

STEREO VIDEO BROADCASTING OVER DVB-H

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

DÖNE BUĞDAYCI

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
ELECTRICAL AND ELECTRONICS ENGINEERING

JANUARY 2012

Approval of the thesis:

STEREO VIDEO BROADCASTING OVER DVB-H

submitted by **DÖNE BUĞDAYCI** in partial fulfillment of the requirements for the degree of **Master of Science in Electrical and Electronics Engineering Department, Middle East Technical University** by,

Prof. Dr. Canan Özgen
Dean, Graduate School of **Natural and Applied Sciences**

Prof. Dr. İsmet Erkmen
Head of Department, **Electrical and Electronics Engineering**

Prof. Dr. Gözde Bozdağı Akar
Supervisor, **Electrical and Electronics Engineering Dept., METU**

Examining Committee Members:

Prof. Dr. Aydın Alatan
Electrical and Electronics Engineering Dept., METU

Prof. Dr. Gözde Bozdağı Akar
Electrical and Electronics Engineering Dept., METU

Assoc. Prof. Dr. Cüneyt Bazlamaçcı
Electrical and Electronics Engineering Dept., METU

Dr. Fatih Kamışlı
Electrical and Electronics Engineering Dept., METU

Dr. Arda Akman
Türk Telekom

Date:

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: DÖNE BUĞDAYCI

Signature :

ABSTRACT

STEREO VIDEO BROADCASTING OVER DVB-H

Buğdaycı, Döne

M.S., Department of Electrical and Electronics Engineering

Supervisor : Prof. Dr. Gözde Bozdağı Akar

January 2012, 77 pages

This thesis proposes a complete framework of an end-to-end transmission of stereo video to mobile devices using DVB-H. Block diagram of the system is presented and operations conducted on the video at each layer are explained. Parameters and methodologies that may make a robust transmission possible are discussed. The transmission performance is analyzed in terms of error robustness under various coding methods, prediction structures, layering and protection strategies for different contents and channel conditions. It also investigates the effect of rate allocation between video quality and protection over robust transmission in erroneous environment. This work provides directive conclusions on the selection of the mentioned parameters and methods.

Keywords: 3D video, DVB-H, MVC, Unequal Protection, FEC

ÖZ

DVB-H ÜZERİNDEN 3B GÖRÜNTÜ YAYINI

Buğdaycı, Döne

Yüksek Lisans, Elektrik-Elektronik Mühendisliği Bölümü

Tez Yöneticisi : Prof. Dr. Gözde Bozdağı Akar

Ocak 2012, 77 sayfa

Bu tez, stereo görüntünün DVB-H üzerinden mobil cihazlara uçtan uça iletiminin bütün bir çerçevesini sunmaktadır. Sistemin blok diyagramı sunulmuş ve her bir ağ katmanında görüntüye uygulanan işlemler açıklanmıştır. Gürbüz iletimi mümkün kılacak parametreler ve yöntemler tartışılmıştır. İletim başarımı, değişik içerikler ve kanal koşulları için çeşitli kodlama yöntemleri, tahmin yapıları, katmanlama ve koruma stratejileri açısından analiz edilmiştir. Aynı zamanda, görüntü kalitesi ile koruma arasındaki oran dağılımının hata içeren ortamlardaki gürbüz iletime etkisini de araştırır. Bu tez bahsedilen parametre ve yöntemlerin seçimi ile ilgili yön gösterici sonuçlar sağlar.

Anahtar Kelimeler: 3B görüntü, DVB-H, MVC, Eşit olmayan koruma, FEC

to my mother

ACKNOWLEDGMENTS

I would like to express my gratitude to my supervisor Prof. Dr. Gözde Bozdağı Akar, for her guidance and encouragement throughout my entire graduate studies. I would like to thank my thesis defence committee members Prof. Dr. Aydın Alatan, Assoc. Prof. Dr. Cüneyt Bazlamaçcı, Dr. Fatih Kamışlı and Dr. Arda Akman for their time and valuable comments on my thesis. I would like to thank my seniors Anıl Aksay and Oğuz Bici for providing the software tools used in the thesis, their help for the tests and for their valuable friendship. I would like to thank Antti Tikanmaki for providing the decapsulator software. I am indebted to my family for their support. This thesis is partially supported by the European Commission within FP7 under Grant 216503 with the acronym MOBILE3DTV.

TABLE OF CONTENTS

ABSTRACT	iv
ÖZ	v
ACKNOWLEDGMENTS	vii
TABLE OF CONTENTS	viii
LIST OF TABLES	xi
LIST OF FIGURES	xii
LIST OF ABBREVIATIONS	xv
CHAPTERS	
1 INTRODUCTION	1
1.1 Scope and Outline of the Thesis	2
2 DVB-H	4
2.1 Introduction	5
2.2 Time Slicing	6
2.3 MPE-FEC	7
2.4 Extensions on Physical Layer	9
3 BASICS OF 3D VIDEO AND CODING	13
3.1 Introduction	13
3.2 3D Display Technologies	15
3.2.1 3D with Glasses	15
3.2.2 Autostereoscopy	17
3.3 3D Video Coding	21
3.3.1 Overview of H.264/AVC	21
3.3.2 Simulcast Coding	24

3.3.3	Frame Compatible Coding	24
3.3.4	2D-plus-depth Coding	26
3.3.5	MultiView Coding	26
4	SYSTEM OVERVIEW	28
4.1	Encoding	29
4.2	Streaming	31
4.2.1	Real-time Transport Protocol	31
4.2.2	User Datagram Protocol	32
4.2.3	Internet Protocol	33
4.3	Encapsulation	35
4.4	Simulation of the Physical Layer and the Mobile Channel	36
4.5	Decapsulation	40
4.6	Frame Loss Concealer	40
4.7	Assembler	41
4.8	Decoding	42
5	SIMULATIONS	43
5.1	Encoding	44
5.2	Channel Simulations	46
5.3	Transmission Simulations	48
5.3.1	Coding Method and Prediction Structure	49
5.3.2	Rate Allocation Between Video Quality and FEC	52
5.3.3	Layering	53
5.3.4	Protection Strategies	54
5.4	Experimental Setup	55
5.5	Results	56
5.5.1	High Error Case	56
5.5.2	Low Error Case	62
6	CONCLUSION	72
	REFERENCES	74
	APPENDICES	

A ENCODING 77

LIST OF TABLES

TABLES

Table 2.1	Physical layer parameters	12
Table 2.2	Corresponding channel bitrates (Mbits/sec) for the possible modulation, code rate and guard interval values	12
Table 4.1	Path delays corresponding to the listed channel modes implemented in the Simulink model	39
Table 4.2	Path gains corresponding to the listed channel modes implemented in the Simulink model	39
Table 5.1	DVB-H physical layer and channel model parameters	47
Table 5.2	The contents used in the transmission tests	55
Table 5.3	Characteristic details of the contents	55
Table 5.4	Parameters and methods that define the experiments	56
Table A.1	Sample configuration file for the encoder	77
Table A.2	Sample configuration file for the encoder continued	78
Table A.3	Sample configuration file for the encoder continued	79

LIST OF FIGURES

FIGURES

Figure 2.1	Usage of DVB-H system within the existing DVB-T structure	4
Figure 2.2	DVB-H time slices and DVB-T services	7
Figure 2.3	Burst parameters	7
Figure 2.4	MPE-FEC Frame, MPE and MPE-FEC Sections, MPEG2-TS packets . . .	8
Figure 2.5	Functional block diagram of the DVB-T physical layer	9
Figure 3.1	Most commonly used stereoscopic 3D techniques	15
Figure 3.2	Typical anaglyph glasses	16
Figure 3.3	Polarized content on the movie screen and the use of polarized glasses . . .	16
Figure 3.4	Polarized glasses	17
Figure 3.5	Shutter glasses	18
Figure 3.6	Autostereoscopic display technologies illustrated: parallax barrier and lenticular lens array	19
Figure 3.7	A prototype mobile phone with electronically switchable 2D/3D autostereoscopic screen developed within the EU FP7 MOBILE3DTV Project	20
Figure 3.8	A commercially available mobile phone with 3D capture and display capabilities, having an autostereoscopic 3D display	20
Figure 3.9	Basic coding structure for H.264/AVC for a macroblock	21
Figure 3.10	Structure of H.264/AVC video encoder	22
Figure 3.11	H.264 Baseline, Main and Extended Profiles	23
Figure 3.12	Left and Right Views Simulcast	24
Figure 3.13	Stereo Interleaving, side-by-side	25
Figure 3.14	Stereo Interleaving, over/under	25

Figure 3.15 Stereo Interleaving schemes, (a) time interleaved, (b) space interleaved over\nder, (c) space interleaved side-by-side	26
Figure 3.16 2D-plus-depth	26
Figure 4.1 Block diagram of the system	28
Figure 4.2 Prediction structures	30
Figure 4.3 Packet structures in the different layers of the network	31
Figure 4.4 RTP Header format	32
Figure 4.5 UDP Header format	33
Figure 4.6 IPv4 Header format	34
Figure 4.7 Block diagram of the software implementation of the end-to-end system . .	35
Figure 4.8 Simulink model of the DVB-H physical layer (transmitter/receiver) and the mobile TU6 Channel	37
Figure 4.9 Channel Model	40
Figure 4.10 Lossy stream including a missing NALU with POC=7 (up) and the stream at the output of the frame concealer with a Skip NALU with the missing POC number inserted (down)	41
Figure 5.1 Layered structure of the simulations	43
Figure 5.2 Joint PSNR vs total bitrate for Heidelberg Simulcast Coding	44
Figure 5.3 Joint PSNR vs total bitrate for Heidelberg MVC IPP Full Coding	45
Figure 5.4 Joint PSNR vs total bitrate for Heidelberg MVC IPP Simplified Coding . .	46
Figure 5.5 Average packet error rate vs channel SNR for the error traces	47
Figure 5.6 Average burst packet error length vs channel SNR for the error traces . . .	48
Figure 5.7 Variance of burst packet error length vs channel SNR for the error traces .	49
Figure 5.8 TS packet loss distribution of the error traces generated for the tests	50
Figure 5.9 Selection of the test videos according to coding method, prediction struc- ture and rate allocation	52
Figure 5.10 Layered transmission schemes	54
Figure 5.11 Transmission results, Left View, Heidelberg, channel SNR=17dB	57

Figure 5.12 Transmission results, Right View, Heidelberg, channel SNR=17dB	59
Figure 5.13 Joint PSNR distribution of the experiments of all the transmission cases, Heidelberg	61
Figure 5.14 Transmission results, Left View, Heidelberg, channel SNR=19dB	64
Figure 5.15 Transmission results, Right View, Heidelberg, channel SNR=19dB	65
Figure 5.16 Transmission results, Rhine, channel SNR=17dB	66
Figure 5.17 Joint PSNR distribution of the experiments of all the transmission cases, Rhine	67
Figure 5.18 Transmission results, Rhine, channel SNR=19dB	68
Figure 5.19 Transmission results, Knights, channel SNR=17dB	69
Figure 5.20 Joint PSNR distribution of the experiments of all the transmission cases, Knights	70
Figure 5.21 Transmission results, Knights, channel SNR=19dB	71

LIST OF ABBREVIATIONS

3D	Three Dimensional	MVC	Multi View Coding
ADT	Application Data Table	NALU	Network Abstraction Layer Unit
AVC	Advanced Video Coding	OFDM	Orthogonal Frequency Division Multiplexing
AWGN	Additive White Gaussian Noise	PLR	Packet Loss Rate
CABAC	Context-Adaptive Binary Adaptive Coding	POC	Picture Order Count
CAVLC	Context-Adaptive Variable Length Coding	PPS	Picture Parameter Set
DCT	Discrete Cosine Transform	PSNR	Peak Signal to Noise Ratio
DIBR	Depth Image Based Rendering	QAM	Quadrature Amplitude Modulation
DVB	Digital Video Broadcasting	QP	Quantization Parameter
DVB-H	Digital Video Broadcasting - Handheld	QPSK	Quadrature Phase-Shift Keying
DVB-NGH	Digital Video Broadcasting - Next Generation Handheld	RD	Rate Distortion
DVB-T	Digital Video Broadcasting - Terrestrial	RS	Reed Solomon
EEP	Equal Error Protection	RTP	Real Time Protocol
FEC	Forward Error Correction	SFN	Single Frequency Network
FFT	Fast Fourier Transform	SNR	Signal to Noise Ratio
FMO	Flexible Macroblock Ordering	SPS	Sequence Parameter Set
GI	Guard Interval	TCP	Transmission Control Protocol
GOP	Group of Pictures	TPS	Transmitter Parameter Signaling
IP	Internet Protocol	TS	Transport Stream
JVT	Joint Video Team	UDP	User Datagram Protocol
LCD	Liquid crystal Display	UEP	Unequal Error Protection
MPE	Multi Protocol Encapsulation	VCEG	Video Coding Experts Group
MPEG	Moving Pictures Expert Group	VCL	Video Coding Layer
MSE	Mean Square Error		

CHAPTER 1

INTRODUCTION

Mobile communication technologies have advanced throughout the last decade from the point of voice communication to data and video communications rapidly. Nowadays, high volume multimedia transmissions are common using state-of-the-art digital communications techniques. In addition, mobile device technologies are now able to handle real-time multimedia communications smoothly. Mobile devices can access multimedia resources through wireless, 3G or broadcast networks where the latter one is referred as mobile-TV. Being aware of the increasing demand for video delivery services for mobile devices, Digital Video Broadcasting (DVB) community developed a standard specifically for broadcasting to mobile devices. In 2008, European Union accepted the Digital Video Broadcasting - Handheld, DVB-H, as the official mobile-TV standard [1]. Today, the physical layer of DVB-H is outdated and a newer standard Digital Video Broadcasting - Terrestrial 2 (DVB-T2) is introduced which covers recent improvements on physical layer technology. As a result, DVB community announced a call for the development of a newer standard addressing the broadcasting of mobile-TV to handheld devices. The new standard is called DVB - Next Generation Handheld, DVB-NGH, and it is planned to be ready by 2015 [2].

During the last decade, 3D video technologies have also evolved significantly. Popularity of 3D contents have risen with increasing number of 3D movie theaters where the illusion of depth is created by the use of special glasses. The increasing demand for 3D contents caught the TV manufacturers attention, creating a 3DTV market. Although many TV's with 3D capability come with glasses, the research on autostereoscopic display technologies, which do not require glasses for the depth perception, is also accelerated.

The mobile device and autostereoscopic display technologies have evolved to a point such

that, these two fields emerged and made the delivery of 3D video services to mobile devices possible. European 3DPhone [3], Mobile3DTV [4] and Korean T-DMB [5] are research projects that worked on the subject. Among these Mobile3DTV project, specifically focused on the delivery of 3D video over DVB-H.

Mobile-TV is more prone to channel errors than conventional roof-top antenna broadcasting due to the introduced mobility. Although there are studies conducted with monoscopic videos, error resilient transmission strategies specifically defined for the delivery of 3D video over DVB-H are required. The delivery of 3D video over DVB-H involves implementation details that have to be addressed for successful delivery.

1.1 Scope and Outline of the Thesis

This thesis presents an end-to-end stereo video broadcasting system over DVB-H. Our aim is to investigate the effects of coding method, prediction structure, layering, protection methods and rate allocation between video quality and FEC rates on the performance of transmission over DVB-H.

Chapter 2 provides information about DVB-H by explaining the key features such as time slicing and MPE-FEC and defining the physical layer extensions brought to DVB-T in order to support mobility and reduce power consumption.

Chapter 3 introduces the basic concepts of 3D video as a review on common technologies. It provides information about common 3D techniques, display technologies and mostly used content representation types.

Chapter 4 provides the system overview and explains each block of the end-to-end system separately by providing implementation details about the simulation software.

The first three sections of chapter 5 are the separate parts of the simulations whereas the last two sections explain the experiments and discuss the results. The chapter starts explaining the preparation of the encoded videos to be used during the transmission tests. Then, the simulations with the channel model are explained by providing some statistical data about the channel error traces obtained from the simulator. In the third section, the steps of the transmission simulations are defined and the methods that are proposed to be tested in the simulations

are explained. The following section explains the experimental setup by defining the parameters of the tests and format of the outputs. Finally, the results of the transmission simulations are analyzed in detail in order to see the affect of coding methods, prediction structures, protection methods and rate allocation between video and protection on the robustness of the transmission.

Chapter 6 is the conclusion of the thesis.

CHAPTER 2

DVB-H

This chapter is about the main features of the DVB-H standard in physical and link layer. Figure 2.1 illustrates the usage of DVB-H services within the existing DVB-T infrastructure. The green labeled components are new features introduced in DVB-H. In the following sections, after an introduction, the two additional link layer features of DVB-H, time slicing and MPE-FEC, are explained. The final section provides an overview of physical layer and explains the additional features.

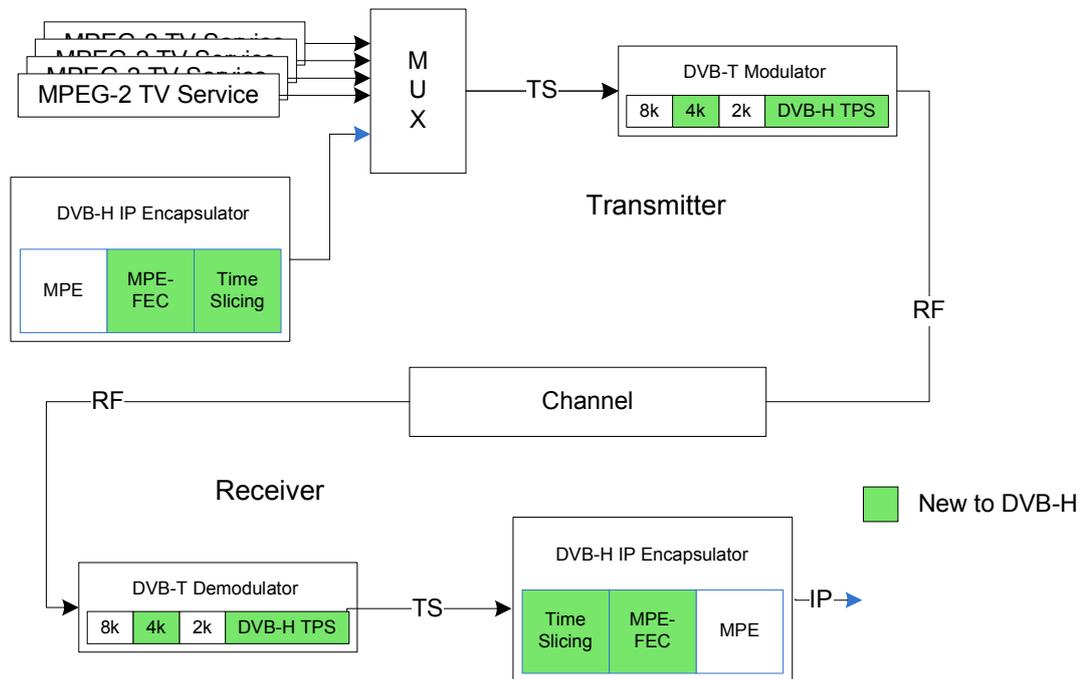


Figure 2.1: Usage of DVB-H system within the existing DVB-T structure

2.1 Introduction

Digital Video Broadcasting-Terrestrial (DVB-T) is the European digital TV standard which is developed within the DVB Project. As the commercial terrestrial digital TV services started to appear on the market in early 2000s, the mobile performance of the DVB-Terrestrial (DVB-T) has been investigated. It appeared that mobile reception required more robust networks than the existing ones planned for fixed rooftop DVB-T reception. The standard is improved further to support the robustness required by mobile reception [6]. Meanwhile, with the advances in technology, handheld devices such as cellular phones and PDAs became a large part of daily life, changing the consumer behaviors. Cellular phones for instance, gone far beyond the voice communication, making high data rate transfers and video communication possible. Hence the DVB Community realized that new specifications addressing the power consumption and reception challenges of handheld devices are required. In 2002, the DVB community started to work on a standard which will allow the delivery of multimedia contents to handheld devices. DVB-T was improved for mobile reception but the aim was not this anymore, it was to specify a standard that will allow digital TV on small, battery powered, handheld devices. By the end of 2004, Digital Video Broadcasting-Handheld (DVB-H) was formally adopted as ETSI standard EN 302 304 [7].

DVB-H is based on the existing DVB-T standard with extensions on the physical layer and two new elements introduced in the link layer to support mobility. Figure 2.1 illustrates the usage of DVB-H services within the existing DVB-T infrastructure. The two additions in the link layer are the time slicing and the additional Forward Error Correction coding (FEC). Time slicing is introduced to reduce power consumption of the receiver device. It allows the transmission of data in periodic bursts as opposed to the continuous transmission in DVB-T. By this way, the receiver is able to power off between two consecutive bursts of the same stream and power up according to the time specifically signaled by *delta_t* method in each burst. It also allows a seamless handover during the change of cells. FEC coding is called the Multi Protocol Encapsulation-Forward Error Correction (MPE-FEC) which is added to improve the noise and Doppler performance and tolerance to impulse interference [6].

Physical layer of DVB-H is based on DVB-T with four new features added to improve mobile reception performance and support backward compatibility. One of them is the new 4K Orthogonal Frequency Division Multiplexing (OFDM) mode which is a trade off between

mobility and Single Frequency Network (SFN) cell size. Second extension is the in-depth interleaver which is an optional alternative to the existing native interleaver. Further, in addition to the existing 6, 7 and 8 MHz channel bandwidths, 5 MHz channel bandwidth is included to be used in nonbroadcast bands. Finally, the presence of DVB-H services and possible use of MPE-FEC is indicated by two additional bits in Transmitter Parameter Signaling (TPS) [6].

In the following sections of this chapter, detailed information about the features new to DVB-H is provided.

2.2 Time Slicing

Time slicing is introduced in order to reduce the average power consumption at the receiver side. The idea in time slicing, is to send the data in a short time period with a significantly higher bitrate than the continuous transmission. This way, the receiver is capable of reducing power consumption by going off in between the delivery of the bursts of a stream and the high bitrate of the burst compensates for the off time. Figure 2.2 shows the DVB-H streams multiplexed with a DVB-T stream, where the x axis represents time and y axis represents bitrate. It represents the idea of time slicing, showing the periodic DVB-H services replacing the continuous transmission of DVB-T. The receiver knows when to expect the next burst from the Δt information present in the current burst [6].

During the off-time in between the reception of bursts, the receiver is able to monitor the neighboring cell activities. The switching between the transport streams is handled in off-time which results in a seamless handover.

Some of the parameters that are used in time slicing are burst duration, burst size, burst bitrate, constant bitrate, off time as shown in Figure 2.3. They are used in power consumption calculations and design of the broadcast streams. Burst duration is in terms of milliseconds whereas burst size is in terms of Megabytes. Formulas and further information can be found in [1].

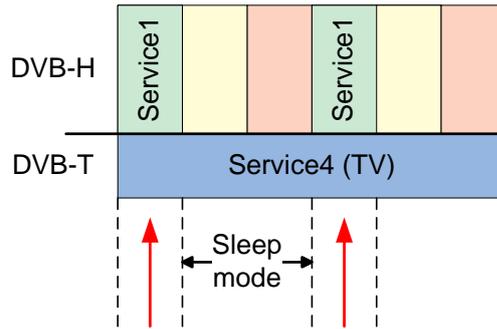


Figure 2.2: DVB-H time slices and DVB-T services

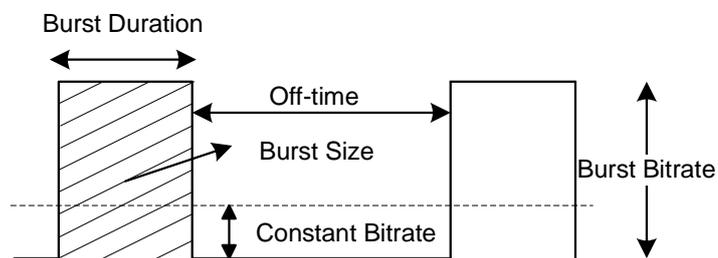


Figure 2.3: Burst parameters

2.3 MPE-FEC

Additional forward error correction is applied in link layer by the use of an MPE-FEC frame illustrated in Figure 2.4. An MPE-FEC frame is a matrix with 255 columns and a variable number of rows, where each position corresponds to an information byte. The number of rows can be any value between 1 and 1024, which makes the maximum size of an MPE-FEC frame 2 Mb atmost. The first 191 columns of the table is dedicated for IP datagram and it is called the *Application Data Table (ADT)*. The remaining 64 columns is dedicated for the parity data and called the *RS Data Table* [1].

IP datagrams are filled into the MPE-FEC frame by starting with the upper left corner of the matrix and going downwards the first column. When an IP datagram ends, the following IP datagram starts right after the previous one. Each position in the matrix has an address that can be used for signaling the section length. After all the IP datagrams have entered the

is generally the case when there is padding in ADT, to obtain the desired effective code rate [1].

After the parity bytes are calculated, each IP datagram is encapsulated into an MPE section and each RS column is encapsulated into an MPE-FEC section. MPE-FEC sections are fixed length and its determined by the number of rows of the MPE-FEC frame whereas MPE section lengths directly depend on IP datagrams. Finally, both MPE nd MPE-FEC sections are encapsulated into MPEG2-TS packets to be used by the physical layer. TS packets are 188 bytes, 4 bytes of which is the TS header and the rest payload [1].

2.4 Extensions on Physical Layer

Functional blocks of the DVB-T physical layer are illustrated in Figure 2.5. The system is compatible with MPEG2-TS packets defined in [8]. This sections provides a brief information about the parts composing the DVB-T physical layer and continues with the new features added in DVB-H. Main blocks of the system are defined in [9] and summarized as follows:

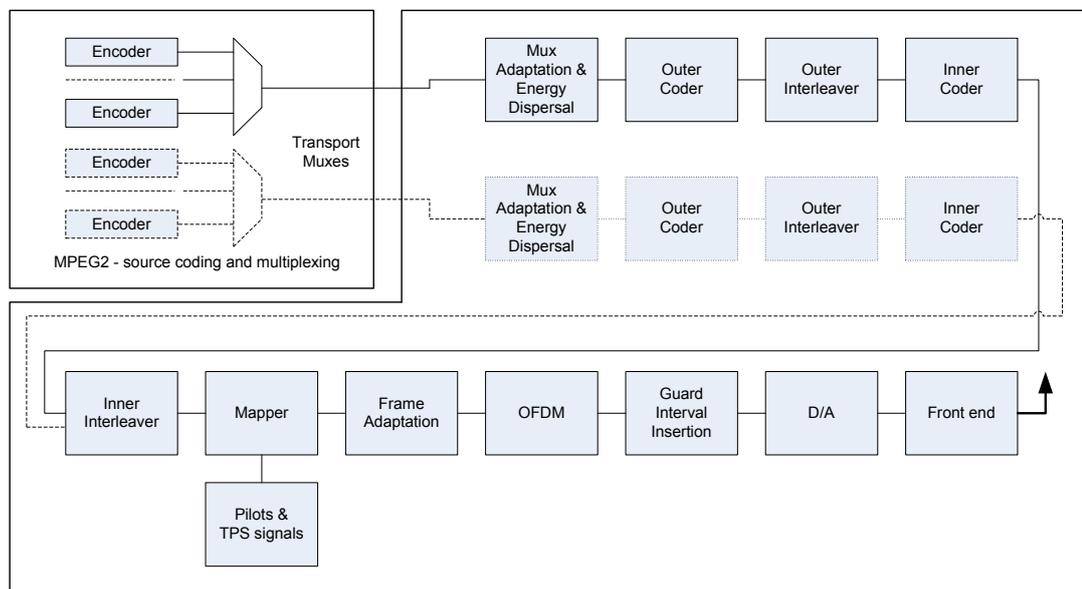


Figure 2.5: Functional block diagram of the DVB-T physical layer

Mux Adaptation & Energy Dispersal The data of the multiplex streams are randomized

through the scrambler with the use of the MPEG2 sync byte (present in every TS packet) and a Pseudo Random Binary Sequence (PRBS) generator.

Outer Coder This is a shortened Reed Solomon code RS (204,188) applied on every input packet, allowing up to 8 bytes of correction. The code and field generator polynomials are similar to the ones used in MPE-FEC.

Outer Interleaver This is a convolutional byte-wise interleaving with depth $I = 12$.

Inner Coder A range of punctured convolutional codes corresponding to different error correction and data rates are allowed. The punctured convolutional codes are based on a mother convolutional code of $1/2$ with 64 states. First the convolutional coding is applied on the input sequence. This is allowed by puncturing operation and code rates of $2/3$, $3/4$, $5/6$ and $7/8$ are possible to achieve after puncturing.

Inner Interleaver The inner interleaving is composed of two parts, first the bit-wise interleaving followed by the symbol interleaving. In bit-wise interleaving, the input is demultiplexed into v sub-streams whose number is determined according to the modulation (QAM) mode used and processed by this number of separate interleavers. The interleaving output consists of v bits each from one of the interleavers. The symbol interleaver takes the input grouped into v bit words and maps them onto the active carriers per OFDM symbol. The number of the active carriers is determined by the FFT mode used (2K or 8K).

Mapping and Modulation Modulation modes used are QPSK, 16-QAM, 64-QAM, non-uniform 16-QAM or non-uniform 64-QAM constellations. Carriers of an OFDM frame are modulated using one of these modes. Mapping bits may take three values being 1, 2, and 4.

OFDM Transmission In OFDM, the transmitted signal is organized in frames each of which contains 68 OFDM symbols. The data to be transmitted are mapped onto the symbols. In addition to the data, an OFDM frame contains scattered pilot cells, continual pilot carriers and TPS carriers. The OFDM modes present in DVB-T are 2K and 8K modes where the numbers correspond to the number of carriers.

DVB-H brings four main extensions on the physical layer of DVB-T in order to compensate for the challenges of time varying wireless channels. One of these extensions is the optional

4K OFDM mode. The significance of the chosen OFDM mode is in the planning of the network. These modes both can be used for single transmitter operation and for Single Frequency Networks (SFNs). The SFN cell size is the largest in 8K mode and lowest in 2K mode. However, the Doppler tolerance of the network is highest in 2K mode and lowest in 8K mode. The additional 4K mode is a trade-off between the network size and the mobility. It has better Doppler tolerance compared to 8K mode and larger cell size compared to 2K mode [6]. Second extension is about the use of the symbol interleaver in 2K and 4K modes. Instead of the native interleaver which is already defined in DVB-T, it is possible to use the in-depth interleaver which interleaves the bits over four or two OFDM symbols for 2K and 4K modes respectively. This brings a flexibility to the symbol interleaver by allowing a 2K or 4K signal to benefit from the memory of the 8K symbol interleaver. As the interleaving depth increases, the reception performance in fading channels where impulse noise is present, also increases. Third, in addition to the 6, 7 and 8 MHz channels, 5 MHz channel bandwidth is defined to be used in nonbroadcast bands. Finally, in order to signal the presence of DVB-H services, possible use of MPE-FEC and to speed up the service discovery, the transmitter parameter signaling is extended [6]. The improvements on TPS are summarized as follows:

- The 4K mode to be used in DVB-H is signalled as an additional transmission mode to the existing 2K and 8K modes.
- The DVB-T hierarchy information is used to specify whether the native or in-depth interleaver is used as the symbol interleaver.
- One formerly unused bit is allocated as the time slicing indicator which signals the presence of at least one time-sliced DVB-H service in the transmission channel.
- One formerly unused bit is allocated as the MPE-FEC indicator to signal the use of MPE-FEC by at least one DVB-H service present in the channel.

Table 2.1 lists the physical layer parameters that affect the channel bitrate and robustness against channel errors. For modulation type, from QPSK to 16QAM, channel bitrate and error probability increases. Table 2.2 lists the corresponding channel bitrates after the selection of modulation type, convolutional code rate and guard interval [9].

Table 2.1: Physical layer parameters

Parameter	Options	Explanation
Modulation	3	QPSK, 16QAM, 64QAM
FFT-size	3	2K, 4K, 8K
In-depth interleaver	2	On / Off (only for 2K and 4K)
Guard Interval	4	1/4, 1/8, 1/16, 1/32
Convolutional code rate	5	1/2, 2/3, 3/4, 5/6, 7/8

Table 2.2: Corresponding channel bitrates (Mbits/sec) for the possible modulation, code rate and guard interval values

Modulation	Code rate	Guard interval			
		1/4	1/8	1/16	1/32
QPSK	1/2	4,98	5,53	5,85	6,03
	2/3	6,64	7,37	7,81	8,04
	3/4	7,46	8,29	8,78	9,05
	5/6	8,29	9,22	9,76	10,05
	7/8	8,71	9,68	10,25	10,56
16-QAM	1/2	9,95	11,06	11,71	12,06
	2/3	13,27	14,75	15,61	16,09
	3/4	14,93	16,59	17,56	18,10
	5/6	16,59	18,43	19,52	20,11
	7/8	17,42	19,35	20,49	21,11
64-QAM	1/2	14,93	16,59	17,56	18,10
	2/3	19,91	22,12	23,42	24,13
	3/4	22,39	24,88	26,35	27,14
	5/6	24,88	27,65	29,27	30,16
	7/8	26,13	29,03	30,74	31,67

CHAPTER 3

BASICS OF 3D VIDEO AND CODING

In the following sections, an overview of stereoscopic 3D display technologies is provided together with the applications of 3D display technologies with and without glasses. The second section is included to explain the commonly used 3D video coding together with advantages and disadvantages according to the application area. [10] introduces four main 3D video coding formats, which are the most commonly used ones currently, and they are summarized in the second section.

3.1 Introduction

Three-dimensional (3D) video enhances the illusion of depth perception. In stereo video, the depth is perceived by the different viewpoints presented to viewer's left and right eyes. In 3D cinema environment, the depth perception is provided by the use of special glasses (anaglyph, polarization, or shutter) which separates the two images displayed in the screen (by either combining separate images from two offset sources or filter offset images from a single source separated to each eye) so that each view is seen by a different eye. On the other hand, in autostereoscopy no glasses are required since in this technology the source splits the images directionally into each eye. The most common implementations of autostereoscopic displays are either parallax-barrier or lenticular array technologies. A parallax-barrier is placed in front of a display in order to allow each eye to see a different set of pixels and therefore to have the depth perception via stereoscopy. The two main drawbacks of this technology are that the resolution of the viewed image is decreased by half and it restricts the viewer to a well-defined position to perceive depth. In lenticular array technology, the spaces and lines of the parallax-barrier is replaced by cylindrical lenses. Lenticular array technology has less

restrictions on the position of the viewer [11].

Among the two display technologies, 3D displays employing glasses are less popular in home environment than the autostereoscopic displays because of the restrictions on the user. Multi-viewpoint autostereoscopic displays provide multiple views ensuring that a user from a specific point sees only two of them. Nowadays, high resolution 3D displays to be used in home environment are available on the market.

The popularity of 3D contents and advances in autostereoscopic displays affected the mobile devices also. Mobile devices are a good candidate of 3D video experience since they are personal. The user can easily adjust the position of the mobile device in order to obtain the depth perception from the certain viewing angle of the autostereoscopic display. Electronically switchable displays allow the use of 2D and 3D on the same device. Various research projects such as the 3D DMB [5], 3DPhone [3] and Mobile3DTV [4], studied the various aspects of delivery of 3D video to mobile devices such as cell phones and PDAs. The system presented in this study is based on two views (stereo video) to be displayed on a lenticular array technology screen.

Stereo video is represented by two formats:

- Stereo format where two separate views of the same scene are present as left and right views,
- 2D-plus-depth format where a single view of the scene is present together with the depth information corresponding the same scene.

Depth information is represented as a sequence of one channel images with the same resolution of the colored view of the scene. In a single frame of the depth sequence, the luminance of the frame changes according to the changes in the depth of the scene. In addition to measuring the depth information by the use of special cameras that works like an infrared LIDAR (laser radar without mechanical scanner) [12], it is possible to obtain the depth information by calculating the disparity map from the two images captured from different angles, a process referred as *depth estimation* [13]. Since autostereoscopic displays rely on two images, an inverse process called *depth image-based rendering (DIBR)* is used to obtain two different images from the single image and the depth information [14].

3.2 3D Display Technologies

Conventional display technologies are monocular meaning that both eyes of the viewer see the same thing captured by a single camera. However in stereo displays, two different views are directed to the left and right eyes of the viewer. These displays generally take two 2D video streams and separate them accordingly. There are mainly two groups of technologies that are used in stereo displays which are conventional displays utilizing glasses and autostereoscopic displays [15]. In the following subsections detailed information about these technologies is presented.

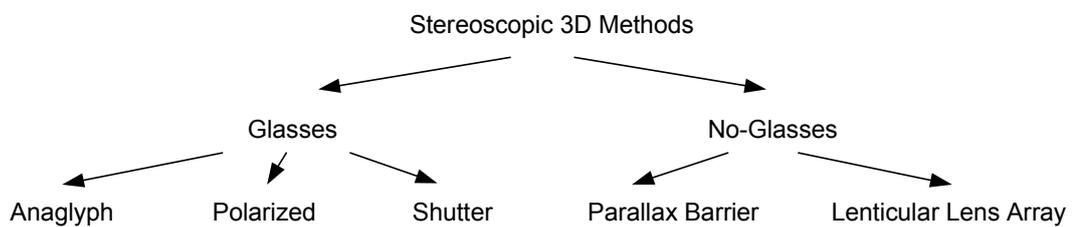


Figure 3.1: Most commonly used stereoscopic 3D techniques

3.2.1 3D with Glasses

There are various types of glasses employing different technologies to direct two different views to each eye. The oldest and most common types of glasses are the anaglyph glasses. An anaglyph video consists of a sequence of two color encoded images superimposed to obtain an overlapped frame [15]. Mostly used color codings are red/blue and red/cyan but combinations such as green/magenta, blue/yellow and orange/yellow also exist. Glasses allow each eye of the viewer to see one of the two superimposed content projected on the screen. Paper-based anaglyph glasses showed in Figure 3.2 are usually distributed in magazines for special 3D broadcasts on television. Plastic, break-resistant anaglyph glasses are used with 3D screens [16].

Currently, most of the movie theaters (large cinema houses and IMAX) showing 3D films use light polarization methods, either linear or circularly polarization. The content is captured by two cameras (or one camera with two lenses) and two projectors containing polarizing

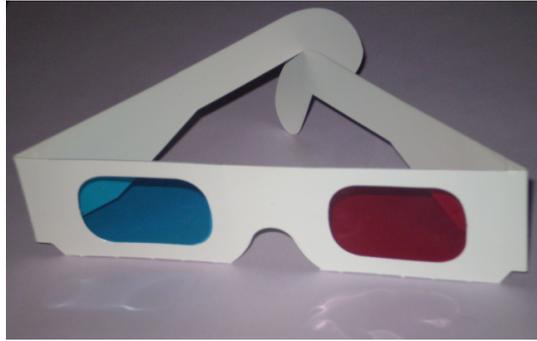


Figure 3.2: Typical anaglyph glasses

filter on their lenses, show both views simultaneously on the screen [15]. In case of linear polarization, a polarizing filter orients the images from the left projector to one plane and the filter on the right projector orients the images to the plane that is perpendicular to the left one. The right and left lenses of the glasses are aligned accordingly so that left eye sees only the left image and right eye sees only the right image. This concept is illustrated in Figure 3.3. The problem with this method is that it requires a certain viewing angle which limits the head movements of the user. It also limits the amount of light reaching the eyes which causes a darker image [16].

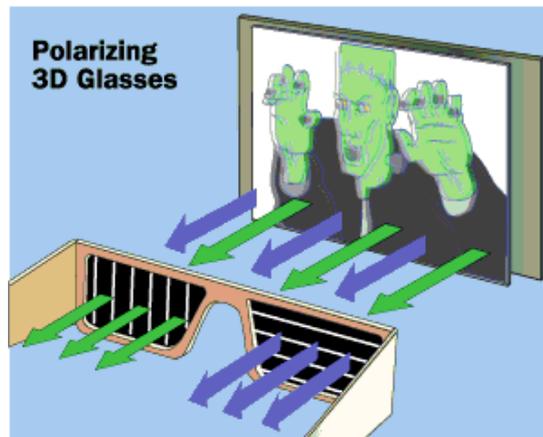


Figure 3.3: Polarized content on the movie screen and the use of polarized glasses

The most commonly employed 3D system is the *RealD Cinema* which utilizes circular polarization to beam the film onto a silver screen [15]. A special filter in front of the projector

converts linearly polarized light into circularly polarized light. The expensive silver screen helps preserve the polarization of the image. Circular polarization is more advantageous to linear polarization since it allows the user to move his/head more naturally without losing the 3D perception. It also eliminates the need for two projectors however, the disadvantage is the costly silver screen required for the circular polarization. An example of glasses used with *RealD Cinema* systems is presented in Figure 3.4 [16].



Figure 3.4: Polarized glasses

Another commonly used 3D projection method is the so-called *interlaced stereo* which is the time multiplexing of the display [15]. In this method, the content is displayed with consecutive left and right views, also known as alternate-frame sequencing, and the viewer uses shuttered glasses. LCD shutter glasses contain liquid-crystal that lets the light through in synchronization with the image displayed on the screen. This method is used by XpanD 3D and earlier IMAX systems. The advantage of it over the circularly polarized systems is that it does not require a silver screen. However, the disadvantage is that it requires expensive shutter glasses which need to be synchronized with the display system through a wired/wireless connection. The glasses which are shown in Figure 3.5, contain necessary electronics and batteries within and therefore are heavier than usual ones [16].

3.2.2 Autostereoscopy

Although 3D with glasses is appreciated in movie theaters where the viewers tolerate the restrictions brought by the glasses for the specific environment and the limited time, in case



Figure 3.5: Shutter glasses

of home entertainment and mobile applications, there is a growing demand for glass-free or *autostereoscopic* displays that can be used out of movie theaters. Autostereoscopic displays eliminates the cost and comfort issues of the glasses. The word "autostereo" implies the automatic nature of the 3D perception lacking the dependency to any legacy device such as the filtered or shuttered glass. Although the word includes other 3D methods such as holographic, volumetric and integral imaging; it is more commonly associated with lenticular or parallax-barrier systems where the separation of the views are via the additional optical elements present on the surface of the displays [15]. Similar to 3DTV systems, mobile applications of 3D also favor autostereoscopic displays due to usability issues. Mobile devices are used in many different environments such as indoors, streets, on the bus, while waiting for a line etc. Therefore, use of glasses is not even possible with them.

In autostereoscopic display technologies, two methods are most commonly employed, namely the parallax-barrier and lenticular lens array [15]. The discovery of parallax-barrier principal is being used over a hundred years now, but the flat-panel implementation has been developed by Sharp in the last decade, introducing the commercial 3D LCD screens [17]. Although Sharp laptops with the 3D LCD screens are obsolete now, there are a number of companies producing similar displays with parallax-barrier technology.

The main difference of a 3D LCD screen from a plain LCD display is that it contains a layer of slits placed in front of the screen, which allows each eye to see a different set of pixels provided that the viewer is sitting in an optimal spot [15]. The concept is illustrated in Figure

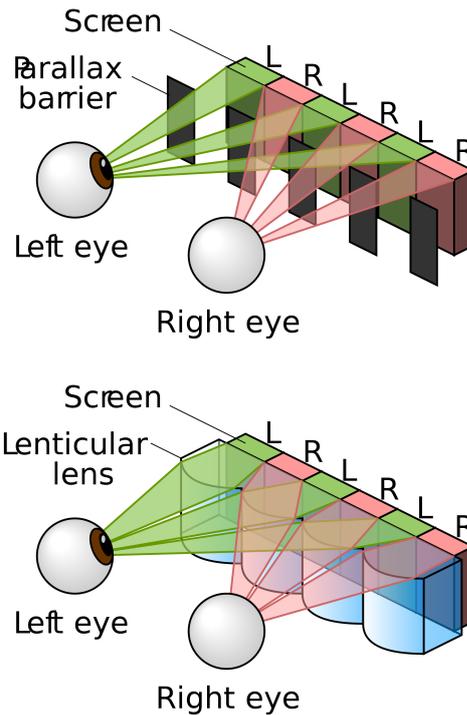


Figure 3.6: Autostereoscopic display technologies illustrated: parallax barrier and lenticular lens array

3.6, where the light is directed to each eye differently by the barrier, creating a sense of depth through the parallax. The two main drawbacks of the technology are the reduction in the horizontal resolution of the image displayed and the restriction on the position of the viewer. On the other hand, it is quite advantageous in terms of cost and ease of use. However, the image quality is very poor in these products. The image quality in newer parallax-barrier displays such as the Nintendo 3DS, HTC Evo 3D (Figure 3.8), and LG Optimus 3D is improved by the use parallax-barrier in front of the backlight. The only drawback is the increase in the production costs [17].

Lenticular array technology is similar to parallax-barrier technology, where the pairs of barrier and space pairs are replaced with a continuous array of tiny lenses. The viewing angle is improved compared to parallax-barrier since the barriers are much narrower in the lens array which is illustrated in Figure 3.6. Philips, improved the challenges in this technology and produced displays under the name WOWvX [17]. There are low cost, attachable parallax-barrier autostereoscopic sheets that can be used with conventional screens such as the 3DeeSlide pro-

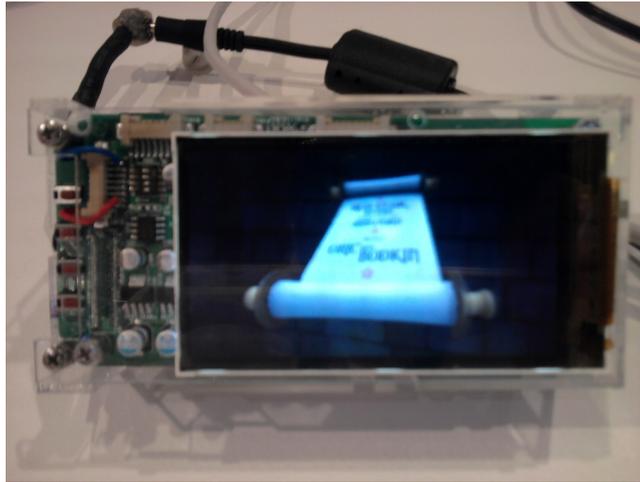


Figure 3.7: A prototype mobile phone with electronically switchable 2D/3D autostereoscopic screen developed within the EU FP7 MOBILE3DTV Project

duced for Iphone. Currently, autostereoscopic display technology is still in development to improve 3D perception quality, reduce and finally remove side effects such as blur, fatigue or dizziness.



Figure 3.8: A commercially available mobile phone with 3D capture and display capabilities, having an autostereoscopic 3D display

3.3 3D Video Coding

Most commonly used 3D video coding formats are simulcast coding, frame-compatible coding, 2D-plus-depth coding and multiview coding. The first three methods are based on the state-of-the-art coding standard H.264/AVC whereas multiview coding is an extension of this standard.

H.264/AVC is used by internet streaming applications, broadcast services and Blue-Ray. It is developed by the ITU-T Video Coding Experts Group (VCEG) together with the International Organization for Standardization (ISO)/International Electrotechnical Commission (IEC) Moving Picture Experts Group (MPEG) within the Joint Video Team (JVT) [18].

3.3.1 Overview of H.264/AVC

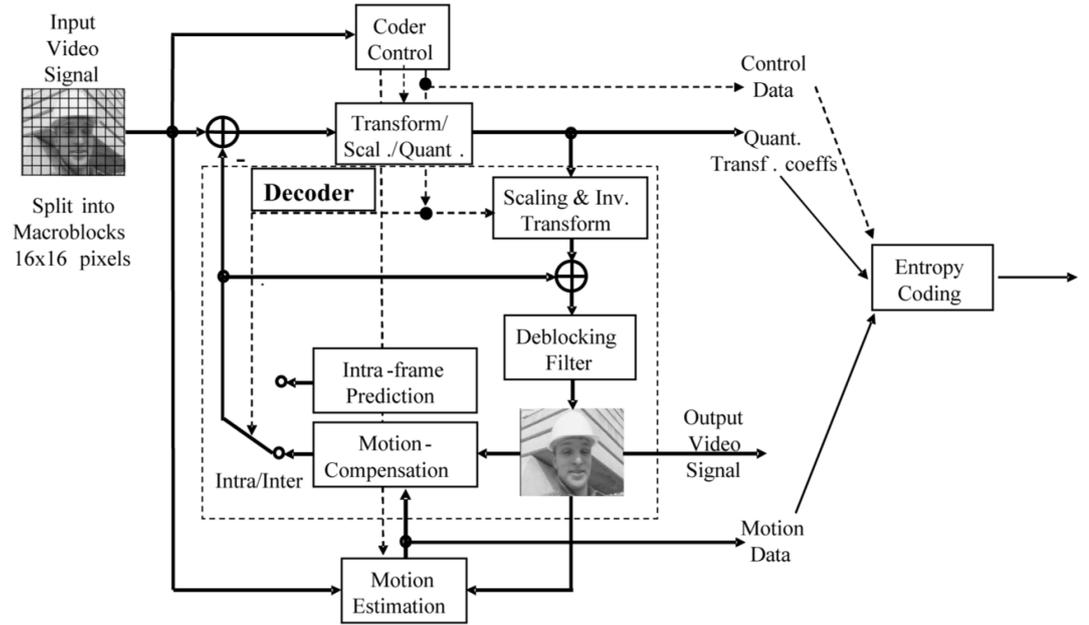


Figure 3.9: Basic coding structure for H.264/AVC for a macroblock

Advanced video coding (AVC) is the standard developed by the Moving Picture Experts Group (MPEG) and Video Coding Experts Group (VCEG) [19]. It is published as MPEG-4 Part 10 and ITU-T Recommendation H.264. It outperforms the earlier standards MPEG-4 and

H.263.

H.264/AVC does not define a CODEC (encoder, decoder pair) but defines the syntax of the encoder output video bitstream and decoder methods [19]. It keeps the same basic functional blocks (prediction, transform, quantization, entropy encoding) but brings improvements within each functional block [19].

Figure 3.9 illustrates the encoding of a single *macroblock*. The output of the Video Coding Layer (VCL) is an encoded bitstream passed to Network Abstraction Layer (NAL) to be packetized. The layered structure of the encoder is illustrated in figure 3.10.

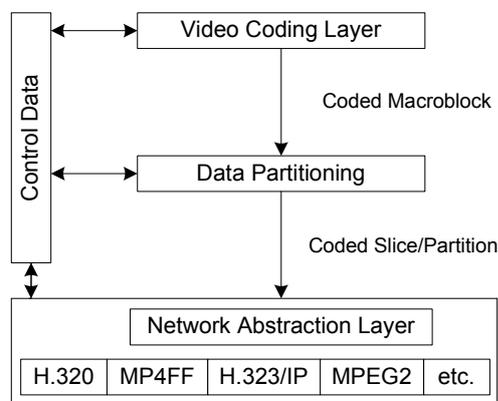


Figure 3.10: Structure of H.264/AVC video encoder

Encoder processes frames in units of *macroblocks* which are 16-by-16 pixel arrays. There are two kinds of prediction modes for a macroblock which are the *intra prediction* and *inter prediction*. Intra prediction is performed for a macroblock from the current frame (previously encoded, decoded and reconstructed macroblocks) and it is marked as an I macroblock. Inter prediction is performed from other frames that are set as *reference pictures* and it is marked as P macroblock. Slices consisting of I macroblocks are referred as I slices and P macroblocks are P slices. Finding a suitable inter prediction is often described as *motion estimation* and subtracting an inter prediction from the current macroblock is called *motion compensation* [19].

Encoder subtracts prediction from the current macroblock to form a *residual*. Block of residuals are transformed by integer transform (approximate for of Discrete Cosine Transform -

DCT) providing a set of *coefficients*. Block of transform coefficients is quantized, according to a quantization parameter (QP). Level of QP determines the compression rate. High Qp results in high compression rate (low video quality) whereas low QP results in low compression rate (high video quality) [19].

Finally, together with the quantized transform coefficients, information to enable the decoder to re-create the prediction, information about the structure of the compressed data and the compression tools used during encoding and information about the complete video sequence are converted into binary codes using Context-Adaptive Variable Length Coding (CAVLC) or Context-Adaptive Binary Arithmetic Coding (CABAC) [19]. Binary codes are packetized into slices and encapsulated into NAL units (NALUs).

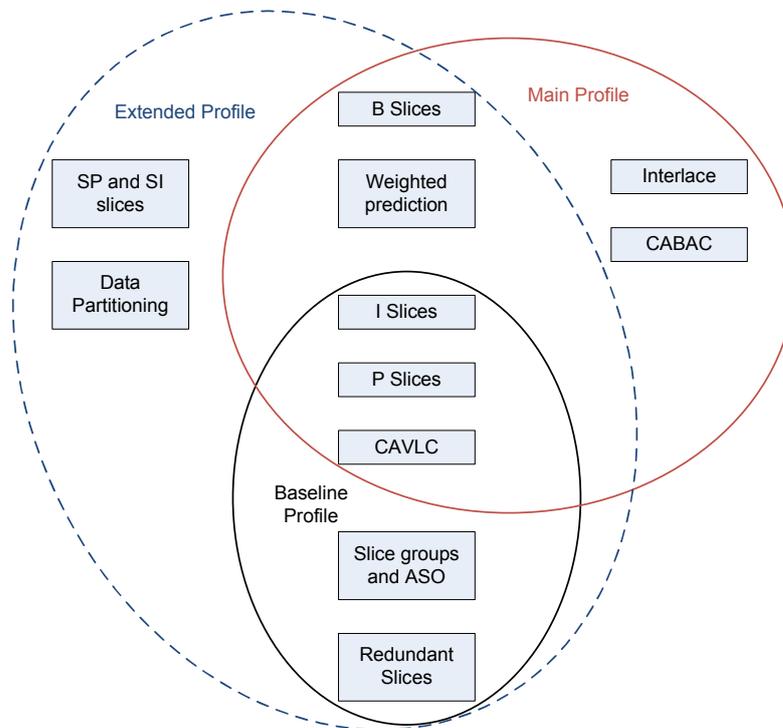


Figure 3.11: H.264 Baseline, Main and Extended Profiles

Figure 3.11 illustrates the main features of H.264 Baseline, Main and Extended profiles. In this thesis, baseline profile is used.

3.3.2 Simulcast Coding

Simulcast refers to simultaneous broadcast, a term used by broadcasters. Simulcast coding is the most straightforward video coding format corresponding to independent coding of each view in stereo or multiview representations. In this coding format, the coding standard to be used is the conventional H.264/AVC [18]. The encoding complexity and delay is minimum because of the independent coding of the views. However, due to the redundancy among the views, coding efficiency is poor. This coding method is backward compatible since each view is encoded as an independent stream allowing a conventional receiver to decode one of the views.

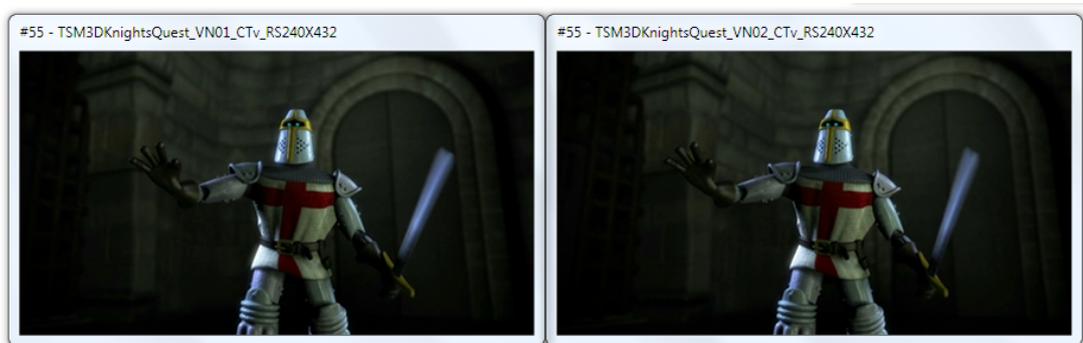


Figure 3.12: Left and Right Views Simulcast

3.3.3 Frame Compatible Coding

Stereo interleaving is also referred as *frame-compatible stereo* format [11]. In stereo video, interleaving of the views is realized by time or spatial multiplexing of the views. In Figure 3.15 (a) is the temporal interleaving where left and right views are multiplexed in time, sending one of each at every time instance. This scheme causes a decrease in frames per second (fps) rate of the video. Spatial multiplexing can be done by either putting two views inside a frame in over\underset format as seen in (b); or in side-by-side format as in (c). Spatial multiplexing causes a decrease in the resolution of the video, by either halving the height or width. Interpolation is used to obtain the left and right views at the receiver side. In order to understand the ordering of the views, H.264/AVC uses stereo SEI message for the signaling. This format, enables the inter-view prediction and signals the use of it by SEI message. Since

it enables inter-view prediction and there is also intra-frame redundancy that can be exploited even the inter-view prediction is off; the coding efficiency is much better than simulcast. On the other hand, this video coding format does not support backward compatibility, since 2D receivers are not able to process SEI message, extract and decode on of the views from the multiplexed frames. Still it can be used for storage of media in discs and other environments for 3D devices.

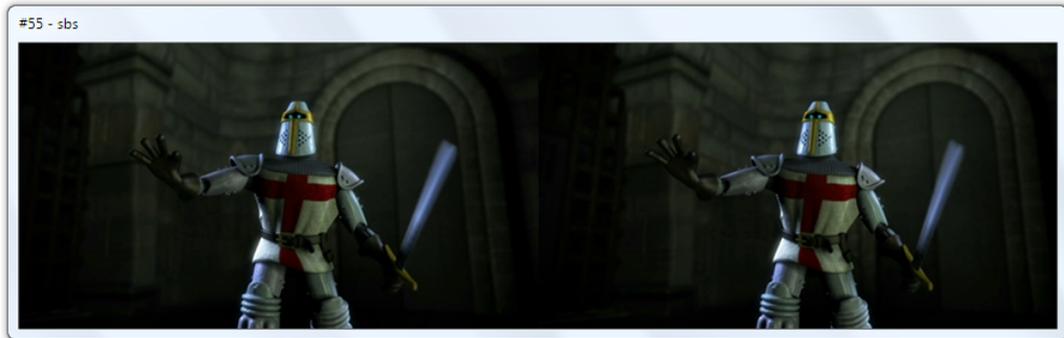


Figure 3.13: Stereo Interleaving, side-by-side



Figure 3.14: Stereo Interleaving, over/under

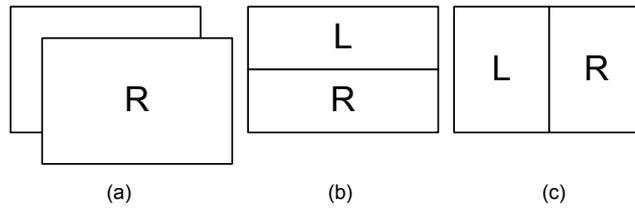


Figure 3.15: Stereo Interleaving schemes, (a) time interleaved, (b) space interleaved over/under, (c) space interleaved side-by-side

3.3.4 2D-plus-depth Coding

This coding method is for the 2D-plus-depth representation. It is backward compatible and similar to simulcast since it also uses independent coding of the 2D and the depth information. 2D receiver are capable of receiving the video while discarding the depth stream. The depth information changes very smoothly, there are too much redundancy between the frames and also inside a frame which results in high compression ratio. In addition to this, it is a one channel sequence which makes the overhead of the second view much smaller compared to other formats. However, the drawback is visual quality decreases when compared to a stereo pair because of the occlusions occurring in the rendering process.



Figure 3.16: 2D-plus-depth

3.3.5 MultiView Coding

This video coding format uses the Multiview Video Coding (MVC) extension of H.264/AVC. It exploits not only the temporal redundancy existing between the frames of a given view

but also the spatial redundancy between the different views of the same scene; resulting in a better compression efficiency than H.264/AVC based simulcast coding. Compression efficiency of MVC is shown in [20] with both objective and subjective quality measuring. MVC is backward compatible since it is solely based on H.264/AVC. One of the views is encoded independently with H.264/AVC and the information regarding the multiview structure is signalled in additional NALUs where NALU type allows the conventional receiver to discard them. Therefore the conventional users receive the independently encoded stream and discard the additional information NALUs while decoding it. The second stream which contains the dependently coded view is also discarded by these users. During the encoding of the second view, inter-view prediction is allowed in addition to temporal prediction within this view. The decoded pictures of other views made available for dependently coded view allowing a wider range of reference pictures for it. All the information regarding the view dependency such as structure of multiview prediction and period of dependently coded frames etc. are signalled in the additional NALUs within the independently coded stream. Still the rate reduction achieved by MVC may not be sufficient compared to available channel bandwidth for systems with large number of views. But in case of stereo video, the rate reduction achieved in comparison to simulcast is significant [11] and the backward compatibility provides advantage over stereo interleaving schemes. It is also preferable to 2D-plus-depth due to the reduced quality of rendered views in the latter one.

CHAPTER 4

SYSTEM OVERVIEW

In this chapter, the framework for end-to-end delivery of stereo video over DVB-H is presented. Figure 4.1 illustrates the system components and the transmission environment. Main processes in an end-to-end video delivery system are the capturing, coding, transmission, decoding and display. For the system presented in this chapter, content capturing is out of context. Section 4.1 is about coding of the contents and explains the details of encoding. Network protocols used down to link layer are explained in Section 4.2 and its subsections. Link and physical layer encapsulation process is described in Section 4.3 followed by Section 4.4 which provides a short information about channel simulation. The last two sections of the chapter provide detailed information about decapsulation and decoding of the received content respectively.

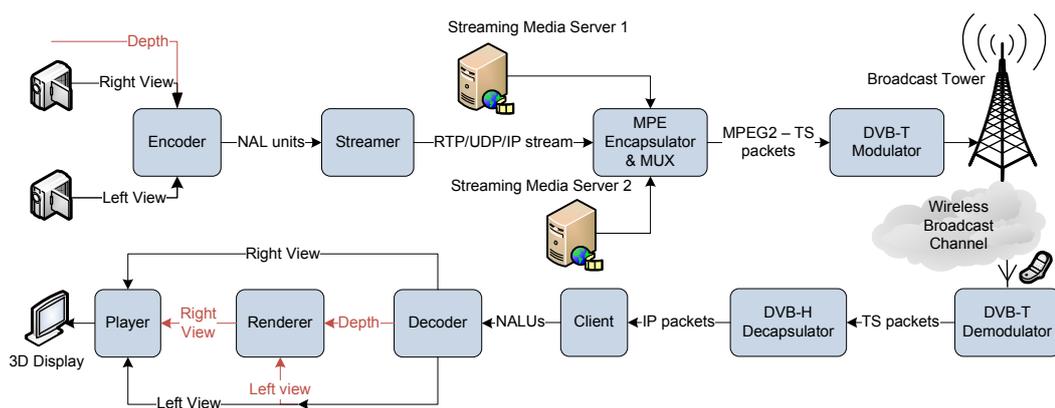


Figure 4.1: Block diagram of the system

4.1 Encoding

In the system, the two representations of 3D video, stereo video and 2D-plus-depth are both implemented. Since these methods have two streams of input videos, the system supports a transmission of two streams simultaneously where they are multiplexed in the link layer of the transmitter and demultiplexed at the link layer of the receiver. This is similar to multiplexing of several streams from different sources in conventional transmission scenario.

In this thesis, the studied 3D video coding formats are simulcast and MVC with only two views (stereo MVC). Stereo interleaving is not included because of backward compatibility issue. Compression performance of simulcast and MVC are compared.

For simulcast coding, inter-frame prediction structure within a view is defined before encoding. For MVC, in addition to this, the inter-view prediction structure must also be defined. There are a number of different prediction structures that can be used with MVC. The advantages and disadvantages of these structures are investigated in [20] [21]. It has been show that use of Hierarchical-B pictures improves the coding efficiency significantly at the cost of increased complexity [21]. Therefore, in this study, use of B-pictures and hierarchical structures are avoided due to the reduced capabilities of mobile devices. Only predicted (P) pictures are allowed within a view with a Group of Pictures (GOP) size of 1. The period of intra-coded (I) pictures is set to 8. Number of reference views is set to 2. This structure is named as IPP since it starts with an I frame and continues with P frames along the intra period. The inter-view prediction structures to be used are defined according to frame types. Figure 4.2 illustrates these structures. The prediction structure which allows prediction from all frames of the reference view is defined as IPP Full prediction and shown in (a). On the other hand, the prediction structure which restricts prediction to only from the I frames (anchor frames) of the reference view is called IPP Simplified (Simp) prediction and shown in (b). IPP Full prediction provides higher compression efficiency with respect to IPP Simplified prediction at the cost of increased complexity and dependency. Performance of the prediction structures under lossy transmission conditions will be investigated in the following chapters.

There are large number of parameters affecting the encoding process. One of the most important parameters of the encoder is the Quantization Parameter, QP. QP controls the quantization since quantizer step size (Qstep) is determined from QP. The output bitrate of the

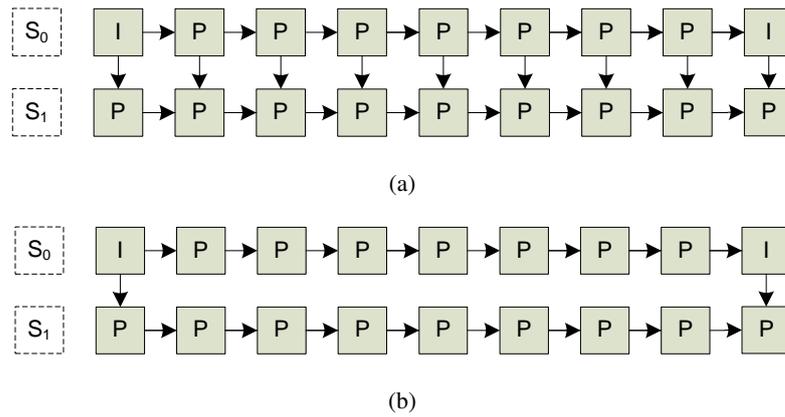


Figure 4.2: Prediction structures S_0 left view, S_1 right view; (a) IPP Full, (b) IPP Simp

encoded video is directly related to QP which determines the quality. QP ranges between 1-51 where the low QP results in high quality whereas high QP results in low quality. When there are two views, the joint quality of the views vary as the quality of each view changes. Therefore, a rate distortion curve is obtained by encoding the views with different QP pairs to obtain the QP pairs that result in the highest quality for a given bitrate. In this thesis, only the effect of QP is included in the study. The error resilient tools embedded in standard H.264/AVC are the data partitioning, slice interleaving, flexible macro block (MB) ordering (FMO), SP/SI frames, reference frame selection, intra block refreshing and redundant slices [3]. SP/SI frames and reference frame selection requires feedback from the decoder, therefore they are not used since DVB-H does not have a feedback channel. Data partitioning, slice interleaving, Flexible Macroblock Ordering (FMO), intra block refreshing and redundant slices are the candidates to be used in MVC. However, none of these tools are implemented in JMVC Reference Software [22] for MVC extension of H.264/AVC. Data partitioning is to group the encoded data according to their priorities in the Video Coding Layer, VCL. An alternative way to implement data partitioning is to group the NALUs carrying different frame types outputted by the VCL. For example, I frames and P frames can be separated to obtain two partitions. This alternative way of data partitioning is employed in order to study the effects of layering on the transmission of 3D video over DVB-H. Detailed information about the implementation of layering is given in following chapter.

4.2 Streaming

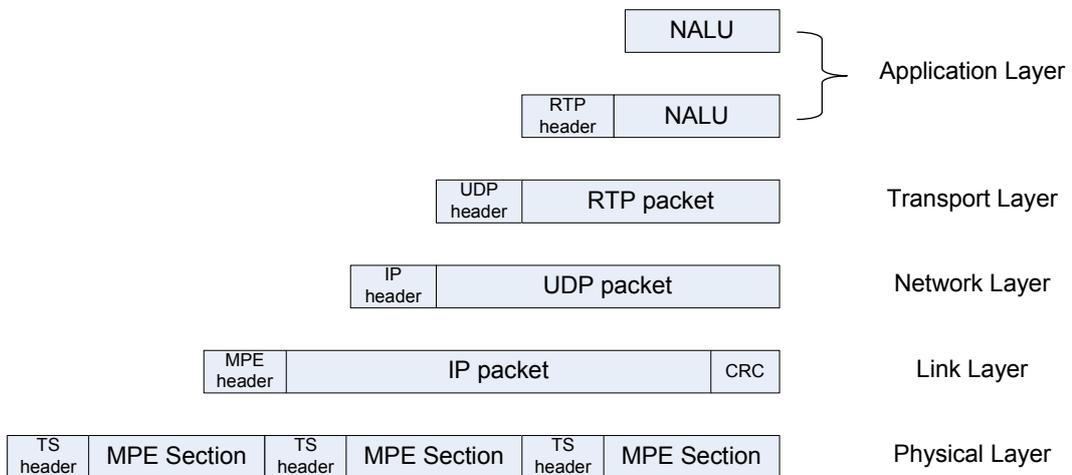


Figure 4.3: Packet structures in the different layers of the network

Encoder outputs are two parallel NALU streams carrying the left and right views. These streams are fed to the streamer which is responsible for the control of the transport. Streamer encapsulates the NAL units into Real Time Transport Protocol (RTP), User Datagram Protocol (UDP) and finally Internet Protocol (IP) datagram and feed them to the encapsulator. In the following subsections details of these protocols are provided.

4.2.1 Real-time Transport Protocol

Real-time Transport Protocol (RTP) is defined for the delivery of packetized audio/video streams over IP networks. It is developed by the Internet Engineering Task Force (IETF) and reported in RFC 3550 [23]. Multimedia streaming applications are often tolerable to packet losses to some extent, whereas delivery of the packets on time is crucial due to real-time characteristics of these applications. RTP compensates for the jitter and out of sequence arrivals that occur during the transmission on IP networks. It also supports data transfer to multiple destinations through multicast [24].

RTP header format is illustrated in Figure 4.4. The most important fields of the RTP header are the payload type, sequence number and timestamp. Payload type (PT - 7 bits) indicates

the data type carried in the RTP packets. Sequence number (SN - 16 bits) is set and used in accordance with RFC 3550. For the single NALU and non-interleaved packetization mode, the sequence number is used to determine decoding order for the NALU. RTP timestamp (32 bits) is set to the sampling timestamp of the content. In case of NAL units such as parameter set and SEI which do not contain timing information, the RTP timestamp is set to the RTP timestamp of the primary coded picture of the access unit in which the NAL unit is included. The clock rate to be used with RTP is 90 kHz.

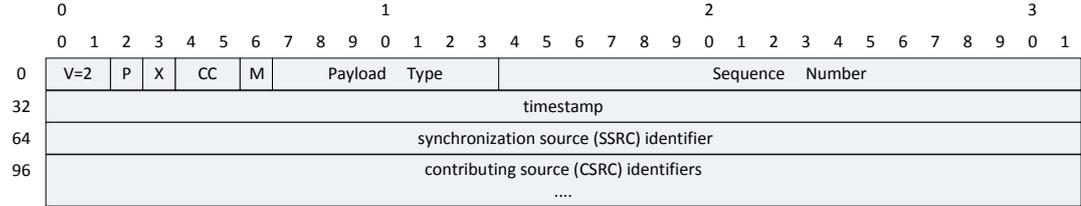


Figure 4.4: RTP Header format

RTP allows packetization of one or more NALUs as the RTP payload. Aggregation of NALUs with identical NALU-time (e.g. slices of a frame) is supported with Single Time Aggregation Packet (STAPs) and aggregation of different NALUs having not necessarily the same NALU-time is possible by means of Multi-Time Aggregation Packets (MTAPs). Depending on the transmission network characteristics, NALUs having sizes larger than maximum transfer unit (MTU) size of the network are fragmented by RTP in order to avoid OSI level 3 network fragmentation. Fragmentation is only allowed for single NAL units (not with the aggregated packets) [25].

4.2.2 User Datagram Protocol

User Datagram Protocol (UDP) is a simple transmission model that works in transport layer of the OSI model. It is formally defined in RFC 768 [26]. It assumes the Internet Protocol (IP) as the underlying protocol and aims to provide a communication environment where for datagrams in a packet-switched network. This protocol has a minimal mechanism where handshaking dialogues which provides the reliability, ordering and data integrity, are avoided. In this sense, it is an unreliable best-effort method with possible loss, duplication and out

of order arrivals of the datagrams. Therefore, time-sensitive applications such as multimedia streaming, often use UDP because packet losses are more tolerable than packet delays in real-time applications. Applications which require data integrity and reliability shall use Transmission Control Protocol (TCP) [24].

UDP provides integrity verification (via checksum) of both the header and payload. However, if transmission reliability is required, it has to be implemented in the upper layers since UDP do not provide any guarantees for message delivery, in fact it does not keep track of the messages sent which makes it unreliable.

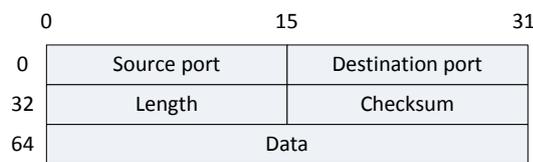


Figure 4.5: UDP Header format

UDP header is a total of 8 bytes long, consisting of 4 fields of 2 bytes each. It is illustrated in Figure 4.5. The first part is the source port which is an optional field and filled with zero in case not used. Depending on the application it may be used to indicate the port of the sending process to which a reply should be addressed in the absence of any other information. Second part is the destination port and third part is the length of the datagram meaning the data plus the header expressed in octets. The last part is the checksum included for error-checking of the header and data. It is calculated using a pseudo-header from the IP header.

4.2.3 Internet Protocol

Internet Protocol (IP) is the main core protocol that makes the internetworking possible in connectionless packet-switched networks. IP is responsible from relaying and routing datagrams across network boundaries. It basically takes the datagrams from the host and deliver them to the destination by the use of addresses. It defines addressing methods and structures to be used during the encapsulation of the datagrams and provides logical location of the addresses by these methods. The most commonly used version is the Internet Protocol version 4 (IPv4) which is the first version of the protocol defined in RFC 791 [27]. Currently IPv6 is

also active and becoming used more and more everyday.

The Internet Protocol is a best-effort delivery method. It assumes that the network is unreliable and is dynamic in terms of availability of the links and nodes. Therefore, the intelligence is in the end nodes whereas routers in between the end points are only responsible from storing and forwarding of the datagrams. Forwarding is done according to the addresses that are kept inside each datagram within the header [24].

	0	4	8	14	17	19	19-31
0	Version	Header Length	Type of service		Total length		
32	Identification				Flags	Fragment offset	
64	Time to live		Protocol		Header Checksum		
96	Source IP address						
128	Destination IP address						
160	Options						
192	Data						

Figure 4.6: IPv4 Header format

The header format of IPv4 is illustrated in Figure 4.6. As explained in [24], IPv4 header consists of 14 fields, 13 of which are required. First part is the Version (4 bits) and the second part is the Internet Header Length (IHL - 4 bits). IHL indicates the number of 32 bit words present in the header and the minimum value is 5 corresponding to 20 bytes. The next field was originally named as Type of Service, then changed to Differentiated Services Code Point (DSCP - 6 bits) by RFC 2474. Explicit Congestion Notification (ECN - 2 bits) is for network congestion and can be used when both parts support it. Following 16-bits is defined as the Total Length which is the size of the whole datagram including header and data, in bytes. Identification field (16 bits) is responsible from identifying the fragments of a larger IP datagram. Flags is a three-bit field that is used to control or identify fragments. Following Fragment offset field (13 bits) specifies the offset of a particular fragment relative to the beginning of the original unfragmented IP datagram. This is measured in eight-byte block units. Time-to-live is an eight-bit field specified in seconds where time units smaller than a second are rounded up to 1 second. It determines the lifetime of a datagram in order to prevent it being forwarded back and forth in the network. Next field is the Protocol (8 bits) which specifies the protocol of the data being carried inside the IP datagram. Error-checking of the header is conducted by the following 16-bit Checksum field. In the network, each hop

computes the checksum of the header and compares it with the value of this field in order to decide whether the packet will be discarded or not. The integrity of the data however is tracked by the upper layer protocol (both UDP and TCP have error-checking mechanisms of their payload). Both Source and Destination IP Address fields are 32-bit long, and they carry the address information of the host and the destination which are assigned according to the addressing methods of Internet Protocol version 4. There is an optional field following the destination address, which allows the use of additional header options which are not very common in nature. Details of these options are specified in the associated RFC. Finally there exists the Data field where the length of the data can be the maximum of a 16 bit word minus the header length which is generally 20 bytes [24].

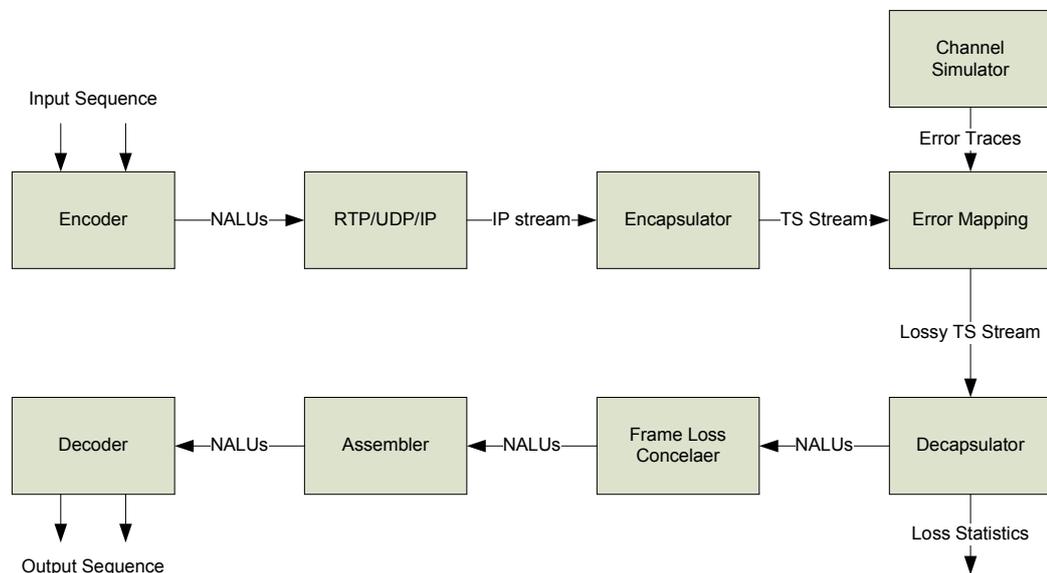


Figure 4.7: Block diagram of the software implementation of the end-to-end system

4.3 Encapsulation

Encapsulator block is the part where the DVB-H specific packetization and error correction coding take place. The implementation of the encapsulator is a modified version of the open source software FATCAPS which is presented in [30]. The modifications involve implementation of missing features of the standard as well as changes to support the 3D streams. One missing feature was the assignment of fixed burst durations to each streaming during mul-

time-multiplexing. Another missing feature was the multiplexing of two streams consecutively with two different identifiers. This software is forward modified in the scope of this thesis in order to be compatible with the MVC encoder output bitstreams, which is a replacement of the old encoder presented in [30]. Transmission of the video sequence in terms of multiple of Group of Pictures (GOP) is realized by checking the NALU frame numbers from the header information of each IP packet and counting the desired number of frames. The operations conducted in link layer are well explained in the DVB-H chapter. The use of two channels is favored in order to apply unequal protection between the left and right video streams. 3D capable receivers should be modified in order to receive two streams instead of tuning into one channel only. This is a simple operation and details are explained under the Decapsulator section.

4.4 Simulation of the Physical Layer and the Mobile Channel

The Simulink model for the physical layer of a DVB-H transmitter, a receiver and the DVH-H wireless channel is presented in [28] and [29]. In this thesis, physical layer and the channel is simulated using this model. Figure 4.8 illustrates the model. The upper chain of the model is the transmitter part. The lower chain is the receiver part which applies the reverse operation to the received signal. In between the two there exists a model of the real channel which is a Rayleigh Fading channel model with Additive White Gaussian Noise. The inner structure of the channel model is illustrated in Figure 4.9.

A random integer generator is used to create the incoming link layer output to be passed to the first block of the physical layer. The link layer output are the MPEG2-TS packets which are 188 bytes of length consisting of 4 bytes of header and 184 bytes of payload data. The TS packets are entered to the structured RS Encoder (204,188) block which generates 16 bytes of parity and adds it to the TS packet, outputting 204 bytes. The following blocks are convolutional interleaver ($l=12$) block, punctured convolutional code block, the DVB inner interleaver, DVB Mary Quadrature Amplitude Modulation (M-QAM) mapper block, Transmitter Parameter Signaling (TPS) pilot insertion block, Orthogonal Frequency Division Multiplexing (OFDM) block and finally the Guard Interval insertion block. After the channel model, we have the receiver chain where the reverse operations are conducted on the received signal. The receiver part assumes perfect channel estimation. It starts with the Guard Interval

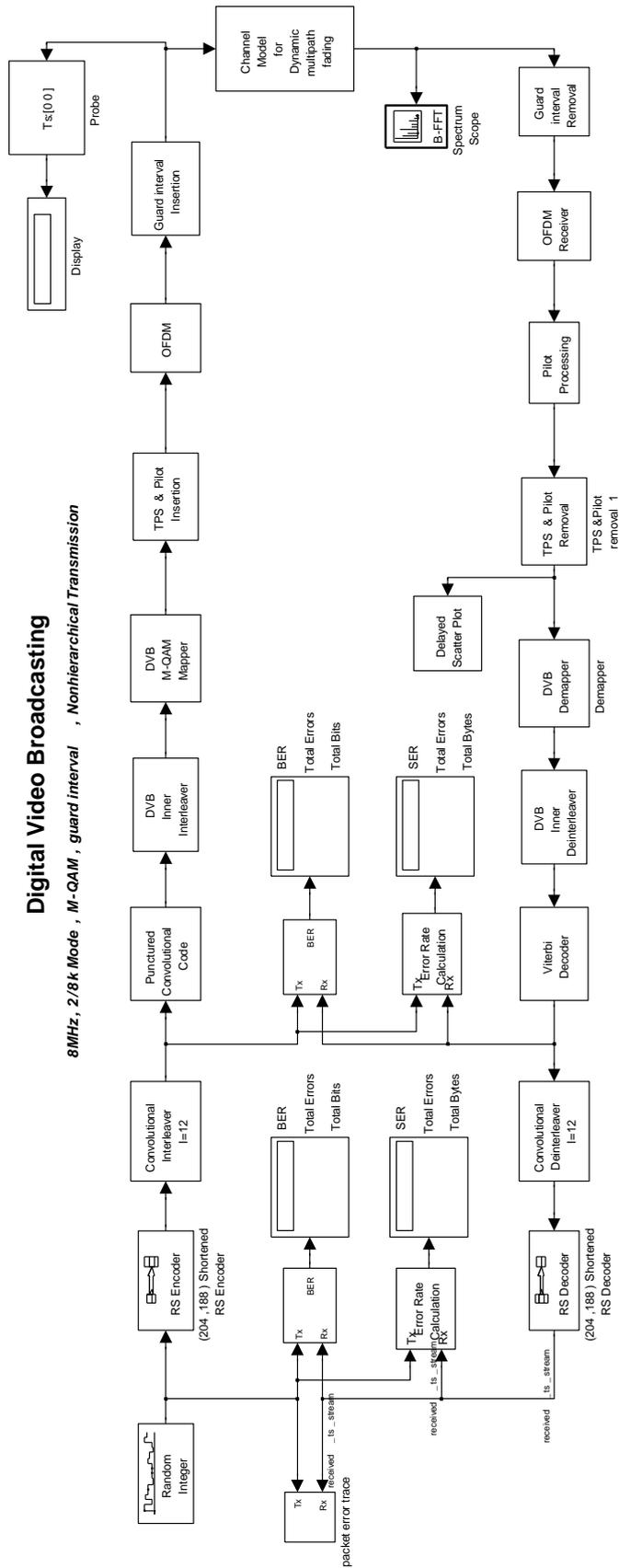


Figure 4.8: Simulink model of the DVB-H physical layer (transmitter/receiver) and the mobile TU6 Channel

Removal block, followed by the OFDM receiver and the pilot processing blocks. The next blocks are the TPS pilot removal block, DVB demapper, DVB inner deinterleaver, Viterbi decoder, convolutional deinterleaver ($l=12$) and finally the RS decoder block. The randomly generated input stream which consists of packets of 188 bytes are compared with the output of the receiver chain. The result of the comparison is marked in the Transport Error Indicator (TEI) field of the TS packet header where a 0 indicates correct reception and 1 indicates an erroneous TS packet. By this way, an error trace of a real transmission is generated and it is used later to map the channel error on the transmission experiments.

The channel is modeled as a Multipath Rayleigh Fading Channel with Additive White Gaussian Noise (AWGN). There a number of channel modes implemented within the Simulink model and the path delays and path gains corresponding these models are given in the Tables 4.1 and 4.2. Each row corresponds to a tap value and each column is a different mode label as listed below:

1. 1 tap flat fading
2. Typical urban 6 taps
3. Typical urban 12 taps
4. Bad urban 6 taps
5. Bad urban 12 taps
6. Rural area 4 taps
7. Rural area 6 taps
8. Hilly terrain 6 taps
9. Hilly terrain 12 taps
10. Indoor commercial
11. Outdoor residential

Table 4.1: Path delays corresponding to the listed channel modes implemented in the Simulink model

Tap	1	2	3	4	5	6	7	8	9	10	11
1	0	0	0	0	0	0	0	0	0	0	0
2		0.2	0.2	0.4	0.2	0.2	0.1	0.2	0.2	0.05	0.45
3		0.6	0.4	1	0.4	0.4	0.2	0.4	0.4	0.15	0.5
4		1.6	0.6	1.6	0.8	0.6	0.3	0.6	0.6	0.225	1.05
5		2.4	0.8	5	1.6		0.4	15	0.8	0.4	3.25
6		5	1.2	6.6	2.2		0.5	17.2	2	0.525	6
7			1.4		3.2				2.4	0.75	8.3
8			1.8		5				15		10
9			2.4		6				15.2		12.05
10			3		7.2				15.8		15
11			3.2		8.2				17.2		
12			5		10				20		

Table 4.2: Path gains corresponding to the listed channel modes implemented in the Simulink model

Tap	1	2	3	4	5	6	7	8	9	10	11
1	0	-3	-4	-3	-7	0	0	0	-10	-4.6	-6
2		0	-3	0	-3	-2	-4	-2	-8	0	-3
3		-2	0	-3	-1	-10	-8	-4	-6	-4.3	0
4		-6	-2	-5	0	-20	-12	-7	-4	-6.5	-1.5
5		-8	-3	-2	-2		-16	-6	0	-3	-4.7
6		-10	-5	-4	-6		-20	-12	0	-15.2	-3
7			-7		-7				-4	-21.7	-12
8			-5		-1				-8		-14.5
9			-6		-2				-9		-17.4
10			-9		-7				-10		-21.7
11			-11		-10				-12		
12			-10		-15				-14		

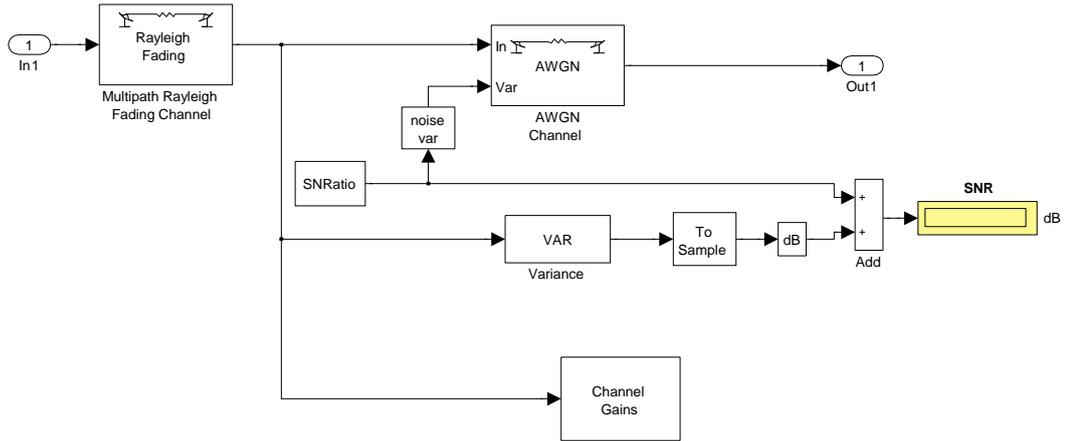


Figure 4.9: Channel Model

4.5 Decapsulation

Decapsulator is the unit where the received stream is processed for error detection and possibly correction. After the error detection and correction at the physical layer, Transport Error Indicator (TEI) bit in the header of the TS packet is set to 0 by the demodulator if it can correct the errors. TS packets are unpacked into MPE and MPE-FEC sections; the MPE frame is filled by the sections according to section erasure method, i.e. contents of the error-free MPE and MPE-FEC sections take their unique places in the MPE-FEC frame table and rest of the frame entries are marked as erasures. RS decoding is done according to the number of entries marked as correct in the MPE-Frame. Since each codeword is calculated row-wise, number of correctly received bytes in a row is checked for error correction. The decapsulator is capable of providing IP, UDP, RTP and H.264 Annex B type outputs. Since there is no operation on IP, UDP and RTP packets, we use the H.264 Annex B output in this system. Software implementation of the decapsulator is obtained from Tampere University of Technology and presented in [30].

4.6 Frame Loss Concealer

This block is added to the end-to-end chain because of the lack of error concealment mechanism in the decoder. The reference encoder and decoder couple used in the system is version

5.0.5 and it is not capable of decoding in case of a missing frame in either of the streams. Modification of the encoder to support error concealment for a missing frame is complicated therefore another solution is utilized for the testing of the offline tests. The solution is the additional frame loss concealer block. This block takes the received streams and check for the missing frames by looking at the Picture Order Count (POC) values of the streams. In case of an existence of a missing frame, it inserts to the stream a fake NALU with the missing POC number where the contents of this NALU indicates a copy of the previous frame. When the decoder takes fake NALU, it does not do any decoding but just copies the previously decoded frame into the reconstructed view. This procedure is illustrated in Figure 4.10. For frames consisting of more than one slice, when some of the slices of a frame are lost, slice copy concealment is used. Note that this kind of loss is different than full frame losses, so the frame copy concealer is not utilized here. The concealment of the lost slices are done in the decoder and it is explained under the Decoder section.

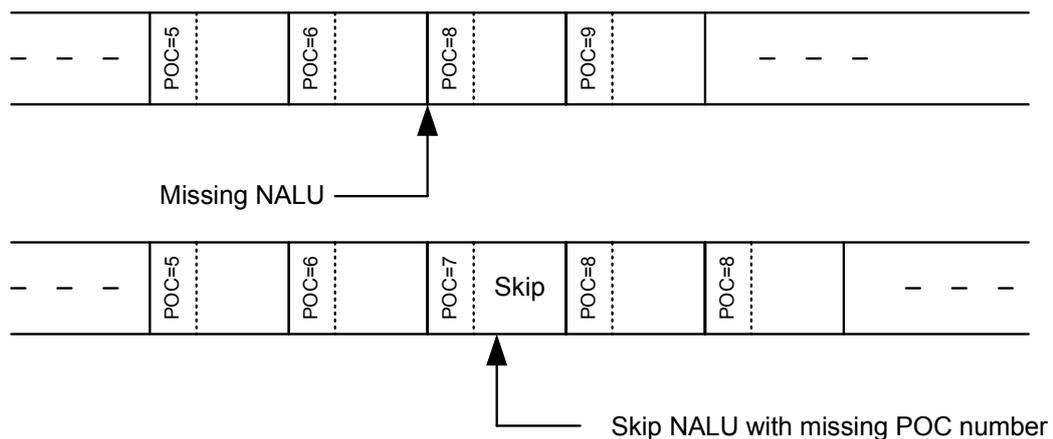


Figure 4.10: Lossy stream including a missing NALU with POC=7 (up) and the stream at the output of the frame concealer with a Skip NALU with the missing POC number inserted (down)

4.7 Assembler

This block is necessary since the reference software decoder requires the input left and right views to be assembled into one single stream. The stream consists of NALUs starting with the identifiers such as the Sequence Parameter Set (SPS) and Picture Parameter Set (PPS) which

contains the parameters that are necessary for the decoder. It is followed by the NALUs of the first frame of left view and then the NALUs of the first frame of right view and continues this way.

4.8 Decoding

Error concealment is not a normative part of H.264/AVC. Two methods for intra and inter error concealment is added into H.264/AVC reference software as informative methods. Intra error concealment uses spatial interpolation using weighted averaging of neighboring pixels. Inter error concealment uses Motion Vector (MV) prediction using neighboring MVs and zero MV with a boundary matching algorithm. However, these concealment strategies are not included in the reference MVC software [22]. In order to achieve error concealment, a modified version of the JMVC 5.0.5 software which utilizes basic frame/slice copy concealment in case of losses which is reported in [31]. In order to implement frame copy concealment, frame loss concealer (bitstream corrector) is used before decoding. Bitstream corrector inserts skip frames into the bitstream for the frames lost. Frame losses can be detected by checking the POC numbers of slices. If all slices of a frame are lost, this module inserts skip frame. Skip frame is a small byte stream consisting of all MacroBlocks (MB) coded by skip mode. By inserting skip frame, decoder is enforced to use frame copy concealment method. The implementation details of slice mode and the slice copy concealment are also reported in [31]. In Hierarchical B-picture coding, the encoding/decoding order of the frames are not same as the display order (i.e. POC number are not consecutive). However using skip MBs or skip frames allow decoder to conceal current frame by using the frame that is first in the prediction buffer of the current frame. The first frame in the prediction buffer is the frame that is closest to the current frame in the encoding order.

CHAPTER 5

SIMULATIONS

There are three distinct operations in the simulations of the end-to-end system. These are the encoding, channel simulation and the transmission simulation. Encoding and channel simulation are conducted separately and outputs of these operations (H.264 streams and channel error traces) are used in the transmission simulation. Simulation environment is illustrated in Figure 5.1. Shaded blocks indicate the offline nature of the operations.

This chapter is consisted of three sections, first one is about the encoding of the data, second one is about the channel simulations and the last one is about the transmission simulations.

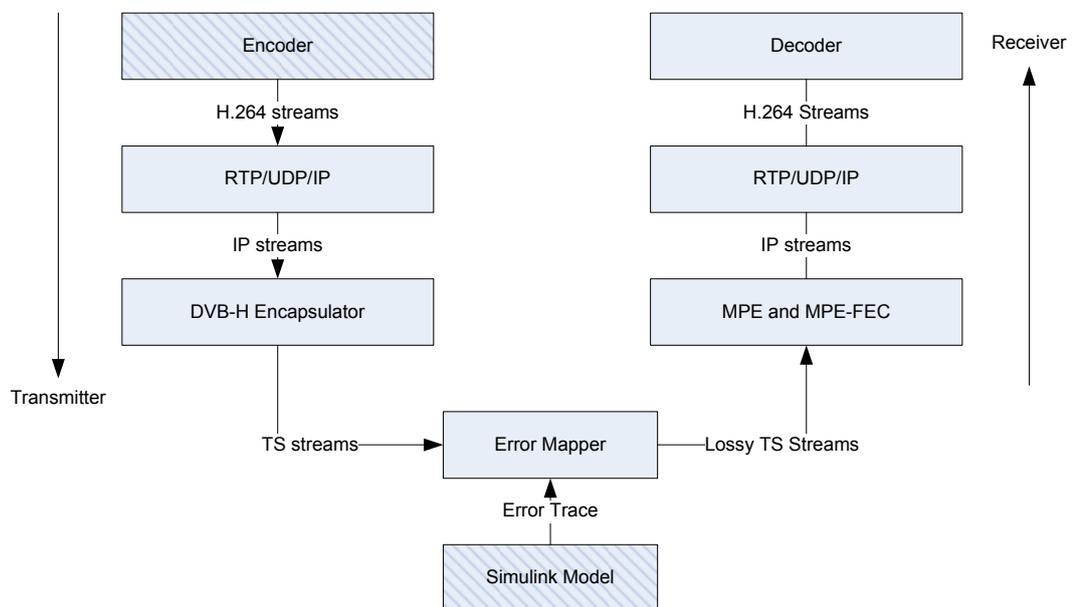


Figure 5.1: Layered structure of the simulations

5.1 Encoding

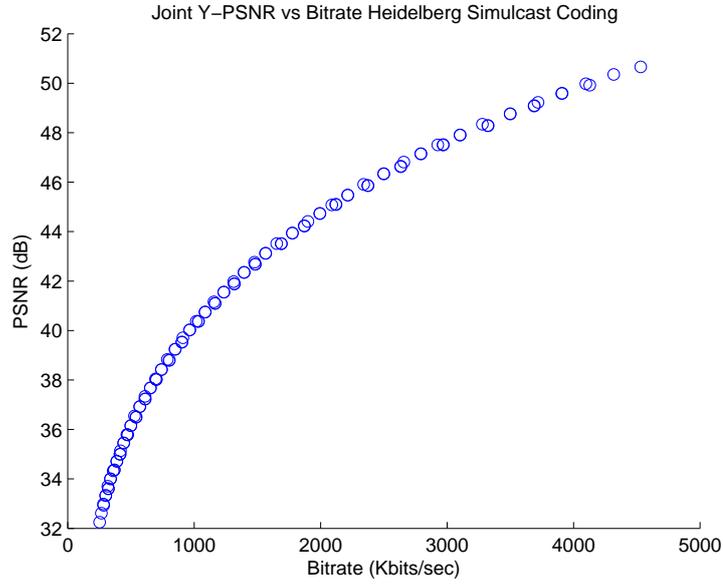


Figure 5.2: Joint PSNR vs total bitrate for Heidelberg Simulcast Coding

The initial step of our simulations is the encoding. The contents are encoded with varying QP in order to obtain a set of videos with a wide range of bitrate and corresponding quality. In chapter 4, the details of the encoder is provided. JMVC Reference software uses Peak Signal-to-Noise Ratio (PSNR) as a quality metric in order to calculate the distortion in a single frame. The quality of a whole video sequence is determined from averaging the PSNR values of each frame in the sequence. PSNR is defined as:

$$PSNR = 10 \cdot \log_{10} \left(\frac{255^2}{MSE} \right) \quad (5.1)$$

where MSE is the Mean Squared Error. MSE is calculated for Y, U and V channels separately where Y contains luminance information while U and V contain chrominance information. U and V pixel values are downsampled by four and grouped together in YUV-420 format which is used in the tests. Each channel information of a pixel is represented by 8-bits, hence have a value in the range 0 - 255. By taking the squared value of the difference of pixel values for the original and the reconstructed images, the error for each pixel is calculated and then

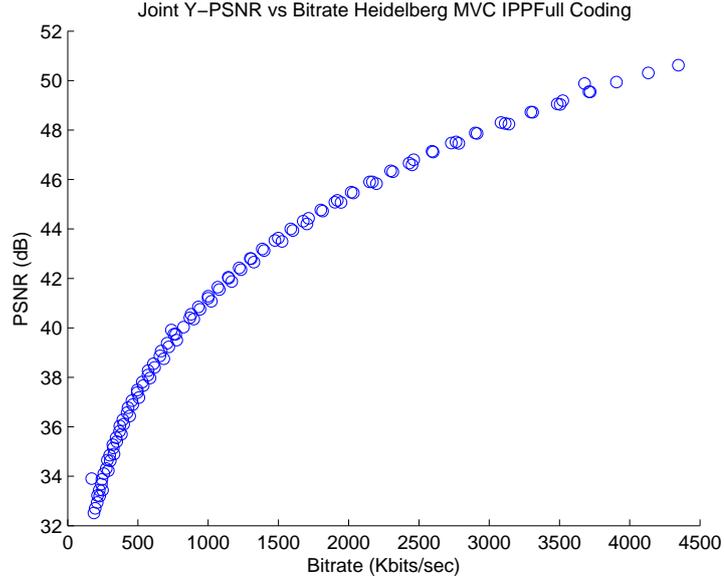


Figure 5.3: Joint PSNR vs total bitrate for Heidelberg MVC IPP Full Coding

these values are added in order to obtain the error for a frame. MSE calculation for a frame is defined as:

$$MSE = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n (I(i, j) - K(i, j))^2 \quad (5.2)$$

where $I(i,j)$ is the original pixel values and $K(i,j)$ are the reconstructed/distorted pixel values. While evaluating the quality of a stereo video, joint PSNR is used which is defined as:

$$PSNR_j = 10 \cdot \log_{10} \left(\frac{255^2}{(MSE_l + MSE_r)/2} \right) \quad (5.3)$$

Joint PSNR is a PSNR based metric for stereo video which includes the error in both views equally to calculate the distortion. In this thesis only PSNR values of luminance data (Y PSNR values) are used in design and comparisons.

Rate distortion curve for a single video is obtained by plotting the distortion change of a set of videos with increasing bitrates in a range. For stereo video, the distortion is measured as the Joint PSNR and the bitrate is the total bitrate of left and right views. The curves are obtained

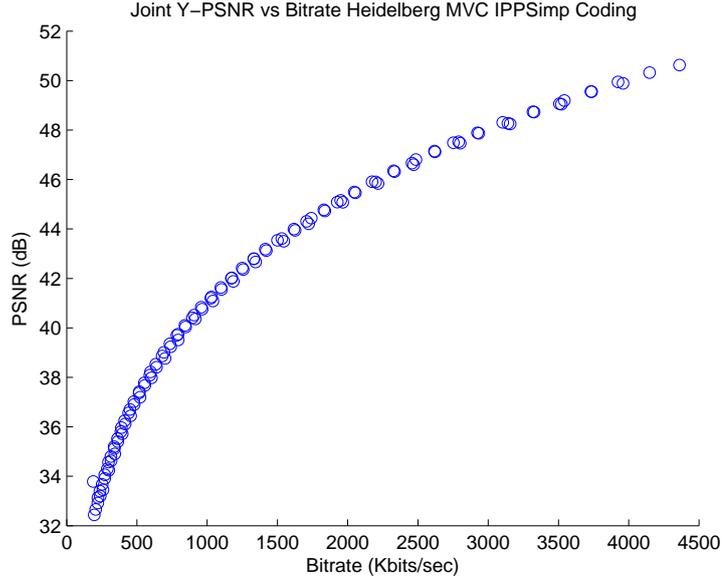


Figure 5.4: Joint PSNR vs total bitrate for Heidelberg MVC IPP Simplified Coding

by varying the Quantization Parameter (QP) of each view. In simulcast coding, left and right views are encoded independently and QP of each view can be varied independently. In case of MVC, due to inter-view dependency between left and right views, combinations of QP pairs are varied jointly. As a result of varying QP values, many QP pairs are obtained. And some of the QP pairs fall under the convex hull, making it an unreasonable choice. When a given bitrate is targeted for a coding scheme, the QP pair which results in the target bitrate with greatest joint PSNR is selected. Similarly, a given joint PSNR is targeted, the QP pair at the target PSNR with smallest bitrate. It has been found that convex hull consists of the QP pairs with a difference of $[-2,2]$. Figures 5.2, 5.3 and 5.4 are the RD curves of Heidelberg video for Simulcast, MVC IPP Full and MVC IPP Simplified codings, obtained by this method.

5.2 Channel Simulations

The parameters of the physical layer of DVB-H is presented in Table 2.1. Selection of these parameters affects bitrate, network coverage, mobility and error robustness of the transmission. In this study, the aim is to use a commonly used parameter set to simulate the DVB-H

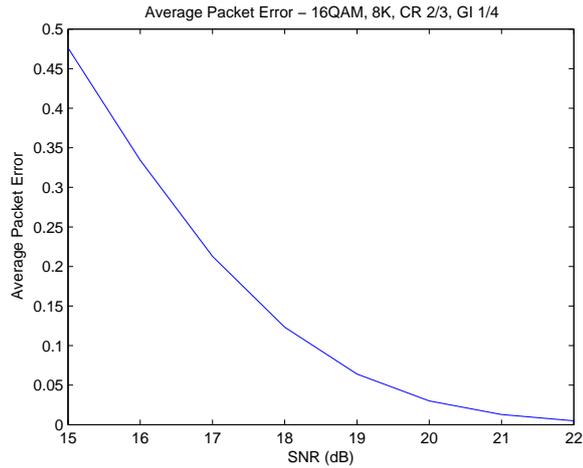


Figure 5.5: Average packet error rate vs channel SNR for the error traces

transmission environment. [32] presents transmission results with QPSK and 16QAM modulations. However, since QPSK provides the lowest bandwidth, broadcasters do not use this modulation. By considering DVB-H applications in several countries, the parameters summarized in Table 5.1 are chosen to be used with the transmissions. Figure 5.5 is the average packet (TS packets) error rate for a range of channel SNRs obtained from the error traces generated by the Simulink model. The TS packet error rate is the ratio of number of erroneously received packets to total number of packets. DVB-H is based on bursty transmission therefore the length of the burst packet errors (errors that are continuous) are important. Figures 5.6 and 5.7 provides the average of packet error burst length and its variance. For channel SNR 17dB, the average length of a burst of errors is around 16 packets where as for channel SNR 19 dB it is around 11 packets.

Table 5.1: DVB-H physical layer and channel model parameters

Modulation	16QAM
Convolutional Code Rate	2/3
Guard Interval	1/4
Carrier Mode	8K
Channel Bandwidth	8 MHz
Channel Model	TU6
Carrier Frequency	666 MHz
Doppler Shift	24 Hz

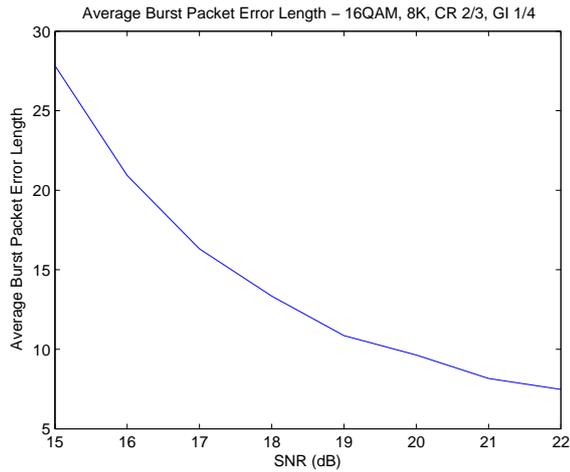


Figure 5.6: Average burst packet error length vs channel SNR for the error traces

Figure 5.8 provides the TS packet loss (error) rate distribution for the channel SNRs 17 dB to 21 dB. It is seen that, for 17 dB, the error rate is centered at 20 percent which is a significantly high amount of error. The error rates decrease rapidly as the channel SNR increases and in case of 21 dB, the channel is almost lossless. In the transmission experiments, 17 dB is taken as the high error channel condition and 19 dB is taken as the low error channel condition.

5.3 Transmission Simulations

In this thesis, a large set of transmission experiments are designed and conducted in order to see the effect of several parameters on the transmission of stereo video over DVB-H channel. The parameters that are going to be studied are as follows:

- Layering
- Coding method
- Prediction Structure (for MVC)
- Protection strategies for left and right views
- Rate allocation among video quality and FEC

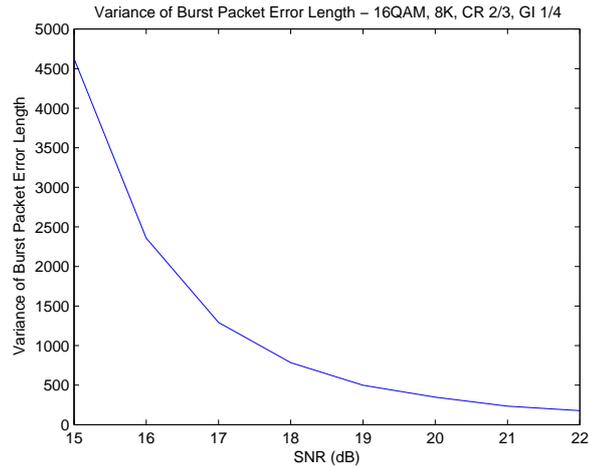


Figure 5.7: Variance of burst packet error length vs channel SNR for the error traces

This section explains the design of the experiments by defining the experimental variables and describing the methods that are going to be tested and compared. In the following subsections, firstly, the implication of coding method and prediction structures into the tests are explained. Next, the design of rate allocation among video quality and FEC, the layering and the protection strategies are explained in order. Finally, the environment of the experiments are drawn by introducing the experimental variables. The results of the experiments are provided and discussed in the following chapter.

5.3.1 Coding Method and Prediction Structure

Compression performance of the two coding methods and the prediction structures are given in [21]. Compression rate of MVC is significantly better than Simulcast coding. The reason is that MVC exploits also the redundancy between the views [20]. Still, Simulcast coding is included in the tests in order to see its transmission performance in comparison to the other coding methods.

When left and right views are encoded with MVC extension (High Profile, Hierarchical B-pictures, CABAC), the simplified and full prediction structures have a very similar compression performance as showed in [21]. However when no B-pictures are allowed and the prediction structure within a view is restricted to IPP, there exist a significant difference. Due to the increased prediction, full prediction structure has a better compression rate compared to

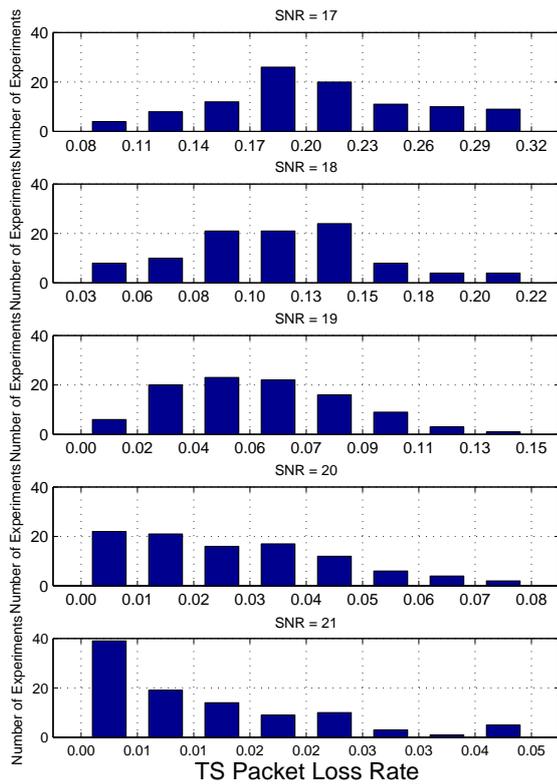


Figure 5.8: TS packet loss distribution of the error traces generated for the tests

simplified structure. However, the increased prediction also increases the dependency of one view to the other, which may result in a poor performance in case of error. Therefore, these two prediction structures are going to be compared in the tests.

When comparing the compression performance of two methods, the higher compression rate is better since it allows the same quality video at a lower bitrate. In case of transmission of these two methods, the lower bitrate video will allow higher protection rate given a constant channel bandwidth or it will be more distorted given a constant protection rate. The reason of the increased distortion is because of the increased dependency. Losing a certain packet will affect more than one area in a frame or frames, causing the error to propagate. In case of stereo video with MVC coding, this dependency is further increased to dependency between the views, making the error propagate also between the views.

The comparison of the coding methods and the prediction structures are based on two criteria:

Equal Bitrate The video encoded with Simulcast coding, having a certain total bitrate (total of left and right view bitrates) is chosen as the base video. Then the videos having the same bitrate with this base video, are chosen from the MVC IPP Full and MVC IPP Simp curves. In this criteria, there are a total of 3 videos for a content, each having the same bitrate and varying quality levels. Among the three, MVC IPP Full, which has the highest compression rate, has the highest joint PSNR value at the chosen bitrate. Simulcast coding has the lowest value and IPP Simplified is in between. Since they have the same video bitrate, they are protected with the same rate at the transmission level.

Equal Quality This time, the videos having the same quality (joint PSNR) value with this base video, are chosen from the MVC IPP Full and MVC IPP Simp curves. In this criteria, there are a total of 2 additional videos for a content, each having the same joint PSNR and having varying bitrate levels. Having the same quality with the Simulcast coded base video, MVC IPP Full has the lowest bitrate, Simulcast has the highest bitrate and MVC IPP Simplified is between the two.

Figure 5.9 illustrates the selection of the test videos for a certain content. The chosen video bitrate to be worked with is 600 Kbps. Therefore, from the RD curve of Simulcast Coding, a QP pair that has the closest bitrate (RD curve is not continuous, the selection is done within a margin) is chosen as the base video, which is the Sim-IPPFull-28,28 (The label IPPFull is irrelevant since there is no inter view prediction for Simulcast coding). The quality of the base video is 37.40 dB. The first selection criteria is the equal bitrate and the corresponding videos are selected according to it. MVC-IPPFull-26,27 and MVC-IPPSimp-26,28 are the equal bitrate associates of the base video having PSNR values 38.55 and 38.20 dB respectively.

The second selection criteria is the equal quality and the corresponding videos are selected according to it. The joint PSNR value of the base video is 37.40 dB and this value is chosen as the target PSNR to extract from the RD curves of the other two methods. The closest values are MVC-IPPFull-27,29 and MVC-IPPSimp-28,28 having 37.50 and 37.40 dB PSNR respectively. Note that these methods are labeled as MVC2 in the graph in order to indicate that they are a second set of videos to be used in comparisons, apart from the first one.

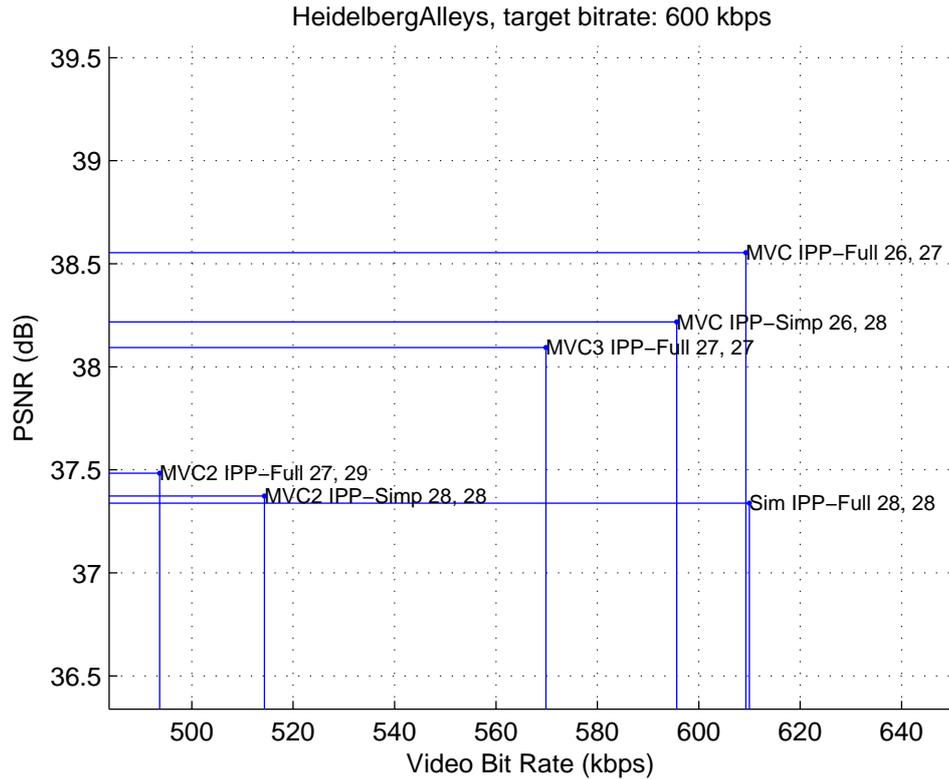


Figure 5.9: Selection of the test videos according to coding method, prediction structure and rate allocation

5.3.2 Rate Allocation Between Video Quality and FEC

As a result of the equal quality and the equal bitrate criteria, there are two sets of videos for MVC IPP Full and MVC-IPP Simplified codings. Using the labels explained above, the only difference between the MVC-IPPFull-26,27 and MVC2-IPPFull-27,29 is the varied QP. In the transmission scenario, for fair comparison, these two videos are allocated with the same channel bandwidth, which is accompanied by varying the protection rate. Hence, MVC-IPPFull-26,27 having a higher video bitrate and quality, is associated with a certain protection rate which is smaller than the protection rate of MVC2-IPPFull-27,29. This is due to the reduced video bitrate and quality which provides more bandwidth for the protection. Therefore these two videos can be compared to see the effect of the rate allocation between the video quality and FEC. In Figure 5.9, there is also a third method labeled MVC3-IPPFull-27,27 which is added to have another sample with a better quality than MVC2 and lower bitrate than MVC. In fact, the IPPFull points constitute a line on which one may find the optimal rate allocation

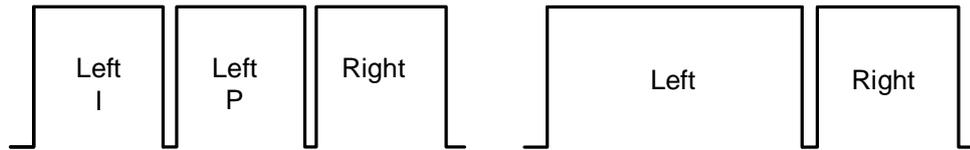
between the video bitrate and FEC for a given channel condition. However this optimization is not in the scope of this thesis.

5.3.3 Layering

Data partitioning is an error resilience tool which allows separating syntax elements according to their importance into different packets of data. It allows the use of unequal error protection (UEP) on the partitions to improve the robustness against errors. It has been shown to be beneficial for mobile video communication in [33]. The dependent structure of multi view coding causes several different importance levels which can be exploited for unequal protection strategies. Therefore the effect of layering is included in tests to be investigated.

The encoding structures used in this thesis are explained in Section 4.1 and illustrated in Figure 4.2. For full prediction structure, both I and P frames of left view are references for the right view, so they are of equal importance in terms of view dependency. However, for simplified prediction structure, only the I frames of the left view are references for the anchor frames of the right view. P frames of left view are of similar importance to P frames of right view in terms of the stereo video quality.

Backward compatibility requires the separation of left and right view frames which results in at least two layers in the transmission. For Simulcast coding and MVC IPP Full prediction coding, the transmissions are conducted over two layers, one consisting of the frames of left view and one consisting of the frames of right view. These two-layered transmissions are labeled as Le-Ri meaning that it is consist of frames of left and right views. The layering comparisons are studied with the IPP simplified prediction cases due to the existence of three importance layers for this case. For MVC with simplified reference prediction structure, right view does not depend on P frames of left view, therefore instead of making the left view frames one layer it can be separated into two layers as I frames and P frames. Hence, for IPP simplified prediction there is an additional three layered transmission scheme for each two layered scenario which enables the comparison of layering effects. The three layered transmission cases are labeled as LeI-LeRest-Ri meaning that the layers consist of I frames of left view, P frames of left view and frames of right view respectively. Figure 5.10 illustrates the implementation of layering in DVB-H physical layer, by means of separate bursts.



(a) Three layered transmission with Left I, Left P, (b) Two layered transmission with Left, Right Right frames frames

Figure 5.10: Layered transmission schemes

5.3.4 Protection Strategies

In the transmission of a compressed video, using forward error correction provides robustness to the channel errors. The disadvantage is the increased transmission bandwidth. In case of a single view, the parameter to be decided is the amount of protection which mainly depends on the error characteristics of the channel. In the transmission tests, rate allocation among video quality and FEC is studied. On the other hand, in case of stereo video, there are more than one burst in the transmission which allows different protection scenarios.

The most straightforward protection scenario is to assign equal protection to both bursts. This is the case for Simulcast coding, since it consists of left and right view data encoded separately resulting in equal importance. However, in case of MVC, there are different importance levels due to the inter-view dependency in the coding of the right view, which makes the unequal protection strategies viable. In the transmission strategy of two layers (left, right views), first of all, an Equal Error Protection (EEP) method is generated using RS(191,256). Then a ratio of the RS columns of right view corresponding to EEP, is transferred to be used as the left view RS columns. This results in different protection rates for left and right view and is referred as Unequal Protection (UEP). For each transmission scenario, six different UEP cases are generated which are labeled with their percentage of transferred RS columns, namely: 5, 10, 15, 20, 25 and 30 %. In case of the three layered (left I frame, left P frames, right frames) transmission, RS column transfer is from the last two layers to the first one.

5.4 Experimental Setup

Three contents are used in the experiments. These are clips of 60 seconds length and can be found in the 3D Video Database of MOBILE3DTV [4]. Table 5.2 provides the resolution and frame rate information of each content. In addition, Table 5.3 summarizes the contents in terms of spatio-temporal characteristics. A sample configuration file used in the encoding of the videos are provided in Appendix. The fixed slice size used in the experiments is set to 1250 bytes in order to avoid RTP fragmentation. Packets with size smaller than 1250 bytes are kept unchanged.

Table 5.2: The contents used in the transmission tests

Content	Width	Height	Fps
HeidelbergAlleys	432	240	12.5
KnightsQuest	432	240	12.5
RhineValleyMoving	432	240	12.5

Table 5.3: Characteristic details of the contents

Content	Characteristics
HeidelbergAlleys	Low Motion, High Detail
KnightsQuest	Computer Generated
RhineValleyMoving	High Camera and Object Motion, Low Detail

Table 5.4 illustrates the main parameters used in the design of the tests. The tests are conducted 100 times using distinct error traces of 17dB and 19 dB channel SNR as high and low error scenarios. There are two coding methods, two prediction structures, two layering structures and seven protection structures. The methods that are going to be compared are as follows:

1. MVC3 - 2 layered - full reference - EEP + 6 UEP
2. MVC2 - 2 layered - full reference - EEP + 6 UEP
3. MVC2 - 3 layered - simplified reference - EEP + 6 UEP
4. MVC2 - 2 layered - simplified reference - EEP + 6 UEP
5. MVC - 3 layered - simplified reference - EEP + 6 UEP
6. MVC - 2 layered - simplified reference - EEP + 6 UEP

7. Sim - 2 layered - EEP
8. MVC - 2 layered - full reference - EEP + 6 UEP

The results are presented in the figures following the same order as above.

Table 5.4: Parameters and methods that define the experiments

Coding methods	Simulcast and MVC
Prediction Structures	IPP Full and IPP Simplified
Slice size	1250 bytes
Number of layers	2 layers (IPP Full, IPP Simplified), 3 Layers (IPP Simplified)
Protection structures	EEP, UEP (6 different cases)
Channel SNR	17 and 19 dB
Number of experiments	100 different error pattern for each transmission

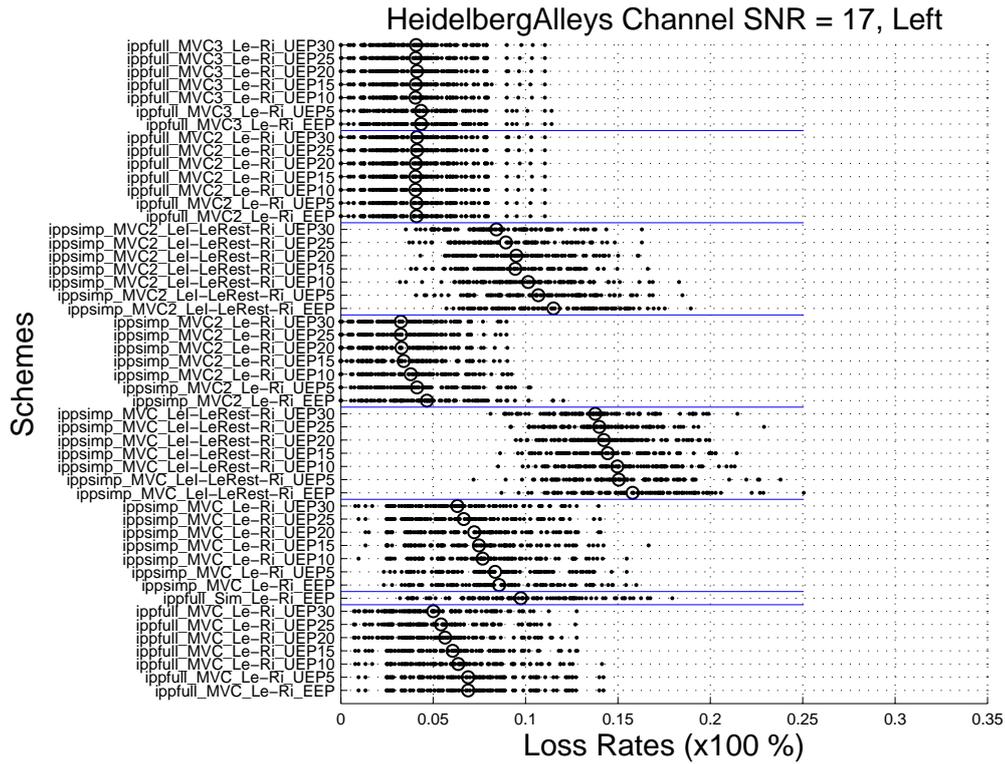
5.5 Results

Results of the transmission tests are presented in a way to indicate the amount of distortion in the video by means of PSNR value in each transmission experiment. In the next subsections, results of the transmissions are presented via figures. In these figures, the y axis has eight different regions each of which corresponds to one of the eight different methods listed above. The x axis represents the PSNR values of the distorted videos obtained from each transmission. The dots correspond to the PSNR value of a single transmitted video obtained from one experiment and the circles correspond to the average of the 100 experiments for that transmission scenario. The discussion of the results are presented separately for high and low error cases.

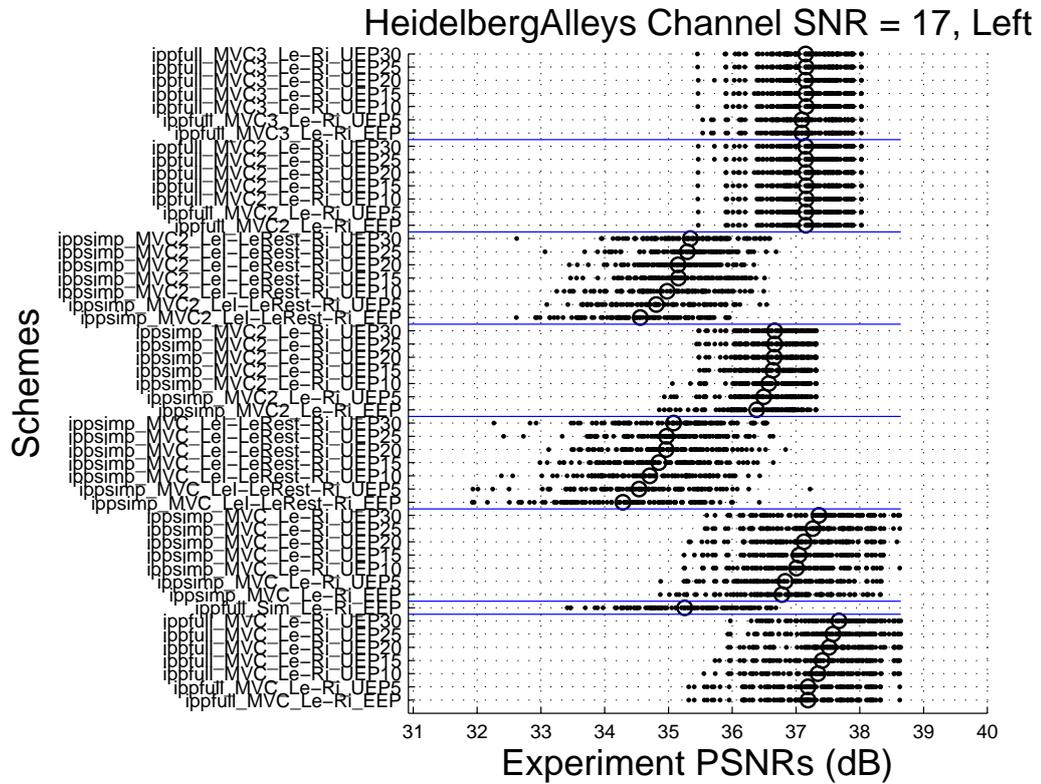
5.5.1 High Error Case

In this study, with the chosen physical layer parameters and channel model, 17 dB channel SNR corresponds to high error case as it is slightly a worse condition than the DVB-H recommendations [1]. Figures 5.11, 5.12 and 5.13(a) are the results of transmissions with 17 dB channel SNR for Heidelberg. The discussion starts with the effect of layering.

There are two different cases that can be used for the comparison of layering effect on trans-



(a) Slice loss rate of all the transmission cases, Left View, Heidelberg, channel SNR=17dB



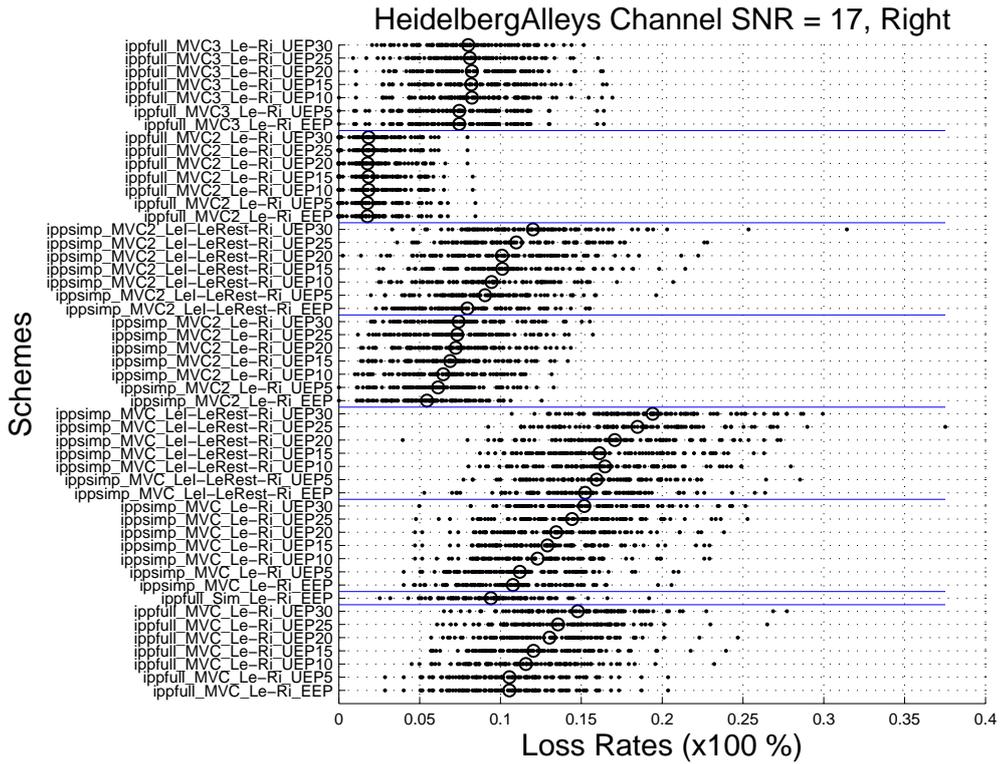
(b) PSNR distribution of the experiments of all the transmission cases, Left View, Heidelberg, channel SNR=17dB

Figure 5.11: Transmission results, Left View, Heidelberg, channel SNR=17dB

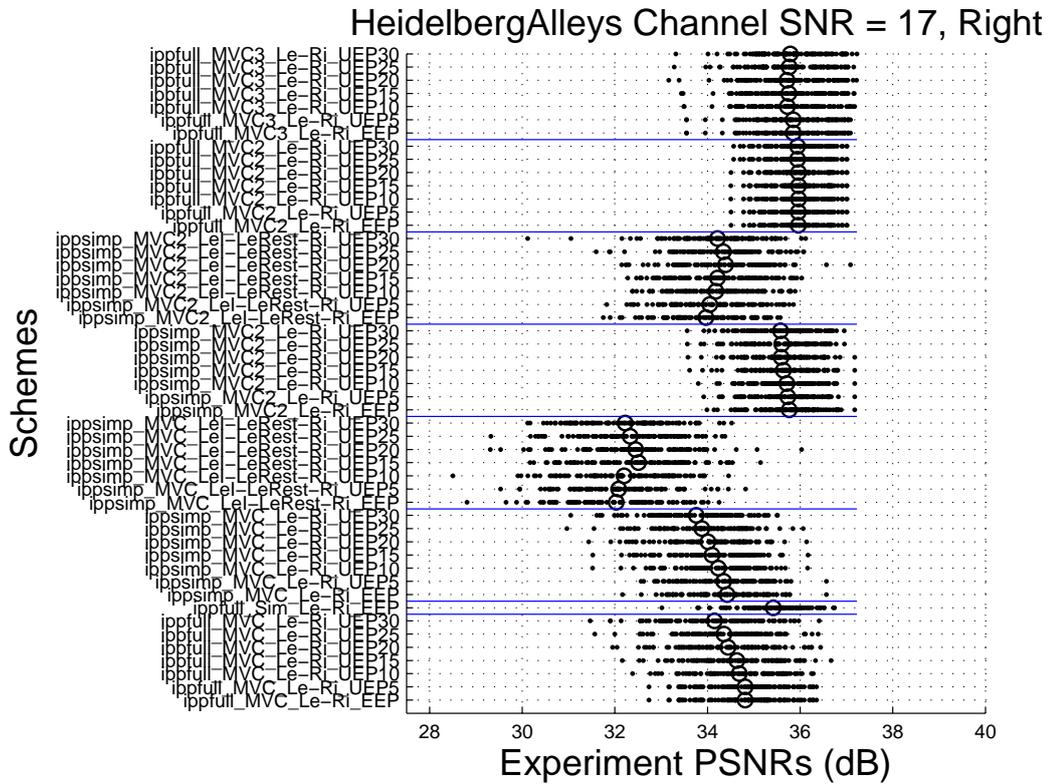
mission. The first one is MVC IPP Simplified Le-Ri (two layers consisting of left and right views) and LeI-LeRest-Ri (three layers consisting of I frames of left view, P frames of left view and right view). The second case, MVC2 IPP Simplified Le-Ri and LeI-LeRest-Ri, is structurally similar to first case, the only difference being the distribution of video quality and FEC rate. In MVC case, the video quality is favored against FEC and in MVC2 case FEC is favored against video quality.

In Figures, transmission cases including IPP simplified prediction structure are the third, fourth, fifth and sixth blocks from the top; third and fourth being the MVC2 cases and fifth and sixth being the MVC cases. Figure 5.11(a) provides the slice loss rates of left view. It is seen that, for both MVC and MVC2 IPP simplified transmissions, slice loss rates of three layered transmissions are significantly higher than those of two layered ones. In fact, the highest average slice loss rate observed in a two layered transmission is still lower than the lowest average slice loss rate observed in the corresponding three layered transmission. Figure 5.11(b) provides the corresponding PSNR distributions and it is in agreement with the slice loss rates. Figure 5.12 provides the slice loss rates and PSNR distributions of the right view. The results are similar to the case of left view, whereas the difference between the two and three layered transmissions are much smaller than the one observed for left view. This shows that, splitting the left view burst into two smaller bursts has an adverse effect on transmission performance in DVB-H. Ideally this would not be expected, however reducing the burst size directly affected the link layer performance of DVB-H. For right view, although the burst size does not change, the protection rate is decreased and this resulted in the increased loss rate. The decrease in PSNR of right view is both from the increased loss rate of right view and due to the decreased PSNR of left view since right view depends on left view I frames. As a result, the joint PSNR given in Figure 5.13(a) of two layered transmission is significantly better than that of three layered transmission cases.

Three layered transmission scenario was considered to employ unequal error protection strategies more efficiently. For left view, unequal protection decreases the loss rate and increases the PSNR for all of the transmission cases except IPP Full MVC3 and MVC2. However, for these cases, the allocation between the video quality is such that all of the protection methods have enough resources to recover the maximum amount of possible errors. This means that these results are saturated and hence should not be considered for the comparison of protec-



(a) Slice loss rate of all the transmission cases, Right View, Heidelberg, channel SNR=17dB



(b) PSNR distribution of the experiments of all the transmission cases, Right view, Heidelberg, channel SNR=17dB

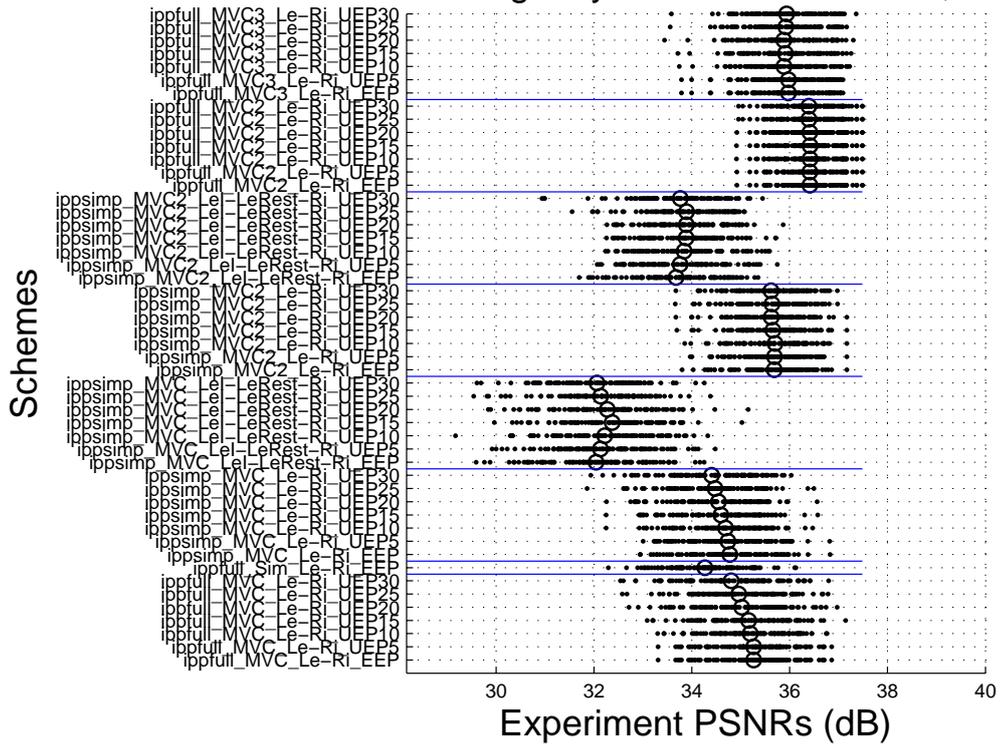
Figure 5.12: Transmission results, Right View, Heidelberg, channel SNR=17dB

tion strategies. The unequal protection methods provide performance improvements in joint PSNR over the equal protection for the three layered transmissions. It has been observed that, increasing the protection of the first burst which contains the I frames of left view, up to 15 percent, improves the joint PSNR. On the other hand for two layered transmissions, although the UEP methods improved the left view PSNR, they cause a slight decrease in right view PSNR. This is the result of decreased protection for right view. Although left view is improved, the increase in slice loss rate of right view causes a decrease in its PSNR values. The resultant effect on joint PSNR can be seen in Figure 5.13(a). According to joint PSNR, the unequal protection does not provide further robustness over equal protection for the two layered transmissions.

In order to evaluate the effect of prediction structure on performance, two layered MVC and MVC2 transmission scenarios are considered. In case of MVC, full reference prediction structure performs slightly better than simplified one. The results are similar for MVC2 where the performance increase is even greater when we consider the simplified and full prediction scenarios. Hence from equal quality and equal bitrate point of views full reference prediction structure provides more robustness independent of the bitrate allocation between quality and FEC.

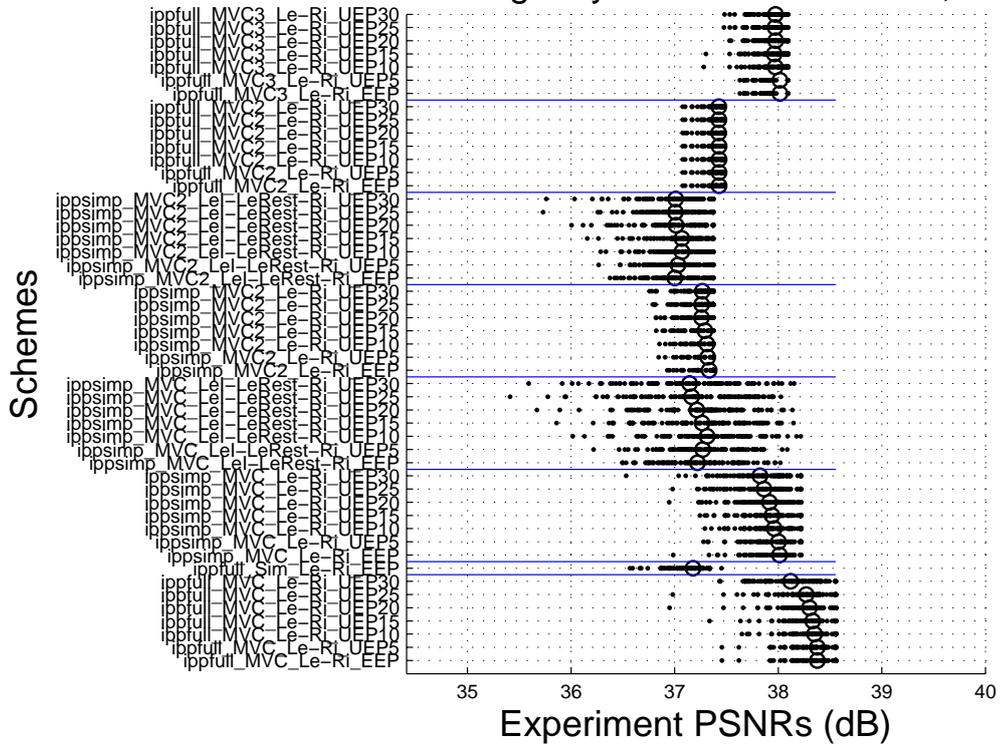
While designing the tests, comparison of methods was based on criteria enforcing equal quality and equal bitrate. Simulcast being taken as the reference, the highest quality video was MVC with full reference structure and the lowest bitrate video was MVC2 with full reference structure. Therefore the optimal allocation of bits among video quality and protection data is expected to be between these two. The performance ordering of these methods according to performance under transmission with channel SNR 17 dB is as MVC2, MVC3 and finally MVC, given in decreasing order. This means that the amount of error in the channel is so high that, higher quality encoding cannot compensate for the losses. The allocation of bits between video quality and protection should be towards increasing protection rates. The data should be compressed as much as possible and protected with FEC. For low channel SNR case (channel SNR 17dB), MVC2 with full reference prediction structure has the best performance regardless of the error protection method used.

HeidelbergAlleys Channel SNR = 17, Joint



(a) Channel SNR=17dB

HeidelbergAlleys Channel SNR = 19, Joint



(b) Channel SNR=19dB

Figure 5.13: Joint PSNR distribution of the experiments of all the transmission cases, Heidelberg

5.5.2 Low Error Case

In this study, with the chosen physical layer parameters and channel model, 19 dB channel SNR corresponds to low error case. Figures 5.14, 5.15 and 5.13(b) are the results of transmissions with 19 dB channel SNR for Heidelberg. The discussion starts with the effect of layering.

Figure 5.14(a) provides the slice loss rates of left view. It is seen that, slice loss rates of three layered transmissions are significantly higher than all the others which are nearly lossless at all. Figure 5.14(b) provides the corresponding PSNR distributions and it is in agreement with the slice loss rates. Figure 5.15 provides the slice loss rates and PSNR distributions of the right view. The results are similar to the case of left view, slice loss rates of two layered IPP simplified transmissions are higher than the ones for left view due to lower protection on right view. For the comparison of two and three layered structures, the argument made in high error case is valid for low error case also. In fact this is independent of channel errors but related to the performance of DVB-H.

In case of left view, only the three layered scenarios has observable errors and the unequal protection methods improve the error recovery successfully. However, for right view there is observable for all the methods except MVC2 IPP full prediction and unequal protection schemes increases the slice loss rates as expected. Therefore for joint PSNR, right view becomes dominant and determines the overall performance which is reduced with unequal protection strategies.

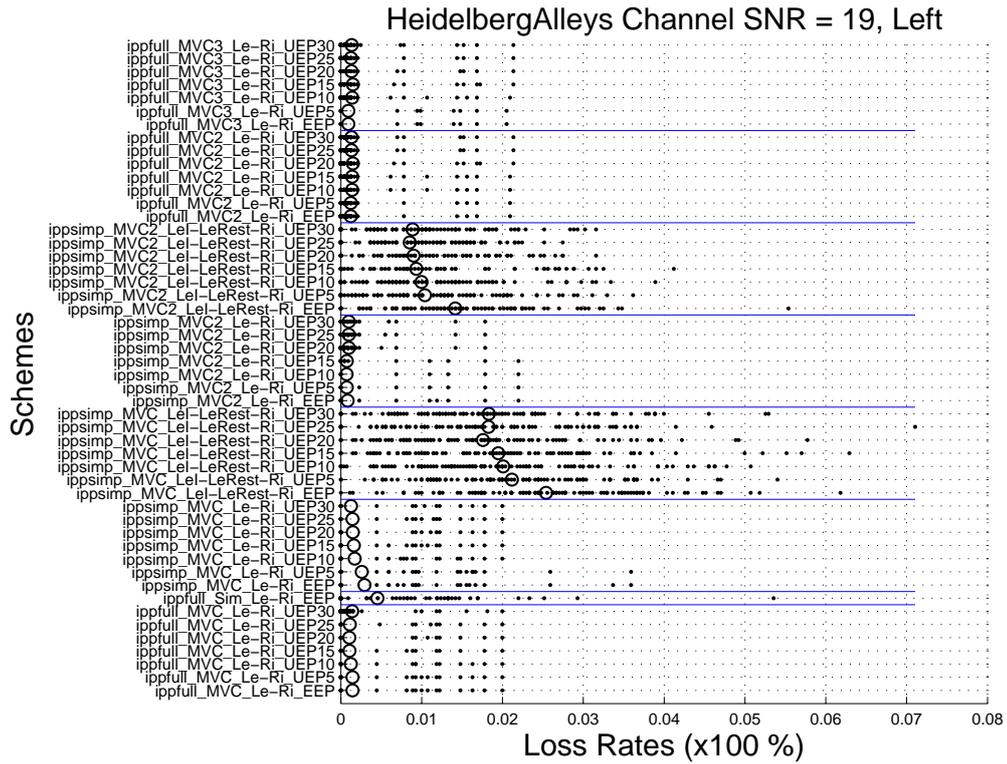
The effect of prediction structure on performance is determined by the compression rate. Since there is low error in the channel, after the recovery of errors, the reconstructed PSNR values become correlated with the original encoding PSNR values. Obviously, full prediction has better results for same bitrate comparisons since it has higher encoding PSNR at the same bitrate and the channel is almost lossless. For equal quality comparisons, the two prediction structures have similar performance since the errors are recovered and encoding quality is the same.

From the rate allocation point of view, in a low error channel with possible FEC protection, favoring the video quality is more reasonable. In comparison of MVC, MVC2 and MVC3, the video with the highest encoding quality performs better. However, we also see that reducing

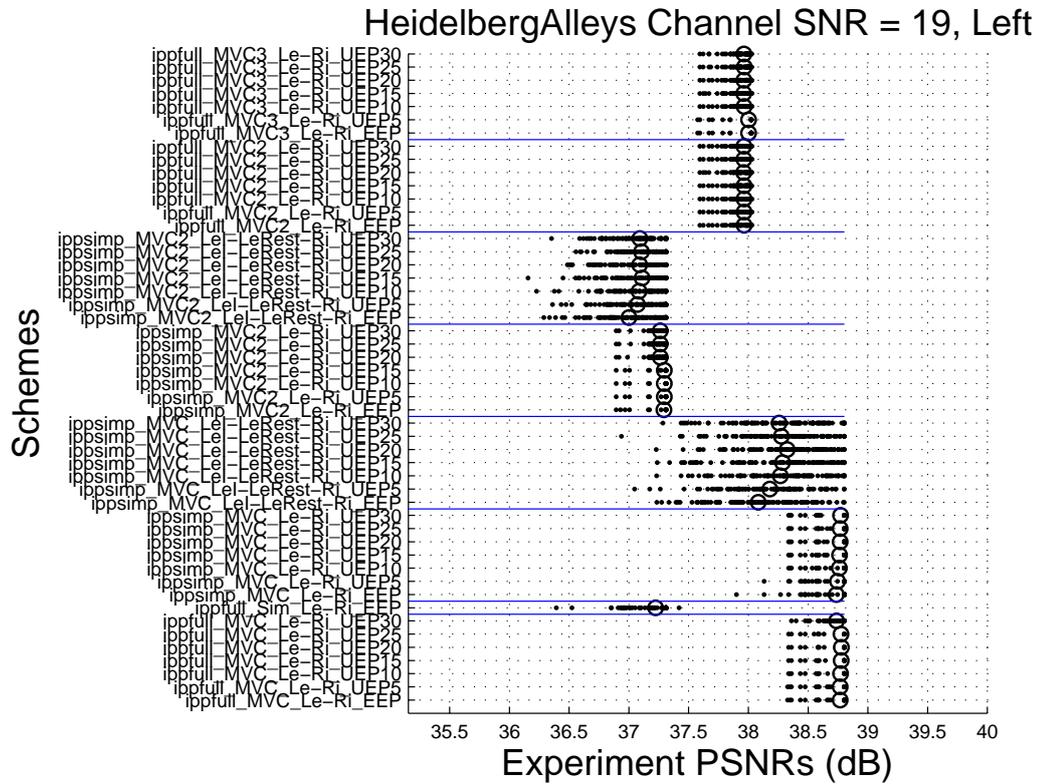
the protection on right view significantly reduces the joint PSNR. This means that the amount of protection used with the video is just right and further increase in the video bitrate (and decrease in FEC bitrate) would not improve the results.

For high channel SNR case (channel SNR 19dB), MVC with full reference prediction structure has the best performance with equal error protection.

The results of the 17 and 19 dB transmissions with the other two contents, RhineValley and KnightQuest, are provided in Figures 5.16, 5.17, 5.18 5.19, 5.20, 5.21 All the discussions made above for the Heidelberg content hold for the other two contents also. Therefore, the results are content independent.

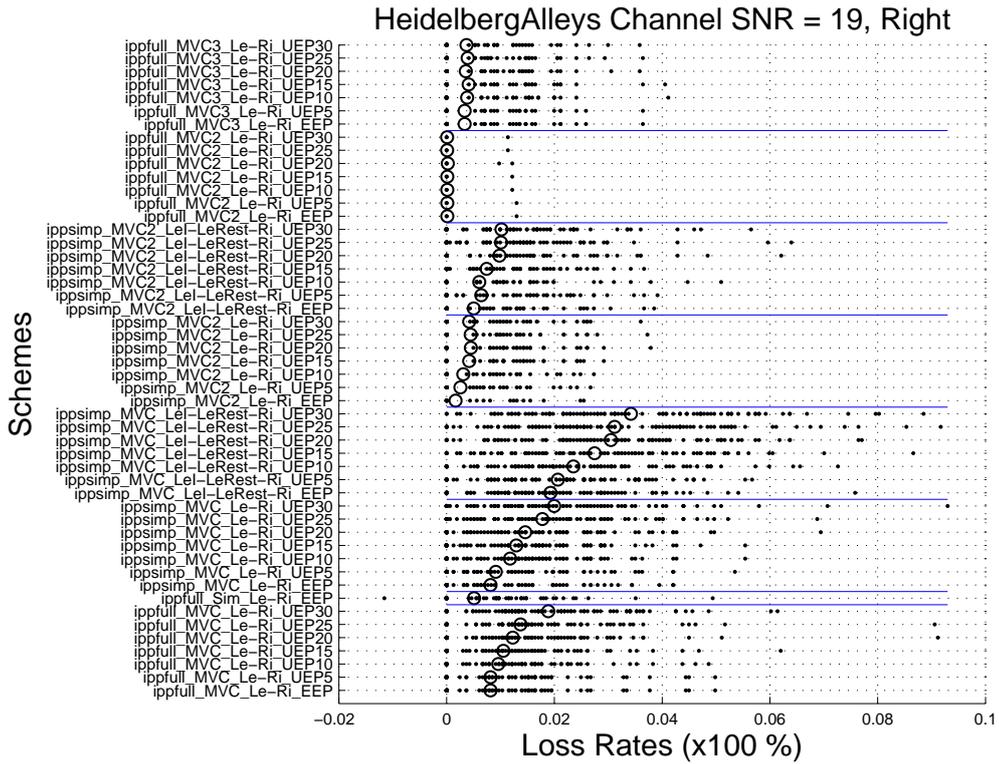


(a) Slice loss rate of all the transmission cases, Left View, Heidelberg, channel SNR=19dB

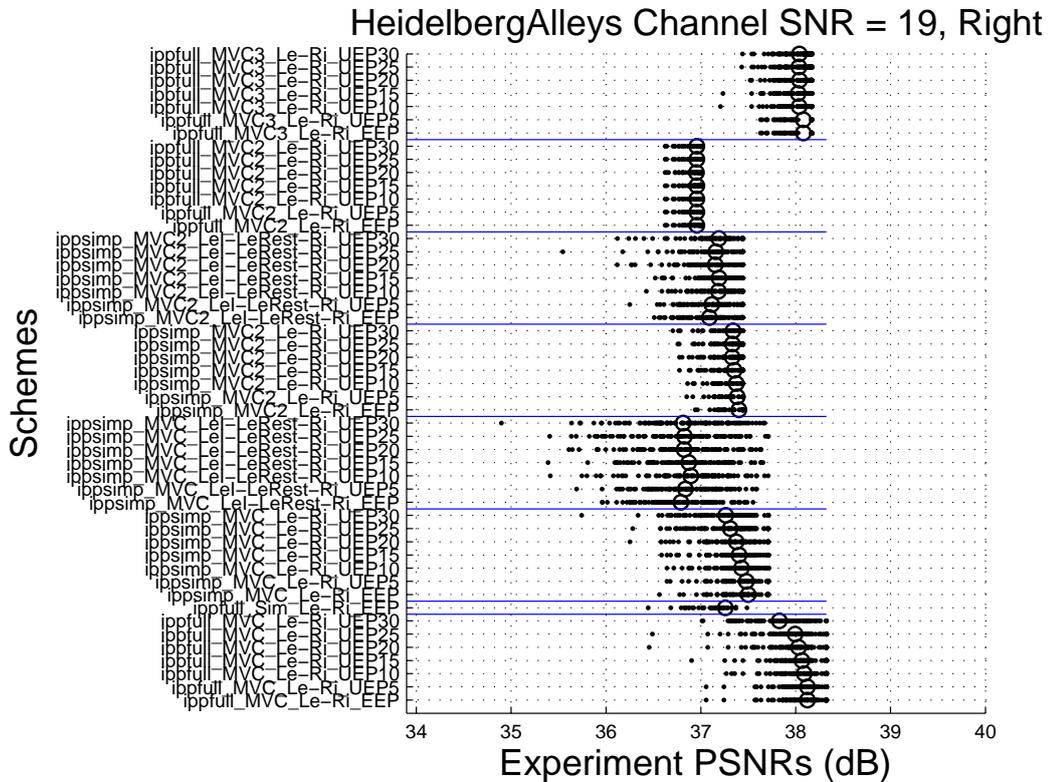


(b) PSNR distribution of the experiments of all the transmission cases, Left View, Heidelberg, channel SNR=19dB

Figure 5.14: Transmission results, Left View, Heidelberg, channel SNR=19dB

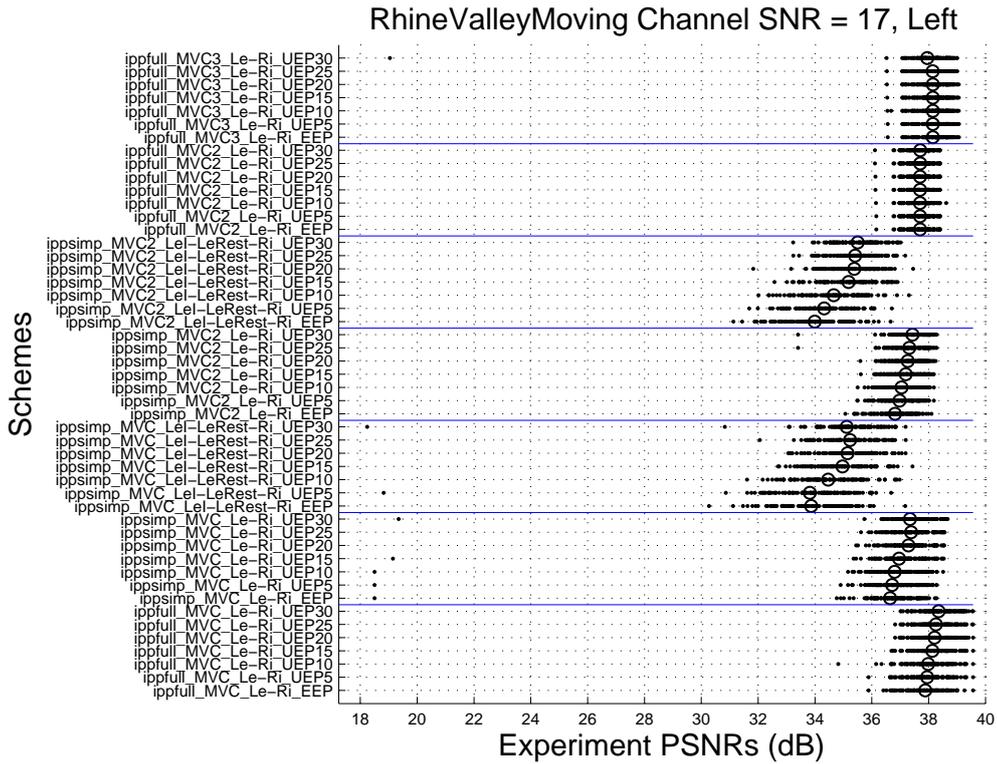


(a) Slice loss rate of all the transmission cases, Right View, Heidelberg, channel SNR=19dB

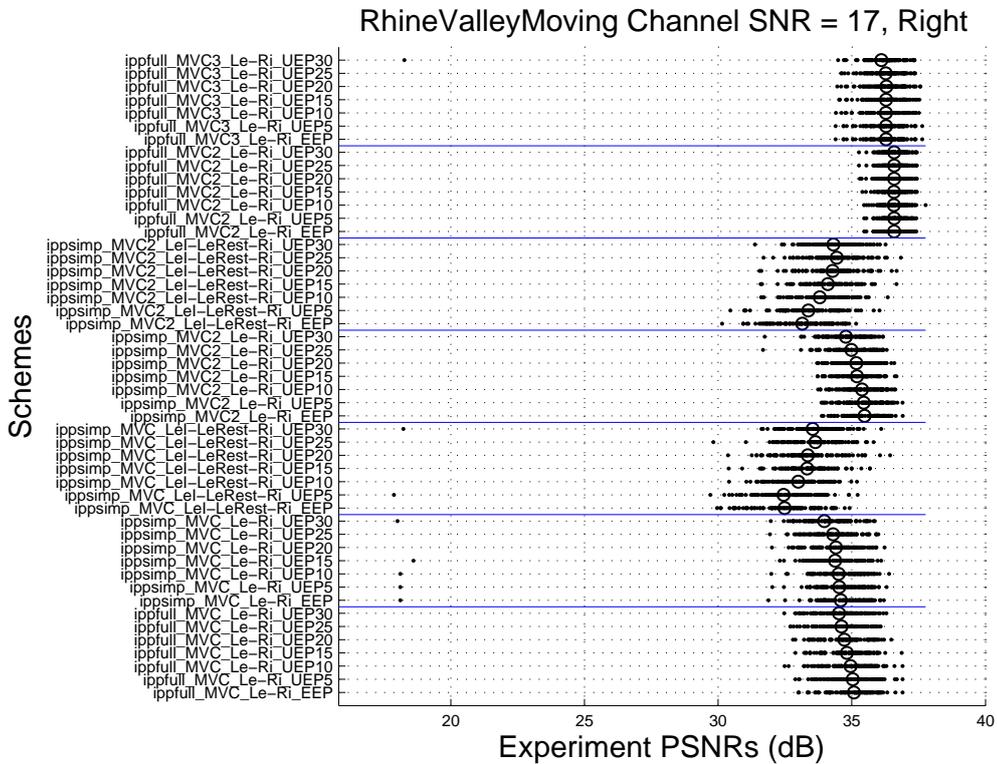


(b) PSNR distribution of the experiments of all the transmission cases, Right view, Heidelberg, channel SNR=19dB

Figure 5.15: Transmission results, Right View, Heidelberg, channel SNR=19dB



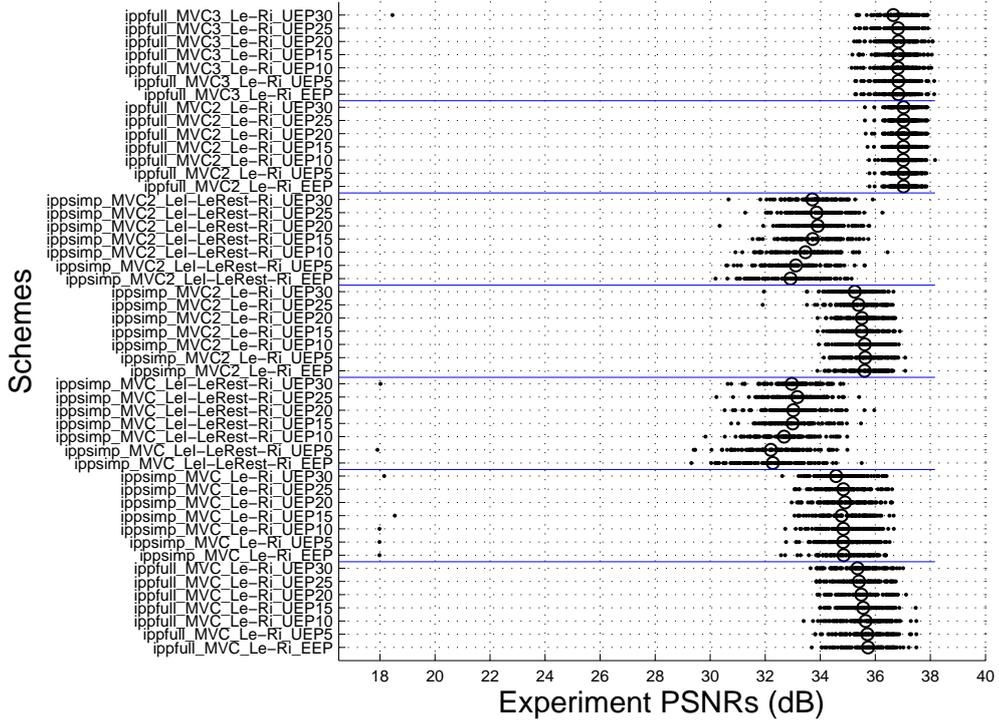
(a) PSNR distribution of the experiments of all the transmission cases, Left View, Rhine, channel SNR=17dB



(b) PSNR distribution of the experiments of all the transmission cases, Right view, Rhine, channel SNR=17dB

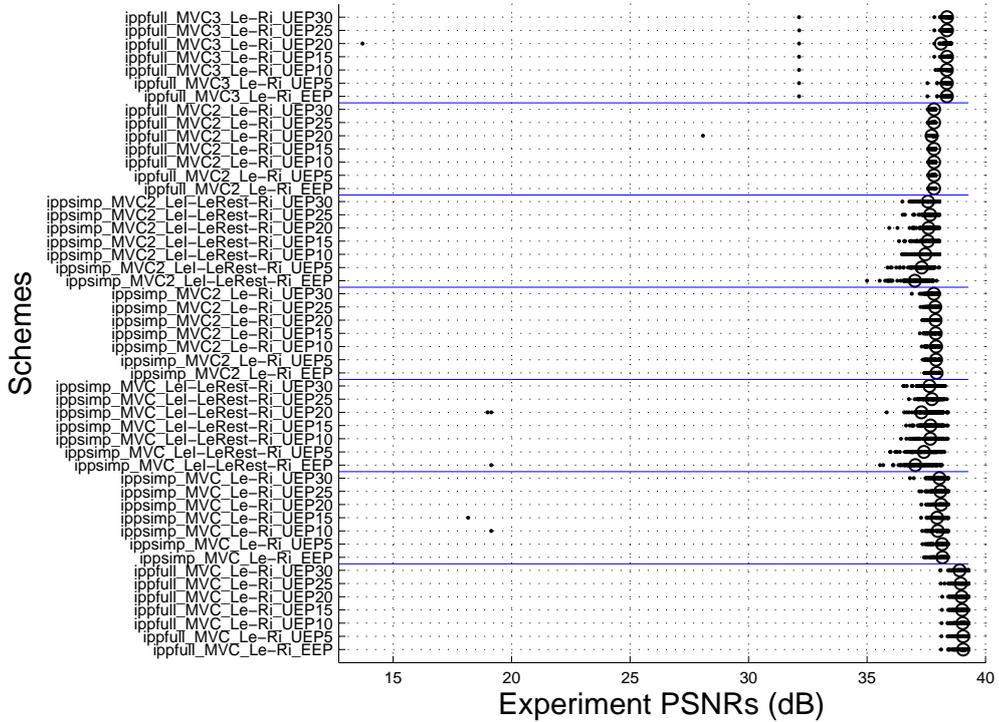
Figure 5.16: Transmission results, Rhine, channel SNR=17dB

RhineValleyMoving Channel SNR = 17, Joint



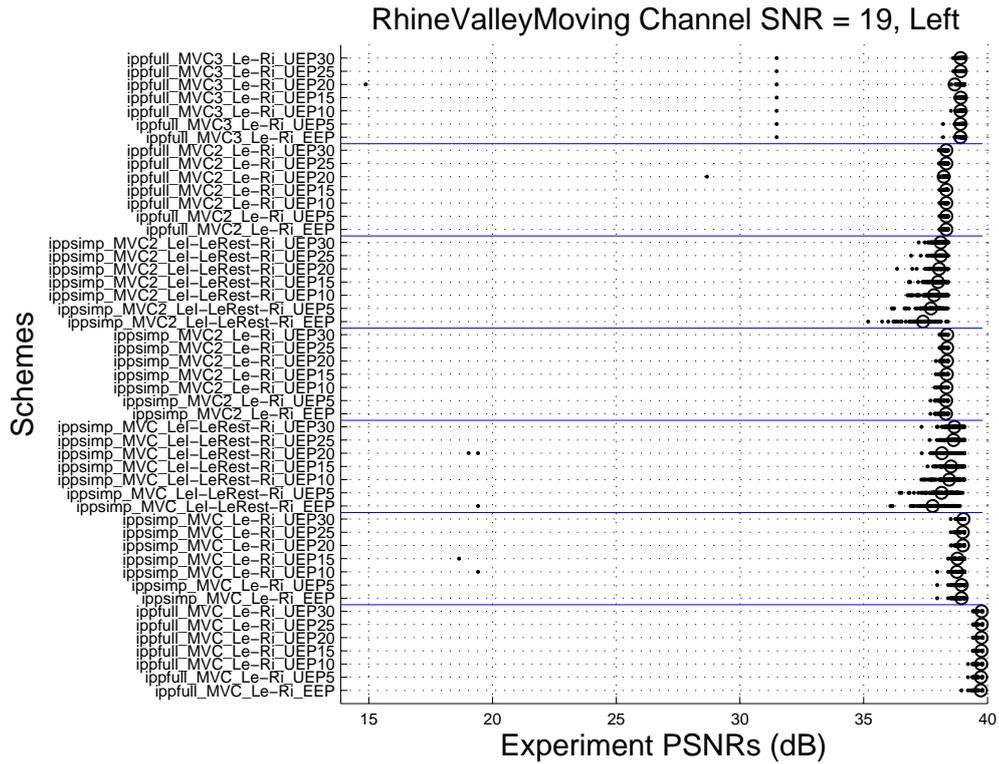
(a) Channel SNR=17dB

RhineValleyMoving Channel SNR = 19, Joint

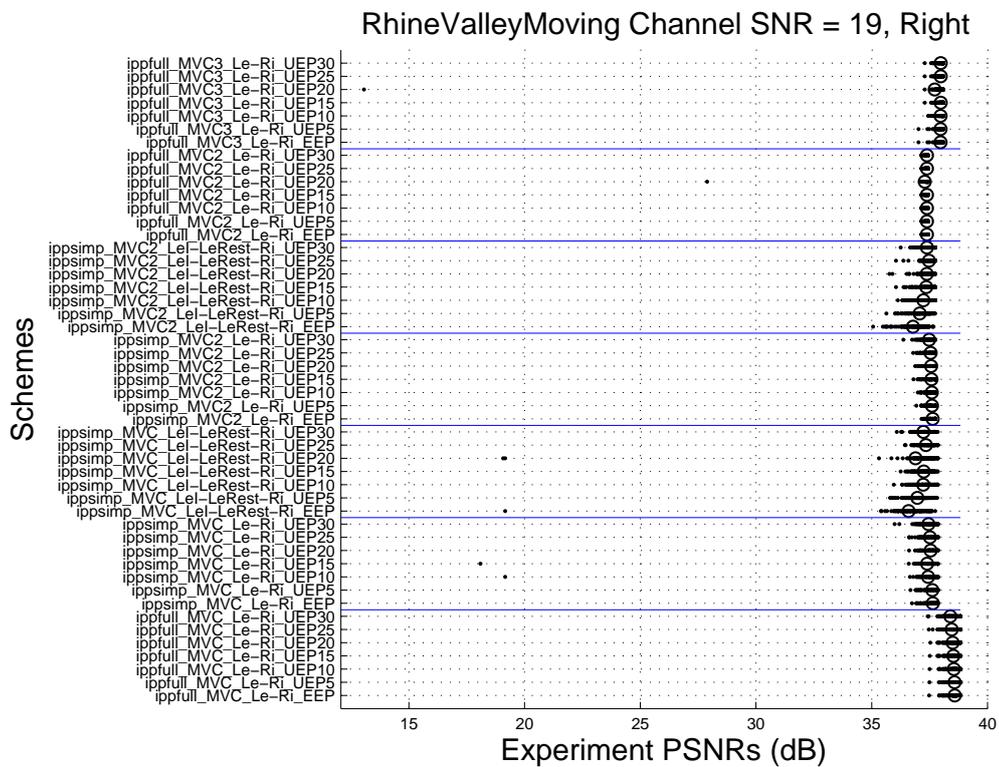


(b) Channel SNR=19dB

Figure 5.17: Joint PSNR distribution of the experiments of all the transmission cases, Rhine

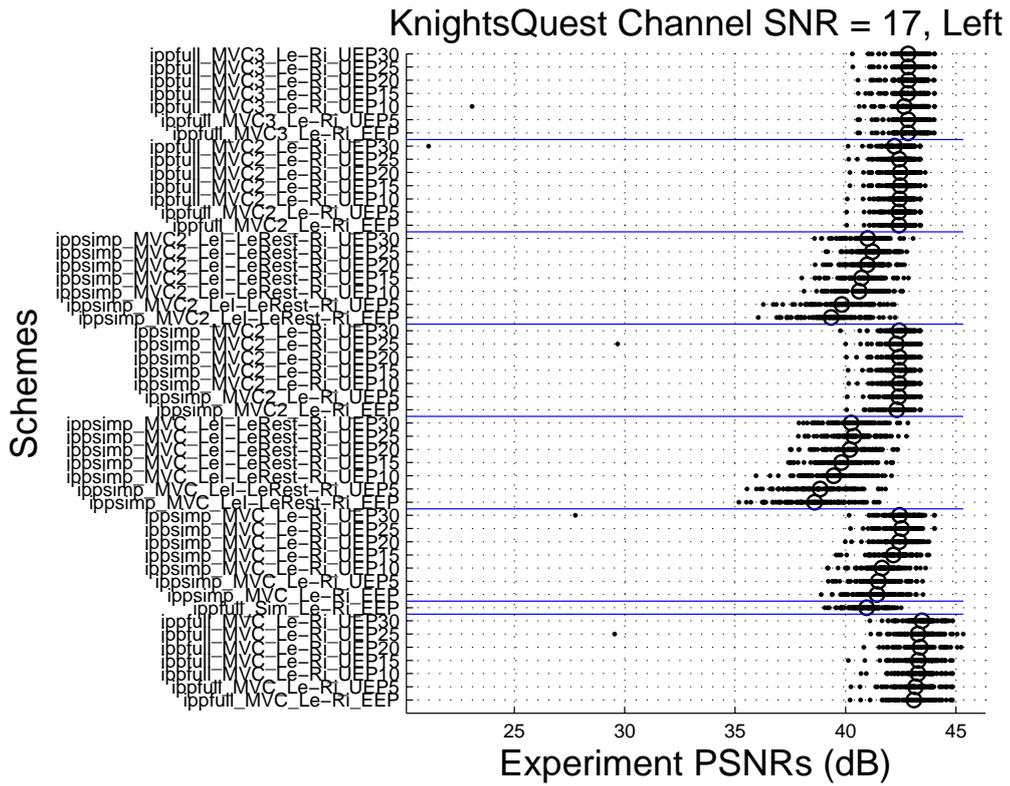


(a) PSNR distribution of the experiments of all the transmission cases, Left View, Rhine, channel SNR=19dB

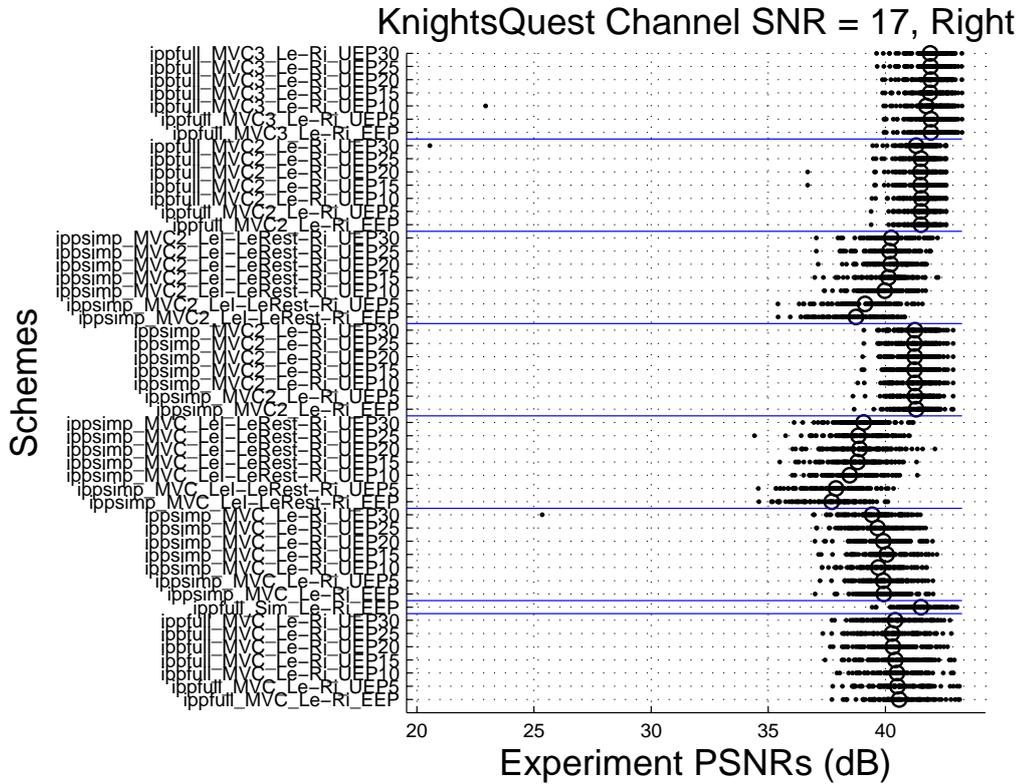


(b) PSNR distribution of the experiments of all the transmission cases, Right view, Rhine, channel SNR=19dB

Figure 5.18: Transmission results, Rhine, channel SNR=19dB



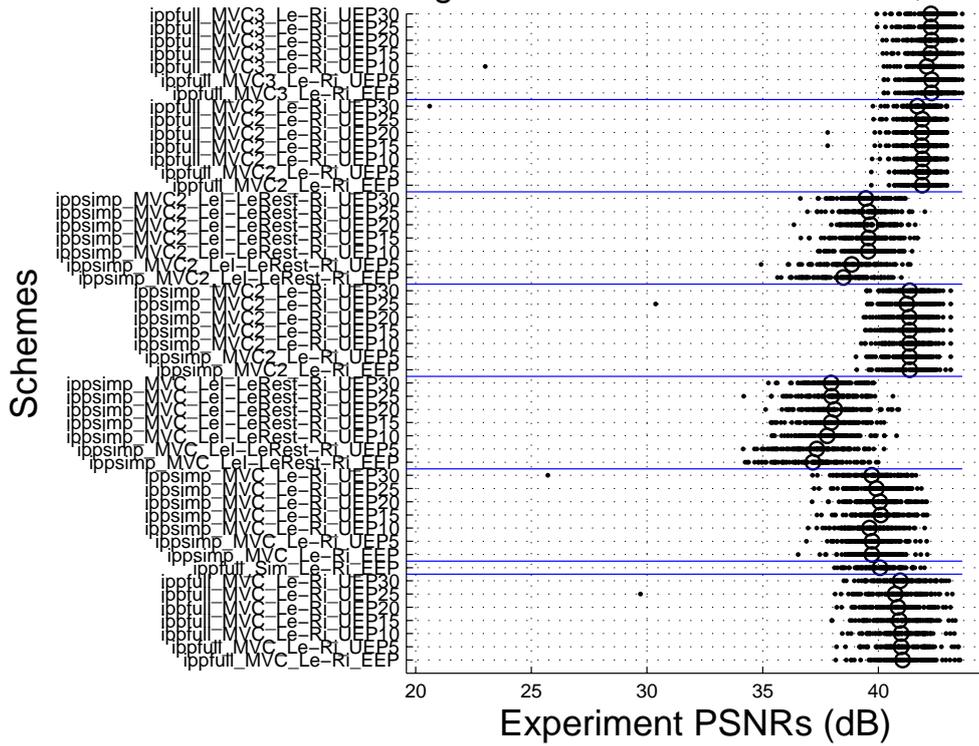
(a) PSNR distribution of the experiments of all the transmission cases, Left View, Knights, channel SNR=17dB



(b) PSNR distribution of the experiments of all the transmission cases, Right view, Knights, channel SNR=17dB

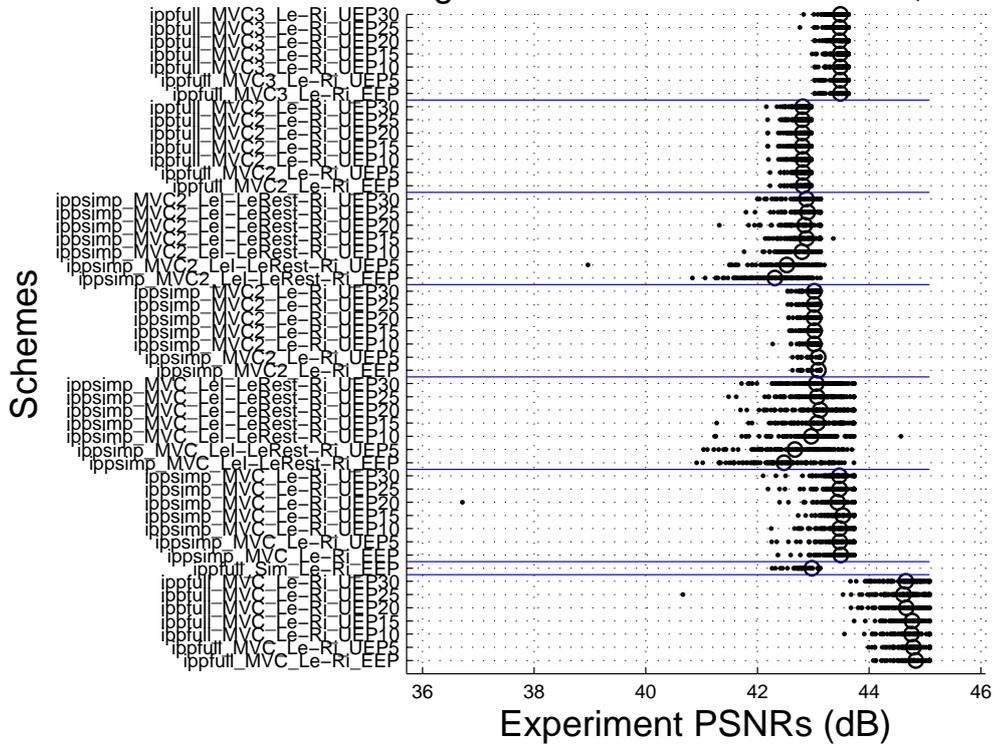
Figure 5.19: Transmission results, Knights, channel SNR=17dB

KnightsQuest Channel SNR = 17, Joint



(a) Channel SNR=17dB

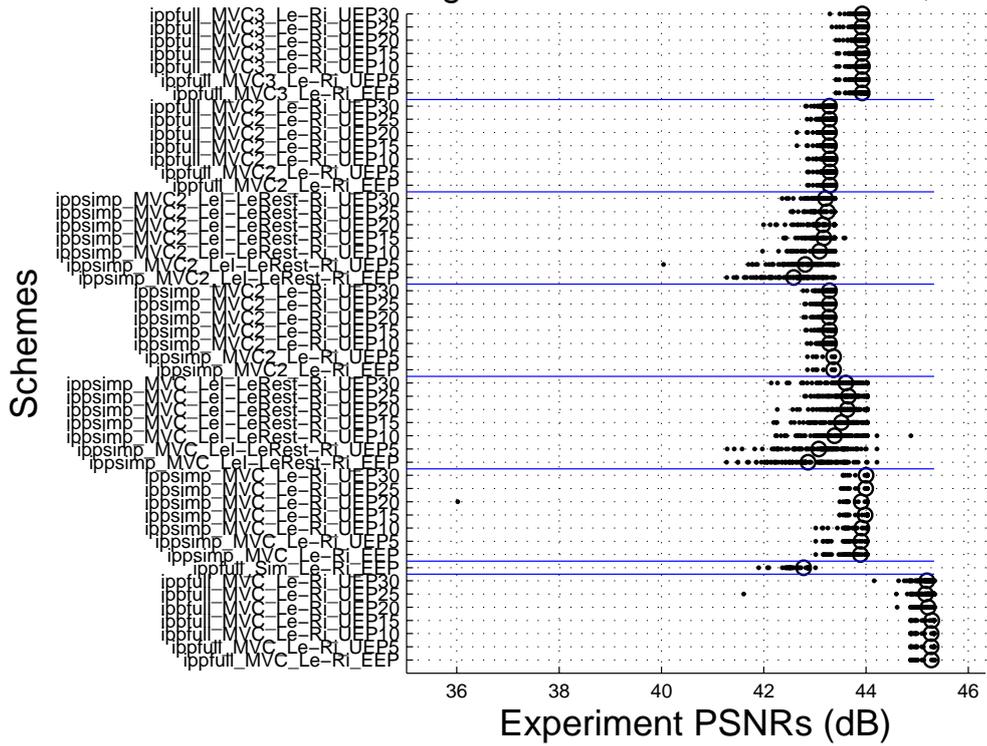
KnightsQuest Channel SNR = 19, Joint



(b) Channel SNR=19dB

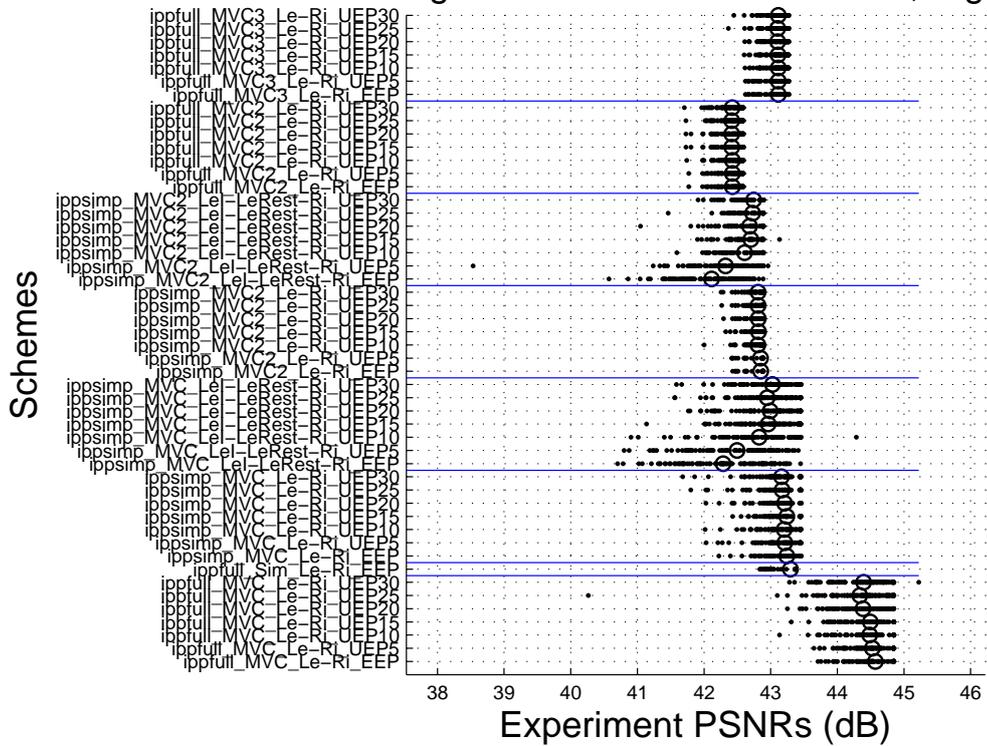
Figure 5.20: Joint PSNR distribution of the experiments of all the transmission cases, Knights

KnightsQuest Channel SNR = 19, Left



(a) PSNR distribution of the experiments of all the transmission cases, Left View, Knights, channel SNR=19dB

KnightsQuest Channel SNR = 19, Right



(b) PSNR distribution of the experiments of all the transmission cases, Right view, Knights, channel SNR=19dB

Figure 5.21: Transmission results, Knights, channel SNR=19dB

CHAPTER 6

CONCLUSION

In this thesis, a complete system that realizes the broadcasting of stereo video over DVB-H is established. The block diagram of the software system is provided and the implementations of the blocks are explained. The parameters and methods that may improve the error resilience of the transmission of stereo video are discussed. The effect of coding method, prediction structure, layering, protection method and rate allocation among video quality and protection are investigated. For this purpose, a wide set of experiments are designed. These experiments are repeated for three different contents and two different channel conditions and 100 channel realizations for each channel condition. Although, previous studies have shown that multi view coding is superior to simulcast coding in transmission, simulcast coding is included in the experiments as a base coding method to compare the other methods with it. The bitrate and corresponding PSNR value of simulcast coding are taken as reference to compare the other methods with Simulcast at the same bitrate and same quality separately. The comparisons are called the equal bitrate and equal quality comparisons. The two prediction methods of MVC, simplified and full reference predictions, are compared with each other and simulcast coding using these two comparison criteria. For both equal quality and equal bitrate comparisons, MVC IPP full prediction structure has the best transmission results in terms of both average PSNR and the PSNR distribution of the experiments. This means that although the dependence of right view on left view is increased, the higher compression rate dominates the performance as it allows for higher protection also. Two different layering scenarios are tested for MVC IPP simplified coding. The layering comparisons are conducted with IPP simplified coding due to presence of three levels of importance among the frames of left and right view.

Independent of the protection method or rate allocation among the video quality and protec-

tion, two layered transmissions are superior to three layered ones with a significant gap in terms of both packet loss and PSNR. In fact, this result is caused by the transmission performance of DVB-H. The burst size is also an important parameter for DVB-H where splitting the left view further into smaller two bursts decreases the recovery probability significantly. Therefore, due to the requirement of backward compatibility, stereo transmission over DVB-H is advised to be done with two bursts consisting of left and right view data.

The unequal protection methods provide performance improvements in joint PSNR over the equal protection for the three layered transmissions. It has been observed that, increasing the protection of the first burst up to 15 percent, improved the joint PSNR. On the other hand, unequal protection methods do not improve the joint PSNR for two layered transmissions. Besides, the three layered transmissions overall perform worse than two layered ones. Therefore, the conclusion is that unequal protection methods does not improve the performance with the implementations proposed in this thesis.

Finally, the rate allocation among the video quality and protection depends on the channel condition, i.e. channel SNR. The conclusions drawn above for the coding methods, prediction structures, layering and protection methods were valid for both channel SNRs. On the other hand, in case of small channel errors, the performance of the transmission is dominated by the video quality which is an expected result since encoding video quality directly determines the best method with no channel error. However, when there is significant error in the channel, such as 20 percent physical layer packet loss, the maximum protection rate possible should be preferred in order to recover the severe packet losses.

REFERENCES

- [1] “ETSI, Digital Video Broadcasting (DVB): DVB-H Implementation Guidelines,” 2009, TR 102 377 V1.3.1.
- [2] www.dvb.org, “DVB - Digital Video Broadcasting.”
- [3] <http://www.3dphone.org/>, “3DPhone Project.”
- [4] www.mobile3dtv.eu, “MOBILE3DTV - Mobile 3DTV Content Delivery Optimization over DVB-H System.”
- [5] S. Cho, N. Hur, J. Kim, K. Yun, and S. Lee, “Carriage of 3D audio-visual services by T-DMB,” *Electronics and Telecommunications Research Institute, Republic of Korea, in Proc ICME*, 2006.
- [6] G. Faria, J. Henriksson, E. Stare, and P. Talmola, “DVB-H: Digital Broadcast Services to Handheld Devices,” *Proceedings of the IEEE*, vol. 94, no. 1, pp. 194–209, 2006.
- [7] “ETSI, Digital Video Broadcasting (DVB): Transmission System for Handheld Terminals (DVB-H),” 2004, EN 302 304 V1.1.1.
- [8] “ISO/IEC 13818 (Parts 1 to 3): Information technology - Generic coding of moving pictures and associated audio information,”
- [9] “ETSI, Digital Video Broadcasting (DVB); Framing structure, channel coding and modulation for digital terrestrial television,” 2009, EN 300 744 V1.6.1.
- [10] A. Vetro, S. Yea, and A. Smolic, “Towards a 3d video format for auto-stereoscopic displays,” in *Proc. SPIE Conference on Applications of Digital Image Processing XXXI*, vol. 7073, September 2008.
- [11] A. Vetro, T. Wiegand, and G. Sullivan, “Overview of the stereo and multiview video coding extensions of the h.264/mpeg-4 avc standard,” *Proceedings of the IEEE*, vol. 99, no. 4, pp. 626 –642, april 2011.
- [12] C. Fehn, P. Kauff, M. O. D. Beeck, F. Ernst, W. IJsselsteijn, M. Pollefeys, L. V. Gool, E. Ofek, and I. Sexton, “An evolutionary and optimised approach on 3d-tv,” in *In Proceedings of International Broadcast Conference*, 2002, pp. 357–365.
- [13] N. Atzpadin, P. Kauff, and O. Schreer, “Stereo analysis by hybrid recursive matching for real-time immersive video conferencing,” *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 14, no. 3, pp. 321 – 334, march 2004.
- [14] C. Fehn, “A 3d-tv system based on video plus depth information,” in *Signals, Systems and Computers, 2003. Conference Record of the Thirty-Seventh Asilomar Conference on*, vol. 2, nov. 2003, pp. 1529 – 1533 Vol.2.

- [15] D. Minoli, *3D Television (3DTV) technology, systems and deployment: rolling out the infrastructure for next generation entertainment*. CRC Press, 2011.
- [16] http://en.wikipedia.org/wiki/3d_film, “Wikipedia Article on 3D film.”
- [17] <http://en.wikipedia.org/wiki/Autostereoscopy>, “Wikipedia Article on Autostereoscopy.”
- [18] T. Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra, “Overview of the h.264/avc video coding standard,” *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 13, no. 7, pp. 560–576, July 2003.
- [19] I. E. G. Richardson, *H.264 and MPEG-4 Video Compression - Video Coding for Next-Generation Multimedia*. Wiley, 2003.
- [20] P. Merkle, A. Smolic, K. Muller, and T. Wiegand, “Efficient prediction structures for multiview video coding,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 11, pp. 1461–1473, 2007.
- [21] A. Aksay and G. Akar, “Evaluation of stereo video coding schemes for mobile devices,” in *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, 2009*, May 2009, pp. 1–4.
- [22] “Joint Video Team of ITU-T VCEG and ISO/IEC MPEG. Reference Software for MVC (JMVC) v.5.0.5.”
- [23] RFC 3550, “Rtp: A transport protocol for real-time applications,” The Internet Engineering Task Force (IETF), 2003.
- [24] A. S. Tanenbaum, *Computer Networks*. Pearson, 2003.
- [25] RFC 3984, “Rtp Payload Format for H.264 Video,” The Internet Engineering Task Force (IETF), 2005.
- [26] J. Postel, “User datagram protocol,” The Internet Engineering Task Force (IETF), RFC 768, 1980.
- [27] RFC 791, “Internet Protocol,” The Internet Engineering Task Force (IETF), 1981.
- [28] M. Oksanen, A. Tikanmaki, A. Gotchev, and I. Defee, “Delivery of 3D Video over DVB-H: Building the Channel,” in *NEM-Summit’08*, 2008.
- [29] R. Calonaci, “Techniche a diversità compatibili con il livello fisico dello standard dvb-t/h,” Master’s thesis, Florence University of Technology, 2007.
- [30] M. O. Bici, A. Aksay, A. Tikanmaki, A. Gotchev, and G. Bozdagi Akar, “Stereo Video Broadcasting Simulation for DVB-H,” in *NEM-Summit’08*, 2008.
- [31] D. Bugdayci, M. O. Bici, A. Aksay, M. Demirtas, G. B. Akar, A. Tikanmäki, and A. Gotchev, “Stereo dvb-h broadcasting system with error resilient tools,” *Mobile3DTV*, Technical Report D3.4, December 2009.
- [32] A. A. Aksay, M. O. Bici, D. Bugdayci, A. Tikanmaki, A. Gotchev, and G. B. Akar, “A Study on the Effect of MPE-FEC for 3D Video Broadcasting over DVB-H,” in *MobiMedia’09*, 2009.

- [33] T. Stockhammer and M. Bystrom, "H.264/AVC Data Partitioning for Mobile Video Communication," in *Image Processing, 2004. ICIP '04. 2004 International Conference on*, vol. 1, oct. 2004, pp. 545 – 548 Vol. 1.

APPENDIX A

ENCODING

Tables A.1, A.2 and A.3 provide an example of the encoder configuration files used.

Table A.1: Sample configuration file for the encoder

General	
InputFile	inYuvs\TSM3DHeidelbergAlleysR_432x240
OutputFile	out264s\TSM3D slice250_MVC_30_31
SourceWidth	432
SourceHeight	240
ReconFile	tmp\Trec
MotionFile	tmp\Tmot
FrameRate	12.5
FramesToBeEncoded	752
Log2MaxFrameNum	10
Log2MaxPocLsb	10
Coding	
SymbolMode	0
FRExt	0
BasisQP	30
BasisQP2Quant	1
SliceArgument	1250

Table A.2: Sample configuration file for the encoder continued

Hierarchical B	
GOPSize	1
IntraPeriod	8
NumberReferenceFrames	2
InterPredPicsFirst	1
DeltaLayer0Quant	0
DeltaLayer1Quant	4
DeltaLayer2Quant	5
DeltaLayer3Quant	6
DeltaLayer4Quant	7
DeltaLayer5Quant	8
PicOrderCntType	0
Motion Search	
SearchMode	4
SearchFuncFullPel	3
SearchFuncSubPel	2
SearchRange	32
BiPredIter	4
IterSearchRange	8
Loop Filter	
LoopFilterDisable	0
LoopFilterAlphaC0Offset	0
LoopFilterBetaOffset	0
Weighted Prediction	
WeightedPrediction	0
WeightedBiprediction	0
Nesting SEI Message	
NestingSEI	0
Snapshot	0
Active View Info SEI Message	
ActiveViewSEI	0
View Scalability Information SEI Message	
ViewScalInfoSEI	0
Multiview Scene Information SEI Message	
Multiview Scene Info SEI	0
MaxDisparity	12

Table A.3: Sample configuration file for the encoder continued

Multiview Acquisition Information SEI Message	
MultiviewAcquisitionInfoSEI	0
AcquisitionInfoFile	Camera_ballroom.cfg
Parallel Decoding Information SEI Message	
PDISEIMessage	0
PDIInitialDelayAnc	0
PDIInitialDelayNonAnc	0
Multiview Coding Parameters	
NumViewsMinusOne	1
ViewOrder	0-1
View_ID	0
Fwd_NumAnchorRefs	0
Bwd_NumAnchorRefs	0
Fwd_NumNonAnchorRefs	0
Bwd_NumNonAnchorRefs	0
View_ID	1
Fwd_NumAnchorRefs	1
Bwd_NumAnchorRefs	0
Fwd_NumNonAnchorRefs	1
Bwd_NumNonAnchorRefs	0
Fwd_AnchorRefs	0
Fwd_NonAnchorRefs	0