where $\rho$ refers to the percentage of zero transform coefficients. The most common distributions used in the above models are the Laplacian and the Cauchy distributions with density functions $f_L(x, \lambda)$ and $f_C(x, \mu)$, respectively, shown below:

$$f_L(x, \lambda) = \frac{1}{2\lambda} \exp\left\{ -\frac{|x|}{\lambda} \right\}, \quad f_C(x, \mu) = \frac{1}{\pi \mu^2} \frac{\mu}{\mu^2 + x^2},$$

(1.16)

where $\lambda$ and $\mu$ are the respective distribution parameters.

In [38], rate and distortion models are developed based on the $\rho$-domain information for MPEG-4 and H.263 coded video content at the macroblock level for rate-control applications. The model assumes a Laplacian distribution of DCT coefficients, expresses the rate as a linear function of $\rho$ and uses the exponential relationship between rate and distortion shown in (1.14). These models are shown below:

$$R(\rho) = \theta(1 - \rho),$$

$$D(\rho) = \sigma^2 e^{-a(1-\rho)},$$

(1.17)

where $\theta$ and $a$ are sequence dependent constants, and $\sigma^2$ is the distribution variance.
Chapter 1. Introduction and Overview

In \[30\], the rate and distortion models are based on approximations of the \( \rho \)-domain behavior and written as the following functions of the quantization step size and sum of absolute transform differences (SATD) of the residual data:

\[
\begin{align*}
R_s(Q_s, SATD(Q_s)) &= \alpha^{\text{SATD}(Q_s)} Q_s^{p_1}, \\
D_s(Q_s, SATD(Q_s)) &= \beta^{\text{SATD}(Q_s)} Q_s^{p_2} + D_{\text{SKIP}}(Q_s),
\end{align*}
\]

where \( D_{\text{SKIP}}(Q_s) \) is the distortion of the SKIP mode macroblocks, and \( \alpha \) and \( \beta \) are sequence dependent model parameters. The exponentials \( p_1 \) and \( p_2 \) are equal to 1 for P and B frames, and equal to 0.8 and 1.2, respectively, for I frames. The disadvantage of this model lies in the parameter extraction which is performed late in the encoding process after transform and quantization. As a result, real-time coder control applications, such as wireless video transmission, could suffer from significant performance degradation due to the delay imposed by the rate-distortion modeling.

In \[35\], video rate and distortion models are derived for H.264/AVC encoded video streams. The models are derived assuming that the residual source samples follow a Laplacian distribution. The distortion is calculated based on the \( q \)-domain analysis which computes the error in quantizing a Laplace distributed source signal. The bit-rate is modeled as a linear combination of the number of non-zero quantized transform coefficients \( N_{nz} \) and the \( \ell_1 \) norm of the quantized transform coefficients \( E_{QTC} \). These two parameters are estimated using the statistical distribution \( f_L(x, \lambda) \) of residual samples. The proposed rate and distortion models are therefore expressed as follows:

\[
\begin{align*}
R_s(Q_s, \lambda) &= \alpha N_{nz} + \beta E_{QTC}, \\
D_s(Q_s, \lambda) &= 2\int_{0}^{\lambda} x^2 f_L(x, \lambda) dx + 2\sum_{i=(i-\frac{5}{6})Q_s}^{(i+\frac{\lambda}{2})Q_s} \int_{(i-\frac{\lambda}{2})Q_s}^{(i+\frac{\lambda}{2})Q_s} (x - iQ_s)^2 f_L(x, \lambda) dx,
\end{align*}
\]

where \( \lambda \) is the Laplacian distribution parameter that can be calculated from the sample variance as such: \( \sigma_x^2 = 2\lambda^2 \). Moreover, the authors derive simplified transform domain equations by relating the distribution of the spatial domain residual samples with their transformed coefficients. This is performed by assuming that the source samples in a 4x4 block are Markovian, and using the correlation coefficient of the spatial domain samples to calculate the transform domain sample variance. However, the work makes several simplifying assumptions especially in the rounding error performed during quantization and they do not address SKIP mode.

1The SKIP mode is an Inter-prediction mode for P-frame macroblocks in which no texture information is encoded. Therefore, the resulting SKIP mode distortion is not that of a quantization error but it is equal to the energy of the residual signal.
Chapter 1. Introduction and Overview

macroblocks which occur very frequently in H.264/AVC encoding.

The work in [36] is similar to the above mentioned work with two main distinctions: the transform coefficients of the residual source samples are assumed to follow a Cauchy distribution $f_C(x, \mu)$, and the bit-rate is estimated using the statistical entropy $H(Q_s)$ of the quantized coefficients. However, several simplifying assumptions are made in terms of the quantization rounding error, the models are based strictly on the distribution of transform coefficients, and the SKIP mode is not addressed in the derivation. The derived rate and distortion models and their simplified versions proposed in this work are shown below:

$$R_s(Q_s, \mu) = H(Q_s) = -\sum_{i=\infty}^{\infty} P(iQ_s) \log_2 (P(iQ_s)),$$

where $P(iQ_s) = \int_{(i-\frac{1}{2})Q_s}^{(i+\frac{1}{2})Q_s} f_C(x, \mu) dx$

$$D_s(Q_s, \mu) = \sum_{i=\infty}^{\infty} \int_{(i-\frac{1}{2})Q_s}^{(i+\frac{1}{2})Q_s} (x - iQ_s)^2 f_C(x, \mu) dx.$$

Simplified models: $R_s(Q_s) \approx aQ_s^{-\alpha}$, $D_s(Q_s) \approx bQ_s^\beta$,

where $a$, $b$, $\alpha$, and $\beta$ are sequence dependent parameters.

Finally, we illustrate a scalable video rate-distortion model presented in [37]. This model was derived for progressive fine granular scalable (PFGS) video coders similar to the type supported by MPEG-4 and the early version of quality scalability that was supported in SVC prior to its exclusion from the standard due to the additional complexity imposed on the decoder.

The model assumes that the residual transform coefficients follow a sum of Laplacian distribution and computes the distortion as the error in quantizing such sources. The rate is derived following the $\rho$-domain analysis in which it can be written as a linear function of the percentage of non-zero transform coefficients. This results in the following rate-distortion behavior:

$$D_s(R_s) = \sigma_x^2 - (a \log^2 R_s + b \log R_s + c) R_s,$$

where $a$, $b$, and $c$ are sequence dependent model parameters.

While the models discussed above are faithful to the assumptions taken, several new features in SVC result in a divergence from existing rate-distortion models, calling for the development of improved models that can accurately capture the runtime rate-distortion behavior. These rate and distortion models can be used in conjunction with rate control and stochastic system models to derive transmission control policies that address the requirements and constraints of wireless video transmission applications. We will undertake these tasks in Chapter 2.
1.5 Overview of Wireless Video Transmission Applications

Wireless video transmission applications can be divided into three main categories: multimedia messaging, streaming, and conversational services [39]. The most popular and widely used application to date is the multimedia messaging service in which a coded video file is downloaded completely by a client and then played by the user equipment (UE). However, with multimedia messaging, the UE requires tremendous storage capacity and the service cannot offer any real-time functionalities. Therefore, the transmitted video is usually restricted to short video clips.

The wireless services that are of interest in this work are streaming and conversational services which we will discuss in detail in what follows.

1.5.1 Streaming Services

Description

Long video presentation sessions, such as mobile TV, require streaming services to deliver the coded video streams to the mobile clients. A typical video streaming system, as described by Chou et al. [40], can be separated into three major components: a streaming server that stores pre-encoded video data, a communication channel over which the encoded video data is transmitted, and several clients with receiver buffers that can store up to a few seconds of received data and then begin playback until the end of the streaming session.

Requirements

Several streaming scenarios can be handled by the aforementioned streaming system. We will address the case of streaming scalable video data to multiple users over a wireless channel that suffers from packet loss, variable delivery delay, and limited capacity. With such a scenario in mind, four challenges arise: protection of the coded bitstream from channel losses, rate allocation for the multiple users under capacity constraint, managing the delay-sensitive video payload, and quality of service (QoS) guarantees in terms of maximum acceptable video distortion for each user.

Fig. 1.5 illustrates the block structure of our generalized video streaming system and its components. The streaming system shown in Fig. 1.5 shows a streaming server with access to \( U \) pre-encoded video streams intended for delivery to \( U \) clients/users. The server receives channel quality information from the link/MAC layer in terms of radio link control (RLC) frame loss rate, estimated channel capacity, and estimated packet delivery delay.
Chapter 1. Introduction and Overview

1.5.2 Conversational Services

Description

Video conversational services allow for real-time applications such as video conferencing and telephony. Conversational services differ from streaming services in that the video traffic is two-way: uplink and downlink. In most wireless networks, separate channels are allocated for the uplink and downlink traffic. Therefore, the problem at hand is similar to the streaming case with two exceptions: the video is encoded and transmitted in real-time, and the buffering period at the decoder is shorter compared to streaming services.

Requirements

The downlink requirements defined for streaming services apply to downlink traffic in conversational services. However, traffic controllers in conversational services require real-time rate and distortion characteristics since the video is being encoded during the transmission session. Moreover, shorter buffering periods result in tighter delay constraints compared to those of streaming services. Fig. 1.6 illustrates an example of a multi-user video conferencing system over HSPA networks.

Uplink requirements vary with the type of wireless network that supports the service. In the case of 3G-based HSPA networks, uplink/downlink traffic shaping is performed at the base station. Mobile ad hoc networks (MANETs) based on IEEE 802.11 WLAN technology, on the other hand, support less centralized control of the uplink/downlink traffic. In this thesis, we focus mainly on the resource allocation problems for centralized video streaming and conversational services.
Chapter 1. Introduction and Overview

1.5.3 Application Layer Forward Error Correction for Video Transmission

The use of application-layer FEC for multimedia transmission has been in use for some time. The effectiveness of application layer unequal error protection (UEP) for scalable video transmission can be seen in [41] and later the concept was expanded by Lee et al. in [42] to bit-error rate adaptive UEP and packet-size for improving the end-to-end video distortion. The popularity of application layer FEC has continued with Schierl et al.’s proposal to use unequal erasure protection (UXP) with the scalable extension of H.264/AVC (SVC) for wireless video broadcasting.

UXP is one example of FEC techniques that are being promoted for the protection of scalable video data from losses during transmission [43]. The effectiveness of UXP for the protection of scalable video has been demonstrated in [44]. UXP is a Forward Error Correction (FEC) scheme that is based on Reed-Solomon (RS) codes. Under UXP, video packets of a GOP are grouped together into a transmission block (TB). Each row of a TB constitutes an RS codeword of length n bytes. Of these n bytes, only k bytes are data bytes (coded video bytes) and t = n − k bytes are parity bytes, where \( \frac{n}{2} < k \leq n \). Each column of a TB is then encapsulated in an RTP packet, which enables UXP to correct up to t packet erasures. In [44], different protection classes are defined for each FGS enhancement layer, higher layers receiving less protection than lower FGS layers and the base layer. Fig. 1.7 illustrates the UXP scheme described above.

However, the UXP schemes proposed in [43],[44] take no consideration for limited channel capacity or packet delivery delays and the issue of UXP delay is very rarely addressed in the

Figure 1.6: Simplified illustration of a video conferencing system over 3G HSDPA/HSUPA network. Every device has the capability of receiving and transmitting a video stream over a downlink / uplink channel, respectively. The channel quality is reported to a central controller located at the base station which in turn dictates the transmission rates to the wireless users.
Figure 1.7: Standard UXP scheme with FGS layer-based classes. Each RTP packet is also fragmented into \( m \) RLC frames. The value of \( m \) depends on the link-layer specifications.
available literature. In fact, to the best of our knowledge, there has been no study for the packet delay and play-out deadline performed for UXP protected video streams.

1.6 Examples of Wireless Transmission Networks

The expensive provisioning of high-rate dedicated radio bearers in traditional constant bit-rate video transmission networks, such as UMTS, has led to a shift towards packet-switched shared IP bearers [45].

Modern video transmission networks such as 3G’s HSPA and IEEE 802.11-based WLANs use shared resources along with flexible transmission bit-rate and error protection adaptation mechanisms to improve the utility and reliability of the system. The transmission parameters adjusted during rate-adaptation (in 3G networks) or link-adaptation (in WLANs) are generally measured at the receiver and are delivered to the transmitter through a feedback mechanism. The transmitter uses this information to determine a specific modulation and coding scheme (MCS) for the next packet to be transmitted. Under good channel conditions, more information bits and spectrally efficient modulation schemes can be used; whereas, under bad channel conditions, link adaptation adds resilience to the transmitted signal, reducing the number of delivered information bits per symbol.

1.6.1 High Speed Packet Access

Emerging 3G systems such as HSDPA and HSUPA rely on centralized quality of service (QoS) provisioning which is achieved through several enhancements implemented at the physical (PHY) and media access control (MAC) layers. These enhancements include adaptive modulation and coding (AMC) techniques that adjust the transmission bit-rate and error protection capabilities (modulation and coding schemes - MCS) based on a channel quality indicator (CQI), hybrid automatic repeat-request (ARQ) to reduce the transmission round-trip times, and a shorter transmission time interval (TTI) that reduces the latency and supports fast scheduling decisions. These functionalities are addressed by the Node B in Fig. 1.6 which is placed closer to the base station and made cognizant of the media and service parameters.

HSDPA Throughput Performance

The system can support multiple video users by allocating different time-slots to different users, such that, time is the resource that is shared among all users. Therefore, at every time-slot, only one user can utilize the channel resources within a cell, and the temporal-share allocation $\tau_u$ of a user $u$ is regarded as the fairness requirement in the system [46]. The link/MAC layer receives feedback from each mobile device reporting the user’s signal to noise ratio (SNR) to
Chapter 1. Introduction and Overview

the base-station which estimates the channel quality. Rate adaptation is then performed by continuously adjusting the modulation and coding parameters every 2 ms, corresponding to the basic high speed downlink shared channel (HS-DSCH) transmit time interval \[47\]. The server receives channel quality information from the link/MAC layer in terms of frame error rate (FER) and estimated channel capacity for each user.

The user equipment (UE) periodically transmits its CQI on an uplink high speed dedicated physical control channel (HS-DPCCH) which also carries fast L1 based packet acknowledgment signaling (Ack/Nack) \[47\]. The HS-DSCH can provide peak data rates that range between 120 kbps and 20 Mbps as a result of variable-rate turbo coding, multi-code operation, and 16QAM modulation. Table 1.1 shows five possible transport format and resource combinations (TFRC) in an HSDPA system corresponding to 10% FER \[47\].

<table>
<thead>
<tr>
<th>TFRC</th>
<th>SNR (dB)</th>
<th>Modulation</th>
<th>Code Rate</th>
<th>Data Rate (15 Codes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-10 – 10</td>
<td>QPSK</td>
<td>1/4</td>
<td>1.8 Mbps</td>
</tr>
<tr>
<td>2</td>
<td>10 – 15</td>
<td>QPSK</td>
<td>1/2</td>
<td>3.32 Mbps</td>
</tr>
<tr>
<td>3</td>
<td>15 – 18</td>
<td>QPSK</td>
<td>3/4</td>
<td>5.3 Mbps</td>
</tr>
<tr>
<td>4</td>
<td>18 – 21</td>
<td>16QAM</td>
<td>1/2</td>
<td>7.2 Mbps</td>
</tr>
<tr>
<td>5</td>
<td>21 – 30</td>
<td>16QAM</td>
<td>3/4</td>
<td>10.7 Mbps</td>
</tr>
</tbody>
</table>

1.6.2 IEEE 802.11-based WLANs

Similar functionalities can be achieved by multi-rate WLANs such as IEEE 802.11a/n and WiMAX (IEEE 802.16) networks with slight variations in the control methodologies. These networks “operate along the same principles of dynamically and flexibly sharing the available resources among users to optimize the data throughput” \[45\]. In 802.11 networks, a user-defined link-adaptation scheme is used to adapt transmission parameters, such as forward error correction (FEC) coding rate and modulation type, in order to counter the variation in wireless link quality and maintain reliable transport services for higher layers.

802.11a/n Throughput Performance

The relationship between the PER, bit error rate (BER), and achievable PHY throughput of IEEE 802.11a/n WLANs is modeled in \[48\]. WLANs typically operate under frequency-selective quasi-static channel conditions. Exact error rate analysis of the PHY under such conditions is a complicated task (see \[49\]). The work in \[48\] bypasses an exact analysis and instead constructs a mapping between the PER and the PHY bit-rate from performance curves obtained through
system simulations under typical channel conditions. The channel BER performance curves are obtained using an elaborate MATLAB model developed at UBC for the 802.11a/n PHY, based on the work in [50] and [51]. Channel responses generated by [51] are fed into the model in [48] to obtain BER performance curves. Using BER curves of all valid combinations of modulation and coding, a look up table is constructed similar to Table 1.2. For simplicity of illustration we have only included the performance curves for the Space Time Block Coding (STBC) modes of a 2x2 802.11n MIMO system. BER curves for all STBC MCS indices for a 2x2 MIMO system can be seen in Fig. 1.8 and Table 1.2. The values of MCS are 0 to 7, with corresponding bit-rates of: \{6.5, 13, 19.5, 26, 39, 52, 58.5, 64\} Mbps.

Figure 1.8: BER performance of STBC modes of an 802.11n MIMO PHY in quasi-static fading channel, perfect channel estimation assumed [48].

1.6.3 Optimal Bit-rate Allocation for Wireless Video Transmission

The general framework adopted in the literature for optimizing the rate allocation of multiple video streams sharing a common resource can be summarized as follows. The streaming system shown in Fig. 1.5 shows a streaming server that has access to $U$ video streams intended for delivery to $U$ clients/users. Let $r_u \in [L_b u, U_b u]$ be the coded video bit-rate of user $u$ bounded below by the base-layer bit-rate $L_b u$, and above by the maximum video bit-rate $U_b u$ which includes all enhancement layers. Each coded video stream can be defined by a rate-distortion relationship $R_u(r_u), D_u(r_u)$ given by equation (1.22). The server receives channel quality information from the link/MAC layer in terms of radio link control (RLC) frame loss.
Table 1.2: Valid MCS Indices for 2x2 802.11n WLAN

<table>
<thead>
<tr>
<th>MCS</th>
<th>Bitrate (Mbps)</th>
<th>Code Rate</th>
<th>SM streams</th>
<th>Modulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>6.5</td>
<td>1/2</td>
<td>1</td>
<td>BPSK</td>
</tr>
<tr>
<td>1</td>
<td>13</td>
<td>1/2</td>
<td>1</td>
<td>QPSK</td>
</tr>
<tr>
<td>2</td>
<td>19.5</td>
<td>3/4</td>
<td>1</td>
<td>QPSK</td>
</tr>
<tr>
<td>3</td>
<td>26</td>
<td>1/2</td>
<td>1</td>
<td>16-QAM</td>
</tr>
<tr>
<td>4</td>
<td>39</td>
<td>3/4</td>
<td>1</td>
<td>16-QAM</td>
</tr>
<tr>
<td>5</td>
<td>52</td>
<td>2/3</td>
<td>1</td>
<td>64-QAM</td>
</tr>
<tr>
<td>6</td>
<td>58.5</td>
<td>3/4</td>
<td>1</td>
<td>64-QAM</td>
</tr>
<tr>
<td>7</td>
<td>64</td>
<td>5/6</td>
<td>1</td>
<td>64-QAM</td>
</tr>
</tbody>
</table>

rate \(p_{rlc}\), estimated channel capacity \(C\), and estimated packet delivery delay \(\bar{\delta}\). This information is used to assign the appropriate error protection (UXP) and video bit-rate allocation.

\[
R_u(r_u) = R_{UXP}(r_u),
D_u(r_u) = D_{enc}(r_u) + D_{loss}(r_u, p_{rlc,u}, \bar{\delta}),
C = \text{capacity},
\]

(1.22)

where \(R_{UXP}(r_u)\) is the protected video bit-rate of user \(u\), respectively, \(D_{enc}(r_u)\) is video encoder distortion, and \(D_{loss}(\cdot)\) is the distortion due to channel loss and written as a function of the video bit-rate \(r_u\), the RTP loss rate \(p_{rtp,u}\) of user \(u\), and the average packet delay across the channel \(\bar{\delta}\).

A generalized media-aware streaming policy can be generated at the server by solving the following constrained optimization problem:

\[
\begin{align*}
\min_{\pi=[r_1, r_2, \ldots, r_U]} & \sum_{u=1}^{U} D_u(r_u) \\
\text{s.t.} & \sum_{u=1}^{U} R_u(r_u) \leq C, \\
& r_u \in [Lb_u, Ub_u]
\end{align*}
\]

(1.23)

where \(D_u(\cdot)\) and \(R_u(\cdot)\) are the cost/objective and constraint functions of user \(u\), respectively, defined in (1.22). Finally, QoS conditions are satisfied by setting the lower bound \(Lb_u\) on the video bit-rate \(r_u\) such that each video user \(u\) is guaranteed the minimum video quality delivered by the base-layer of the coded scalable video stream. Existing cross layer optimization solutions are based on variants of the optimization problem defined in (1.23). In the following chapters, we will give detailed descriptions and point out the limitations of these solutions. Moreover, we will propose more effective resource allocation techniques that jointly optimize the video bit-rate, error protection, and latency control algorithms in pre-encoded and real-time streaming scenarios.
1.7 Thesis Contributions and Organization

1.7.1 Summary of Contributions

The main contributions of this thesis are summarized as follows:

- We develop analytical rate and distortion models that capture the dynamic behavior of scalable coded video content using the CGS/MGS feature in SVC. These models are based on the assumption that the DCT coefficients follow a Laplacian distribution. In the cases when the true coefficient distribution diverges from the Laplacian assumption, we propose linear scaling of the Laplacian parameter \( \lambda \) and show that such scaling results in an accurate estimate of the rate and distortion behavior.

- We present simplified rate and distortion models with parameters that can be derived from a limited number of empirical samples. We show that the simplified model performance approaches the analytically derived model performance with an acceptable loss in accuracy.

- We propose a rate-minimized unequal erasure protection (UXP) scheme based on the UXP scheme proposed by Liebl et al. [43] and adopted for SVC in [44]. The proposed rate-minimized scheme balances the tradeoff between increasing the protection overhead and reducing the UXP failure rate.

- We derive an analytical expression for packet delay and play-out deadline of UXP protected scalable video. A group of coded video frames (GOP) is embedded in a single UXP transmission block (TB) thus changing the play-out deadline of the protected packets. This modified packet sequence alters the regular transport stream structure. Therefore, we derive an upper bound for the number of link-layer (RLC) frames allowed per TB as a function of the RLC frame delivery delay.

- We derive a loss-distortion model for hierarchical predictive video coders using picture copy error concealment. A loss-distortion video model is essential for media-aware streaming policies which require extensive knowledge of the video encoding structure since it reflects the sensitivity of the decoded video stream to packet losses affecting the bitstream during transmission.

- We analyze the performance of delay-aware, capacity-aware, and rate-minimized UXP in conjunction with optimized rate-allocation compared to simpler streaming strategies. Although delay and capacity constrained rate-allocation requires more computation, the bulk of the computation workload is performed off-line and does not affect the run-time performance of the system.
Chapter 1. Introduction and Overview

- We show that a UXP protected rate-constrained optimization problem is non-convex, to which respect we propose an approximate solution using convex-relaxation and projection onto the feasible set.

- We show that the joint optimization of traffic control and link adaptation processes considerably improves the received scalable video quality of all video users and enhances the channel bandwidth utilization in a multi-user wireless video streaming system. Moreover, the joint optimization determines the optimal time-share among the multiple users that maximize the defined objective.

- We demonstrate that the joint link-adaptation and traffic-control optimization problem is a non-convex constrained optimization problem with combined discrete and continuous variables. Moreover, we propose an algorithm that first solves a continuous relaxation of the discrete optimization to find the optimal PHY rate points, and next solves for the packet drop ratios using the discrete PHY rate points that lie in the vicinity of the optimal solution of the continuously relaxed problem.

1.7.2 Thesis Organization

This is a manuscript-based thesis which follows the specifications required by the University of British Columbia for this format. In addition to this introductory chapter, the thesis includes three chapters which were originally prepared for journal publication and have been slightly modified in order to offer a logical progression in the thesis. The final chapter discusses the conclusions and directions for future work.

The remaining chapters in this thesis are organized as follows:

In Chapter 2, we derive analytical rate and distortion models for video bitstreams coded using the single layer and CGS/MGS scalable extension of H.264/AVC. We then present simplified empirical models that approximate the analytically derived model performance and demonstrate the usefulness of these models in a real-time video streaming scenario.

Chapter 3 considers a multi-user video streaming scenario with pre-encoded scalable video streams in a 3G-type environment. In this chapter, we derive a loss-distortion model for hierarchical predictive coders and an analytical expression for the packet delay and play-out deadline of UXP protected scalable video. We also analyze the performance of delay-aware, capacity-aware, and rate-minimized UXP in conjunction with optimized rate-allocation compared to simpler streaming strategies.

In Chapter 4, we present two media-aware scalable video transmission schemes over multi-rate PHY IEEE 802.11 WLANs. The first approach, called Intelligent Link Adaptation (ILA),
optimizes the allocation of scalable video enhancement layers under service time share constraints. The second approach jointly optimizes the operation of link-adaptation and a traffic control module in order to add a selective packet dropping functionality to the overall streaming system.

Finally, we draw our conclusions in Chapter 5 wrapping up the contributions of this thesis and proposing several suggestions for future research in this topic.

Note that the notations used in the different chapters are independent of each other.
Chapter 1. Introduction and Overview

1.8 References


[7] Universal Mobile Telecommunications System (UMTS); Multimedia Broadcast/Multicast Service (MBMS); Protocols and Codecs, European Telecommunications Standards Institute, March 2005.


The references included in this chapter are generic references. More specific references to the subsequent chapters will follow at the end of each chapter.
Chapter 1. Introduction and Overview


Chapter 1. Introduction and Overview


Chapter 1. Introduction and Overview


Chapter 1. Introduction and Overview


Chapter 2

Rate and Distortion Modeling of CGS/MGS Coded Scalable Video Content

2.1 Introduction

The advent of video applications into the wireless scene has called for highly media-aware coder control and streaming algorithms to manage the difficulties of wireless transmission. Capacity restrictions, heterogeneous device and network capabilities, and error prone transmissions are just some of the problems resulting from the characteristics of modern video communication systems, to which scalable video coding (SVC) offers an attractive solution [1].

Video rate and distortion modeling lies at the core of model-based coder control algorithms, whereby a coded video packet is abstracted in terms of its rate $R_s$ and the distortion $D_s$ of the resulting coded picture [2]. The rate is defined as the size in bytes of the video packet and the distortion is often measured in terms of the mean-square-error (MSE) or peak-signal-to-noise-ratio (PSNR) between the uncoded original and the reconstructed picture. Rate-distortion (R-D) models are used in real-time coder and transmission control algorithms to predict the rate and distortion of a video packet prior to the completion of the encoding process. However, scalability and encoder complexity add a variety of parameters that render traditional rate and distortion models inaccurate. For instance, the interdependency between rate-distortion optimized (RDO) motion estimation and rate-control in H.264/AVC is described as a “chicken and egg” dilemma [3]. Therefore, it is important that the encoder parameters used in rate-distortion modeling be accessible as early as possible in the encoding process. Fig. 2.1 shows a simplified block diagram of a video encoder with two possible parameter extraction points, the first right after the prediction stage, and the second after transform and quantization.

1 A version of this chapter will be submitted for publication. H. Mansour, P. Nasiopoulos, and V. Krishnamurthy, “Rate and Distortion Modeling of CGS/MGS coded Scalable Video Content.”
Chapter 2. Rate and Distortion Modeling of CGS/MGS Coded Scalable Video Content

Figure 2.1: Simplified encoder block diagram showing the possible extraction points for rate-distortion model parameters.

A more pressing complication arises from the development of scalability features in the state of the art scalable video coding standard. In SVC, quality scalability is achieved using medium-grained or coarse-grained scalability (MGS/CGS) where scalable enhancement packets deliver quality refinements to a preceding layer representation by re-quantizing the residual signal using a smaller quantization step size and encoding only the quantization refinements in the enhancement layer packets \(^1\). Moreover, MGS/CGS enhancement layer packets are coded using the key-picture concept where the coding loop is closed at the highest scalability layer for key-pictures only, while the loop is closed at the base-layer quality for the remaining pictures. This approach achieves a tradeoff between the coding efficiency of enhancement layers and the drift at the decoder \(^1\). These new features in SVC result in a divergence from existing rate-distortion models, calling for the development of improved models that can accurately capture the runtime rate-distortion behavior.

2.1.1 Main Contributions

The main contributions of this chapter can be summarized as follows:

1. We develop analytical rate and distortion models that capture the dynamic behavior of scalable coded video content using the CGS/MGS feature in SVC. These models are based on the assumption that the DCT coefficients follow a Laplacian distribution. In the cases when the true coefficient distribution diverges from the Laplacian assumption, we propose linear scaling of the Laplacian parameter \(\lambda\) to compensate for error in the assumption and show that such scaling results in an accurate estimate of the rate and distortion behavior.

2. We present simplified rate and distortion models with parameters that can be derived from a limited number of empirical samples. We show that the simplified model performance approaches the analytically derived model performance with an acceptable loss in accuracy.
3. We present a real-time joint bit-rate and error protection allocation scheme for multiple scalable video streams sharing a single downlink channel. Using the simplified rate and distortion models, we formulate a global optimization problem that minimizes the expected sum of the video frame distortions of all users by adjusting the encoding quality at the base- and enhancement-layers as well as the application layer error protection overhead used to combat packet losses.

The remainder of this chapter is organized as follows. In Section 2.2 we develop analytical rate and distortion models that capture the behavior of CGS/MGS coded scalable video bitstreams. Section 2.3 presents simplified empirical models that approximate the performance of the analytical derivations. We illustrate in Section 2.4 the benefit of using real-time scalable video modeling in a multi-user video streaming scenario. We finally draw our conclusions in Section 2.5.

2.2 Analytical Derivation of the Rate-Distortion Behavior

The objective of a video encoder is to represent a sequence of images in as few bits as possible. Modern video encoders achieve this task through four stages depicted in Fig 2.1. The first stage is Intra/Inter prediction which captures the spatial and temporal correlations resulting in a pixel domain residual signal. The second stage transforms the residual signal in order to collect the energy of the signal in fewer samples than the pixel domain residual signal, resulting in a sparser representation of the residual. The third stage quantizes the transform coefficients to reduce its dynamic range. Finally, a fourth stage translates the quantized coefficients into a binary bitstream through entropy coding.

2.2.1 Distribution of Residual Samples

The purpose of rate and distortion models is to estimate the number of bits required to encode a video frame and the corresponding error that results from the lossy encoding process. This encoding error arises from the quantization process and other rounding errors that are performed on the transform coefficients. We follow an approach to rate-distortion modeling based on the statistical distributions of the residual signal and its transformed coefficients.

Several models have been presented in the literature that attempt to model the residual transform coefficients [2][4-6]. These include the generalized Gaussian, Laplacian, and Cauchy distributions. The most widely used distribution is the Laplacian which we will adopt in this work for its accuracy, simplicity, and analytical tractability. Moreover, our experiments have shown that the distribution of pixel domain residual samples can also be approximated by the Laplacian distribution, an observation supported by [6].
Let $f_X(x, \lambda)$ be the zero-mean Laplacian density function given by
\[
f_X(x, \lambda) = \frac{1}{2\lambda} \exp \left( -\frac{|x|}{\lambda} \right),
\] (2.1)
where $\lambda$ is a parameter that fully determines the Laplacian distribution. We will use the maximum likelihood (ML) estimator to find the value of $\lambda$ that fits our data. Given a sample data set $X$ of size $N$, where $x_i \in X, i = 1, 2, \cdots N$, the ML estimator of $\lambda$ is given by
\[
\lambda = \frac{1}{N} \sum_{i=1}^{N} |x_i|.
\] (2.2)

It is easy to see from (2.2) that the ML of $\lambda$ corresponds to the residual mean absolute difference (MAD) in the case of the pixel domain samples and the sum of absolute transform difference (SATD) in the case of the transform domain samples.

Fig. 2.2 demonstrates the accuracy of the ML Laplacian fit to the histogram of selected Intra- and Inter- coded frames from the Foreman sequence.

### 2.2.2 Base-layer Models

We start by deriving the rate and distortion models for base-layer frames encoded using an H.264/AVC compliant encoder.
Quantization Step Size Calculation

Let $X$ be a 4x4 block of pixels in the residual frame $I_r$ obtained after Intra/Inter prediction. $X$ is then subjected to the integer cosine transform defined in the H.264/AVC standard which results in the transformed block $\hat{X}$ such that,

$$\hat{X} = H X H^T,$$

where $H$ is the forward integer transform defined in (1.3).

The transform block is then quantized using the process described in (1.10) which can be re-written as follows:

$$\hat{X}_q = \text{sign}(\hat{X}) \left\lfloor \frac{\mid \hat{X} \mid}{M(Q_m)2^{-15-2Q_e} + f} \right\rfloor,$$
$$\hat{X}_r = \hat{X}_q \cdot S(Q_m)2^{Q_e},$$

where $\cdot$ is the Hadamard product which denotes elementwise multiplication, $Q$ is the quantization parameter, $Q_m = Q \mod 6$, and $Q_e = \lfloor Q/6 \rfloor$. Note that in the above equation and in the remainder of this thesis, the addition of a scalar to a matrix denotes adding the scalar to each element of the matrix.

Recall that the quantization and reconstruction matrices are given by the following relationship:

$$M(Q_m) \cdot S(Q_m) \cdot V \approx 2^{21},$$

where $V$ is the matrix defined in (1.7). Then combining (2.4) and (2.5) we get the following quantization and true reconstruction equations

$$\tilde{\hat{X}}_q = \text{sign}(\tilde{\hat{X}}) \left\lfloor \frac{\mid \tilde{\hat{X}} \mid}{V S(Q_m)2^{Q_e-6} + f} \right\rfloor,$$
$$\tilde{\hat{X}} = \tilde{\hat{X}}_q \cdot V S(Q_m)2^{Q_e-6},$$

where $\tilde{\hat{X}}$ is the true reconstruction of the quantized coefficient block $\hat{X}_q$. It can be seen from (2.4), (2.6), and (1.12) that

$$\tilde{\hat{X}} = \hat{X}_r \cdot V 2^{-6},$$

which is a natural relationship since $H_i$ is not the true inverse but a scaled inverse of the forward transform $H$.

Consequently, (2.6) shows that the quantization step size $Q_s$ of Eq. (1.8) employed in the H.264/AVC quantization process is a 4x4 matrix given by

$$Q_s = V \cdot S(Q_m)2^{Q_e-6},$$
such that, different transform coefficients are subjected to a different quantization step size. For example, a quantization parameter value \( Q = 26 \) results in the following quantization step matrix

\[
Q_s = \begin{bmatrix}
52 & 80 & 52 & 80 \\
80 & 125 & 80 & 125 \\
52 & 80 & 52 & 80 \\
80 & 125 & 80 & 125
\end{bmatrix}.
\]

\[\text{(2.9)}\]

**Distortion Model**

The video distortion \( D_s \) of a video frame is measured in terms of the MSE between the original picture and its reconstruction in the pixel domain. This can be expressed as the expected value of the \( \ell_2 \) norm of the encoding noise as shown below:

\[
D_s = E \left[ \|X - \hat{X}\|_2^2 \right] = D_c + D_{\text{SKIP}},
\]

where \( \hat{X} \) is the decoded video signal in the pixel domain, \( D_c \) is the distortion due to quantizing the transformed residual samples, and \( D_{\text{SKIP}} \) is the distortion due to SKIP macroblocks.

The encoding distortion results from the quantization error and associated integer rounding operations performed in the transform domain. To estimate the transform domain distortion \( \hat{D}_c \), we compute \( q \)-domain error of quantizing a Laplacian distributed source signal with a quantization step \( Q_s \) and a dead-zone parameter \( f \). This error is estimated by the expected value of the \( \ell_2 \) norm of the quantization noise \( \|\hat{X} - \tilde{X}\|_2^2 \).

Let \( u \) and \( v \) be the row and column indices in a 4x4 block, and let \( \mathbb{P}(iQ_s(u,v)) \) be the probability of having a quantized coefficient value equal to \( iQ_s(u,v) \). Assuming that every transform coefficient \( \hat{X}(u,v) \) is sampled from a Laplacian distribution with parameter \( \hat{\lambda} \), the quantization error \( D_c(q_s,\hat{\lambda}) \) is given by the following expression:

\[
\hat{D}_c(q_s,\hat{\lambda}) = E \left[ \|\hat{X} - \tilde{X}\|_2^2 \right] = \sum_{i=-\infty}^{\infty} \mathbb{P}(iQ_s)(\hat{X} - iq_s)^2
\]

\[
= 2\int_0^{(1-f)q_s} x^2 f_X(x,\hat{\lambda}) dx + 2\sum_{i=1}^{\infty} \int_{(i-1)q_s}^{(i+1-f)q_s} (x - iq_s)^2 f_X(x,\hat{\lambda}) dx,
\]

\[
= 2\hat{\lambda}^2 + \left(2f - 1\right)q_s^2 - 2\hat{\lambda}q_s \sum_{i=1}^{\infty} e^{-\left((i-1)fq_s/\hat{\lambda} \right)},
\]

\[\text{(2.11)}\]

where \( q_s = Q_s(u,v) \) is the quantization step corresponding to the location of the transform coefficient in the 4x4 block since \( Q_s \) is a non-uniform matrix, and \( f_X(x,\hat{\lambda}) \) is the Laplacian
density function defined in (2.1). The value of \( f \) is equal to \( \frac{1}{3} \) for Intra-coded blocks and \( \frac{1}{6} \) for Inter-coded blocks [7]. Note that in the above formulation, the transform coefficient distribution varies with the location of the coefficient in the 4x4 matrix \( \hat{X} \). Therefore, the transform domain quantization error would be equal to the average of the quantization errors of each coefficient as shown below:

\[
\hat{D}_c(Q_s) = \frac{1}{16} \sum_{u=1}^{4} \sum_{v=1}^{4} \hat{D}_c(Q_s(u, v), \hat{\lambda}(u, v)).
\]  

(2.12)

The distortion calculated in (2.11) reflects the error in quantizing an integer transform coefficient \( \hat{X} \). For an energy preserving transform, such as the regular DCT, the quantization error of transform coefficients would be equal to the error exhibited by the reconstructed signal. However, the integer transform defined in H.264/AVC is not energy preserving since it is an integer approximation of the DCT. In [8], the authors find a scaling matrix \( W \) which compensates for the different norms of the forward and scaled-inverse integer transforms. To derive the matrix \( W \), let

\[
G = H^{-1}T H^{-1},
\]

then

\[
W = \sqrt{GEG^T} = \begin{bmatrix}
1/4 & 1/\sqrt{40} & 1/4 & 1/\sqrt{40} \\
1/\sqrt{40} & 1/10 & 1/\sqrt{40} & 1/10 \\
1/4 & 1/\sqrt{40} & 1/4 & 1/\sqrt{40} \\
1/\sqrt{40} & 1/10 & 1/\sqrt{40} & 1/10
\end{bmatrix},
\]  

(2.13)

where \( E \) is a 4x4 matrix of 1s. Consequently, the pixel domain encoder distortion can be calculated using the transform domain quantized coefficients according to the following:

\[
E \left[ \| X - \hat{X} \|_2^2 \right] = E \left[ \| \left( \hat{X} - \hat{\hat{X}} \right) \cdot W \|_2^2 \right] \\
= \sum_{i=-\infty}^{\infty} \int_{(i-f)Q_s}^{(i+1-f)Q_s} \left( (\hat{X} - iQ_s) \cdot W \right)^2 f_\hat{X}(x, \hat{\lambda}) dx \\
= \sum_{i=-\infty}^{\infty} \int_{(i-f)Q_w}^{(i+1-f)Q_w} \left( \hat{Y} - iQ_w \right)^2 f_{\hat{Y}}(y, \hat{\lambda}) dy,
\]  

(2.14)

where \( i = \left\lfloor \frac{\hat{X}}{Q_s} + f \right\rfloor = \left\lfloor \frac{\hat{X}W}{Q_w} + f \right\rfloor \), \( \hat{Y} = \hat{X} \cdot W \) are the newly scaled transform coefficients, \( Q_w = Q_s \cdot W \) is a uniform quantization matrix, and \( f_{\hat{Y}}(y, \hat{\lambda}) \) is the density function of the scaled transform coefficients \( \hat{Y} \).

The formulation in (2.14) simplifies the distortion calculation in (2.11) in that all scaled transform coefficients belong to the same distribution \( f_{\hat{Y}}(y, \hat{\lambda}) \) and are subjected to the same quantization step \( Q_w \). Moreover, the scaled transform coefficients approximate the true DCT coefficients of the residual samples which leads to the conclusion that the transform domain quantization error calculated in (2.14) is equal to the pixel domain distortion.
Model parameter estimation using pixel domain samples: The distortion model derived above assumes that the DCT transform coefficients are Laplace distributed. However, the true distribution of coefficients is closer to the generalized Gaussian distribution with a shape parameter < 1. Therefore, to compensate the error in the Laplacian assumption and estimate the model parameter using pixel domain residual samples, we estimate the Laplacian parameter $\hat{\lambda}_Y$ as a linear function of the pixel domain parameter as follows:

$$\hat{\lambda}_Y = a\lambda_X + b,$$

where $\lambda_X$ is the pixel-domain Laplacian distribution parameter, and $a$ and $b$ are sequence dependent constants that can be calculated from a limited number of frames and updated using linear regression during the encoding process.

Fig. 2.3 illustrates the performance of the proposed distortion model compared to actual encoding of the six reference video sequences: Foreman, Bus, Football, City, Soccer, and Mobile.

Figure 2.3: Performance of the derived distortion model for six reference video sequences encoded at QPs of 38, 32, and 26.
Chapter 2. Rate and Distortion Modeling of CGS/MGS Coded Scalable Video Content

Bit-Rate Model

The bit-rate of a coded video frame is equal to the number of bits required to code the header and control information plus the number of bits required to encode the quantized levels of the transform coefficients. Let \( R_s \) be the size in bits of a coded video frame, then for a quantization step size \( Q_w \), \( R_s \) is expressed as:

\[
R_s(Q_w) = R_c(Q_w) + R_h,
\]

where \( R_c \) and \( R_h \) are the number bits required to code the texture and header information, respectively.

In [2], a model is proposed to estimate the video frame header size as a linear function of the number of zero and non-zero motion vectors. Other models have used the CAVLC coding assumptions to make the same estimation. We will not address the header bit-rate estimation problem since existing models are accurate enough. Our focus will be on estimation of the bit-rate of quantized transform coefficients of the residual signal.

Previous works that estimate the rate of encoding the quantized transform coefficients \( R_c(Q_w) \) have modeled the bit-rate as a linear function of the \( \ell_0 \) and \( \ell_1 \) norms \([6] [2]\). These norms are calculated using the distribution of integer transform coefficients. In \([5]\), the authors calculate the entropy of the quantized coefficients assuming a Cauchy distributed source. Since H.364/AVC uses a highly sophisticated entropy coder such as CABAC, we will also use the entropy of the quantized transform coefficients to model the residual bit-rate. The distinction in our work lies in our choosing a Laplacian distribution to model the scaled transform coefficients. Moreover, the use of entropy to model the residual bit-rate reduces the number of model parameters to those used to estimate the scaled transform coefficient MAD \( \hat{\lambda}_Y \).

Let \( iQ_w \) be the quantized coefficient value with probability \( \mathbb{P}(iQ_w) \) which is calculated as follows:

\[
\mathbb{P}(iQ_w) = \frac{(i+1-f)Q_w}{(i-f)Q_w} f_Y(y, \hat{\lambda}_Y)dy = \begin{cases} 
1 - e^{-(1-f)Q_w/\hat{\lambda}_Y}, & i = 0 \\
\frac{1}{2} e^{-(|i|-f)Q_w/\hat{\lambda}_Y} \left( 1 - e^{-Q_w/\hat{\lambda}_Y} \right), & i = \pm 1, \pm 2, \ldots 
\end{cases}
\]

Then the entropy of the quantized coefficients \( H(Q_w) \) for a quantization step \( Q_w \) is given by

\[
H(Q_w) = - \sum_{i=-\infty}^{\infty} \mathbb{P}(iQ_w) \log_2(\mathbb{P}(iQ_w)) \\
= -(1 - e^{-(1-f)Q_w/\hat{\lambda}_Y}) \log_2(1 - e^{-(1-f)Q_w/\hat{\lambda}_Y}) \\
- e^{-(1-f)Q_w/\hat{\lambda}_Y} \left[ \log_2(1 - e^{-Q_w/\hat{\lambda}_Y}) - 1 + \frac{Q_w}{\hat{\lambda}_Y \ln 2} \left( f - \frac{1}{1-e^{-Q_w/\hat{\lambda}_Y}} \right) \right]
\]

(2.18)
Chapter 2. Rate and Distortion Modeling of CGS/MGS Coded Scalable Video Content

The quantized coefficient entropy $H(Q_w)$ determines the number of bits/pixel required to encode a source signal with a distribution $f_{\hat{Y}}(y, \lambda_{\hat{Y}})$. Therefore, the coded frame size is given by the sum of the coded residual frame bits $R_cQ_w$ and the header bits $R_h$

$$R_s(Q_w) = NH(Q_w) + R_h,$$  \hspace{1cm} (2.19)

where $N$ is the number of pixels per video frame. Note that the rate estimation using the quantized signal entropy assumes that the entropy encoder employed by the video encoder can meet these requirements. Fortunately, the CABAC entropy coder supported by H.264/AVC meets these requirements. Moreover, the CABAC coder takes advantage of the dependency between adjacent samples, a feature which is not supported in our entropy calculation. The entropy function derived above assumes that the quantized coefficients are coded independently which results in an upper bound on the real frame bit-rate. Therefore, to compensate for the error in this estimation, we further subject the entropy calculation to linear scaling to match the actual coded video frame rate. This linear scaling is sequence dependent but remains constant over the entire duration of each video sequence.

Fig. 2.4 illustrates the accuracy of the rate model for the six different reference sequences.

![Figure 2.4: Performance of the derived rate model for six reference video sequences encoded at QPs of 38, 32, and 26.](image)
Comparison with Existing Models

We will compare the performance of our proposed models with that of existing models described in Chapter 1. Of these, we are interested in the real-time rate and distortion models proposed in [2, 5, 6].

These models illustrate three directions to rate-distortion modeling. In [2], the SATD of the transform coefficients as well as the quantization step size are the two main parameters used in the modeling. We will refer to these models by SATD. In [5], the transform coefficients are assumed to be Cauchy distributed. We will refer to these models as Cauchy. In [6], the transform coefficients are assumed to follow a Laplacian distribution as shown in (1.19). The distortion is derived as the error in quantizing a Laplacian distributed source using a scalar quantizer which parallels our derived model. The bit-rate is estimated as a weighted sum of the \(\ell_0\) and \(\ell_1\) norms of the quantized transform coefficients given by \(N_{nz}\) and \(E_{QTC}\), respectively.

Our derivations have shown that the entropy of a quantized Laplacian source derived in (2.18) includes the \(N_{nz}\) and \(E_{QTC}\) terms as shown below:

\[
H(Q_w) = -(1 - e^{-(1-f)Q_w/\lambda_Y}) \log_2(1 - e^{-(1-f)Q_w/\lambda_Y}) \\
- e^{-(1-f)Q_w/\lambda_Y} \left[ \log_2(1 - e^{-Q_w/\lambda_Y}) - 1 \right] + N_{nz} f \ln 2 + E_{QTC}(1 - e^{-Q_w/\lambda_Y}).
\]

Therefore, the models presented in [6] are simplified versions of our derived models where the simplification lies in the rate estimation and the assuming a fixed rounding operator \(f = 1/6\).

Table 2.1 shows the root mean squared error (RMSE) of the rate and distortion estimation between the actual video bit-rate and PSNR, and the estimated bit-rate and PSNR given by our proposed model, the SATD model, and the Cauchy model. We have excluded the Cauchy model in the distortion comparison because our regeneration of the results for distortion estimation presented in [5] did not reflect the results reported in the mentioned paper.

The results shown in Table 2.1 emphasize the accuracy of our proposed models especially with distortion estimation. The minimum RMSE value for each sequence is shown in bold font. It can be seen from the table that our proposed rate model has a similar, if not slightly better performance, than the SATD rate model. Meanwhile, our proposed distortion model outperforms SATD in all cases. It is worth mentioning that one advantage of our proposed models is that they are based on pixel-domain statistics whereas the SATD models are based on transform domain statistics which require additional computational complexity.

Concerning the Cauchy distribution based models, we have found several derivation errors made in [5]. These include finding the closed form solutions of Cauchy based entropy and the Laplacian based entropy shown in that paper. These errors might explain the reported improvement over the Laplacian based. However, our experiments have shown that the Cauchy-
Table 2.1: Comparison of Rate (Size in Bits) and Distortion (PSNR) Estimation in terms of RMSE

<table>
<thead>
<tr>
<th>Sequence</th>
<th>QP</th>
<th>Rate</th>
<th>Distortion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Proposed SATD</td>
<td>Cauchy</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Foreman</td>
<td>38</td>
<td>193.8</td>
<td>238.5</td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>336.7</td>
<td>329.1</td>
</tr>
<tr>
<td></td>
<td>26</td>
<td>415.6</td>
<td>647.5</td>
</tr>
<tr>
<td>Bus</td>
<td>38</td>
<td>286.1</td>
<td>103.8</td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>322.1</td>
<td>320.1</td>
</tr>
<tr>
<td></td>
<td>26</td>
<td>393.4</td>
<td>1130.4</td>
</tr>
<tr>
<td>Football</td>
<td>38</td>
<td>438.8</td>
<td>414.4</td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>723.6</td>
<td>1023.4</td>
</tr>
<tr>
<td></td>
<td>26</td>
<td>870.6</td>
<td>1355.9</td>
</tr>
<tr>
<td>City</td>
<td>38</td>
<td>48.5</td>
<td>64.7</td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>359.3</td>
<td>121.7</td>
</tr>
<tr>
<td></td>
<td>26</td>
<td>148.3</td>
<td>192.8</td>
</tr>
<tr>
<td>Soccer</td>
<td>38</td>
<td>255.5</td>
<td>292.2</td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>517.3</td>
<td>2159.2</td>
</tr>
<tr>
<td></td>
<td>26</td>
<td>705.4</td>
<td>902.7</td>
</tr>
<tr>
<td>Mobile</td>
<td>38</td>
<td>175.1</td>
<td>128</td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>711</td>
<td>367</td>
</tr>
<tr>
<td></td>
<td>26</td>
<td>645.1</td>
<td>13274</td>
</tr>
</tbody>
</table>

based models still failed to outperform our proposed models even after correcting the Cauchy derivations. Moreover, the simplified Cauchy models presented in [5] are an exact match to the SATD models presented in [2].

### 2.2.3 Enhancement-layer Models

In CGS/MGS scalable coding mode, a base layer representation of the video sequence is encoded at a QP value $Q_1$ and an enhancement layer representation is coded at a QP value $Q_2$, such that, $Q_2 < Q_1$. Assuming that no refinement motion vectors are encoded in the enhancement layer, then the un-coded pixel-domain residual samples $X$ and their corresponding transform coefficients $\hat{Y}$ are the same for both layers.

Let $\hat{Y}_1$ be the reconstructed transform coefficients of the base-layer representation after quantization with a step size $Q_{w,1}$. The enhancement layer representation of the same frame contains the quantization of the base layer quantization-noise signal ($\hat{Y} - \hat{Y}_1$) using the step size $Q_{w,2}$, where $Q_{w,2} < Q_{w,1}$. 

43
Distortion Model

Let $\hat{Y}_2$ be the reconstructed transform coefficient value after decoding the base and enhancement layer representations of a video frame. Then the enhancement layer quantization noise $\hat{D}_{c,2}$ is given by the following expression:

$$\hat{D}_{c,2}(Q_w,1,\hat{\lambda}) = \|\hat{Y} - \hat{Y}_2\|_2^2 = \|\hat{Y} - (iQ_w,1 + jQ_w,2)\|_2^2,$$

(2.21)

where $i$ and $j$ are the quantization levels of the base and enhancement layer representations, respectively.

Assume that the scaled transform coefficients follow the Laplacian distribution $f_{\hat{Y}}(y,\hat{\lambda}\hat{Y})$. Then the enhancement layer distortion can be estimated as the expected value of the $\ell_2$ norm of the refinement quantization noise shown in (2.21). This can be calculated as follows:

$$\hat{D}_{c,2}(Q_w,1,\hat{\lambda}) = \sum_{i=-\infty}^{\infty} \mathbb{P}(iQ_w,1) \sum_{j=-\infty}^{\infty} \mathbb{P}(jQ_w,2|iQ_w,1)(\hat{Y} - jQ_w,2 - iQ_w,1)^2,$$

(2.22)

where $\mathbb{P}(iQ_w,1)$ is the base layer quantized coefficient probability given in (2.17), and $\mathbb{P}(jQ_w,2|iQ_w,1)$ is the enhancement layer quantized refinement probability conditioned on the base layer quantization and expressed as follows:

$$\mathbb{P}(jQ_w,2|iQ_w,1) = \int_{(j-f)Q_w,2+iQ_w,1}^{(j+1-f)Q_w,2+iQ_w,1} f_{\hat{Y}}(y,\hat{\lambda}\hat{Y}) dy.$$

(2.23)

Consequently, the closed form of the enhancement layer distortion is expressed as follows:

$$\hat{D}_{c,2}(Q_w,1,\hat{\lambda}) = \hat{D}_{c}(Q_w,2,\hat{\lambda}) \left[ 1 - e^{-(1-f)Q_w,1/\hat{\lambda}\hat{Y}} \frac{1 - e^{-Q_w,1/\hat{\lambda}\hat{Y}}}{1 - e^{-2Q_w,1/\hat{\lambda}\hat{Y}}} \right],$$

(2.24)

where $\hat{D}_{c}(Q_w,2,\hat{\lambda})$ is the single layer distortion derived in (2.14) for a quantization step size $Q_w,2$.

Fig. 2.3 illustrates the PSNR performance of the proposed distortion model for the CGS layer compared to actual encoding of the six reference video sequences.

Rate Model

Following the analysis for the base-layer bit-rate, the enhancement layer bit-rate is the sum of the enhancement layer header bits plus the number of bits required to encode the residual signal.
Rate and Distortion Modeling of CGS/MGS Coded Scalable Video Content

Chapter 2. Rate and Distortion Modeling of CGS/MGS Coded Scalable Video Content

Figure 2.5: Performance of the derived scalable distortion model for the enhancement layer distortion of six reference video sequences with a base layer encoded at a QP = 38 and CGS enhancement layer encoded at a QP = 32.

quantization refinement. We estimate the enhancement layer quantization refinement using the conditional entropy $H(Q_{w,2}|Q_{w,1})$ of the quantized base-layer quantization noise ($\hat{Y} - \tilde{Y}_1$). The combined stream (base layer + CGS) bit-rate can then be written as a function of the joint entropy $H(Q_{w,2},Q_{w,1})$ shown below:

$$H(Q_{w,2},Q_{w,1}) = H(Q_{w,1}) + H(Q_{w,2}|Q_{w,1}),$$

(2.25)

where $H(Q_{w,1})$ is the base layer quantized coefficient entropy derived in (2.18). In what follows, we derive the conditional entropy term.

$$H(Q_{w,2}|Q_{w,1}) = -\sum_{i=-\infty}^{\infty} \mathbb{P}(iQ_{w,1}) \sum_{j=-\infty}^{\infty} \mathbb{P}(jQ_{w,2}|iQ_{w,1}) \log_2(\mathbb{P}(jQ_{w,2}|iQ_{w,1}))$$

$$= 1 - e^{-(1-f)Q_{w,1}/\lambda} \left[ H(Q_{w,2}) + \frac{Q_{w,1}}{\lambda} \ln 2 \frac{1-e^{-(3-f)Q_{w,1}/\lambda}}{1-e^{-2Q_{w,1}/\lambda}} \right],$$

(2.26)

where $H(Q_{w,2})$ is the single layer quantized coefficient entropy in (2.18) evaluated at the quantization step $Q_{w,2}$. 

45
Therefore, the total scalable frame bit-rate can now be written as follows:

\[ R_s(Q_{w,1}, Q_{w,2}) = N \left( H(Q_{w,1}) + H(Q_{w,2} | Q_{w,1}) \right) + R_{h,1} + R_{h,2}. \]  

(2.27)

Fig. 2.6 illustrates the accuracy of the rate model for the six different reference sequences.

Figure 2.6: Performance of the derived scalable rate model for six reference video sequences with the base layer encoded at QP = 38 and CGS enhancement layer encoded at QP = 32.

### 2.3 Empirical Modeling of the Rate-Distortion Behavior

In this section, we propose simplified rate and distortion models that are derived from empirical data estimates. In these models, we use the quantization parameter (QP) and the mean absolute difference (MAD) of the residual signal \( \lambda(Q_{P_0}) \), obtained at extraction point (1) of Fig. 2.1, given an initial quantization parameter value \( Q_{P_0} \). These empirical models can accurately estimate the following:

1. The MAD of the residual signal \( \lambda(Q_{P_i}) \) at any target \( Q_{P_i} \) different from \( Q_{P_0} \). The value of \( Q_{P_0} \) can be set to the minimum acceptable video quality guaranteed by the QoS. The prediction error \( \lambda(Q_{P_i}) \) is then used to estimate the rate and distortion of the coded
2. The decoded picture distortion measured in terms of the luminance PSNR (Y-PSNR) at any target $QP_t$. The distortion $D$ is expressed as a function of the estimated prediction error $\lambda(QP_t)$ and the quantization parameter $QP_t$.

3. The coded packet rate measured in terms of the size in bytes of the MGS enhancement layer packets at any target $QP_t$. For base layer packets, the traditional quadratic rate model and its linear approximation used for rate control in [9] remains valid. For MGS enhancement packets, we have found that the rate $R$ is a linear function of the enhancement layer distortion $D$.

### 2.3.1 Prediction Error Estimation Model

In H.264/AVC and base-layer SVC, the macroblock (MB) prediction mode is chosen after a rate-distortion optimization (RDO) process, such that the chosen mode minimizes the following Lagrangian cost function

$$J(m, Q_s) = D(m, Q_s) + \mu(Q_s)R(m, Q_s), \quad (2.28)$$

where $m$ is the macroblock coding mode, $Q_s$ is the quantization step size, $D$ is the prediction distortion measured in terms of the sum of absolute differences (SAD) or sum of squared differences (SSD) between the original and prediction signals, $R$ is the number of bits required to encode the MB using the specific mode, and $\mu$ is the Lagrange multiplier given as a function of QP [10].

The rate control algorithm used in H.264/AVC allows for the estimation of the prediction MAD $\tilde{\lambda}$ of a frame using the following linear model:

$$\tilde{\lambda}_t = \alpha \tilde{\lambda}_{t-1} + \beta, \quad (2.29)$$

where $\tilde{\lambda}_{t-1}$ is the actual MAD of the previous frame, and $\alpha$ and $\beta$ are model parameters that are initialized to 1 and 0, respectively. These parameters are updated by a linear regression method similar to that of MPEG-4 Q2 after coding each picture [9].

The model described in (2.29) is sufficient when both the current and previous frames have the same QP value. However, the QP values of real-time encoded MGS/CGS video frames change in both the base and enhancement layers. Let $q_t, q_{t-1} \in \{0, 1, \ldots, 51\}$ be the current and previous QP values, respectively. It is evident from (2.28) that the resulting residual signal, and consequently the residual MAD, cannot be calculated before the completion of the RDO process which is a function of QP. Consequently, encoders would have to perform a new RDO process.
to update $\lambda(q_{t-1})$ for different values of $q_t$, which is very costly in terms of computational resources and encoding delay. The relationship that estimates $\tilde{\lambda}(q_t)$ given the residual MAD at an initial $q_{t-1}$ is shown below:

$$\tilde{\lambda}(q_t) = \tilde{\lambda}(q_{t-1})2^{a(q_t-q_{t-1})},$$

(2.30)

where $a$ is a model parameter typically valued around 0.07 for most sequences. The model shown in (2.30) allows an encoder to accurately estimate the prediction error at different QP values using only a single prediction run at an initial $q_{t-1}$. The estimated $\lambda(q_t)$ is then used in the rate and distortion estimation models. As a result, optimal encoding can be performed using a total of only two prediction runs, the first run at the initial $q_{t-1}$ and the second run at the optimal target $q_t$.

### 2.3.2 Distortion Model

The Laplacian-based distortion model derived in (2.11) can be simplified by observing the variation of $D_s$ as a function of $\lambda$ and QP independently. Our observations have shown that we can estimate the MSE as a linear function of $\lambda$ and an exponential function of QP as shown in Fig. 2.7.

![Figure 2.7](image-url)

Figure 2.7: Relationship of the derived MSE with respect to the parameter $\lambda$ and the quantization parameter QP. The figures show a linear relationship of the MSE with $\lambda$ and an exponential relationship with the QP.

As a result, we approximate the distortion measure using the following expression:

$$D_s(\lambda, QP) = a\lambda e^{bQP+c},$$

(2.31)
where \(a\), \(b\), and \(c\) are sequence dependent constants.

Let \(q_b\) and \(q_e\) be the base and enhancement layer QP values, respectively, and let \(\tilde{\lambda}_b = \tilde{\lambda}(q_b)\) and \(\tilde{\lambda}_e = \tilde{\lambda}(q_e)\) be the respective prediction MAD estimates. For an H.264/AVC coded video stream and H.264/AVC compliant base-layer SVC stream, we propose the following distortion expression:

\[
D_b(\tilde{\lambda}, q_b) = b_1 \log_{10} \left( (\tilde{\lambda}_b)^r + 1 \right) \cdot q_b + b_2, \tag{2.32}
\]

where \(r\) is a value that depends on the frame type, \(b_1\) and \(b_2\) are sequence-dependent model parameters typically valued at 0.52 and 47, respectively. The parameters \(b_1\) and \(b_2\) can be refined for each sequence during encoding. The value of \(r\) depends on the frame type, such that, \(r = 1\) for Inter-coded frames and \(r \approx \frac{5}{6}\) for Intra-coded frames.

For MGS/CGS scalability in SVC, the term \(\tilde{\lambda}(q_b)\) in (2.32) is simply replaced by the estimate of the quality refined prediction error \(\tilde{\lambda}(q_e)\), thus resulting in the following MGS/CGS distortion model:

\[
D_e(\tilde{\lambda}, q_e) = b_1 \log_{10} \left( (\tilde{\lambda}_e)^r + 1 \right) \cdot q_e + b_2, \tag{2.33}
\]

where \(\tilde{\lambda}_e = \tilde{\lambda}_b 2^{a(q_e - q_b)}\).

### 2.3.3 Rate Model

In this section, we present the rate model for base and MGS enhancement layer SVC coded video frames. The presented model is based on the work proposed in [2] which we extended to incorporate MGS/CGS scalability. The H.264/AVC compatible base layer in SVC can be expressed as follows:

\[
R_b(\tilde{\lambda}, Q_b) = c_1(\tilde{\lambda}_b)^r/Q_b, = c_2(\tilde{\lambda}_b)^r2^{-q_b/6} \tag{2.34}
\]

where \(c_1\) is a model parameter, \(Q_b\) is the quantization step size at the base layer, and \(r\) is the same power factor described in (2.32). Note that instead of the lookup table conversion from quantization parameter to quantization step described in the H.264/AVC standard, we approximate the conversion as follows:

\[
Q = 0.625 \cdot 2^{q_b/6}.
\]

We extend the model in (2.34) to incorporate MGS/CGS packets which only contain refinements on the quantization of residual texture information [1]. Therefore, we express the
Chapter 2. Rate and Distortion Modeling of CGS/MGS Coded Scalable Video Content

enhancement layer bit-rate as follows:

\[ R_e(\tilde{\lambda}, q_b) = c_3(\tilde{\lambda})^{r_2 2^{-q_e/6}}. \]  \hspace{1cm} (2.35)

To better illustrate the performance of the simplified models we show model behavior for the Foreman, Bus, Soccer, and Mobile sequences in Fig. 2.8.

![Figure 2.8: Comparison between the modeled distortion and rate estimates and the actual data.](image)

2.4 Real-Time Joint Rate and Protection Allocation for Multi-User Scalable Video Streaming

To demonstrate the effectiveness of the above models for video transmission, we present a real-time joint bit-rate and error protection allocation scheme for multiple scalable video streams sharing a single downlink channel. High speed downlink packet access (HSDPA) systems allow for multiple live video streams to share a common downlink channel among multiple mobile users. However, the unreliable nature of the wireless link results in packet losses and fluctuations in the available channel capacity. This calls for flexible error protection and rate control strategies implemented at the video encoders that can respond to the variation in channel conditions.
Figure 2.9: Schematic diagram of the live multi-user scalable video streaming environment. The mobile users share the same downlink channel. The channel SNR is reported to the streaming server which accordingly adjusts the encoding quality of the scalable video sources to satisfy the corresponding throughput constraint.

2.4.1 System Overview

We consider a live multi-cast streaming setup over high speed downlink packet access (HSDPA) networks. Let \( U \) be the number of video sources intended for delivery to \( U \) mobile users/clients. Fig. 2.9 illustrates the general structure of the multi-user scalable video streaming system described above. We assume that the video sources have quality scalable encoders, such as the CGS/MGS scalability features of SVC [11], with the capability of adjusting the encoding quality of the base and enhancement layers at every video frame instant. The encoded video streams are transmitted over a shared wireless link with varying quality reflected in terms of the channel signal to noise ratio (SNR). A streaming server receives channel feedback in terms of channel throughput \( C \) and packet loss ratio \( p_c \), and accordingly allocates for each source encoder the target base and enhancement layer bit-rates, as well as the level of application layer error protection needed to combat the packet losses encountered during transmission.

Channel Model

We consider a wireless 3G system similar to HSDPA in which the physical layer can choose between several modulation and coding schemes (MCS) in order to adapt the transmission according to channel quality information (CQI) received from the mobile device. The system can support multiple video users by allocating different time-slots to different users, such that, time is the resource that is shared among all users. Therefore, at every time-slot, only one user can utilize the channel resources within a cell, and the temporal-share allocation \( \tau_u \) of user \( u \)
is regarded as the fairness requirement in the system. The link/MAC layer receives feedback from each mobile device reporting the user’s SNR to the base-station which estimates the channel quality. Link adaptation is then performed by continuously adjusting the modulation and coding parameters every 2 ms, corresponding to the basic HS-DSCH transmit time interval (TTI) \([12]\). The server receives channel quality information from the link/MAC layer in terms of packet loss rate \(p_c\) and estimated channel capacity for each user.

Table 2.2 shows an example of the possible channel capacities and packet loss rates achievable in an HSDPA system. We consider channel capacities of 1.8 Mbps, 3.6 Mbps, 5.3 Mbps, and 7.2 Mbps which correspond to the transport format and resource combinations (TFRC) of an HSDPA system with \(p_c = 10\%\) as defined in \([12]\). We have extended this list to include additional channel coding rates that achieve loss rates of 2\% and 4\% as shown in Table 2.2.

Table 2.2: Extended List of Modulation and Coding Schemes

<table>
<thead>
<tr>
<th>TFRC</th>
<th>SNR-range (dB)</th>
<th>Modulation</th>
<th>FER</th>
<th>Data Rate (15 Codes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-10 – 10</td>
<td>QPSK</td>
<td>2%</td>
<td>1.52 Mbps</td>
</tr>
<tr>
<td>2</td>
<td>4%</td>
<td></td>
<td></td>
<td>1.63 Mbps</td>
</tr>
<tr>
<td>3</td>
<td>10%</td>
<td></td>
<td></td>
<td>1.8 Mbps</td>
</tr>
<tr>
<td>4</td>
<td>10 – 15</td>
<td>QPSK</td>
<td>2%</td>
<td>3.32 Mbps</td>
</tr>
<tr>
<td>5</td>
<td>4%</td>
<td></td>
<td></td>
<td>3.43 Mbps</td>
</tr>
<tr>
<td>6</td>
<td>10%</td>
<td></td>
<td></td>
<td>3.6 Mbps</td>
</tr>
<tr>
<td>7</td>
<td>15 – 18</td>
<td>QPSK</td>
<td>2%</td>
<td>5.12 Mbps</td>
</tr>
<tr>
<td>8</td>
<td>4%</td>
<td></td>
<td></td>
<td>5.23 Mbps</td>
</tr>
<tr>
<td>9</td>
<td>10%</td>
<td></td>
<td></td>
<td>5.3 Mbps</td>
</tr>
<tr>
<td>10</td>
<td>18 – 21</td>
<td>16QAM</td>
<td>2%</td>
<td>6.64 Mbps</td>
</tr>
<tr>
<td>11</td>
<td>4%</td>
<td></td>
<td></td>
<td>6.86 Mbps</td>
</tr>
<tr>
<td>12</td>
<td>10%</td>
<td></td>
<td></td>
<td>7.2 Mbps</td>
</tr>
</tbody>
</table>

Moreover, HSDPA implements adaptive modulation and coding to switch between different TFRCs depending on the reported SNR levels. We assume the channel to be slowly fading with a channel state described by the following Gauss-Markov model:

\[ h_{t+1} = ah_t + w_t, \]  \hspace{1cm} (2.36)

where \(h_t\) is the channel state at TTI \(t\), \(a \in [0, 1]\) is a correlation parameter that approaches 1, and \(w_t \sim N(0, (1 - a^2)\sigma_h^2)\) is white Gaussian noise.
Frame-level UXP

The effectiveness of UXP for the protection of scalable video has been demonstrated in [13]. UXP is a Forward Error Correction (FEC) scheme that is based on Reed-Solomon (RS) codes. Unlike the traditional approach to UXP, where all coded video packets of a group of pictures (GOP) are grouped together into a transmission block (TB), we propose to implement UXP on an individual coded picture level. Therefore, a TB will only contain the base and CGS/MGS enhancement layer packets of a single coded picture.

Each row of a TB constitutes an \((n, k)\) RS codeword, where \(n\) is the length of the codeword in bytes, \(k\) is the number of data bytes (coded video bytes), and \(n-k\) bytes are parity bytes, where \(\frac{n}{2} < k \leq n\). Each column of a TB is then encapsulated in an RTP packet, which enables UXP to correct up to \(n-k\) packet erasures. Different protection classes are defined for each enhancement layer packet, higher layers receiving less protection than lower layers. The UXP failure rate \(P_x\) is then given by:

\[
P_x(n, k, p_c) = \sum_{i=n-k+1}^{n} \binom{n}{i} p_c^i (1-p_c)^{n-i}.
\]

The expression for \(P\) in (2.37) is equivalent to the complementary CDF of the binomial distribution \(B(n, p_c)\) which can be approximated by the continuous complementary CDF of the corresponding normal approximation \(N(np_c, np_c(1-p_c))\) shown below:

\[
P_x(n, k, p_c) = \frac{1}{2} \left[ 1 - \text{erf}\left(\frac{n-k-\mu}{\sqrt{2}\sigma^2}\right) \right],
\]

where \(\mu = np_c\) and \(\sigma^2 = np_c(1-p_c)\), \(\text{erf}(.)\) is the Gauss error function.

2.4.2 Joint Rate and Protection Allocation for Multiple Scalable Video Streams

In this section, we formulate the global optimization problem that attempts to jointly allocate the appropriate rate and UXP protection for all users while minimizing the expected video distortion.

Definition of the Optimization Objective

The objective of the global optimization problem is to minimize the sum of expected video distortions for all users. This corresponds to maximizing the sum of the PSNR estimates given by equations (2.32) and (2.33).
Let $\bar{D}^u$ be the estimated video distortion of user $u$ defined as a function of the base and enhancement layer QPs $q_b^u, q_e^u$, UXP protection parameters $k_b^u, k_e^u$, and channel loss rate $p_c$. We express $\bar{D}^u$ as:

$$\bar{D}^u(q_b^u, q_e^u, k_b^u, k_e^u, p_c, \tilde{m}^u) = P_e(1 - P_b)D_b + (1 - P_e)(1 - P_b)D_e + P_bD_{EC},$$

(2.39)

where $P_b = P_x(n, k_b^u, p_c)$, $P_e = P_x(n, k_e^u, p_c)$, and $P_x$, $D_b$, $D_e$ are given by (2.38), (2.32) and (2.33), respectively. $D_{EC}$ is the error concealment distortion which we will derive in detail in Chapter 3.

The global objective function $f(.)$ is then given by:

$$f(q_b, k_b, q_e, k_e, \tilde{m}, p_c) = \sum_{u=1}^U \bar{D}^u(q_b^u, q_e^u, k_b^u, k_e^u, p_c, \tilde{m}^u),$$

(2.40)

where $q_b, k_b, q_e, k_e$ and $\tilde{m}$ are vectors with $U$ elements corresponding to the user parameters, e.g. $q_b = [q_1^b ... q_U^b]^T$.

**System Constraints**

The live multi-cast system under consideration is constrained primarily by the channel capacity $C$, which varies according to the perceived channel SNR.

Let $\bar{R}^u$ be the total video bit-rate of user $u$ incorporating base- and enhancement-layer packets and UXP overhead. We express $\bar{R}^u$ as:

$$\bar{R}^u(q_b^u, q_e^u, k_b^u, k_e^u, n, p_c, \tilde{m}^u) = \frac{n}{k_b^u} R_b + \frac{n}{k_e^u} R_e,$$

(2.41)

where $R_b$ and $R_e$ are given by (2.34) and (2.35), respectively.

The primary constraint function $l(.)$ can then be expressed as follows:

$$l(q_b, k_b, q_e, k_e, \tilde{m}, p_c) = \sum_{u=1}^U \bar{R}^u(q_b^u, q_e^u, k_b^u, k_e^u, n, p_c, \tilde{m}^u).$$

(2.42)

**Global Optimization Problem**

We formulate a global optimization problem in this section that maximizes the sum of expected video PSNRs subject to the channel capacity constraint and variable bounds. The optimization variables are identified as the base and enhancement layer quantization parameters $q_b$ and $q_e$, and the UXP coding rates controlled by $k_b$ and $k_e$. 

54
Chapter 2. Rate and Distortion Modeling of CGS/MGS Coded Scalable Video Content

The QP range is given by \( \{0, \ldots, q_0\} \), where 0 corresponds to the best video quality, \( q_0 \) corresponds to the maximum allowable quantization parameter that satisfies the QoS guarantees. Moreover, the enhancement layer QP should always be smaller than the base layer QP in order to provide the quality refinement that the MGS/CGS layer was intended for.

The range of the UXP coding rate is given by the parameters \( k_b \) and \( k_e \) which are bounded by the interval \( \left[ \frac{n}{2} + 1, n \right] \). In [13], we developed a rate-minimized UXP allocation scheme that selects the protection levels for all UXP classes prior to the rate allocation. This separate stage was necessary since the function \( P_x \) of (2.38) is non-convex over the full range of \( k \in \left[ \frac{n}{2} + 1, n \right] \). Fig. 2.10 illustrates the non-convex shape of the \( P_x \) over the full range of \( k \). However, such an approach results in sub-optimal rate and protection allocation. We found that we can jointly optimize for rate and error protection by narrowing down the range of \( k \) to a small convex interval. Since \( P_x \) is either 0 or 1 for most of the range of \( k \), and a recovery failure rate greater than 50% is simply useless, we have narrowed down the feasible range of \( k \) to the interval where \( 0 < P_x \leq 0.5 \). This interval corresponds to the following bounds for \( k \)

\[
n - \mu - \text{erf}^{-1}(0.99)\sqrt{2\sigma^2} \leq k \leq n - \mu,
\]

where \( \mu \) and \( \sigma^2 \) are the mean and variance of the approximate normal distribution function discussed in Section 2.4.1. The new feasible region is now convex and corresponds to the highlighted area in Fig. 2.10.

![Figure 2.10: Plot of the UXP failure rate function \( P_x(.) \) showing the non-convex shape of the function over the full range of \( k \) and the convex region over the reduced feasible interval.](image-url)
Thus, we define the global constrained optimization problem as follows:

\[
\begin{align*}
\max_{q_b, k_b, q_e, k_e} & \quad f(q_b, k_b, q_e, k_e, \tilde{m}, p_c) \\
\text{subject to} & \quad l(q_b, k_b, q_e, k_e, \tilde{m}, p_c) \leq C, \\
& \quad 0 \leq q^u_b \leq q_0, \\
& \quad 0 \leq q^u_e \leq q^u_b - 6, \\
& \quad n - \mu - 2\sqrt{2\sigma^2} \leq k^u_e < k^u_b \leq n - \mu, \\
& \quad \text{for all } u \in \{1, 2, \ldots, U\},
\end{align*}
\] (2.43)

where \(\text{erf}^{-1}(0.99) \approx 2\).

### 2.4.3 Simulation Results

In this section, we compare the performance of our proposed streaming approach to existing streaming schemes. For our comparisons, we used the JSVM-9-8 reference software \([15]\) to encode 7 reference video sequences in CIF resolution: Foreman, Football, Bus, Mobile, Soccer, Crew, and City. Each sequence is composed of 150 frames encoded in SVC, with an H.264/AVC compatible base layer, a GOP size of one, an IDR period of 32 frames, a frame rate of 30 fps, and minimum QoS guarantee corresponding to \(q_0 = 38\).

We follow the same channel model described in Section 2.4.1 with 7 mobile clients/users each receiving a different video sequence and having similar channel SNRs. We simulate a slowly fading channel described by equation (2.36) with a correlation factor \(a = 0.9\). Fig. 2.11 illustrates the simulated SNR values and the corresponding achievable peak data rates for a period of 450 TTIs. The figure shows how the adaptive modulation and coding process maps the perceived SNR values to the achievable peak data rates according to the TFRCs defined in Table 3.1 for a channel loss rate \(p_c = 10\%\).

For our comparison, we define the following three real-time streaming approaches which cover the different centralized streaming approaches available in the literature:

1. Joint Base/Enhancement Rate-Protection Allocation (proposed): This is our proposed approach as defined by equation (2.43).

2. Enhancement-Only Rate-Protection Allocation: This is the traditional approach in which the base layer quality and protection is fixed to the guaranteed QoS quality and the optimization is performed strictly on the enhancement layer rate and protection allocation. This approach corresponds to solving (2.43) while fixing \(q^u_b = q_0\) and \(k^u_b = n - \mu - 2\sqrt{2\sigma^2}\) for all \(u\).

3. Joint Base/Enh. Rate Allocation: In this approach, we preset the base and enhancement
layer protection levels and optimize for the base and enhancement layer rate-allocation. This approach corresponds to solving (2.43) while fixing the base and enhancement layer protection parameters $k_u^b$ and $k_u^e$. To compare the performance of this approach, we set the base-layer protection to full recovery $P_b \approx 0$, and we consider two enhancement layer recovery rates: $P_e = 0.05$ and $P_e = 0.1$.

Fig. 2.12 shows a comparison in average PSNR performance between the four streaming schemes described above. We observe that our proposed approach significantly outperforms the remaining schemes with improvements averaging around 1.2 dB and ranging between 0.35 dB to 1.9 dB over the closest contender, streaming approach 2) defined above. Moreover, Fig. 2.12 shows that the improvement in performance is higher when the channel is under-provisioned. This is due to the improved rate and protection allocation structure defined in (2.43). The results also emphasize the importance of jointly optimizing for error protection allocation as can be seen by the improvement in performance of our proposed scheme over that of streaming approach 3) defined above.
In this chapter, we derived single layer and scalable video rate and distortion models based on an assumption that the transform coefficients follow a Laplacian distribution. The models are customized to capture the behavior of scalable video bitstream encoded using the CGS feature of SVC. Moreover, the models compensate for errors in the distribution assumptions by linearly scaling the Laplacian distribution parameter $\lambda$. Furthermore, we present simplified approximations of the derived models that allow for a run-time calculation of sequence dependent model constants. Our models use the mean absolute difference (MAD) of the residual signal extracted early in the encoding process and the quantization parameter (QP) values to estimate the residual MAD, rate, and distortion of a coded video frame at any QP value and for either base-layer and MGS layer packets. We have also shown performance evaluations that demonstrate that our proposed models accurately estimate the aforementioned metrics.

Moreover, we presented a real-time joint bit-rate and error protection allocation scheme for streaming multiple live-encoded scalable video streams over a shared downlink channel with limited capacity. This scheme utilizes the derived video rate and distortion estimation models and solves a global constrained optimization problem to allocate the appropriate base and enhancement layer parameters according to the available network resources. Performance evaluations show that our proposed scheme results in significant improvements averaging 1.2dB in PSNR over existing streaming strategies.

In the next chapter, we address the problem of streaming pre-encoded scalable video content to multiple video receivers over a 3GPP type network. We also analyze the performance of media-aware multi-user video streaming strategies in capacity limited wireless channels suffering from latency problems and packet losses.
2.6 References


Chapter 2. Rate and Distortion Modeling of CGS/MGS Coded Scalable Video Content


Chapter 3

Channel Aware Multi-User Scalable Video Streaming over Lossy Under-Provisioned Channels

3.1 Introduction

The advances in wireless and mobile network technologies are enabling a multitude of multimedia streaming applications. These advances are triggering an accelerated growth in wireless video streaming applications that are bandwidth-intensive, delay-sensitive, and loss-tolerant [1]. Broadband ad-hoc wireless networks and high-throughput mobile 3G networks, such as IEEE 802.11-based wireless local area networks (WLAN) and high-speed downlink packet access (HSDPA) respectively, are being deployed at an expanding pace and the number of subscribers to video streaming services is on the increase [2] [3]. However, with the expanding availability of streaming services comes the demand for streaming and scheduling policies that manage the large data volume.

We propose a delay and capacity constrained multi-user scalable video streaming scenario and analyze its performance compared to several media-aware scalable video streaming scenarios over lossy channels. We develop an optimized unequal erasure protection (UXP) approach, based on the scheme proposed in [4], that minimizes the overall protected video bit-rate while guaranteeing the necessary error protection. This improved UXP protection, together with a capacity and delay aware rate allocation demonstrate superior performance compared to other streaming schemes that do not utilize rate optimization and/or channel delay knowledge.

Moreover, we analyze the properties of our proposed optimization problem in terms of convexity and complexity. Our preliminary observations indicate that using unequal error protection renders any optimization problem with bit-rate constraint as non-convex since the bit-rate constraint equation is piecewise concave as illustrated in Fig. 3.1.

Figure 3.1: The plot shows that unequal error protection (UEP) results naturally in a non-convex optimization problem as witnessed by the concavity of the UEP function compared to linearity of equal error protection (EEP). The protected bit-rate region is bounded by the available channel capacity.

3.1.1 Summary of Contributions

The main contributions of this chapter are summarized as follows:

1. We propose a rate-minimized UXP scheme based on the UXP scheme proposed by Liebl at. al [5] and adopted for SVC in [4]. The proposed rate-minimized scheme balances the tradeoff between increasing the protection overhead and reducing the UXP failure rate.

2. We derive an analytical expression for packet delay and play-out deadline of UXP protected scalable video. A group of coded video frames (GOP) is embedded in a single UXP transmission block (TB) thus changing the play-out deadline of the protected packets. This modified packet sequence alters the regular transport stream structure. Therefore, we derive an upper bound for the number of radio link control (RLC) frames allowed per TB as a function of the RLC frame delivery delay.

3. We derive a loss-distortion model for hierarchical predictive video coders using picture
copy error concealment. A loss-distortion video model is essential for media-aware streaming policies which require extensive knowledge of the video encoding structure since it reflects the sensitivity of the decoded video stream to packet losses affecting the bitstream during transmission.

4. We analyze the performance of delay-aware, capacity-aware, and rate-minimized UXP in conjunction with optimized rate-allocation compared to simpler streaming strategies. Although delay and capacity constrained rate-allocation requires more computation, the bulk of the computation workload is performed off-line and does not affect the run-time performance of the system.

5. We show that the UXP protected rate-constrained optimization problem is non-convex, to which respect we propose an approximate solution using convex-relaxation and projection onto the feasible set.

The remainder of this chapter is organized as follows. Section 3.2 presents a review of existing proposals for multi-user video streaming. We present in Section 3.3 the different components of a multi-user video streaming system in terms of unequal error protection, packet delay analysis and multi-user rate allocation. Section 3.4 discusses our proposed adaptive unequal erasure protection scheme. We formulate the packet delay analysis of UXP protected video streams in section 3.5. Section 3.6 defines a new loss-distortion model for hierarchically predictive video coders. We formulate the rate allocation problem in Section 3.7 and discuss the convexity of the problem in Section 3.8. The performance of the proposed media-aware multi-user video streaming algorithms are analyzed in Section 3.9, and finally we present our conclusions in Section 3.10.

3.2 Literature Review

Extensive research has been done in this field starting with the pioneering work of Chou et al. [6] in which the problem of rate-distortion optimized (RaDiO) streaming of an entire video presentation is reduced to the problem of error-cost optimized transmission of an isolated data unit.

To our knowledge, the most closely related work is that in [3] and [7]. In [3], Pahlawatta et al. investigate channel-aware resource allocation and packet scheduling for the transmission of scalable video over wireless networks such as HSDPA and IEEE 802.16. They present a gradient-based packet scheduling scheme that uses channel quality information and the gradient of a cost function derived from the received video distortion and the decoder error concealment. However, no application layer FEC or delay analysis are considered, which are major components of the
work presented in this chapter. Zhang et al. address in \cite{7} resource allocation for scalable video transmission in wireless 3G networks by proposing a distortion-minimized bit allocation scheme with hybrid unequal error protection (UEP) and delay-constrained automatic repeat request (ARQ), which dynamically adapts to the estimated time-varying network conditions. The work presented in \cite{7}, however, does not address the case of multiple video users and the delay factor included in the work is limited to setting the maximum number of retransmissions in ARQ.

More recent work on multi-user video streaming can be found in \cite{8} and \cite{9} which propose distributed approaches to packet-based streaming, based on priority-queuing, that utilize local information feedback acquired from various network horizons to maximize the decoded video quality of multiple users simultaneously streaming over the same multi-hop wireless network. While these works provide an extensive analysis of the multi-hop wireless video streaming paradigm, the main focus of this chapter is in analyzing the effects of application layer FEC, accurate distortion modeling, and delay-aware dynamic rate allocation on the performance of wireless video streaming applications, namely in 3G high speed packet access (HSPA) environments where centralized resource allocation is valid and computationally inexpensive. Moreover, it is possible to combine our findings with those of \cite{8} and \cite{9} for a more comprehensive analysis of cross-layer optimized wireless video streaming solutions.

Other related work is found in \cite{10} and the references included therein, where video streaming is extensively studied over wireless ad-hoc networks and the concept of RaDiO is extended to congestion distortion optimized streaming (CoDiO). While these works constitute a comprehensive study of wireless video streaming over ad-hoc networks, there is no work, to our knowledge, that compares the performance of these algorithms, especially in conjunction with UXP. However, the main conclusions of the aforementioned works that we wish to draw upon are the importance of delay analysis and delay-aware rate control in multi-user video streaming. More sophisticated streaming algorithms are presented in \cite{11} and \cite{12} which employ stochastic dynamic programming to minimize the end-to-end distortion while considering congestion, channel state, and queuing delay. \cite{11} further extends this idea using a priori stochastic models for the source and channel. The models are used to solve for an optimal control policy that minimizes the average coding-distortion off-line, and reducing the on-line computational cost to simply identify the source-channel state.

Application-layer FEC for multimedia transmission has been in use for some time. The effectiveness of application layer UEP for scalable video transmission can be seen in \cite{13} where bit-error-rate adaptive UEP and packet-size are employed to improve the end-to-end video distortion. The popularity of application layer FEC has continued with Schierl et al.’s proposal to use UXP with the scalable extension of H.264/AVC (SVC) for wireless video broadcasting \cite{4}. However, no consideration is taken for limited channel capacity or packet delivery delays and
Chapter 3. Channel Aware Multi-User Scalable Video Streaming over Lossy Under- Provisioned Channels

Figure 3.2: General structure of the proposed video streaming system over 3G wireless networks. The link layer feeds back the estimated channel capacity $C$, RLC loss rate $p$, and the packet delay $d$ to the streaming server. The rate and distortion functions, $R_u$ and $D_u$, are defined in (3.21).

The issue of UXP delay is very rarely addressed in the available literature. In fact, to the best of our knowledge, there has been no study for the packet delay and play-out deadline performed for UXP protected video streams. More recent work studying SVC transmission over 3G networks is found in [14] and [15] which focus on the buffer management and scheduling policies for SVC video over 3G systems such as HSDPA.

Finally, for network-aware multiuser resource allocation with fairness considerations, we refer the reader to the recently developed game theoretic approach in [16] which utilizes a pricing mechanism for message exchanges between users that reach Nash equilibrium generating budget balanced allocations that satisfy voluntary participation.

3.3 Video Streaming System Overview

In this section, we present the different components of a multi-user video streaming system in terms of video rate-distortion characteristics, and the underlying architecture and limitations of the transmission channel. A video streaming system, in general, is comprised of a server which stores pre-encoded video files and several clients that request these files from the server. The files are delivered to the client across a channel that can have time-varying throughput and suffer from latency problems and packet losses.

Fig. 3.2 illustrates the block structure of our generalized video streaming system and its components. The streaming system shown in Fig. 3.2 shows a streaming server with access to $U$ pre-encoded video streams intended for delivery to $U$ clients/users. The server receives channel quality information from the link/MAC layer in terms of RLC frame loss rate, estimated channel capacity, and estimated packet delivery delay. We define the following system parameters that
characterize the multi-user video streaming problem:

- $D_u$, the estimated video distortion perceived at the decoder of user $u$.
- $R_u$, the estimated video bit-rate of user $u$ including UXP overhead and RLC retransmissions.
- $P_u$, vector containing the UXP failure rates of all the scalable video classes of user $u$.
- $C$, shared channel capacity by all $U$ users.
- $p$ and $\delta$, RLC frame loss rate and packet delivery delay reported by the MAC layer.

In order to address the delay-sensitive multi-user video streaming problem we divide the task into four categories: application layer error protection, packet delay restrictions, estimation of the decoded video distortion, and multi-user bit-rate allocation.

### 3.4 Application Layer Error Protection

We present in this section a rate-minimized UXP scheme that allows a scalable coded video stream to recover from packet losses without violating the channel capacity constraints.

#### 3.4.1 Rate-Minimized Adaptive UXP Allocation

In this subsection, we develop a rate-minimized UXP protection scheme that balances the tradeoff between protection overhead and the UXP failure rate. The main drawback in the standard UXP scheme proposed in [1] is that all NALUs of the same FGS layer will have the same protection level. However, SVC allows for the assignment of up to 64 quality levels to FGS NALUs, since not all NALUs of a certain FGS level necessarily contribute the same improvement to the video distortion [17]. For this reason, we have developed a UXP allocation scheme that divides the UXP classes according to NALU priority level instead of the FGS level. Moreover, our objective is to reduce the protection overhead while maintaining the same video distortion as the standard UXP.

Let $q \in Q$, where $Q = \{0, 1, 2, \ldots, 63\}$ be the NALU quality level index, and $r_u(q)$ be the bit-rate of level $q$ NALUs. Also, let $P(\cdot)$ be a function that gives the code failure rate of a UXP class $q$. We find the dimension $k_q \in K, K = \left\{ \frac{n}{2} + 1, \ldots, n \right\}$, for codewords of class $q$ by minimizing the following regularized cost function:

$$k^*_q = \arg \min_{k_q \in K} \sum_{q=1}^{n} r_u(q) + \omega_q P(k_q, p, \delta), \quad (3.1)$$
where $\omega_q$ is the regularizing parameter. Experiments have shown that $\omega_q$ can be expressed in terms of the relative distortion contributed by all packets belonging to the priority level $q$. The UXP failure rate $P(.)$ is given by

$$P(k_q,p_{\text{rtp}}) = \sum_{i=n-k_q+1}^{n} \binom{n}{i} p_{\text{rtp}}^i (1-p_{\text{rtp}})^{n-i}$$

Equation (3.2) shows the UXP failure rate equation as a binomial complementary CDF under the assumption taken in [5] that the RTP packets are lost independently. This UXP failure rate can also be formulated as a function $P_u(r_u)$ which constitutes a real-valued, piecewise-constant, increasing function of $r_u$ with range $[0,1]$. $P_u(r_u)$ is increasing because the UXP scheme allocates higher protection to the base layer than FGS enhancement layers which translates into a lower failure rate for the base layer compared to enhancement layer failure rates.

As a result of the approach described above, we can write the UXP code dimension $k_u$ and protected video bit-rate $R_{\text{UXP}}$ as functions of the unprotected video bit-rate $r_u$, such that, $k_u(r_u)$ is an integer valued, piecewise-constant, increasing function of $r_u$ with range $[\frac{n}{2}, n]$, and $R_{\text{UXP}}$ is given by

$$R_{\text{UXP}}(r_u) = \int_{L_u}^{r_u} \frac{n}{k_u(x)} dx.$$  

The benefit of this modified protection allocation scheme arises in capacity limited channels where part of the encoded video bitstream is most probably not going to be transmitted. In such cases, low priority data are allocated little or no protection bits since they will not be transmitted by the streaming server.

### 3.5 Packet Delay Restrictions in Video Streaming

In this section, we represent the wireless channel in terms of its variable components and constraints. An emerging application is video streaming over ad-hoc and 3G/3.5G wireless mobile networks such as High Speed Down-link Packet Access (HSDPA) [1-3]. Typically, such networks suffer from channel fading which leads to packet loss and queuing delays. In the following channel abstraction, we will follow the protocol stack architecture described in VCEG-N80 [18] for 3GPP network architecture. The link-layer performs basic error protection in terms of non-persistent ARQ with retransmission deadlines. The video payload is carried over RTP/UDP/IP, where each RTP packet is fragmented into $m$ RLC frames. We assume...
that the link-layer reports the following fields to the streaming server:

\[
S_{\text{rlc}} = \text{the radio-frame size},
\]
\[
p = \text{steady-state RLC loss rate/probability},
\]
\[
\delta = \text{transmission delay}
\]  

(3.4)

Moreover, we assume that the streaming server has enough information to be able to predict the channel state and state-transition probabilities.

Let \( p_{\text{rtp}} \) be the RTP loss probability, \( S_{\text{rtp}} \) be the size of an RTP packet, and \( N_{\text{max}} \) the maximum number of allowable RLC retransmissions. We consider a correlated fading channel model as in [7,19] where RLC frame loss can be approximated by a two state Markov chain defined by the transition matrix

\[
M = \begin{bmatrix}
y & 1 - y \\
1 - z & z
\end{bmatrix},
\]

where \( y \) and \( 1 - y \) are the probabilities that the current RLC transmission was successful given that the previous transmission was successful or unsuccessful, respectively. Consequently, we can then express \( p_{\text{rtp}} \) in terms of the steady-state RLC error rate \( y \) and \( z \) as:

\[
p_{\text{rtp}} = 1 - \left[ 1 - \left( \frac{1 - y}{2 - y - z} \right) \right]^{\frac{S_{\text{rtp}}}{S_{\text{rlc}}}},
\]

where \( A \) is a constant representing the average number of RLC transmissions and retransmissions and is bounded by \( 1 < A \leq 1 + N_{\text{max}} \).

### 3.5.1 Packet Delay Analysis

Let \( \tau_{s,1} \) be the sender time at which the first link-layer frame of an RTP encapsulated NALU is transmitted. The frame is received by the client at time \( \tau_{r,1} \) after a delay \( \delta_1 \), such that

\[
\tau_{r,1} = \tau_{s,1} + \delta_1.
\]

(3.5)

The second frame is transmitted after one Transmission Time Interval (TTI) at time \( \tau_{s,2} = \tau_{s,1} + \xi \) and received at \( \tau_{r,2} = \tau_{s,2} + \delta_2 \), where \( \xi \) is the TTI and \( \delta_2 \neq \delta_1 \).

In the case where a single NALU is encapsulated in single RTP packet, then \( m \) RLC frames have to be sent to transmit the NALU, as shown in Fig. 3.3. Let \( \Lambda_l \) be the total delay of NALU \( l \) defined as the total time between receiving the first RLC frame and the last frame, and let \( \tilde{\delta} \)
be the average delay for all RLC frames sent and $\tilde{\delta}_i = \delta_i - \bar{\delta}$, then $\Lambda_l$ is expressed as

$$
\Lambda_l = \tau_{r,m} - \tau_{r,1}
= \sum_{i=2}^{m} (\xi + \delta_i - \delta_{i-1})
= (m - 1)\xi + \tilde{\delta}_m - \tilde{\delta}_1
$$

(3.6)

The expression for the delay in equation (3.6) is useful for expressing worst case values for the delay in terms of its mean and standard deviation.

### 3.5.2 Play-out Deadline and the Maximum NALU Size

We derive the upper bound on the maximum number of RLC frames allowed per NALU in this subsection. When the size of a video NALU is larger than the link-layer RLC frame size, the NALU is partitioned into $m$ RLC frames. The NALU delay comes into effect after the first NALU is decoded. A streaming server should attempt to deliver all RLC frames of a NALU before the play-out deadline of the NALU, otherwise, it should avoid sending those RLC frames which would be received on time.

Let $f$ be the video display frame rate, $\tilde{\delta}$ be the delay shift off the mean, and $\xi$ be the transmission time interval. The number of RLC frames $m_l$ per NALU $l$ is therefore bounded above by

$$
m_l \leq \frac{(l - 1)(\frac{1}{f}) - \sum_{i=2}^{l-1} (m_i \times \xi) + \tilde{\delta}_0 - \tilde{\delta}_{m_l}}{\xi}
$$

(3.7)

The derivation of the above constraint is shown below.

Every transmitted NALU $l$ has a time deadline by which it should be played at the decoder.
Chapter 3. Channel Aware Multi-User Scalable Video Streaming over Lossy Under-Provisioned Channels

This deadline, called play-out deadline and represented by \( \tau_{DTS,l} \), is affected by the video display frame rate \( f \) and the previous packet delay \( \Delta_{l-1} \). Let \( \tau_{R1} \) be the time at which the first NALU is fully received, i.e. \( \tau_{R1} = \tau_{r,m1} \) = time at which the last RLC frame of the first NALU is received at the client. The play-out deadline of the second NALU is therefore automatically set to

\[
\tau_{DTS,2} = \tau_{R1} + \frac{1}{f}. \tag{3.8}
\]

The delay of the second NALU \( \Delta_2 \) is then expressed as

\[
\Delta_2 = \tau_{r,m2} - \tau_{R1} = m \times \xi + \tilde{\delta}_{m2} - \tilde{\delta}_0, \tag{3.9}
\]

where \( \tilde{\delta}_0 = \tilde{\delta}_{m1} = \tilde{\delta}_{R1} \), and \( \tau_{r,m2} \) is the time at which the last RLC frame of the second NALU is received. For NALU 2 to meet its play-out deadline, \( \tau_{r,m2} \leq \tau_{DTS,2} \), which translates to \( \Delta_2 \leq \frac{1}{f} \).

Similarly, \( \tau_{DTS,3} = \tau_{R1} + \frac{2}{f} \) and hence,

\[
\Delta_3 = \tau_{r,m3} - \tau_{r,m2} = m_3 \times \xi + \tilde{\delta}_{m3} - \tilde{\delta}_{m2}, \tag{3.10}
\]

and

\[
\tau_{r,m3} \leq \tau_{R1} + 2\left(\frac{1}{f}\right). \tag{3.11}
\]

From equations (3.9), (3.10) and (3.11), we get

\[
\Delta_3 \leq 2\left(\frac{1}{f}\right) - \Delta_2,
\]

which results in the constraint on the number of RLC frames in NALU 3 expressed in terms of \( \tilde{\delta} \) of the last frame in NALU 1 and the last frame in NALU 3 as shown below

\[
m_3 \leq \frac{2\left(\frac{1}{f}\right) - m_2 \times \xi + \tilde{\delta}_0 - \tilde{\delta}_{m3}}{\xi}.
\]

Therefore, a general expression for the constraint on the number of RLC frames in a NALU \( l \) is written as in (3.7).

3.5.3 UXP Delay Analysis

We expand the delay analysis described above to the UXP case and illustrate the modifications in this subsection. Under UXP, the basic video payload unit is not NALU anymore but the Transmission Block (TB) which contains \( n \) RTP packets, each of which is fragmented into \( m \)
Chapter 3. Channel Aware Multi-User Scalable Video Streaming over Lossy Under-Provisioned Channels

RLC frames. The NALUs of a single GOP are aggregated into one TB in addition to the protection/parity symbols, as described above. Moreover, no NALU can be decoded before the entire TB is received at the decoder. Therefore, the new constraint on the number RLC frames per RTP packet of a TB is written as

$$m_t \leq \Psi_t = \frac{(t-1)(N_f) - \sum_{i=2}^{t-1} (m_i \times n \times \xi) + \delta_{m_0} - \delta_{m_t}}{n \times \xi},$$  \hspace{1cm} (3.12)

where $m_i$ is the number of RLC frames per RTP packet in TB $i$, $\xi$ is the TTI length in seconds, and $i$ is a summation index. The derivation of this constraint is shown below.

Let $t$ be the TB index in a streaming session, such that $t \in \{0, 1, ..., T\}$, where $T$ is the total number of GOPs in the streaming session. Consider a fixed number of NALUs $N$ per GOP, the play-out deadline of TB 1 can be written as

$$\tau_{DTS,1} = \tau_{m_0} + \frac{N}{f},$$  \hspace{1cm} (3.13)

where $\tau_{m_0}$ is the time at which the last RLC frame of TB 0 arrives at the receiver. Following the analysis of Section 3.5, the delay of TB 1 $\Lambda_1$ can then be written as

$$\Lambda_1 = \tau_{m_1} - \tau_{m_0} = n \times m_1 \times \xi + \tilde{\delta}_{m_1} - \tilde{\delta}_{m_0},$$  \hspace{1cm} (3.14)

and the general delay constraint equation of TB $t$ is expressed as

$$\Lambda_t \leq (t-1) \frac{N}{f} - \Lambda_{t-1}.$$  \hspace{1cm} (3.15)

We can then extend this analysis to write the constraint $\Psi_t$ on the number of RLC frames per RTP packet in TB $t$ as in (3.12).

Equation (3.12) above appropriately formulates the delay constraint $\Psi_t$ in terms of the delay jitter $\tilde{\delta}$, which is beneficial in the case when the server has statistical knowledge of the channel delay. In such a case, the optimization can be performed for a greedy constraint where the term $\tilde{\delta}_{m_0} - \tilde{\delta}_{m_t} = -2\sigma$, where $\sigma$ is the standard deviation of the packet-delay distribution. This worst case scenario arises when the last RLC frame of the first TB 0 arrives one standard deviation $\sigma$ sooner than the mean delay $\bar{\delta}$, i.e. $\delta_{m_0} = \delta_0 = \bar{\delta} - \sigma$, and the last RLC frame of the current TB $t$ arrives with a delay $\delta_{m_t} = \bar{\delta} + \sigma$.

In the event that the expected delay of a NALU exceeds its play-out deadline, the server should not send the NALU since the decoder will have assumed the packet to be lost and already used a concealment strategy to compensate for the lost packet. We will therefore move to discuss loss-distortion modeling in the following section. Loss-distortion modeling captures
the sensitivity of a video codec to packet loss.

![Diagram of dependency graphs for non-scalable H.264/AVC and scalable SVC](image)

Figure 3.4: Comparison between the acyclic dependency graphs of non-scalable H.264/AVC (top) and scalable SVC (bottom). The SVC dependency graph shows the hierarchical predictive structure of an SVC coder. This prediction structure plays an important role in loss-distortion modeling as discussed in Section 3.6.

### 3.6 Loss-Distortion Modeling

This section defines a new loss-distortion model for hierarchically predictive video coders such as SVC, the scalable extension of H.264/AVC. A video loss-distortion model is essential for media-aware streaming policies which require extensive knowledge of the video encoding structure since it reflects the sensitivity of the decoded video stream to packet losses affecting the bitstream during transmission. Rigorous work has been done to find an accurate loss-distortion model for hybrid video codecs, such as, H.263 and H.264/AVC [20][21]. However, due to the difference in coding structure of hierarchically predictive coders illustrated in Fig. 3.4, a new framework is required to accurately model the packet loss sensitivity.

#### 3.6.1 Hierarchical Prediction Structure and the Picture Copy (PC) Concealment in SVC

In this subsection, we describe the prediction structure and error concealment strategy in SVC which affect the loss-distortion model, and consequently the estimate of the video distortion at the streaming server. The hierarchical prediction structure imposes a coding dependency between the B-frames of a GOP. The benefits of this dependency are realized in the improvement to the coding efficiency of the encoder compared to non-hierarchical H.264/AVC [22]. In the
follows discussion, we consider an SVC encoded stream with four dyadic hierarchy stages and 8 frames per GOP for tractability.

Figure 3.5: (a) Dyadic hierarchical structure of two GOPs with 4 temporal scalability levels. The frame loss dependency among B-frames can be seen in the groups of empty frames where the lost frames are those that have an X underneath. The remaining empty frames are affected during decoding due to their dependency on the lost frames. (b) Example of an SVC base layer coding sequence with a fixed GOP size G and an intra-refresh period $T_{\text{refresh}}$. The $\Delta$s indicate the mean square error between the specified frames.

Fig. 3.5 (a) shows the dependency of lower level hierarchical B-frames on the higher level frames, where frames $B_2$, $B_7$ and $B_{12}$ are assumed to be lost. As a result, the loss of frame $B_2$ causes the degradation of frames $B_1$ and $B_3$. The loss of $B_7$ will only affect frame 7 since it has no dependencies. On the other hand, if an I or a P frame (key-frames) is lost, then all dependent B-frames and all subsequent frames are affected until an Intra-refresh (IDR) frame is received.

In order to recover from these losses, decoders generally use basic error concealment algorithms to limit the error propagation due to frame loss. The picture copy (PC) concealment algorithm used in SVC replaces each lost frame by the previous temporal picture in the higher hierarchical level [17]. For instance, if the frame $B_2$ in Figure 3.5 (a) is lost, it is concealed using frame $I_0$. I and P frame loss concealment can only be performed using previously decoded I or P frames. Therefore, the loss-distortion can be separable into a distortion due to key-frame...
loss and a distortion due to B-frame loss expressed as

\[
D_{\text{base,loss}}(P_{\text{loss}}) = \left( \frac{1}{G}D_{\text{key,loss}} + \frac{G-1}{G}D_{\text{B,loss}} \right) \times P_{\text{loss}},
\]

(3.16)

where \(P_{\text{loss}}\) is the base layer packet loss rate, \(G\) is the GOP size, and \(D_{\text{B,loss}}\) and \(D_{\text{key,loss}}\) are defined by (3.17) and (3.19), respectively, in the following subsection.

### 3.6.2 Proposed Loss-Distortion Model for SVC

Consider a base layer SVC stream with constant integer GOP size \(G\), dyadic hierarchical decomposition, and an IDR period \(T_{IDR} = M \times G\) frames. Fig. 3.5(b) shows a coded sequence with the above description.

Let \(\triangle_i, i = \{1, 2, 3, ...G\}\) denote the concealment mismatch in terms of the mean square error (MSE) between the lost picture and the replacement picture, where \(i\) is the distance between a candidate lost frame and its concealment frame, as shown in Fig. 3.5(b). The distortion caused by the loss of a frame amounts to the mismatch in the concealment strategy used to replace the lost frame in addition to the propagation error that arises from the prediction dependency between the coded frames. For tractability, and without loss of generality, we consider the GOP size \(G = 8\) and four temporal/hierarchical levels.

#### Distortion Due to B-frame Loss

Consider the hierarchical structure shown in Fig. 3.5(a) where the key picture is at level 0, one B-frame at level 1, two B-frames at level 2... etc. Let \(D_{Bi}, i = 1, 2, 3\) correspond to the average concealment distortion induced by the loss of a level \(i\) frame, and let \(\nu\) be the discount factor associated with the propagation error to each lower level. The loss-distortion due to hierarchical B-frame loss in a GOP is then given by the expected value of the B-frame concealment distortion is expressed as

\[
D_{\text{B,loss}} = P_{B1}D_{B1} + P_{B2}D_{B2} + P_{B3}D_{B3},
\]

(3.17)

where

\[
\begin{align*}
P_{Bi} & = \frac{2^{i-1}}{2^i-1} \text{ for uniform NALU losses}, \\
D_{B3} & = \triangle_1, \quad D_{B2} = \triangle_2 + 2 \times (\triangle_2/\nu), \\
D_{B1} & = \triangle_4 + 2 \times (\triangle_4/\nu) + 4 \times (\triangle_4/\nu^2).
\end{align*}
\]

(3.18)

Experimental tests run on 2 different video sequences (Foreman and Bus) simulating random losses independently to each level show that, on average, the propagation error is reduced by
3 dB in PSNR as we move farther away from the lost frame, therefore, a typical value of $\nu$ would be $\nu = 10^{3/20} \approx 1.5$. Moreover, our model assumes a simplified uniform NALU loss rate since experiments with non-uniform link-layer frame loss models have resulted in no significant reduction in model accuracy.

**Distortion Due to Key-picture (I/P-frame) Loss**

The distortion caused by the loss of a key picture propagates to multiple GOPs. Let $D_{P_j}, j \in \{0, 1, \ldots, M - 1\}$ be the loss-distortion of key-frames in an IDR period of $M$ GOPs, where $P_0$ indicates an I-frame, and $P_1, P_2, \ldots, P_{M-1}$ the remaining P-frames.

The distortion caused by the loss of a key-frame is equal to the concealment mismatch from the last correctly received key-frame plus the error propagating into the dependent and subsequent frames. Let $\triangle_G$ approximate the concealment mismatch if the previous key picture is correctly decoded, and let $2\triangle_G$ be the mismatch if the previous key picture is lost. Consequently, the distortion due to the loss of a key-picture $D_{key,loss}$ formulates to

$$D_{key,loss} = \sum_{j=0}^{M-1} \frac{1}{M} D_{P_j} = G\triangle_G \left( (1 + p) + \frac{1}{M} \sum_{j=0}^{M-1} \sum_{i=1}^{M-j-1} (1 - i\gamma_j) \right)$$

where $\gamma_j = \frac{1}{M-j}$, $p$ is the probability that the previous key picture is lost, and $\gamma_j$ is a discount in the error signal caused by INTRA update macroblocks and loop-filtering in the subsequent key-pictures. We assume in the above formulation that the B-frames in each GOP suffer the same degradation as the key picture.

To verify our model, we encoded the two sequences Bus and Foreman in SVC with different Intra-refresh periods of 24, 36, and 48 frames and a GOP size of 8 frames. The encoded sequences were subjected to 3%, 5%, and 10% packet loss rates (PLRs) using the SVC/AVC loss simulator available from [23]. The effectiveness of our loss-distortion model is demonstrated in Fig. 3.6 which also show the error-bounds in the simulations for a 95% confidence interval around the displayed average values.

The plots show that the model matches the performance of the Bus sequence better than the Foreman sequence. This is due to the homogeneous level of motion exhibited in the Bus sequence compared to the varying motion in the Foreman sequence. Note that in our tests, we used time averaged parameter $\triangle$ for the sensitivity of a video stream to the concealment mismatch. For more dynamic video sequences, the parameter $\triangle$ can be updated in real time to deliver a more accurate representation.
Chapter 3. Channel Aware Multi-User Scalable Video Streaming over Lossy Under-Provisioned Channels

Figure 3.6: The figures show the performance of the loss-distortion model, derived in Section 3.6, in estimating the MSE of the sequences Foreman and Bus with packet losses of 3%, 5%, and 10% as a function of the Intra-refresh period. Error bounds are shown for a 95% confidence interval.

3.7 Multi-User Bit-rate Allocation

In this section, we discuss the problem of multi-user bit-rate allocation under stationary, non-stationary, and delay aware scenarios. The streaming system shown in Fig. 3.2 shows a streaming server that has access to $U$ pre-encoded video streams intended for delivery to $U$ clients/users. The server receives channel quality information from the link/MAC layer in terms of RLC frame loss rate, estimated channel capacity, and estimated packet delivery delay.

Let $u \in U$, where $U = \{1, 2, ...U\}$, be the video/user index. Each FGS scalable video indexed by $u$ is characterized by a bit-rate $r_u \in [L_u, H_u]$, where $L_u$ is the base-layer bit-rate and $H_u$ is the maximum video bit-rate which includes all enhancement layers. The video is also characterized by the encoder distortion $D_{\text{enc}}(r_u)$ which can be written as a function of the bit-rate $r_u$, given by the general rate-distortion model proposed in [20] and expressed as:

$$D_{\text{enc}}(r_u) = \alpha_u + \frac{\theta_u}{r_u - \beta_u}, \ r_u \neq \beta_u$$  \hspace{1cm} (3.20)

the parameters $\alpha_u$, $\theta_u$, and $\beta_u$ are video content and encoder dependant.

The packet losses encountered during video transmission as a result of channel fading and congestion calls for error protection and rate allocation measures to be taken by the server. It is therefore insufficient to characterize a video stream $u$ in terms of the pair $\{r_u, D_{\text{enc}}(r_u)\}$. 

76
Let $p$ be the RLC packet loss rate, we define the pair $\{R_u(r_u), D_u(r_u)\}$ as the total video bit-rate including UXP protection and the decoded video distortion of user $u$ after channel loss, respectively, such that

$$
R_u(r_u) = \left( 1 + \sum_{j=0}^{N_{\text{max}}} j(1-p)p^j \right) R_{\text{UXP}}(r_u)
$$

$$
D_u(r_u) = D_{\text{enc}}(r_u) + D_{\text{loss}}(r_u) + D_{\text{base,loss}}
\begin{align*}
&= \alpha_u + \frac{\theta_u}{(x-\beta_u)^2} \int_{r_u}^{r_{u+1}} P_u(x) \, dx + D_{\text{base,loss}} \\
&= \int_{r_u}^{r_{u+1}} \left( \frac{\theta_u}{(x-\beta_u)^2} \right) P_u(x) \, dx + D_{\text{base,loss}} \tag{3.21}
\end{align*}
$$

where $R_{\text{UXP}}(r_u)$ is the protected video bit-rate defined in (3.3), $\sum_{j=0}^{N_{\text{max}}} j(1-p)p^j$ is the channel redundancy due to MAC retransmissions, $D_{\text{enc}}(r_u)$ is the video encoder distortion defined by equation (3.20), $D_{\text{loss}}(r_u)$ is the distortion due to channel loss, $P_u(x)$ is the probability that UXP protection fails, and $D_{\text{base,loss}}$ is the base layer loss-distortion. The functions $R_{\text{UXP}}(r_u)$ and $P_u(x)$ depend on the error protection scheme used, as discussed in Section 3.6.1 for the UXP case. The loss-distortion model $D_{\text{base,loss}}$ is defined in (3.16) of Section 3.6.

The expression for $D_{\text{loss}}(r_u)$ in Eq. (3.21) is a result of application layer UXP, where base layer frames receive higher protection than enhancement layer frames. Consequently, the failure of UXP for some enhancement layer will only cause an increase in distortion equal to that contributed by the lost packets in the enhancement layer. To illustrate the derivation of $D_{\text{loss}}(r_u)$, we drop the user index $u$ and we let $e$ be the enhancement layer index. The enhancement layer occupies a bit-rate ranging from $r_e$ to $r_{e+1}$, $r_e < r_{e+1}$. The improvement in distortion resulting from the correct reception of layer $e$ corresponds to $D_{\text{enc}}(r_e) - D_{\text{enc}}(r_{e+1})$, where $D_{\text{enc}}$ is defined in (3.20). However, since packets in an enhancement layer can have different UXP failure rates as discussed in Section 3.4, the increase in distortion due to the loss of layer $e$ is written as

$$
d_{\text{loss}}(e) = \int_{r_e}^{r_{e+1}} \left( \frac{\theta}{(x-\beta)^2} \right) P_u(x) \, dx,
\begin{align*}
\text{where } \frac{\theta}{(x-\beta)^2} &= \frac{\partial}{\partial x} (D_{\text{enc}}(r_e) - D_{\text{enc}}(x)). \tag{3.22}
\end{align*}
$$

In a multi-user video streaming system, clients compete for channel capacity to receive the best video quality. We assume that all users share the same down-link channel and contest for the full range of the channel capacity $C$. In what follows, we formulate the rate-allocation problem as distortion-minimized capacity-constrained optimization problem and leave the discussion of the solution and proof that UXP renders a rate constrained optimization problem non-convex to Section 3.8.
3.7.1 Stationary Rate-Distortion Behavior

Assuming that the video R-D relationship is stationary over the duration of the streaming session, the server has to solve the following constrained optimization problem at the beginning of the streaming interval:

\[
\begin{align*}
\min_{\bf r} & = [r_1,r_2,...,r_U]' \sum_{u=1}^{U} D_u(r_u) \\
\text{s.t.} & \sum_{u=1}^{U} R_u(r_u) \leq C, \\
& L < r < H
\end{align*}
\] (3.23)

where \(D_u\) and \(R_u\) are the distortion and rate functions defined in (3.21), \(r\) is a \((U \times 1)\) vector containing the extracted video bit-rates \(r_u\) of user \(u\) and bounded below by the base-layer rates \(L = [L_1,...L_U]'\) and above by the total rates which include all enhancement layers \(H = [H_1,...H_U]'\).

It is worth noting that the constrained optimization problem defined in (3.23) satisfies the QoS conditions of delivering a minimum guaranteed video quality for all users. This is established by setting a lower bound \(L_u\) on the video extraction rate \(r_u\) which corresponds to the base-layer bit-rate of stream \(u\).

3.7.2 Non-Stationary Rate-Distortion Behavior

In the non-stationary case, several solutions have been proposed to handle the R-D variations by optimizing the streaming based on a time-varying rate-distortion (R-D) cost function for each individual NALU [12,24]. Such solutions, however, do not apply to a UXP protected video streaming situation. In the UXP context, the most basic time-localized block is the transmission block (TB). Therefore, we propose to solve a time-localized optimization problem for the streaming of every TB. The streaming session is thus divided into \(T\) rate extraction intervals, and the solution to the multi-user streaming problem becomes a multi-interval streaming policy \(\pi\), where \(\pi = [r_1,r_2,...,r_T]'\) is a \((U \times T)\) matrix defining extracted video bit-rates \(r_{u,t}\) of user \(u\) at interval \(t\).

Let \(t\) be the time-interval index at which the video user rate is modified, which also corresponds to the TB index. Although, the video rate-distortion (R-D) model of (3.20) captures the time-averaged R-D relationship of a coded video sequence, this model still holds for the time-localized R-D description while the upper and lower bounds on \(r_{u,t}\) vary with \(t\). Therefore,
Chapter 3. Channel Aware Multi-User Scalable Video Streaming over Lossy Under-Provisioned Channels

(3.23) is modified to incorporate the interval index as follows:

$$\min_{r_t=[r_{1,t}, r_{2,t}, \ldots, r_{U,t}]} \sum_{u=1}^{U} D_u(r_{u,t}, p_{rtp,u,t})$$

$s.t. \sum_{u=1}^{U} R_u(r_{u,t}, p) \leq C_t,$

$$L_t < r_t < H_t$$

which is solved $T$ times, once for each TB interval. The result is the policy matrix $\pi = [r_1, r_2, \ldots, r_T]'$. This scheme is also flexible enough to adapt to fluctuating channel capacity given an accurate capacity estimate. Note that the solutions to the constrained optimization problems shown in (3.23) and (3.24) are studied in Section 3.8.

3.7.3 Delay-Aware Non-Stationary Behavior

The streaming problems that we have presented so far, formulated as capacity-constrained optimization problems, do not include the effect of the play-out deadline of TBs as a result of the RLC frame queuing delays. The effect of delay can be included in our constrained optimization setup by forcing additional constraints on the number of allowed RLC frames $m_t^{(u)}$ per RTP packet in TB $t$ of stream $u$.

The value of $m_t^{(u)}$ is a function of the size of the RTP packet which, in turn, is a function of the size of the TB $t$. We can express this relationship as follows:

$$m_t^{(u)}(r_u) = \frac{R_{UXP}(r_{u,t}) \times N}{n \times S_{rlc} \times f}.$$ 

Thus, the modified capacity and delay constrained optimization problem is written as follows:

$$\min_{r_t=[r_{1,t}, r_{2,t}, \ldots, r_{U,t}]} \sum_{u=1}^{U} D_u(r_{u,t})$$

$s.t. \sum_{u=1}^{U} R_u(r_{u,t}) \leq C_t,$

$$m_t \leq \Psi_t$$

where $m_t = [m_t^{(1)}, m_t^{(2)}, \ldots, m_t^{(U)}]'$, and $\Psi_t = [\psi_t^{(1)}, \psi_t^{(2)}, \ldots, \psi_t^{(U)}]'$.

3.8 On the Non-convexity of the Rate Allocation Problem

In this section, we study the nature of the constrained optimization problems described above in terms of convexity and complexity of solution. The constrained optimization problems described
above can be generalized in terms of the following problem:

\[
\begin{align*}
\min_{x=[x_1, x_2, \ldots, x_U]} & \quad \sum_{u=1}^{U} D_u(x_u) \\
\text{s.t.} & \quad \sum_{u=1}^{U} R_u(x_u) \leq C, \\
& \quad m_u(x_u) \leq \psi_u, \text{ for all } u \in \{1, 2, \ldots U\}.
\end{align*}
\]

(3.26)

Note that \(D_{\text{base,loss}}\) is a constant and can therefore be eliminated from the convexity analysis of the objective function \(D_u(x_u)\). Equation (3.27) lists the components of each of the objective and constraint functions as well as their relation to the terms previously defined in this chapter.

**Objective:**

\[
\begin{align*}
D_u(x_u) &= d_u(x_u) + \int_{L_u}^{x_u} p_u(s) g'_u(s) ds, \quad L_u \leq x_u \leq H_u \\
d_u(x_u) &= \alpha_u + \frac{\theta_u}{x_u - \theta_u}, \\
g'_u(x_u) &= \partial (d_u(L_u) - d_u(x_u))/\partial x_u, \\
P_u(x_u) &= \text{increasing piecewise constant function of } x_u \text{ with range } [0, 1].
\end{align*}
\]

(3.27)

**Constraints:**

\[
\begin{align*}
R_u(x_u) &= (1 + a) \int_{L_u}^{x_u} h_u(s) ds \\
m_u(x_u) &= b \int_{L_u}^{x_u} h_u(s) ds, \\
h_u(x_u) &= \frac{n}{\kappa_u(x_u)} \\
&= \text{decreasing piecewise constant function of } x_u \text{ with range } [1, 2],
\end{align*}
\]

(3.28)

where \(a\) and \(b\) are positive constants \(<1\).

Since \(d_u(x_u)\) is, by definition, a monotonically decreasing convex function of \(x_u\) over the interval \(L_u \leq x_u \leq H_u\), with \(d''_u(x_u) < 0\) and \(d''_u(x_u) > 0\), the first and second derivatives of \(f_u(x_u)\) are shown below:

\[
\begin{align*}
f'_u(x_u) &= d'_u(x_u) + p_u(x_u) g'_u(x_u), \\
&= (1 - p_u(x_u))d''_u(x_u) \leq 0
\end{align*}
\]

(3.29)

\[
\begin{align*}
f''_u(x_u) &= d''_u(x_u) + p'_u(x_u) g'_u(x_u) + p_u(x_u) g''_u(x_u) \\
&= (1 - p_u(x_u))d''_u(x_u) - p'_u(x_u)d'_u(x_u) \geq 0
\end{align*}
\]

with equality at \(p_u(x_u) = 1\). Thus, the objective function \(f_u(x_u)\) is only once differentiable,
The constraint functions $R_u(x_u)$ and $m_u(x_u)$, on the other hand, which are scaled versions of the same function $e_u(x_u)$ shown in (3.30) below, are concave piecewise linear.

$$e_u(x_u) = \int_{L_u}^{x_u} h_u(s) ds$$
$$e'_u(x_u) = h_u(x_u) > 0$$
$$e''_u(x_u) = h'_u(x_u) \leq 0.$$  \hspace{1cm} (3.30)

As a result, we are faced with a non-convex, non-differentiable optimization problem. One method of solving this nonconvex problem is to use augmented Lagrangian algorithms. We refer the reader to [25] for proof of convergence.

As an alternative to augmented Lagrangian, we propose the following method: we linearize the constraints, taking the non-feasible linear envelope of the constraint functions as shown in Fig. 3.7, which yields a convex relaxation to the non-convex optimization problem.

![Figure 3.7: Example of convex relaxation using the linear envelope of the UEP constraint function. The optimal solution $x^*$ to the convex-relaxed problem is non-feasible, therefore, we achieve a feasible solution by projecting onto the feasible non-convex set using the projection operator $P(x^*)$ defined by $P(u) = \arg \min_x \|x - u\|_2^2$ s.t. $x \in \mathcal{X}$, where $\mathcal{X}$ is the feasible set.](image)

The solution to the convex relaxed problem can be easily obtained using standard gradient descent algorithms, in particular we have used quadratic programming (QP) [25]. This solution, however, falls outside the feasible non-convex set. A feasible solution is therefore obtained by projecting the solution of the convex-relaxed problem onto the non-convex set of the original constrained optimization problem described in (3.20). Note that this method does
not guarantee an optimal global solution to the non-convex problem, however, it is a good approximate feasible solution. Our experiments have shown that the solution delivered from convex-relaxation followed by projection onto the feasible set falls close to that of the augmented Lagrangian, while requiring fewer iterations and guaranteeing feasibility. Moreover, the convex-relaxed problem we are solving is reduced to a quadratic programming problem with linear constraints (QPLC). This problem can be solved in polynomial time, where the dimensionality is given by the number of users $U$.

### 3.9 Simulations and Results

In this section, we study the performance of different rate-allocation schemes, which we formulated as the constrained optimization problems of Section 3.7, for multiple video streams transmitted over a lossy channel with different channel capacities and packet loss rates.

For our performance testing, we encoded and prioritized 9 video sequences: Bus, Foreman, Crew, Mobile, Harbor, City, Football, Ice, and Soccer in CIF resolution at 15 frames per second in SVC using JSVM6 available from [26]. The encoded streams are composed of an H.264/AVC compatible base layer and 2 FGS enhancement layers.

#### 3.9.1 Scheme Definition

Below, we define five streaming schemes based on the description of the general streaming system shown in Fig. 3.2, differing in UXP allocation and the rate extraction:

**Fixed UXP Fixed Rate (FUFR)**

The UXP encoder at the server allocates the RS parity symbols for each video stream, such that, the base layer and 2 FGS enhancement layers have maximum PLRs of 0.001%, 1%, and 5%, respectively. Rate extraction is performed by dividing the bit-rate range $[L_u, H_b]$ of each video stream into 1000 steps and then incrementing the base-layer rate $L_u$ by one step-size for each user until the capacity constraint is satisfied. No consideration is given to the delay constraint or fluctuation of video bit-rate over the duration of the streaming session.

**Optimized UXP, Fixed Rate (OUFR)**

The UXP encoder uses the optimized UXP approach described in Section 3.4.1 to allocate the RS parity symbols. Rate extraction is performed similar to the FUFR scheme. This scheme is also un-aware of any delay constraints or video bit-rate fluctuations.
Chapter 3. Channel Aware Multi-User Scalable Video Streaming over Lossy Under-Provisioned Channels

Optimized UXP, Optimized Rate (OUOR)

In this scheme we use the optimized UXP approach illustrated in Section 3.4.1 and the time-averaged rate allocation scheme described in Section 3.7.1. Delay constraints and video bit-rate fluctuation are not considered.

Multi-interval Optimized UXP, Optimized Rate (MOUOR)

The UXP encoder performs optimal UXP allocation at each interval $t$ using an estimated $p_{rtpt}$ value. The video bit-rate fluctuation is taken into consideration, and a multi-interval rate allocation based on the approach in Section 3.7.2 is performed for each of the $T$ intervals to produce the streaming policy matrix $\Pi$. This scheme is still blind to any delay constraints.

Delay-aware Multi-interval Optimized UXP, Optimized Rate (DMOUOR)

This scheme is similar to the MOUOR, with the distinction that it considers the delay constraints in the rate-allocation procedure as described in Section 3.7.3. Moreover, UXP allocation is adjusted at each interval $t$ for optimal error-protection and rate-extraction.

![Figure 3.8](image)

(a) Y-PSNR vs channel capacity
(b) Y-PSNR vs RLC packet loss rate

Figure 3.8: Comparison of the Y-PSNR performance between the four streaming schemes FUFR, OUFR, OUOR, and MOUOR without considering the effect of packet delay as described in Section 3.7.2.
Chapter 3. Channel Aware Multi-User Scalable Video Streaming over Lossy Under-Provisioned Channels

3.9.2 Performance Testing

In this subsection, we compare the performance of each of the aforementioned streaming schemes in terms of the end-to-end distortion. We assume that the application layer UXP and rate allocation is performed at the streaming server prior to the beginning of the streaming session. The RLC loss rate is received through feed-back from the link-layer. The UXP codeword length is set to $n = 128$ bytes with the ability to embed one or more GOPs in a single Transmission Block (TB) in order to ensure that the RTP packet size is close to an integer multiple of the RLC frame size. This helps in reducing the number of stuffing bits used at the MAC/PHY layers.

In order to test the performance of each of the aforementioned streaming schemes, we simulate the lossy channel using the 3GPP off-line simulator available from [27], listed in the “Common conditions for SVC error resilience testing” [28]. Moreover, we have added a UXP encoder and decoder to simulate the specifications listed in [5]. The UXP encoder aggregates all NALUs (base-layer and FGS) belonging to a single GOP into one TB and encapsulates each column of the TB in a single RTP packet according to the “RTP Payload Format for H.264 Video” [29]. We have performed several modifications to the software in [27] in order to add the effect of packet delay and channel capacity on the streaming system. In the case of packet delay, the 3GPP simulator drops all RTP packets belonging to a TB $l$ that exceed the play-out

Figure 3.9: Comparison of the Y-PSNR performance between the four streaming schemes FUFR, OUFR, OUOR, and DMOUOR with the effect of packet delay as described in Section 3.7.3.
Chapter 3. Channel Aware Multi-User Scalable Video Streaming over Lossy Under-Provisioned Channels

deadline of TB $l$.

We simulate link-layer frame loss rates of 1%, 3%, and 10% as provided by the first three radio bearer bit-error traces in [27]. We also consider channel capacities of 3.6 Mbps, 5.3 Mbps, 7.2 Mbps, and 10.7 Mbps which correspond to four feasible channel states in an HSDPA system as listed in Table I of [30]. These four states correspond to the modulation and coding schemes shown in Table 3.1.

Table 3.1: HSDPA Modulation and Coding Schemes

<table>
<thead>
<tr>
<th>Modulation</th>
<th>Code Rate</th>
<th>Data Rate (15 Codes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>QPSK</td>
<td>1/2</td>
<td>3.6 Mbps</td>
</tr>
<tr>
<td>QPSK</td>
<td>3/4</td>
<td>5.3 Mbps</td>
</tr>
<tr>
<td>16QAM</td>
<td>1/4</td>
<td>7.2 Mbps</td>
</tr>
<tr>
<td>16QAM</td>
<td>3/4</td>
<td>10.7 Mbps</td>
</tr>
</tbody>
</table>

As a performance metric, we use the PSNR calculated using the average luminance mean-square-error (MSE) of all 9 streams as shown below:

$$PSNR = 10 \log_{10} \left( \frac{(255)^2}{\sum_{u=1}^{9} MSE_u / 9} \right).$$

Performance without the Delay Effect

We first compare the performance of the four schemes FUFR, OUFR, OUOR, and MOUOR without considering the effect of packet delay. This is to simulate the performance of each of these schemes in situations where packet delay is not a factor. Fig. 3.8 shows the Y-PSNR of the four streaming schemes with respect to the channel capacity and the RLC packet loss rate, respectively. In Fig. 3.8(a), the four schemes are tested using the network simulator with 10% RLC loss rate for four channel capacity state. The figure shows that OUOR outperforms the remaining schemes in all four channel states with an improvement of 0.43 dB at 7.2 Mbps over the nearest scheme. In Fig. 3.8(b), we fix the channel capacity at 3.6 Mbps and vary the link-layer RLC frame loss rate between 1%, 3%, and 10%.

The plots shown in Fig. 3.8 demonstrate the importance of using intelligent rate-allocation strategies in conjunction with optimal UXP protection. Under no delay constraints, both OUOR and MOUOR exhibit better performance relative to the simple FUFR and OUFR schemes. Moreover, the performance improvement of MOUOR, which updates $p_{rtp}$ and the video R-D characteristics at each interval $t$, peaks at intermediate capacity and packet loss rate (PLR) levels. Under the 3.6 Mbps capacity constraint, very few FGS enhancement are transmitted which results in relatively similar performance with the non-optimal streaming schemes. On
the other hand, a 10.7 Mbps capacity constraint is too loose resulting in the transmission of almost all FGS packets available for all the video sequences. These results show that under active constraints that do not approach the acceptable lower bounds of the video bit-rates, non-stationary optimized rate allocation delivers significantly improved performance approaching 0.77dB in Y-PSNR over stationary optimized rate allocation.

Performance with the Delay Effect

We next introduce the effect of packet delay and enforce the TB play-deadlines described in Section 3.7.3. Under a play-out deadline constraint of 533 ms, we compare the performance of each of the four schemes FUFR, OUFR, OUOR, and DMOUOR for varying capacities and varying PLRs as shown in Fig. 3.9 (a) and (b), respectively. Both figures show that DMOUOR achieves significant gains over the other schemes with maximum gains of 2dB for the 10% loss rate and 10.7 Mbps capacity (Fig. 3.9 (a)) and 1.65dB for 10% PLR and 3.6Mbps capacity (Fig. 3.9 (b)).

The performance gains exhibited by the delay-aware DMOUOR scheme as shown in Fig. 3.9 emphasize the importance and relevance of using delay-aware streaming strategy when transmitting delay-sensitive multimedia content. As shown, the improvement reaches a significant 2dB in Y-PSNR for 7.2 and 10.7 Mbps compared to the OUOR scheme which also achieves considerable gain over the non-optimal FUFR and OUFR. It is worth noting that, contrary to expectation, the performance of FUFR and OUFR in the delay sensitive scenario remains constant, if not diminishes, as the channel capacity increases. This results from the non-optimal rate allocation strategy used in these two schemes. Although the channel capacity increases, the rate allocation method used attempts to maximize throughput by embedding more FGS enhancement NALUs but ends up overloading the network. The smarter rate-allocating used in the OUOR scheme results in a relatively better performance than FUFR and OUFR, however, it is no match for the delay-aware DMOUOR scheme. Moreover, by comparing the performance of the DMOUOR scheme, which uses non-stationary R-D characteristics and GOP-based estimation of channel parameters, to the performance of the OUOR scheme, that uses time-averaged estimates of these parameters, we can see that accurate estimation of the system parameters can significantly improve the performance of the streaming system. Fig. 3.10 illustrates the effect of the delay-aware, non-stationary allocation on individual stream performance compared to stationary, delay-unaware scenarios.

Next, we compare the performance of the DMOUOR, OUOR, and OUFR schemes under different delay constraints. Fig. 3.11 demonstrates the improved performance of DMOUOR over the other schemes for different values of the video play-out deadline. The tests are performed for play-out deadlines of 266, 400, 533, 666, and 800 msec with a 5.3 Mbps channel capacity and
Figure 3.10: Comparison of the Y-PSNR performance between the four schemes DMOUOR, OUOR, OUFR, and FUFR for individual video sequences under 3.6 Mbps, 5.3 Mbps, and 7.2 Mbps capacity and 10% PLR.

Figure 3.11: Comparison of the Y-PSNR performance between the three schemes DMOUOR, OUOR, and OUFR for different play-out deadlines over a channel with 5.3 Mbps capacity and 3% PLR.
3% packet loss rate. The figure shows that the improvement of DMOUOR is more significant for tighter delay constraints.

3.10 Conclusion

In this chapter, we have analyzed the performance of five multi-user scalable video streaming strategies in a simulated wireless network scenario which suffers from limited capacity and packet loss. For our analysis, we proposed a new error protection approach using UXP and derived a loss-distortion model for hierarchical predictive coders. We have also derived an analytical expression for the delay and play-out deadlines for UXP protected video streams and incorporated this expression in a delay-aware rate-constrained streaming scheme (DMOUOR). The relevance of delay-sensitivity for wireless video streaming applications has been well emphasized in the literature [12], which makes our work all the more critical. Finally, the performance gains of 1.65 dB to 2 dB in Y-PSNR demonstrated by our delay-aware DMOUOR scheme compared to delay un-aware schemes come at the cost of increased off-line computational complexity which does not affect the run-time performance of the streaming system. In the next chapter we address real-time and pre-encoded multi-user scalable video streaming over IEEE 802.11-based WLANs.
3.11 References


Chapter 3. Channel Aware Multi-User Scalable Video Streaming over Lossy Under-Provisioned Channels


Chapter 3. Channel Aware Multi-User Scalable Video Streaming over Lossy Under-Provisioned Channels


Chapter 4

Dynamic Resource Allocation for MGS H.264/AVC Video Transmission over Link-Adaptive Networks

4.1 Introduction

The rapid growth of broadband wireless services, along with the advancements in video compression technology, will enable future high quality digital video broadcast and telephony applications. While the promise of high quality video applications seems real, there are still challenges that must be addressed before such applications are efficiently deployed. In particular, the variation in wireless link quality has to be taken into account when designing video streaming systems. The variation in wireless channel quality is usually handled using a link adaptation scheme that maintains a certain level of link reliability under varying channel conditions. The use of link adaptation results in multi-rate operation, where lower physical (PHY) layer transmission rates are used to achieve higher reliability under bad channel conditions. With multi-rate operation, the throughput available to each video stream may vary considerably over time, making it impossible to deliver the video stream without significant over-reservation. To counter this situation scalable video coding is used, where a single video sequence is encoded into a base layer and several enhancement layers. For continuity of video playback, it is enough

\footnote{A version of this chapter has been accepted for publication. H. Mansour, Y. P. Fallah, P. Nasiopoulos, V. Krishnamurthy, “Dynamic Resource Allocation for MGS H.264/AVC Video Transmission over Link-Adaptive Networks,” to appear in the IEEE Transactions on Multimedia. 
to receive the base layer, whereas the enhancement layers are only used to increase the quality of the rendered video. When the available throughput decreases, the higher enhancement layers are dropped. This method is the conventional solution for handling capacity variation using scalable video. A traffic control module may be used for this purpose.

To achieve higher efficiency, the operation of the traffic control module and the link adaptation scheme can be jointly optimized. In this chapter, we offer a new traffic control algorithm as well as an optimization framework that achieve the maximum possible video quality for multiple stations that are using Medium Grain Scalable (MGS) video. The work in this chapter differs from the previous works in that it offers a packet drop mechanism tied to an intelligent link adaptation scheme, and it jointly optimizes the operation of these schemes under capacity constraints in a multirate wireless network, for multiple users. We also use a new rate distortion model for MGS video in the optimization framework. Our proposed solution is specifically useful for digital video broadcast services, where a base station (BS) in a wireless domain distributes pre-encoded video content to multiple users in the domain (Fig. 4.1). The presented algorithms will be deployed in the base station. Intensive packet processing is not required in the proposed method. The algorithm adjusts the parameters of the link adaptation and packet drop schemes dynamically, on a per GOP (Group of Pictures) or per frame basis. When H.264 scalable video coding is used, several sub-streams of encoded units are produced for a single video sequence. Packets of each stream are tagged and treated differently in the PHY (link adaptation) and MAC or network layers (packet dropping scheme).

For scalable video coding we consider the medium grain scalable version of the H.264/AVC standard [1], known as SVC [2,3]. The methods presented in this chapter are applicable to all packet based multirate networks such as IEEE 802.11 or IEEE 802.16 [4–6]. For demonstration purposes we consider the physical layer of the IEEE 802.11 Wireless Local Area Network (WLAN).

4.1.1 Main Contributions

The main contributions of this chapter can be listed as follows:

1. We show that the joint optimization of traffic control and link adaptation processes considerably improves the received scalable video quality of all video users and enhances the channel bandwidth utilization in a multi-user wireless video streaming system.

2. We formalize the resource allocation problem statement showing the differences in optimization objectives between existing solutions and our proposed solution. Our proposed technique solves for the PHY modulation $C_l^{(u)}$ and truncation percentage $x_l^{(u)}$, for all video layers $l$ of users $u$, that maximize the expected sum of video PSNRs $D^u$ while satisfying
the service temporal share constraint $\eta$. Moreover, the joint optimization determines the optimal time-share among the multiple users that maximize the defined objective.

3. We show that the joint link-adaptation and traffic-control optimization problem is a non-convex constrained optimization problem with combined discrete and continuous variables. Moreover, we propose an algorithm that first solves a continuous relaxation of the discrete optimization to find the optimal PHY rate points, and next solves for the packet drop ratios using the discrete PHY rate points that lie in the vicinity of the optimal solution of the continuously relaxed problem.

4.1.2 Overview of Related Work

Transmission of scalable video over different types of wireless networks has been studied in several recent works [7,12]. These works consider different aspects and issues of transmitting SVC over wireless networks. The work in [7] presents an optimal solution for transmission of scalable video traffic over multiple input multiple output (MIMO) systems, through optimizing quantization parameters, GOP (Group of Picture) size and channel coding and symbol constellation for a simplified MIMO system. The work presented in [8] attempts to improve the received video quality in a 3G wireless network by optimizing for power and coding rate allocation. Another method relevant to this topic is the work presented in [9], which focuses on employing spatial multiplexing feature of a MIMO system to better deliver scalable video.
This method proposes an adaptive channel selection for scalable video transmission over MIMO channels where partial channel information is derived and used to improve the system performance. In another recent work, [10], an adaptive scheme is provided for assigning different FEC codes to packets of each layer and controlling the transmission rate by dropping higher layers of video. The method reported in [11] presents a schemes for transmission of wavelet based scalable video over WLANs. This method is different from the other works in the literature that consider the more general H.264 based scalable video. In another relevant work, [12] proposes a rate distortion optimized method for transmission of scalable video over CDMA networks.

Our proposed solution differs from existing works in that it offers a combined link adaptation and MAC/Network layer packet drop scheme that maximizes the collective quality of the delivered video to multiple users. We present an optimization framework that adjusts the operation of the link adaptation algorithm in the PHY, as well as the network traffic control (packet drop) module in the MAC or network layers, under bandwidth (temporal fairness) constraints. Our method targets multirate wireless networks such as 802.11 and 802.16, and uses H.264 MGS scalable video. Our proposed solution is specifically useful for digital video broadcast services, where a base station in a wireless domain distributes pre-encoded video content to multiple users. The presented algorithm will be deployed in the base station and does not require intensive packet processing or application layer involvement. We list the main contributions of our work in the following subsection.

The remainder of this chapter is organized as follows: In Section 4.2 we describe the wireless video transmission system and formalize the resource allocation problem. We develop in Section 4.3 the joint link-adaptation and traffic-control optimization problem as a non-convex constrained optimization problem and present an algorithm that achieves the optimal solution. The performance of the proposed joint resource allocation scheme is analyzed in Section 4.4, and finally we present our conclusions in Section 4.5.

### 4.2 General Overview of the Resource Allocation Problem

In this section, we describe the different components of the video streaming system, formalize the problem at hand and discuss the existing solutions and their limitations. We consider a wireless video streaming system in which a transmitter (base-station or access-point) serves multiple video users/clients. The video streaming service is allocated a fixed share of the bandwidth that the transmitter can support, such that all video users compete for a portion of the allocated bandwidth share. Digital video broadcast services are an example of this operation scenario. Fig. 4.1 illustrates an overview of the transmission system described above.
4.2.1 System Overview

Broadband wireless networks are known to be susceptible to variations in wireless link quality. To counter the variation in link quality, a user-defined link adaptation scheme is used in wireless networks such as IEEE 802.11 and IEEE 802.16 networks. This mechanism is also known as rate adaptation in other standards (e.g., 3GPP). The link adaptation scheme adapts transmission parameters to dynamic channel conditions, based on some measured parameters. These parameters are generally measured at the receiver and are delivered to the transmitter through a feedback mechanism. The transmitter uses this information to determine a specific modulation and coding scheme (MCS) for the next packet to be transmitted. Under good channel conditions, more information bits and spectrally efficient modulation schemes can be used; whereas, under bad channel conditions, link adaptation adds resilience to the transmitted signal, reducing the number of delivered information bits per symbol.

Several transmitter parameters can be modified by the link adaptive scheme. The forward error correction (FEC) coding rate and modulation type are the most common. In recent years, several other parameters have also been incorporated into link adaptation, such as transmitter power, cyclic prefix length and switching between space time block coding (STBC), spatial multiplexing (SM) and beam-forming in MIMO systems.

Link adaptation adds an additional constraint to networks that have to deal with limited resources. In communications scenarios where the available capacity is limited, the nodes connected to a link with limited capacity have to restrict the amount of traffic that passes through that link. We call the module that performs this task "traffic control module". This module works in conjunction with the scheduling and admission control schemes which ensure that resources are properly assigned to flows. With link adaptive wireless networks, the task of traffic control becomes more complicated. In this article we assume the existence of the scheduling and admission control modules and focus on the design of the traffic control module instead. The details of how scheduling is done in multi-rate networks are given in [13]. For the wireless access networks considered in this article, the traffic control module is deployed in the BS of the network. The BS is usually connected to the internet or other networks through higher capacity or wired links; nevertheless, traffic control can be performed on any outgoing link (to stations or to the backbone).

4.2.2 Problem Description

Link adaptation and dynamic variation of capacity in wireless networks necessitates dynamic traffic control, which in turn requires the applications to be tolerant of variable throughput. For video applications, scalable video will be used to adapt the video bit-rate to the variable
available throughput. With SVC, the simplest solution to adapt to decreased throughput is to drop the higher enhancement layers.

To formulate the problem of dynamic traffic control we consider the scenario described earlier, where a fixed portion of the available bandwidth is dedicated to the video service and is shared by multiple stations. Let \( u \) be the video user index and let \( l \in \{0, 1, ...L\} \) be the scalable video layer index, such that, layer 0 corresponds to the base layer and layer \( L \) the highest (lowest priority) enhancement layer. Let \( r_i^u \) be the estimated video bit-rate required by layer \( l \) of user \( u \), and let \( C_i^u \) be the PHY transmission rate allocated for layer \( l \) of user \( u \). We define the temporal share \( \tau_i^u \) occupied by video layer \( l \) of user \( u \) as follows:

\[
\tau_i^u = \frac{r_i^u}{C_i^u}. \tag{4.1}
\]

In this chapter, we assume that the scheduling algorithm is able to assign a fair share of the bandwidth, or a fair time share, to each user (video stream) or an aggregate of users \([13]\). The fair scheduler provides the total time share \( \eta \) for all video users (aggregate of their traffic) that share the streaming service. Within the aggregate service the time shares are distributed according to the method presented in this chapter. The aim here is to maximize the total quality of the delivered video under the time share constraint for the entire video service. Here we propose an optimization framework that achieves this goal. The proposed method provides significant improvement over existing mechanisms, which are described in the next section.

4.2.3 Overview of Existing Resource Allocation Solutions

The 802.11 or 802.16 standards do not mandate a specific link adaptation scheme. The most common mechanism of selecting between the different transmission rates is to use SNR thresholds for the different transmission rates that meet a maximum error constraint, usually 10% PER for packet length of 1000 bytes \([4]\). The highest transmission rate that meets the maximum PER requirement is selected. To adapt the scalable video traffic to rate changes, different schemes may be used, as explained below.

**Conventional Layer Drop**

The conventional and simple solution for adapting the bit-rate of the scalable video to the available throughput is to drop the higher layers of video, until the remaining traffic fits in the available throughput for the stream. We call this method “conventional layer drop” (CLD), in
Chapter 4. Dynamic Resource Allocation for MGS H.264/AVC Video Transmission over Link-Adaptive Networks

this chapter. The CLD method can be formulated as follows:

$$\max_{C^u, L} \hat{L} \quad \text{s.t.} \quad \sum_{l=0}^{\hat{L}} \tau_l^u(C_l^u) \leq \eta^u,$$

and $p_l(C_l^u) \leq 10\%$ \hspace{1cm} (4.2)

where $\hat{L} \in \{0, 1, \ldots, L\}$ is the highest admissible video layer, $L$ is the maximum number of available video layers, $p_l$ is the packet error probability associated with the selected transmission rate $C_l^u$ for user $u$, and $\eta^u$ is the temporal share of user $u$.

With CLD, the transmission rate of all video layers belonging to each user will be the same. The transmission rates are set by the link adaptation scheme regardless of the type of traffic. The objective of CLD is to fit as many layers as possible in the available bandwidth or time share for the flow.

**Intelligent Link Adaptation**

A more intelligent method to is to replace the conventional link adaptation, and layer drop schemes, with a combined link adaptation and layer drop scheme that assigns different PHY rates to each layer and drops the enhancement layers that do not fit in the available throughput for the stream. This method is called Intelligent Link Adaptation (ILA) and has been studied in detail in \cite{14}. ILA is based on per-stream traffic control, and aims at improving the video quality of users individually, and on a long term average basis. We formulate the ILA scheme as shown below:

$$\min_{L, C^u} \bar{D}^u(\hat{L}, C^u) \quad \text{s.t.} \quad \sum_{l=0}^{\hat{L}} \tau_l^u(C_l^u) \leq \eta^u,$$

where $\bar{D}^u(\hat{L}, C^u)$ is the time-averaged video distortion model given as a function of the maximum number of admissible video layers $\hat{L}$ and the PHY transmission rate $C^u$.

The solution to this problem consists of a combinatorial search over all achievable PHY rate points (and corresponding PER points) to choose the set that minimizes the video distortion function $\bar{D}^u$. Since ILA considers video streams individually, it can be used at any source that transmits the video, i.e., the base station or the subscriber stations.

**4.2.4 Brief Overview of the Proposed Solution**

In a multi-user system, combining the resources assigned to all users will allow dynamic allocation of these resources on a short term basis, which results in more efficient use of the limited capacity and higher overall video quality. Our proposed technique solves for the PHY modulation $C_l^u$ and truncation percentage $x_l^u$, for all video layers $l$ of users $u$, that maximize
the expected sum of video PSNRs $D^u$ while satisfying the service temporal share constraint $\eta$. Moreover, the joint optimization determines the optimal time-share among the multiple users that maximize the defined objective. A detailed description of this technique is presented in Section 4.3.

The rate-distortion (R-D) model used in the ILA scheme is based on the model proposed in [15] and extended in [14]. This model has its shortcomings in that it cannot be used to predict the R-D behavior of the next video frame or group of frames (GOF). Therefore, these models can only be used when long term average quality is concerned. In this chapter we will use the real-time rate and distortion models developed in Chapter 2 for MGS scalable coded video streams which can be used in online or batch traffic management and resource allocation algorithms on short term (GOF size) resource allocation decision making. Note that the ILA scheme minimizes the sum of video MSEs whereas the algorithm proposed in Section 4.3 maximizes the sum of video PSNRs.

Figure 4.2: The Joint operation of the Link-Adaptation and Traffic-Control modules in our proposed framework. The algorithm combines the resources ($\vec{R}, N, \vec{\tau}$) assigned to all users allowing dynamic allocation of these resources on a short term basis. This results in more efficient use of the limited capacity $\eta$ and higher overall received video quality $\vec{D}$ of all transmitted streams that share the same wireless channel.
4.3 Proposed Resource Allocation Scheme

In this section, we present a new link-adaptation scheme for real-time resource allocation of multiple MGS scalable video streams. The objective of our optimization framework is to maximize the total quality of multiple video streams that share the same wireless channel, while maintaining the total load on the network constant, by adjusting the PHY modulation and video layer truncation percentages.

4.3.1 Joint Link Adaptation and MAC/Network Traffic Control

We propose a joint link adaptation and traffic control scheme which strives to optimize the use of the assigned time share (capacity) by achieving maximum video quality given the channel conditions. The optimization framework specifies the PHY transmission rate of each video layer (for the GOF duration), which is then enforced by the link adaptation module. The optimization framework also specifies the percentage of the packets that should be dropped from each layer of each video stream, this decision is then enforced by the Network or MAC layer traffic control module which selectively blocks or drops some packets from each video layer. Fig. 4.2 illustrates the interaction between the link adaptation and traffic control modules.

We consider a wireless system where \( U \) scalable video (MGS) users/clients are served over a single wireless channel with a streaming service constraint \( \eta \), where \( \eta \leq 1 \). For simplicity of illustration and without loss of generality, we assume that each video stream \( u \) is composed of a base-layer and \( L^u = 2 \) CGS enhancement layers. The scalable video streams are characterized in terms of their rate-distortion parameters: \( \{ \tilde{d}^u_0, \tilde{d}^u_1, \tilde{d}^u_2 \} \) and \( \{ \tilde{r}^u_0, \tilde{r}^u_1, \tilde{r}^u_2 \} \) given by equations (2.32) – (2.35) above. The rate and distortion parameters are calculated for fixed values of base and enhancement layer QPs.

In a multirate wireless physical layer considered here, different PHY rates are achieved by changing the modulation and coding schemes, resulting in different packet loss ratios under a given SNR \( \gamma \). We denote such loss ratio as \( p_l = f(C_l, \gamma) \), for each video layer assigned a PHY rate \( C_l \). The relationship between PHY rate and PER in a MIMO PHY is not straightforward. Fortunately, we can bypass an exact analysis and construct a one-to-one mapping between the values of \( C^u \) and \( P^u \) from the BER performance curves shown in Fig. 4.3. These curves are obtained through system simulations under typical channel conditions similar to the work done in [14]. BER performance curves for different conditions and networks can be used too, and the formulation of the problem does not depend on these curves, as long as there exists a one to one mapping of \( C^u \) and \( P^u \) values. The relationship between \( C_l \) and \( p_l \) can be seen in Fig. 4.4.

Let \( t \) and \( G \) be the GOF index and GOF size, respectively. We define \( \tilde{D}^l_t \) to be the predicted
average PSNR for GOF $t$ of video stream $u$, such that

$$\hat{D}_t(P) = p_0 d_{EC} + (1 - p_0)\tilde{d}_0$$
$$\quad + (1 - p_1)(1 - p_0)(\tilde{d}_1 - \tilde{d}_0)$$
$$\quad + (1 - p_2)(1 - p_1)(1 - p_0)(\tilde{d}_2 - \tilde{d}_1)$$

(4.4)

where $d_{EC}$ is the distortion due to the error concealment mismatch, and $P = f(C, \gamma) = [p_0 p_1 p_2]^T$ is the packet error rate (PER) vector. Note that we dropped the superscript $u$ from the expression above for clarity of illustration.

The predicted temporal time share of a GOF $t$ of stream $u$, denoted as $\tilde{\tau}_t^u$, can be expressed as follows:

$$\tilde{\tau}_t(C) = \frac{\tilde{r}_0}{C_0} + \frac{\tilde{r}_1}{C_1} + \frac{\tilde{r}_2}{C_2}.$$ 

(4.5)

We propose to employ a packet dropping/stopping stage in the proposed scheme that blocks the transmission of some packets of each layer, effectively reducing the bit-rate of the layer.

Let $x_l$ be the percentage of blocked packets of layer $l$, resulting in an effective video layer bit-rate $\tilde{R}_l = \tilde{\tau}_l(1 - x_l)$. Therefore, we can rewrite the video stream’s predicted PSNR and
corresponding time share as shown below:

\[
\tilde{D}_t(X, C) = p_0 d_{EC} + (1 - p_0) \tilde{d}_0 \\
+ (1 - x_1)(1 - p_1)(1 - p_0)(\tilde{d}_1 - \tilde{d}_0) \\
+ (1 - x_2) \prod_{l=0}^{2}(1 - p_l)(\tilde{d}_2 - \tilde{d}_1)
\]

\[
\tilde{\tau}_t(X, C) = \frac{\tilde{r}_0}{C_0} + (1 - x_1)\frac{\tilde{r}_1}{C_1} + (1 - x_2)\frac{\tilde{r}_2}{C_2}
\]

(4.6)

where \( l \) is the video layer index, and \( C_l \) is the PHY transmission rate of layer \( l \). The video layer dependency is accounted for in the above formulation by setting \( x_1 \leq x_2 \), such that, if \( x_1 \) of the packets are dropped from layer 1 then their dependent packets are also dropped from layer 2. Notice that we do not assign a drop rate for the base layer to avoid the costly quality degradation associated with error concealment mismatch and to satisfy quality of service (QoS) guarantees.

The new multi-user resource allocation problem can now be formulated as the following constrained optimization problem with variables \( X \) and \( C \):

\[
\max_{X, C} \sum_{u=1}^{U} \tilde{D}_t^u(X^u, C^u)
\]

subject to

\[
\sum_{u=1}^{U} \tilde{\tau}_t^u(X^u, C^u) \leq \eta_t,
\]

\( C_0^{u} \leq C_1^{u} \leq C_2^{u} \),

\( 0 \leq x_1^{u} \leq x_2^{u} \leq 1 \) for all \( u \),

(4.7)

where \( X = [X^1 X^2 ... X^U] \), \( X^u = [x_1^u x_2^u]^T \), \( C = [C^1 C^2 ... C^U] \), and \( C^u = [C_0^u C_1^u C_2^u]^T \) defined in (4.6), \( u \in \{1, 2, ... U\} \) is the user index.

The constrained optimization problem defined in (4.7) is a mixed integer non-linear programming problem whose solution is non-trivial. The variables \( X^u \in [0,1] \) are continuous, while \( C^u \in C \) are discrete and \( C = \{C_0, C_1, ..., C_M\} \) is the set of feasible PHY transmission rates associated with the current channel SNR, and \( M \) is the cardinality of \( C \). In the next subsection, we will discuss our proposed solution to the problem in Eq. (4.7).

### 4.3.2 Solution of the Global Optimization Problem

Mixed integer programming problems are typically NP-hard. Therefore, we develop an algorithm in this section to solve the above mentioned problem (defined in (4.7)) by first solving the continuously-relaxed problem illustrated below, then projecting the solution of the continuous relaxation on the set of achievable PHY rate points and solving for the layer drop rates \( X^u \)
that satisfy the temporal share constraint $\eta$.

**Continuous relaxation of the optimization problem**

The continuous relaxation of the discrete constrained optimization problem of equation (4.7) can be performed by replacing the discrete variable $C^u$ with a continuous valued variable $W^u = [w_0^u w_1^u w_2^u]^T \in \mathcal{W}^3$, where $\mathcal{W} = [C_0, ..., C_M]$, and finding the continuous approximation of the function $P(C^u, \gamma^u)$, where $\gamma^u$ is the channel SNR of user $u$.

We re-write the continuously relaxed constraint and objective functions as follows:

$$
\tilde{D}_t(W) = \rho(w_0) d_E + (1 - \rho(w_0)) \tilde{d}_0 + (1 - \rho(w_1))(1 - \rho(w_0))(\tilde{d}_1 - \tilde{d}_0) + \prod_{l=0}^{2}(1 - \rho(w_l))(\tilde{d}_{2} - \tilde{d}_{1})
$$

and

$$
\tilde{r}_t(W) = \tilde{r}_w^0 + \tilde{r}_w^1 + \tilde{r}_w^2,$$

where $\rho(W^u, \gamma^u)$ is the continuous approximation of the PER function $P(C^u, \gamma^u)$.

**Normal Approximation of the PER Curves**

The PER vs PHY transmission rate relationship is given by a discrete set of achievable points that are a direct consequence of the channel SNR and the employed modulation and coding scheme. We have found that this relationship can be modeled using the continuous complementary CDF of the normal distribution $N(np_c, np_c(1 - p_c))$ shown below:

$$
\rho(n, C, p_c) = \frac{1}{2} \left[ 1 - \text{erf} \left( \frac{n - C - \mu}{\sqrt{2}\sigma^2} \right) \right],
$$

where $C$ is the PHY transmission rate, $\mu = np_c$ and $\sigma^2 = np_c(1 - p_c)$, $\text{erf}(.)$ is the Gauss error function, and $n$ and $p_c$ are model parameters calculated offline.

Table 4.1 shows an example of the valid MCS indices for a 2x2 802.11n WLAN channel used in the construction of the mapping between $C^u$ and $P^u$.

Fig. 4.4 shows the achievable PER vs PHY rate curves and their continuous approximation using the gaussian complimentary CDF for different channel SNRs. The packet error rate values are calculated for an average packet length of 1000 bytes. It can be seen from the figure that the function $\rho$ of (4.9) is non-convex over the full range of $W$. Fig. 4.5 illustrates the non-convex shape of $\rho$ over the full range of $W$, which renders the constrained optimization problem non-convex. Since $\rho$ is either 0 or 1 for most of the range of $W$, and a PER rate greater than 50% is simply useless, we have narrowed down the feasible range of $W$ to the convex interval
Table 4.1: Valid MCS Indices for 2x2 802.11n WLAN

<table>
<thead>
<tr>
<th>MCS</th>
<th>Bitrate (Mbps)</th>
<th>Code Rate</th>
<th>SM streams</th>
<th>Modulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>6.5</td>
<td>1/2</td>
<td>1</td>
<td>BPSK</td>
</tr>
<tr>
<td>1</td>
<td>13</td>
<td>1/2</td>
<td>1</td>
<td>QPSK</td>
</tr>
<tr>
<td>2</td>
<td>19.5</td>
<td>3/4</td>
<td>1</td>
<td>QPSK</td>
</tr>
<tr>
<td>3</td>
<td>26</td>
<td>1/2</td>
<td>1</td>
<td>16-QAM</td>
</tr>
<tr>
<td>4</td>
<td>39</td>
<td>3/4</td>
<td>1</td>
<td>16-QAM</td>
</tr>
<tr>
<td>5</td>
<td>52</td>
<td>2/3</td>
<td>1</td>
<td>64-QAM</td>
</tr>
<tr>
<td>6</td>
<td>58.5</td>
<td>3/4</td>
<td>1</td>
<td>64-QAM</td>
</tr>
<tr>
<td>7</td>
<td>65</td>
<td>5/6</td>
<td>1</td>
<td>64-QAM</td>
</tr>
</tbody>
</table>

Figure 4.4: Example of the achievable packet error rate (PER) curves and their continuous approximation as a function of the PHY transmission rates for different SNRs.

where $0 < \rho \leq 0.5$. This interval corresponds to the following bounds for $W$:

$$n - \mu - \text{erf}^{-1}(0.99)\sqrt{2\sigma^2} \leq W \leq n - \mu, \quad (4.10)$$

where $\mu$ and $\sigma^2$ are the mean and variance of the approximate normal distribution function shown in equation (4.9). The new feasible region is now convex and corresponds to the highlighted area in Fig. 4.5.

Joint Link-Adaptation and Traffic-Control (JLATC) Algorithm

For simplicity of presentation, we describe the algorithm for a three layer scenario. The continuously-relaxed constrained optimization problem can now be written as shown below:
Chapter 4. Dynamic Resource Allocation for MGS H.264/AVC Video Transmission over Link-Adaptive Networks

Figure 4.5: Plot of the function $\rho(.)$ showing the non-convex shape of the function over the full range of $W$ and the convex region over the reduced feasible interval.

\[
\begin{align*}
\max_{W} & \quad \sum_{u=1}^{U} \tilde{D}_u(W^u) \\
\text{subject to} & \quad \sum_{u=1}^{U} \tilde{\tau}_u(W^u) \leq \eta_t, \\
& \quad n - \mu - 2\sqrt{2\sigma^2} \leq w^u_l \leq n - \mu, \\
& \quad w^u_0 \leq w^u_l \leq w^u_2, \\
& \quad \text{for all } u \in \{1,2...U\}, l \in \{0,1,2\},
\end{align*}
\]  

(4.11)

where $W = [W^1W^2...W^U], W^u = [w^u_0w^u_1w^u_2]^T$ is the continuous valued PHY rate defined in (4.8).

The solution to Eq. (4.11) is very straightforward, since the problem is convex, and results in the continuous valued PHY rate allocation matrix $W^*$.

The optimal rates $W^*$, however, are not realizable since the transmitter can only use the predefined set of discrete PHY rates $C$. Moreover, each optimal rate point $w^*_u$ falls between two achievable PHY rate points in $C$. Therefore, for each optimal rate vector $W^u = [w^u_0w^u_1w^u_2]^T$ of user $u$, we propose the following approach:

1: Find the rate points $C'_l$ and $C''_l$, for all $0 \leq l \leq 2$, such that $C'_l \leq w^u_l \leq C''_l$, and $C'_l, C''_l \in C$ are the discrete achievable PHY rate points adjacent to $w^*_u$.

2: Form the set of vectors

\[
\hat{C} = \begin{bmatrix}
\hat{C}_0 \\
\hat{C}_1 \\
\hat{C}_2
\end{bmatrix},
\]

where $\hat{C}_0 = C'_0$, $\hat{C}_1 \in \{C'_1, C''_1\}$, and $\hat{C}_2 \in \{C'_2, C''_2\}$.  

105
3: For every vector $\hat{C}$, find the layer drop rates $\hat{X}^u$ that solve the following equation:

$$\max_{X} \sum_{u=1}^{U} \tilde{D}_t^u(\hat{C}, \hat{X}^u)$$

subject to $\sum_{u=1}^{U} \tilde{\tau}_t^u(\hat{C}, \hat{X}^u) \leq \eta_t$, and $0 \leq x_1^u \leq x_2^u \leq 1$ for all $u$. 

$$\text{(4.12)}$$

4: Choose the set of mapped PHY rate points $C^u$ and layer drop rates $X^u$ that results in the maximum value for the objective function in (4.12) as shown below:

$$\{C^u, X^u\} = \arg \max_{\{C, X\}} \tilde{D}_t^u(\hat{C}, \hat{X}).$$

Figure 4.6: Example of the mapping process involved in discretizing the continuous valued optimal rate points $w_l^*$ to the achievable discrete PHY rates $C_l'$ or $C_l''$, where $l \in \{0, 1, 2\}$. 

Fig. 4.6 illustrates the mapping process from the optimal rate vector $W^*$ to the achievable PHY rate points. Note that $w_0^*$ is mapped to the lower rate point $C_0'$ to avoid a possible increase in PER for the base layer, while $w_1^*$ and $w_2^*$ can each be mapped to two possible rate points $C_1', C_1''$ and $C_2', C_2''$, respectively. 

This algorithm, in effect, narrows down the combinatorial search range limiting it to the achievable PHY rate points that are adjacent to the optimal solution given by $W^*$. 

106
Figure 4.7: Comparison in average Y-PSNR performance between the three resource allocation schemes: JLATC (proposed), ILA, and CLD. The performance of each scheme is plotted versus different channel SNRs and the following service temporal share values: (a) $\eta = 0.4$, (b) $\eta = 0.5$, (c) $\eta = 0.65$. 

107
4.4 Performance Evaluation

In this section, we compare the performance of our proposed streaming approach to existing streaming schemes. For our comparisons, we used the JSVM-9-8 reference software [16] to encode 7 reference video sequences in CIF resolution: Foreman, Football, Bus, Mobile, Soccer, Crew, and City. Each sequence is composed of 150 frames encoded in SVC, with an H.264/AVC compatible base layer, and two MGS enhancement layers. The coded video streams are composed of I and P frames only, with an IDR period of 32 frames, a frame rate of 30 fps, and minimum QoS guarantee corresponding to the base-layer quantization parameter value of $q_0 = 38$. The first and second MGS enhancement layers have QP values of 32 and 26, respectively.

The channel BER performance curves were obtained using an elaborate MATLAB model developed at UBC for the 802.11a/n PHY, based on the work in [17] and [18]. For the IEEE 802.11n model Channel Model D of [17] was used. A more accurate doubly exponentially decaying multi-cluster PDP is also used. In [18], a MATLAB implementation of MIMO channels based on [17] is provided. This program is used to generate channel responses which are fed into our model to obtain BER performance curves. Using bit error rate (BER) curves of all valid combinations of modulation and coding, a look up table is constructed. This look up table is used to generate one-to-one mapping between the PHY rate and PER which are approximated using the normal cumulative CDF curve as described earlier in Section 4.3.2. For simplicity of simulations and experiments we have only included the performance curves for the Space Time Block Coding (STBC) modes of a 2X2 802.11n MIMO system. This does not have any effect on the generality of the optimization problem, and the results are easily extendable. BER curves for all STBC MCS indices for a 2x2 MIMO system can be seen in Fig. 4.3 and Table 4.1. The values of MCS are 0 to 7, with corresponding bitrates of: \{6.5, 13, 19.5, 26, 39, 52, 58.5, 64\} Mbps.

Furthermore, we assume that video frames are already packetized and no intrusive packet processing is done. This is of particular importance, since most intermediate network devices, such as WiMAX base stations, should avoid deep packet processing. Instead, we use all available options in the MAC and PHY layers (which are closely monitored and controlled by base and subscriber stations) to provide the best possible service.

To demonstrate the advantages of our proposed Joint Link Adaptation and Traffic Control (JLATC) algorithm, we compare its performance with that of the ILA and CLD schemes. Resource allocation using each of the three schemes is performed for every group of eight pictures. Therefore, the ILA model parameters are calculated for every GOF of size $G = 8$ and stored at the transmitter. A major drawback for the ILA scheme is that the model parameters of future GOFs cannot be predicted from the parameters of the current GOF. As a result, the
ILA scheme suffers in real-time applications and can only use the initial model estimates for the entire streaming session. Our JLATC, on the other hand, requires only the model parameters of the first GOF which it uses consequently to predict and update the model parameter of future GOFs.

For our performance evaluations we use the video luminance peak signal-to-noise ratio (Y-PSNR) as the picture quality metric, and compare the received video PSNR from each resource allocation scheme for different channel SNR values and different bandwidth budgets $\eta$ for the streaming service. Fig. 4.7 illustrates the results of the comparison in terms of the average Y-PSNR for all seven streams. The figure shows that for a temporal share $\eta = 0.4$ which corresponds to tight bandwidth constraints, the proposed JLATC scheme significantly outperforms its closest contender, the ILA scheme, by an average of 1.2 dB for different channel SNR values. The performance gain peaks at 2.3 dB in average Y-PSNR in the case where the channel SNR = 12.5 dB. In the case where the channel is not congested the plots show that both the JLATC and ILA schemes approach the maximum achievable video PSNR that is given by the video stream encoding. The effect of the three resource allocation schemes on the performance of
each individual stream can also be seen in Fig. 4.8.

The JLATC performs better than the ILA scheme because it can utilize the available channel bandwidth, given by the temporal share $\eta$, more efficiently than the ILA scheme. Fig. 4.9 demonstrates the improved temporal share utilization of our proposed JLATC scheme over the ILA and CLD schemes. The figure shows the temporal share utilization and corresponding average video PSNR for every GOF of streamed video frames.

If we consider the temporal share utilization and corresponding PSNR performance of the three schemes shown in Fig. 4.9 (a) and (b), it can be seen that the JLATC scheme results in considerable improvement in PSNR over the ILA scheme for a small improvement in temporal share utilization. Such behavior arises from the different PHY rate allocation schemes involved in each scheme. The ILA scheme solves the combinatorial optimization problem defined in (4.3) which provides the admitted video layers with high error protection sacrificing the higher video layers. The JLATC scheme performs the PHY rate allocation by solving the continuously-relaxed problem and then mapping the optimal solution to the achievable PHY rate points. This procedure results in a different PHY rate allocation than the ILA scheme which has clearly shown to produce better video PSNR performance. The reason for such performance is that when the ILA scheme cannot admit a video layer or is forced to allocate a PHY rate with high PER to the video layer, a large portion of the available temporal share is used to provide more error protection for the lower video layers. This problem is avoided by the JLATC scheme by enforcing the constraint given by Eq. (4.10) which limits the feasible interval of the PHY rate to the convex region of the PER curves.
Figure 4.9: Comparison of bandwidth utilization and corresponding video Y-PSNR between the three resource allocation schemes: JLATC (proposed), ILA, and CLD. The plots correspond to a streaming temporal share budget $\eta = 0.5$ and channel SNR = 12.5 dB for (a) and (b), and SNR = 15 dB for (c) and (d).
4.5 Conclusion

In this chapter, we show that jointly optimizing the link-adaptation and traffic control of a multi-user wireless video streaming system considerably improves the received video quality for all video users and enhances the channel bandwidth utilization. Multi-rate networks such as IEEE 802.11 or 802.16 use link adaptation mechanisms in the physical layer to adjust the modulation and coding schemes that control the reliability of transmission for varying channel conditions. Conventional schemes to link-adaptation and quality control drop the enhancement layers of scalable coded video streams in order to match the reduced PHY bit-rate required to achieve higher reliability. We propose a new resource allocation scheme that takes advantage of the robust and efficient delivery provided by scalable coded video streams to jointly optimize the link-adaptation scheme in the PHY layer with a new traffic control module in the network or MAC layer. The traffic control module operates by dropping the necessary percentage of packets from enhancement layers that cannot be fully transmitted, in order to fully utilize the available transmission bandwidth. We show that the resulting joint optimization problem is a non-convex constrained optimization problem with combined discrete and continuous variables. Since, the solution to this problem is non-trivial, we propose an algorithm that first solves a continuous relaxation of the discrete optimization to find the optimal PHY rate points, and next solves for the packet drop ratios using the discrete PHY rate points that lie in the vicinity of the optimal solution of the continuously relaxed problem. Moreover, the algorithm defines operational intervals that maintain the convexity of the continuously relaxed problem. Performance evaluations comparing our proposed scheme with existing solutions show that the proposed approach results in significant gains in terms of average video PSNR that can reach 3 dB in some cases for different channel SNRs and different bandwidth budgets.
4.6 References


Chapter 5

Discussion and Conclusions

5.1 Significance of the Research

In the introduction we discussed how the emergence of digital video coding during the last two decades has revolutionized the manner in which different sectors of human society interact. Coupled with a leap in wireless data transmission technology, mobile digital video content is allowing us to experience the dawn of an era of video mobility in which the moving picture can accompany us everywhere in our day to day activities. The impact of these services has opened a global market for new multimedia services stretching from the entertainment industry to the business sector to creating a virtual arena for political debate and news reporting.

In this thesis, we addressed different aspects of wireless video transmission of scalable video content to multiple users over under-provisioned links. We approached the problem from a media-aware resource allocation perspective, in which the video payload is characterized in terms of its bit-rate and expected distortion at the receiver.

We started by deriving analytical rate and distortion models that characterize the dynamic behavior of scalable coded video content. In Chapter 2, we derived analytical rate and distortion models for video bitstreams coded using the single layer and CGS/MGS scalable extension of H.264/AVC. We then presented simplified empirical models that approximate the analytically derived model performance and demonstrated the usefulness of these models in a real-time video streaming scenario.

Next, we addressed the problem of multi-user streaming of pre-encoded scalable video content over 3G networks. In Chapter 3, we derived a loss-distortion model for hierarchical predictive coders and an analytical expression for the packet delay and play-out deadline of UXP protected scalable video. We also analyzed the performance of delay-aware, capacity-aware, and rate-minimized UXP in conjunction with optimized rate-allocation compared to simpler streaming strategies.

Finally, we considered the transmission of pre-encoded and real-time scalable video content over IEEE 802.11-based WLANs. We approached this task as a service time constrained optimization which has the advantage over throughput fairness in that the variation in PHY transmission rate for one station does not affect the other stations. In Chapter 4, we pre-
Chapter 5. Discussion and Conclusions

Presented two media-aware scalable video transmission schemes over multi-rate PHY IEEE 802.11 WLANs. The first approach, called Intelligent Link Adaptation (ILA), optimizes the allocation of scalable video enhancement layers under service time share constraints. The second approach, which supersedes ILA, jointly optimizes the operation of link-adaptation and a traffic control module in order to add a selective packet dropping functionality to the overall streaming system.

5.2 Potential Applications

Digital video transmission applications include multimedia messaging, video streaming, and video conferencing over a wide range of modern transmission and storage systems. These include mobile TV, internet video streaming over wired and wireless networks, and SD and HD TV broadcast to DVD and Blu-ray disc players [1]. Our research addresses some of the problems faced by digital video transmission that arise from the diversity of the new digital video services and heterogeneity of network infrastructures. These networks include broadband ad hoc wireless networks and high-throughput mobile 3G/3.5G networks, such as, IEEE 802.11-based wireless local area networks (WLAN) [2,3], high-speed packet access (HSPA) [4-6] networks, multimedia broadcast and multicast services (MBMS) [7], and the next generation of HSPA long term evolution (LTE) [8] networks.

Other potential applications arise from the analysis and development of video rate and distortion models. In addition to their demonstrated effective use in video transmission policies, rate and distortion models can be used for video rate control, enhancement of the rate-distortion optimization process, and the improvement of the transform and quantization schemes in future video codecs.

Another application of R-D modeling is the compression of High Dynamic Range (HDR) video content with backward compatibility constraints. HDR imaging is a representation discipline which matches the human visual system capability of processing much wider ranges of light intensities than current display systems [9]. In order to exhibit the HDR content on existing, namely low-dynamic-range (LDR), display devices, the dynamic range of the HDR image should be reduced using a lossy process called tone-mapping. A distorted version of the HDR sequence can then be recovered from its LDR representation using an inverse tone-mapping process. We are currently investigating optimal tone-mapping methods that use video distortion modeling to generate an LDR representation. The aim is to find the best LDR representation, which after video compression, would deliver the best possible inverse tone-mapped HDR representation of the original HDR video sequence.
Chapter 5. Discussion and Conclusions

5.3 Contributions

The main contributions of this thesis are summarized as follows:

- We developed analytical rate and distortion models that capture the dynamic behavior of scalable coded video content using the CGS/MGS feature in SVC. These models are based on the assumption that the DCT coefficients follow a Laplacian distribution. In the cases when the true coefficient distribution diverges from the Laplacian assumption, we propose linear scaling of the Laplacian parameter \( \lambda \) to compensate for error in the assumption and show that such scaling results in an accurate estimate of the rate and distortion behavior.

We also presented simplified rate and distortion models with parameters that can be derived from a limited number of empirical samples. We show that the simplified model performance approaches the analytically derived model performance with an acceptable loss in accuracy.

- We proposed a rate-minimized unequal erasure protection (UXP) scheme based on the UXP scheme proposed by Liebl et al. \[10\] and adopted for SVC in \[11\]. The proposed rate-minimized scheme balances the tradeoff between increasing the protection overhead and reducing the UXP failure rate.

We derived an analytical expression for packet delay and play-out deadline of UXP protected scalable video. A group of coded video frames (GOP) is embedded in a single UXP transmission block (TB) thus changing the play-out deadline of the protected packets. This modified packet sequence alters the regular transport stream structure. Therefore, we derived an upper bound for the number of radio link control (RLC) frames allowed per TB as a function of the RLC frame delivery delay.

We derived a loss-distortion model for hierarchical predictive video coders using picture copy error concealment. A loss-distortion video model is essential for media-aware streaming policies which require extensive knowledge of the video encoding structure since it reflects the sensitivity of the decoded video stream to packet losses affecting the bitstream during transmission.

We analyzed the performance of delay-aware, capacity-aware, and rate-minimized UXP in conjunction with optimized rate-allocation compared to simpler streaming strategies. Although delay and capacity constrained rate-allocation requires more computation, the bulk of the computation workload is performed off-line and does not affect the run-time performance of the system.
• We showed that a UXP protected rate-constrained optimization problem is non-convex, to which respect we propose an approximate solution using convex-relaxation and projection onto the feasible set.

We showed that the joint optimization of traffic control and link adaptation processes considerably improves the received scalable video quality of all video users and enhances the channel bandwidth utilization in a multi-user wireless video streaming system. Moreover, the joint optimization determines the optimal time-share among the multiple users that maximize the defined objective.

We demonstrated that the joint link-adaptation and traffic-control optimization problem is a non-convex constrained optimization problem with combined discrete and continuous variables. Moreover, we proposed an algorithm that first solves a continuous relaxation of the discrete optimization to find the optimal PHY rate points, and next solves for the packet drop ratios using the discrete PHY rate points that lie in the vicinity of the optimal solution of the continuously relaxed problem.

Note that the main difficulty in some of the problem formulations we addressed lies in the complexity of the optimization problems that arise. These problems range from non-convexity issues in some cases to mixed integer programming problems in other cases. In all of these situations, we have devised fast algorithms that either achieve optimality or achieve a close approximation of the optimal solution through continuous relaxation, convex relaxation, and reduction of the dimensionality of integer programming problems.

5.4 Future Research

We envision several ideas with which this research can be extended in order to improve and complement the current schemes.

5.4.1 Enhanced Rate-Distortion Modeling

In the case of rate-distortion modeling, it would be worth investigating the effectiveness of deriving rate and distortion models for generalized Gaussian distributed sources. Moreover, the H.264/AVC entropy coder takes advantage of the correlation between adjacent transform coefficient samples, while our current models assume that all samples are coded independently. One idea would be to attempt to incorporate these coding dependencies in the calculation of the quantized coefficient entropy. There are other features supported by the Fidelity Range Extension (FRExt) of H.264/AVC which we do not incorporate in our current models. One of
these features is the use of an 8x8 block integer transform instead of the 4x4 block transform, which affects the transform coefficient distribution.

5.4.2 Increasing Video Streaming Complexity

The video transmission scenarios presented in Chapters 2, 3, and 4 can be further expanded to capture additional feature of mobile wireless networks. The complexity of channel models used in our research can always be increased to include the many more of the variables that play a role in real-world scenarios. These could include considering the effects of channel access contention, dynamic channel quality prediction, and multiple link and routing problems. Of these issues, we are currently investigating stochastic dynamic programming approaches for scalable video buffer management in a cognitive radio setting [12]. Moreover, several issues such as radio link buffer management, allocation of spread spectrum codes, allocation of hybrid FDMA/TDMA slots, and the use of rateless fountain codes for application layer forward error correction such as Raptor codes [13] can also be addressed.
5.5 References


[7] *Universal Mobile Telecommunications System (UMTS); Multimedia Broadcast/Multicast Service (MBMS); Protocols and Codecs*, European Telecommunications Standards Institute, March 2005.


Chapter 5. Discussion and Conclusions

