UNIVERSITY OF SOUTHAMPTON
FACULTY OF PHYSICAL SCIENCE AND ENGINEERING
SCHOOL OF ELECTRONICS AND COMPUTER SCIENCE

Joint Source-Channel Decoding and Unequal Error Protection Aided Video Transmission

by

Chuan Zhu

A thesis submitted for the award of Doctor of Philosophy
at the University of Southampton

July 1, 2016

SUPERVISOR:
Prof. Lajos Hanzo
FREng, FIEEE, FIEE, Dsc, EIC IEEE Press
Chair in Telecommunications, Head of Group
Dr. Rong Zhang
PhD, MIEEE
Department of Electronics and Computer Science
University of Southampton
Southampton SO17 1BJ
United Kingdom

Dedicated to my family

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

Faculty of Physical Sciences and Engineering

School of Electronics and Computer Science

Doctor of Philosophy

Joint Source-Channel Decoding and Unequal Error Protection Aided Video Transmission

by Chuan Zhu

Shannon's source and channel-coding separation theorem has been the basic philosophy followed by most state-of-the-art wireless transceivers, where all source signals are assumed to have the same importance and are protected equally by the channel encoder. However, due to its assumption of using idealized Gaussian channels, potentially infinite encoding/decoding delay as well as complexity, Shannon's source and channel-coding theorem is not strictly applicable in practical wireless scenarios. This is because it is almost impossible to remove all the redundancy of the video source, therefore there exits a certain amount of residual correlation. Moreover, real-life applications are often required to transmit source signals associated with unequal importance, such as the base-layer and enhancement-layer of layered video codecs. Hence joint source-channel coding (JSCC) was proposed for wireless scenarios by jointly exploiting the residual redundancy of the video source and the intentionally imposed redundancy of unequal error protection (UEP) techniques.

A video codec suitable for lossless video compression and iterative source-channel decoding (ISCD) is proposed by further developing the Markov random field (MRF) soft source decoder for the transmission of video sequences, rather than still video pictures. More explicitly, we used low complexity frame-differencing (FD) for removing the temporal-domain inter-frame redundancy. The low-complexity FD technique does not rely on power-thirsty motion-vector-based motion-compensation and as its other benefit, it does not require the transmission of motion vectors. However, this results in more residual redundancy than the more complex motion-vector based motion-compensation. Furthermore, variable-length code (VrLC) is used for removing the residual spatial redundancy of the FD signal, while exploiting the correlations amongst the FD pixels within the current frame with the aid of our MRF model based soft-in-soft-out (SISO) module. By adopting the MRF for modelling of the video pixels, we can infer maximum *a posteriori* (MAP) image estimates from the 2-D spatial redundancy between the video pixels, while simpler models like the Markov chain can only explore the 1-D spatial redundancy. Although the estimation

of the MRF parameters is a challenge compared to the 1-D Markov models, we conceived novel extrinsic information transfer (EXIT)-chart-based estimation methods, which are shown to be effective. Moreover, a novel three-stage ISCD structure is proposed, which outperforms the two-stage architecture. Furthermore, we examined the convergence of the three-stage iterative decoding process using 3D EXIT charts. The performance of our system operating both with and without FD is compared to our benchmarker schemes.

In support of inter-layer forward error correction (IL-FEC) coded layered video transmission, we conceived an adaptive truncated HARQ (ATHARQ) scheme for minimizing the video distortion under the constraint of a given total number of transmission time slots. More specifically, we investigated the merits of IL-FEC schemes in the context of truncated HARQ (THARQ) transmission schemes, where the packet scheduling arrangements were carefully designed for exploiting the specific characteristics of each IL-FEC coded packet. Furthermore, we developed a method of on-line optimization for our IL-ATHARQ transmission scheme, in order to find the most appropriate FEC code rate distribution among the video layers that reduced the video distortion. Type-I HARQ relying on Convolutional Codes (CC) was used for simplicity, because our focus was on the design of the scheduling schemes. The performance of our IL-ATHARQ scheme as well as of the rate-optimized IL-ATHARQ scheme using a RSC codec were compared to the benchmarkers using different video sequences, followed by characterizing both the effects of the delay as well as of the channel quality prediction errors on the attainable system performance.

Finally, we conceived an UEP scheme for layered video transmission in the downlink of a visible light communication system, explicitly, we proposed a hierarchical colour-shift keying (HCSK) modulation scheme based on the standardized colour-shift keying (CSK), which is capable of conveying inter-dependent layers of video signals. More specifically, we proposed the Type I and Type II HCSK arrangements, which can be flexibly configured by according to the channel quality, video quality, etc., where the Type II constellation allows us to use a wider range of FERs for the higher layers upon varying the constellation-shaping parameter δ_l. Our simulation results show that Type II HCSK provides a high grade of flexibility in terms of both its system configuration and optimization. Furthermore, we provided a detailed design example for the employment of HCSK transmitting scalable video sources with the aid of a RSC code. An optimisation method was conceived in order to enhance the UEP and to improve the quality of the received video. The performance of our optimised M-HCSK-RSC video transmission system using different HCSK constellation sizes was compared to the relevant benchmarker schemes using different video sequences.

Acknowledgements

First and foremost, I would like to express my gratitude to Professor Lajos Hanzo for his valuable guidance and advice throughout my research. His enthusiasm, inspiration and perseverance not only cultivate in me to become a researcher, but also benefit me in my life. His willingness to give his time so generously has been very much appreciated. It would not be possible to finish this thesis without his friendly and generous help.

I would also like to thank the co-authors of my papers: Lajos Hanzo, Yongkai Huo, Rong Zhang, Robert G. Maunder, Mohammed El-Hajjar, Chen Dong, Junyi Jiang, Hua Sun, Bo Zhang for their willingness to help, discussions, knowledge.

I must also acknowledge all my colleagues in the Communications group for their help and friendship. Special thanks go to Chao Xu, Soon Xin Ng, Dandan Liang, Abdulah Jeza Aljohani, Kent Tsz Kan Cheung, Tong Bai, Yichuan Li, Xin Zuo, Tao Wang and all others not mentioned here.

Finally, I must give thanks to my family, particularly my parents and my girlfriend, for their unconditional love, support and care. Without you, I would not be where I am today.

Contents

List of Publications

Journal Paper

1. **C. Zhu, Y. Huo, J. Jiang, H. Sun, C. Dong, R. Zhang and L. Hanzo**, Hierarchical Colour-Shift-Keying Aided Layered Video Streaming for the Visible Light Downlink, IEEE Access, in press.
2. **C. Zhu, Y. Huo, B. Zhang, R. Zhang, M. El-Hajjar and L. Hanzo**, Adaptive Truncated HARQ Aided Layered Video Streaming Relying On Inter-Layer FEC Coding, IEEE Transactions on Vehicular Technology, vol. 65, no. 3, pp. 1506-1521, March 2016.
3. **Y. Huo, C. Zhu and L. Hanzo**, Spatio-temporal iterative source-channel decoding aided video transmission, IEEE Transactions on Vehicular Technology, vol.62, no.4, pp.1597,1609, May 2013.

Conference Paper

1. **C. Zhu, Y. Huo and L. Hanzo**, Iterative Joint Source-Channel Decoding Aided Transmission of Losslessly Compressed Video, 2013 IEEE Wireless Communications and Networking Conference (WCNC), Shanghai, 2013, pp. 4582-4587.

List of Symbols

Chapter 2. Markov Random Field Model Assisted Joint Source-Channel Decoding

N	The number of frames consisted in the uncompressed video sequence
n	The video frame index
N_{ip}	The number of consecutive inter-frame-coded frames in a specific group of pictures
$\boldsymbol{F}_n$	The n^{th} frame in the uncompressed video sequence
R, C	The vertical and horizontal pixel-size of a video frame
r, c	The vertical and horizontal pixel-size of a macroblock
L	The total number of sub-blocks in $\boldsymbol{F}_n$
$\boldsymbol{U}_l^n$	The one-dimensional vector scanned from the l th macroblock from the n th frame
$u_{l,m}^n$	The random variable that represents the m th source symbol / pixel value from $\boldsymbol{U}_l^n$
i_U	One realization of $u_{l,m}^n$, which belongs to the legitimate symbol / pixel set $\mathbf{I}_U = \{0, 1, \cdots, 2^K - 1\}$
K	The number of bits used to represent i_U
$\hat{\boldsymbol{F}}_n$	The prediction of $\boldsymbol{F}_n$
$\hat{\boldsymbol{U}}_l^n$	The vector extracted from the l th macroblock of the predicted frame $\hat{\boldsymbol{F}}_n$, where $\hat{\boldsymbol{U}}_l^n = \left(\hat{u}_{l,1}^n, ..., \hat{u}_{l,M}^n\right)$
$\tilde{\boldsymbol{U}}_l^{n-1}$	The vector generated from the previously reconstructed $(n-1)^{st}$ frame $\tilde{\boldsymbol{F}}_{n-1}$
$\boldsymbol{E}_l^n$	The FD vector
$e_{l,m}^n$	The m th element of $\boldsymbol{E}_l^n$
$\boldsymbol{W}_l^n$	The quantized version of $\boldsymbol{E}_l^n$
$\boldsymbol{V}_l^n$	The output bit vector of the VrLC encoder
J_l^n	The vector length of $\boldsymbol{V}_l^n$
$\boldsymbol{X}_g^n$	The vector of L_{π_2} bits from the VrLC encoder's output buffer
$\boldsymbol{Y}_g^n$	The encoded vector of interleaved $\boldsymbol{X}_g^n$, which has a length of L_{π_3}

$T_{MRF}\left(\cdot\right)$	The transfer function of $I_{E,MRF}$ vs $I_{A,MRF}$
T	The 'temperature' MRF parameter
δ	The exponent in the potential function of Eq. (2.7)
N_{MRF}	The number of inner iteration loops used by the MRF soft decoder
I_{inner}	The number of inner iteration loops between the FEC decoder and the VLC decoder
I_{outer}	The number of inner iteration loops between the MRF decoder and the VLC decoder
L_{c_av}	The average VrLC codeword length
R_s	The coding rate of the source encoder
C_R	The compression ratio of the source encoder, where $C_R = 1/R_s$
R_n	The 'natural' code rate (NCR) of the uncompressed video
R_c	The channel coding rate
$R_{\mathbf{total}}$	The total coding rate in the system depicted in Fig. 2.1

Chapter 3. Adaptive Truncated HARQ Aided Layered Video Streaming Relying On Inter-Layer FEC Coding

N_T	The limit of the total number of transmission time slots (TS)
N_L	The total number of video layers
L_n	The n th layer of the compressed video stream
$\lvert h_n \rvert$	The amplitude of the channel at TS n, $1 \leq n \leq N_T$
A_n	$A_n = \lvert h_n \rvert^2$
$\left\lvert \widetilde{h}_n \right\rvert$	The prediction of $\lvert h_n \rvert$ as described in Section 3.2.3
ϵ_i	The video distortion due to the corruption or absence of layer i, which is measured using the peak signal-to-noise ratio (PSNR), $0 \leq i < N_L$
SNR_n	The N_L-element vector, which represents the SNR values of the signals in the receiver buffers after the n-th transmission. In other words, $SNR_{n,i}$, which is the i-th element of SNR_n, represents the SNR of the signals in the receiver buffer as defined in Section 3.2.2, $1 \leq n \leq N_T$, $0 \leq i < N_L$
$SNR_n\left(\cdot\right)$	The N_L-element vector, which represents the predicted value of SNR_n, depending on both SNR_{n-1}, as well as on the channel conditions and on the scheduling decisions, etc.
ℓ_i	The length of the bitstream of layer i, $0 \leq i < N_L$
R	The overall coding rate of the system
$\boldsymbol{r}$	The vector including the FEC coding rates of all the layers, where r_i is the coding rate of layer i, $0 \leq i < N_L$

$\mathcal{D}$	The decision-set including all the possible choices concerning which particular layer to transmit, where $\mathcal{D} = \{d_1, d_2, \ldots, d_{N_L}\}$. The decision vector d_k has N_L binary elements, where $d_{k,i}$ is defined as $d_{k,i} = \begin{cases} 1 & k = i \\ 0 & k \neq i \end{cases}$, $0 \leq i < N_L$, $0 \leq k < N_L$, which means that the k-th layer is chosen for transmission
D_n	The actual decisions adopted for transmitting at TS n, where we have $D_n \in \mathcal{D}$
$p_{n,i}\left(d_k\right)$	The packet error ratio (PER) of layer L_i at TS n, using decision d_k, when layer L_{i-1} is correctly decoded
δ	The estimation error tolerance threshold defined in Eq. (3.16)
T_p	The lookup-table (LUT) used for the estimation of the PER

Chapter 4. Hierarchical Colour-Shift-Keying Aided Layered Video Streaming

M	Used to define M-ary CSK modulation
$\mathbf{b}$	The bit sequence to be transmitted
$\mathbf{b}_s$	The bit sequence representing a symbol
N_b	The number of bits in a symbol, which is defined by $N_b = \log_2 M$
$\mathbf{s}$	The CSK symbol, which can be represented by $\mathbf{s} = [s_i, s_j, s_k]$, $\mathbf{s} \in \mathcal{S}$
$\hat{\mathbf{s}}$	The estimate of $\mathbf{s}$
$\mathcal{S}$	$\mathcal{S} = \{\mathbf{s}_0, \mathbf{s}_1, \ldots, \mathbf{s}_{M-1}\}$ is the collection of all the M legitimate symbols in M-ary CSK
$\mathbf{s}_m$	The m th legitimate symbol in M-ary CSK , $\mathbf{s}_m \in \mathcal{S}$
$\mathbf{q}$	The 2-dimensional vector representing s, which can be expressed as $\mathbf{q} = [x, y]$
$\mathbf{v}_i$, $\mathbf{v}_j$, $\mathbf{v}_k$	The three vertices of the triangular area of the legitimate CSK symbols mapped into the CIE1931 coordinates of $\mathbf{v}_i = [x_i, y_i]$, $\mathbf{v}_j = [x_j, y_j]$ and $\mathbf{v}_k = [x_k, y_k]$
$\mathbf{c}$	The centroid of the triangular area of the legitimate CSK symbols
$\mathbf{H}$	The (3×3)-element matrix, where each element represents the channel gain between a specific light source and a PD
$\mathbf{n}$	The AWGN noise vector represented by $\mathbf{n} = [n_i, n_j, n_k]$
σ^2	The total power of $\mathbf{n}$
γ_o	The optical SNR defined as $\gamma_o = E_s/\sigma^2$
E_s	The average of the power of all possible M-ary CSK symbols, which is defined as $E_s = \mathbb{E}\left\{\lvert\lvert\mathbf{s}\rvert\rvert^2\right\}$
N_L	The total number of M-HCSK layers in each symbol defined as $N_L = N_b/2$
l	Index of the l-th M-HCSK layer
$\mathbf{b}_{sL_l}$	The 2 bits that correspond to the l-th M-HCSK layer, where $l \in \{0, 1, \ldots, N_L - 1\}$
$\mathbf{d}_p$	The basic vectors (BVs) defined in Eq. (4.18), where $p \in \{i, j, k\}$

h	The index of the M-HCSK sub-areas, $h \in \{0,1,2,3\}$
$\mathbf{v}_{p,h}$	The vertices of each of the 4 areas of layer $l = 1$, where $p \in \{i,j,k\}$, $h \in \{0,1,2,3\}$
$\mathbf{d}_{p,h}$	The BVs of each of the 4 areas of layer $l = 1$, where $p \in \{i,j,k\}$, $h \in \{0,1,2,3\}$
$\mathbf{c}_h$	The centroids of each of the 4 areas of layer $l = 1$, where $h \in \{0,1,2,3\}$
α	The scaling factor defined in Eq. (4.21)
δ_0	The parameter used for describing the scaling factor α, which is defined in Eq. (4.21)
δ_l	The value of δ_0 of Eq. (4.21), when generating the constellation of Layer l.
N_v	The number of layers consisted in the compressed video stream
Lv_n	The n th layer of the compressed video stream
ℓv_n	The length of the n-th layer Lv_n
Bf_l	The l th buffer used in Fig. 4.11b, $l \in \{0,\ldots,N_L-1\}$
ℓ_{Bf}	The bit size of the buffer Bf_l
N_T	The total number of M-HCSK frame types
p_n	The conditional FER associated with each video layer's FER
ϵ_n	The distortion associated with each video layer
$p_{t,l}$	The FER of the l-th layer of the t-th M-HCSK frame
$\delta_{t,l}$	The δ_l used for the l-th layer of the t-th M-HCSK frame, where $l \in \{0,1,\ldots,N_L-2\}$

Chapter 1

Introduction

1.1 Historical Overview of Video Compression

Over the recent years we have seen dramatic growth in the number of personal video capturing devices that are shipped with either laptop computers and gaming consoles, or in the form of smart phones. With that growth comes the trend of the so-called “We Media” [1], which ends up in increasingly content-rich videos that are streamed from diverse sources and shared as well as viewed everyday over the internet. In order to prepare raw/uncompressed video for the convenience of storage or transmission, video codecs [2] come to rescue by exploring the correlations and removing the redundancy, hence providing compressed video sequences requiring a reduced file storage and transmission bandwidth. Video codecs can be categorized by whether they are using lossy or lossless compression. Video codecs using lossy compression produce compressed video sequences that cannot be flawlessly reconstructed to be replicas of the original uncompressed video, because some of the psychosocially masked details are removed for the sake of achieving a high compression ratio. These video codecs include the H.120, H.261, H.263, H.264/AVC, MPEG-2 and MPEG-4 schemes, which will be detailed in Section 1.1.1. Video codecs using lossless compression, on the other hand, exactly reproduce the original video sequence by reconstructing it from the compressed video, as detailed in Section 1.1.2.

1.1.1 Brief Review of Lossy Video Compression

In Table 1.1, we listed the major video coding standards and below we will briefly detail them as follows:

H.120

H.120 [5], as the first digital video coding standard, was standardized by the International Telecommunication Union (ITU). H.120 adopted basic video compression techniques including pixel-by-pixel differential pulse code modulation (DPCM),

Lossy Video Codecs	Year	Institution(s)
H.120	1984-1988	International Telecommunication Union (ITU)
H.261	1988-1993	
Motion-JPEG	1992-1998+	Joint Photographic Experts Group [3] (JPEG)
MPEG-1	1993-1998+	Moving Picture Experts Group (MPEG)
MPEG-2/H.262	1994-1998+	International Organization for Standardization (ISO) and ITU
DV	1994-1995+	International Electrotechnical Commission (IEC)
H.263	1995-2005	ITU Telecommunication Standardization Sector [4] (ITU-T)
RealVideo	1997-2008+	RealNetworks
MPEG-4 SP/ASP	1998-2004+	MPEG
WMV9/VC-1 /SMPTE 421M	2005-2006+	The society of Motion Picture and Television Engineers (SMPTE) 421M
H.264/MPEG-4 part 10/AVC	2003-2012	ITU-T Video Coding Experts Group (VCEG) and the ISO/IEC JTC1 MPEG
Audio and Video Coding Standard	2003-2005	Initiated by the government of China
HEVC/MPEG-H Part 2/H.265	2012-2015+	ISO/IEC MPEG and ITU-T VCEG

Table 1.1: Video compression standards

scalar quantization and variable-length coding (VrLC). Its evolved version employed motion compensation relying on background prediction.

H.261

The ITU H.261 standard [6] is the successor of H.120 and it is the first practical digital video compression standard. The H.261 standard was designed for transmitting video streams at bit rates between 40 kbit/s and 2 Mbit/s over Integrated Services Digital Networks (ISDN). It only supports two types of video resolutions, namely the 352×288-pixel Common Intermediate Format [6] (CIF) and 176×144-pixel Quarter Common Intermediate Format [6] (QCIF). H. 261 introduced a basic processing unit referred to as a macroblock, instead of the pixel-by-pixel based processing of H.120. It also employed new techniques, such as the discrete cosine transform (DCT). The so-called hybrid-coding based framework of H. 261 combined motion-compensation (MC) with DCT-based MC-error residual (MCER) compression. This combination has also been adopted by the subsequent video compression standards.

Motion-JPEG

The Joint Photographic Experts Group [3] (JPEG) standardized the so-called Motion-JPEG scheme in 1992, which was used by video codecs in web browsers, digital cameras, webcams, video cameras and video editors. Motion-JPEG only uses intraframe compression without interframe prediction, therefore results in lower compression ratios, while imposing lower processing and memory requirements on the hardware devices.

MPEG-1

In 1992, influenced by the H.261 standard, the MPEG-1 standard [7] was proposed by the Moving Picture Experts Group (MPEG) to compress both video and audio at rates up to about 1.5 Mbit/s in order to be used for Video Compact Discs (VCD) and digital cable/satellite TV. The MPEG-1 standard supports the compression of either CIF-format or 352 × 240-pixel videos. MPEG-1 downsamples the images and uses picture rates of only 24–30 Hz in order to meet the low data rate supported by VCD. As a result, the output video quality did not exhibit much improvement over the conventional video cassette recorder.

MPEG-2/H.262

MPEG-2/H.262 was developed in 1993, in order to support applications in over-the-air digital television broadcasting and in the Digital Versatile Disc (DVD) Video standard at higher data rates. Typically, MPEG-2 outperforms MPEG-1 at the bit-rate of 3 Mbit/s and above. Furthermore, MPEG-2 provides support for interlaced video, which is the format used by analog broadcast TV systems. This was not supported by MPEG-1. MPEG-2 also supports higher video resolutions, such as (720 × 576), (720 × 480), or the resolution of 1920 × 1080, which is used by the so-called high-definition (HD) video applications.

DV

The Digital Video (DV), standardized by the International Electrotechnical Commission (IEC) in 1994 for supporting video camera recorders, is another video coding specification [8] that only uses intraframe coding. DV video uses a frame-by-frame based compression combined with DCT and operating at a fixed bitrate of 25Mbit/s. Compared to other video codecs using intraframe compression, DV outperforms Motion-JPEG and has similar performance as the above-mentioned MPEG-2/H.262 in its Intra-frame (I) coding mode.

H.263

The H.263 ITU-Tstandard (1995) was developed by the ITU Telecommunication Standardization Sector [4] (ITU-T) in order to support video conferencing and cell phones. Apart from the video formats supported by the older H.261 codec, namely QCIF and CIF, H.263 supports additional formats including SQCIF (128×96), 4CIF

(704 × 576), and 16CIF (1408 × 1152). Half pixel precision is used for motion compensation at a quadrupled complexity, whereas H.261 used a less complex, but more coarse motion-resolution and a loop filter. Some parts of the hierarchical structure of the datastream are now optional, hence the codec can be configured for operating at a lower datarate or in a more error-resilient mode. H.263 is targeted at low bitrates, and outperforms all prior standards at all bit rates. At very low bit rates, video quality is better compared to MPEG-2 / H.262. Additional annexes were added during the further development, resulting in the H.263v2/H.263+ standards ratified in 1998 and the H.263v3/H.263++ completed in 2000. Annexes I to T were added in the former for enhancing the encoding efficiency and robustness against data loss, and annexes U to X were added in the latter one.

RealVideo

First released as early as 1997, RealVideo is a popular video compression format even at the time of writing. RealNetworks, the company that developed RealVideo, made it proprietary. Since its first release, RealVideo has had 10 versions, each associated with a specific format. RealVideo actually adopted H.263 as the video codec before version 8, but for versions 9 and 10 it switched to RealNetworks' own proprietary video codec. RealMedia-formatted video is suitable for streaming over networks, and it is applicable to lip-synchronised live television.

MPEG-4 SP/ASP

Introduced in 1998, MPEG-4 [2] defined the compression of both audio and video signals. MPEG-4 consists of several parts, each covering a certain aspect of the overall specification. Additionally, each part of MPEG-4 consists multiple levels and profiles, which are used for describing the set of capabilities for a subset of applications. More specifically, Simple Profile (SP) is compatible with H.261 and its performance is comparable to H.263. On the other hand, the Advanced Simple Profile (ASP) provides a higher compression efficiency and it is suitable for DVD players. MPEG-4 SP/ASP includes numerous novel coding concepts such as interactive graphics, object and shape coding, wavelet-based still image coding, face modeling, scalable coding and 3D graphics. MPEG-4 improved both the coding efficiency as well as the error resilience of MPEG-2.

WMV9/VC-1/SMPTE 421M

The Windows Media Video 9 (WMV9) format originated from a proprietary video compression format owned by Microsoft. In 2006 it was made an open standard by the Society of Motion Picture and Television Engineers (SMPTE), which become known as SMPTE 421M [9] or VC-1. WMV9, together with H.262/MPEG-2 Part 2 and H.264/MPEG-4 AVC, was adopted for HD DVD and for Blu-ray video discs. WMV9 uses 4x4-pixel blocks instead of the more conventional 8x8-pixel blocks, and its compression efficiency is slightly inferior to that of H.264 but it is less computational demanding.

H.264/MPEG-4 part 10/AVC

The H.264 standard [10] was developed by the Joint Video Team (JVT), namely the partnership of the ITU-T Video Coding Experts Group (VCEG) as well as of the ISO/IEC JTC1 MPEG. It was named as H.264 by ITU-T, while it was described in MPEG-4 part 10 as the Advanced Video Coding (AVC) profile. Since its first release in 2003, H.264 has became one of the most popular video compression standards at the time of writing. H.264 has been adopted by Blu-ray Discs as one of the compression methods, but it has also been used for Internet streaming, HDTV broadcasts over terrestrial (Advanced Television Systems Committee standards, ISDB-T, DVB-T or DVB-T2), cable (DVB-C), and satellite (DVB-S and DVB-S2).

H.264 is capable of high-efficiency compression at half the bit-rate at the same video quality compared to its predecessors, as a benefit of new techniques, such as improved multi-picture-based prediction, context-adaptive binary arithmetic coding (CABAC) and context-adaptive variable-lengths coding (CAVrLC), etc. H. 264 is also suitable for transmission over various network, due to its employment of the Network Abstraction Layer (NAL) definition, data partitioning (DP), flexible macroblock ordering (FMO) and as a benefit of employing redundant slices (RS).

Several extensions have been developed for H.264 in order to increase its functionality, including the Scalable Video Coding [11] (SVC) designed for creating different-resolution sub-bitstreams from the bitstream of the H.264 standard, and the Multiview Video Coding scheme (MVC) [12] designed for encoding several views together.

Audio and Video Coding Standard

Standardized as the Chinese compression standard in 2005, the Audio Video Standard (AVS) [13] has been a major competitor to the popular video standards, such as MPEG-4 and H.264. The AVS scheme achieves a similar performance to the H.264 standard, at a lower computational complexity.

HEVC/MPEG-H Part 2/H.265

H.265 [14,15], also known as ISO/IEC 23008-2 MPEG-H Part 2 and High Efficiency Video Coding (HEVC), is being developed by the Joint Collaborative Team on Video Coding (JCT-VC), which is formed by the ISO/IEC MPEG and ITU-T VCEG teams.

H.265 constitutes a substantial leap over H.264 in terms of its compression efficiency. At the same video quality, H.265 is capable of reducing the bitrate by a factor of two compared to H.264/MPEG-4 AVC. More specifically, H.265 is capable of reducing the bitrate by 64% compared to H.264 using Ultra High Definition (UHD) videos at a resolution of 8192 × 4320 [16]. H.265 applied numerous technical improvements to H.264. For instance, the concept of coding tree units (CTUs) includes block structures of up to 64x64 samples, which are used in order to provide the option of variable-sized partitioning of the video picture [15]. H.265 also supports more computationally-intensive solutions for the sake of achieving improvements. More specifically, intra-frame prediction in HEVC specifies 33 directional modes, which is substantially more flexible than the previous 8 directional modes. Longer interpolation filters having more taps are used for motion compensation for the sake of achieving an improved precision. Finally, HEVC increased the number of bits used for representing the motion vectors (MVs) [15].

Several important extensions of H.265 have also been developed. More specifically, the SVC and MVC extensions were provided in Version 2 of H.265 standardized in October 2014. 3D-HEVC extensions conceived for 3D video were described in Version 3 released in April 2015.

1.1.2 Introduction of Lossless Video Compression

The video coding standards introduced in Section 1.1.1 can be configured for achieving perceptually ’lossless’ compression, implying that the difference between the decompressed video and the original video sequence can hardly be noticed by the human eye. However, the lossless video compression codecs we are reviewing in this section refer to those achieving exact lossless compression, which implies that the recovered video is exactly identical for every pixel compared to the original video. The latter lossless video codecs have not received such wide adoption as the video codecs of Section 1.1.1, mainly because their compression ratio is as low as 2 to 3, while perceptually lossless compression is capable of achieving 10 to 1000. Lossless video codecs are mainly used for specific applications [17] requiring a high fidelity, namely medical diagnosis, fine art archiving and military purposes, especially in those applications that require re-encoding or editing, where lossless encoding is preferable in order to prevent coding error accumulation.

The basic design concept of the lossless video coding schemes is shown in Figure 1.1. Instead of the lossy DCT and the quantization of the motion prediction error after prediction, lossless video coding opts for using reversible transforms for compressing the prediction error without any loss of information. Many lossless video codecs have been developed [18]. In Table 1.2, we listed a few representatives of this family and describe them as follows:

Huffyuv

The first version of the Huffyuv [19] lossless video codec was developed by Ben Rudiak-Gould in 2000 for compressing both the YUY2 and RGB colour spaces. As its name suggests, Huffyuv uses a Huffman encoder for compressing the prediction

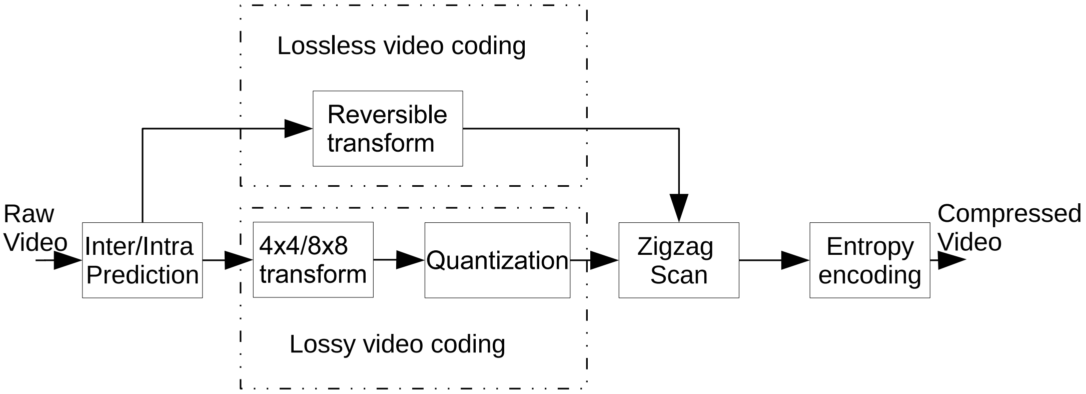


Figure 1.1: Simplified conceptual schematic diagram of lossy and lossless video coding arrangements.

Lossy Video Codecs	Year	Author(s) / Institution(s)
Huffyuv [19]	2000-2002	Ben Rudiak-Gould
FFV1 [20]	2003-2013	Michael Niedermayer, FFmpeg team
Arithyuv [21]	2005-2005	Julien Muchembled
MSU Lossless Video Codec [22]	2004-2005	Graphics & Media Lab Video Group, Moscow State University
YULS [23]	2004-2007	Maxim Smirnov / YUVsoft
Lagarith [24]	2004-2011	Ben Greenwood
H.264 lossless [25]	2004-2012	ITU-T VCEG and the ISO/IEC JTC1 MPEG

Table 1.2: Video compression standards

error. Huffyuv offers a limited compression ratio at a relatively low complexity [26]. It was then developed further to another lossless video codec referred to as Lagarith, which improves the attainable compression ratio.

FFV1

Developed and maintained as the free, open-source project FFmpeg, FFV1 (FF video codec 1) [20] has been the most balanced and most flexible lossless video codec [26]. As an intra-frame video codec, FFV1 relies on the so-called median prediction, which predicts the value of a pixel by its neighbouring pixels. For entropy coding, variable length coding or arithmetic coding is used. FFV1 has been integrated into many commercial software applications [20], as a benefit of its low complexity and relatively good compression performance.

Arithyuv

Arithyuv [21] is another lossless video codec, which supports the compression of the YUY2 colorspace. Interlaced video sources are supported. Arithyuv also adopted median prediction and used an assembly-optimized implementation of arithmetic coding. Compared to its predecessor Huffyuv, Arithyuv offers a higher compression ratio at an increased complexity [26].

Moscow State University (MSU) Lossless Video Codec

The MSU Lossless Video Codec [22] was developed by the Graphics & Media Lab Video Group of Moscow State University (MSU). This team also contributed to the evaluation and comparison of the existing lossless video codecs . This codec supports multiple input formats, such as YUY2, YUYV, YV12, RGB24 (RGB) and RGB32. Interframe coding was introduced in order to provide a high compression ratio. This codec ranks the second in the MSU lossless video codec comparison report issued in 2007 [22].

YULS

YULS [23] is a lossless video codec developed by the YUVsoft Corp., which was formed from the MSU Video Group. It is a high-complexity lossless video codec, which achieves the best compression ratio among all known lossless video codecs [26]. Color formats including RGB24, YV12 and YUY2 are supported by YULS, and a maximum video resolution of 1024 by 768 can be supported. However, owing to its high complexity, it is not suitable for real-time video capturing.

Lagarith

Lagarith [24] is an off-shoot from the development of Huffyuv and improved Huffyuv by increasing the compression ratio. RGB24, RGB32, RGBA, YUY2, and YV12 color formats are supported by Lagarith. Apart from using median prediction, Lagarith introduced modified Run Length Encoding plus Arithmetic compression to Huffyuv, in order to further reduce the residual redundancy in the resultant bit stream. Since only intra-frame coding is used, video editing including cutting, joining and seeking is straight-forward. Given its high compression ratio when using RGB24 format video, Lagarith is considered to be the most suitable for video capture and video editing [26].

H.264 lossless

In 2004, H.264 included the lossless video coding technique in the H.264 fidelity range extensions (FRExt) [25], which simply skips the DCT and its quantization in order to losslessly compress the intra/inter prediction errors. This regime was tested using the x264 codec [27] implementation of H.264 lossless coding by comparing it to other lossless codecs, demonstrating that the lossless H.264 codec has a mediocre compression ratio. However, x264 significantly improved its speed by utilizing the so-called multi-threading techniques [26].

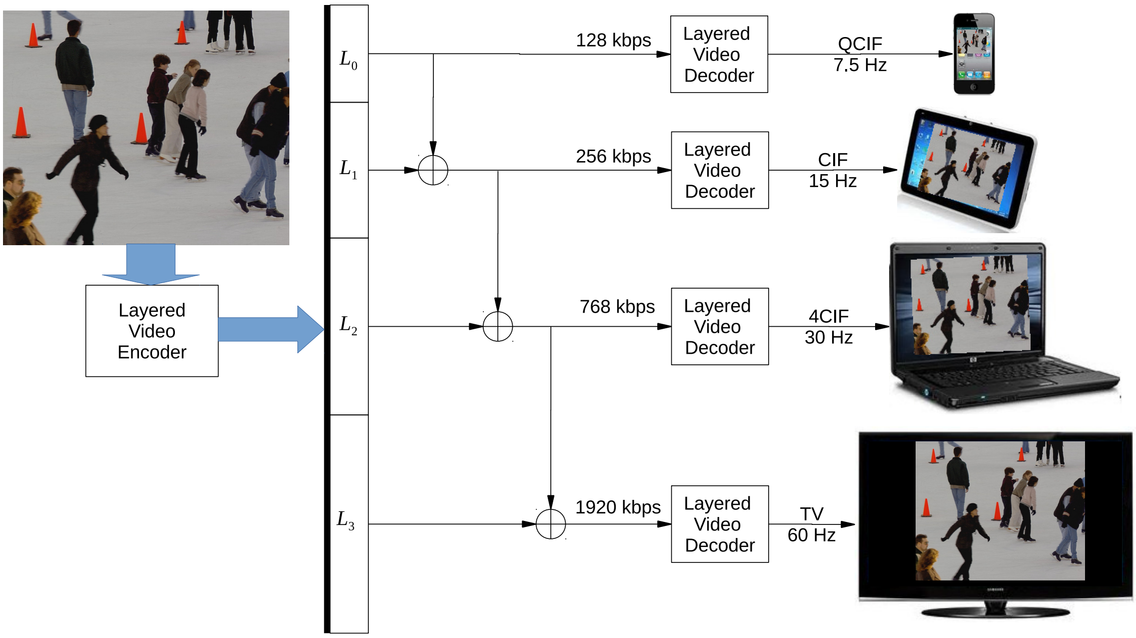


Figure 1.2: Streaming and decoding structure of a layered video [28]

1.1.3 Layered Video Coding Techniques

Layered video coding [28, 29, 30] is a popular video compression scheme designed for heterogeneous networking and multiple terminal problems. Explicitly, layered video encoding creates multiple video streams containing multiple layers of different importance, where the base-layer (BL) represents the lowest-resolution video having the lowest bitrate, while the enhancement-layers gradually increase the bitrate and hence the video quality. Therefore, layered video coding is capable of providing various terminals with different video quality/resolution. Furthermore, it is also capable of supporting progressive reception of video streams according to the specific network and channel conditions of the different users.

The schematic of a layered video arrangement is shown in Figure 1.2, where we can see that the original video sequence is encoded by a layered video encoder into multiple layers L_0, L_1,... L_n $(n = 3)$. Explicitly, L_0 is the base layer, which is the most important one and the rest of the layers L_1,... L_n are the enhancement layers, which can gradually improve the video quality attained. The dependency of the scalable video layers on each other can be expressed as $L_0 \Leftarrow L_1 \Leftarrow \ldots \Leftarrow L_n$, where each item on the right of the $\Leftarrow$ symbol depends on all the items to the left of it. In order to successfully decode the video stream using the i-th layer, the decoder has to invoke the information from all the previous layers, namely L_0, L_1,... L_{i-1}. As shown in Figure 1.2, the four different terminals on the right have different video quality requirements, and they typically rely on networks associated with different transmission bandwidths. The BL has for example a bitrate of 128 kbits per second (kbps) in order to provide the mobile terminal with QCIF-resolution video at 7.5 frames-per-second (FPS). When combined with the EL bitstreams, the resultant bitstream has a higher data-rate and naturally can be decoded into video sequences of higher resolution, of higher perceptual-quality and higher frame-rates. For example, the tablet computer might require the CIF-format video sequence at 15 FPS, the laptop may need the 4CIF format video sequence at 30 FPS, and the HD-TV requires a higher resolution at 60 FPS. Accordingly, we have the bitstream combinations of $\{L_0, L_1\}$, $\{L_0, L_1, L_2\}$ and $\{L_0, L_1, L_2, L_3\}$ in order to support the decoder of each terminal.

Layered video coding has been standardized by several major video coding standards [31, 11, 10, 32, 33], namely the MPEG-2, H.264/AVC and H.265/HEVC. Here we give a brief introduction to three forms of standardized layered video, namely to the data partitioning mode of H.264/AVC, to the scalable extensions of H.264/AVC and H.265/HEVC, and to the multiview extension of MPEG-2, H.264/AVC and H.265/HEVC.

The video data partitioning (DP) scheme [10] of H.264/AVC generates three partitions of different error-sensitivity per video slice, namely type A, type B and type C partition [34]. The resultant three partitions represents contents of different semantic importances. The type A partition contains important bitstreams such as header information, including macroblock (MB) types, quantization parameters and motion vectors. The type B and type C partitions are also referred to as the intra-frame-partition and inter-frame-partition, respectively. The type B partition contains intra-frame-coded information, including the coded block patterns (CPB) and intra-frame-coded coefficients, while the type C partition consists of the inter-frame-CBPs and the inter-frame coded coefficients. The "BL" of the data partitioning layers refer to the type A partition, because it exhibits the highest importance owing to the dependence of both type B and type C partitions on it. On the other hand, the type B and type C partition rely on the type A partition for their successful decoding, but they do not depend on each other. Hence the type B and type C streams may be treated as "EL"s, since they depend on the BL and they refine the quality of the decoded result, although they do not strictly follow the hierarchy shown in Figure 1.2.

The SVC scheme [11] was specified in the Annex G extension of the H.264/AVC video compression standard [10]. SVC is capable of generating an encoded stream containing multiple interdependent layers, where the layers of lower importance can be discarded in order to tailor the bit-rate according to the specific user-requirements and/or channel quality, as shown in Figure 1.2. Upon decoding all the layers, similar compression ratio and quality can be achieved as that of the H.264/AVC codec. SVC has also been adopted in the scalability extension of the H.265/HEVC standard, which is referred to as SHVC [35, 36]. By using different subsets or the entire SVC-generated layers for decoding, the video can be recovered at various perceptional qualities. There are several techniques of differentiating the perceptional qualities in order to generate the appropriate layers for progressive refinement, using different scalable video techniques summarized as follows:

1. **Temporal Scalability**: By using temporal scalability, H.264 SVC drops those video frames that are less dependent on the hierarchical B frame structure in order to decrease the frame frequency [11]. Dropping every other frame halves the frame-rate. For example, the frame rate of 60 FPS of a certain original video can be scaled down to 30 FPS, 15 FPS or 7.5 FPS.
2. **Spatial Scalability**: Spatial scalability [11] provides video frame sequences having different frame resolutions. The BL provides the decodable video with the lowest frame resolution, whilst using more ELs for decoding, higher resolutions gradually reaching the original video resolution can be obtained. Apart from using motion compensation and intra-prediction for each layer, H.264 SVC invoked inter-layer prediction using the lower layers for predicting the higher EL layers in order to increase the coding efficiency for the sake of achieving improved rate-distortion (RD) properties.
3. **Quality-Based Scalability**: While producing video layers associated with the same frame-rate and resolution, quality-based scalability [11] can be used to provide videos with different frame qualities. There are two typical types of quality scalability techniques that are used in H.264 SVC, namely coarse grain scalability (CGS) and medium grain scalability (MGS):
   (a) Coarse Grain Scalability: Coarse Grain Scalability is also considered to be the main quality scalability technique [11]. By using Coarse Grain Scalability, the motion-compensated residual signal is re-quantized with successively smaller quantization step sizes in order to achieve gradually increasing quality/fidelity. Up to eight different-quality layers can be provided by using Coarse Grain Scalability and the quality of the video can be successively increased by involving more ELs in the decoding process.
   (b) Medium Grain Scalability [11]: Compared to Coarse Grain Scalability, Medium Grain Scalability provides a finer granularity level of quality based scalability. Each Coarse Grain Scalability enhancement layer can be partitioned into several Medium Grain Scalability layers so that each of the Medium Grain Scalability layers can then be dropped to reduce the bitrate and the quality. Medium Grain Scalability processes the video frame by dividing and grouping the DCT coefficients according to their importance after the transform coding of macroblocks. As a result, up to 16 Medium Grain Scalability layers can be generated from a given enhancement layer.
4. **Combined Scalability** [11, 37]: The aforementioned three scalabilities can also be jointly supported by H.264 SVC. Different combinations of the frame frequencies, spatial resolutions and of video frame qualities (achieved with the

aid of Coarse Grain Scalability or Medium Grain Scalability) can be extracted from a specific encoded bit stream. To identify the network abstraction layer units (NALUs) in H.264 [10], the knowledge of their dependency hierarchy and of the types of scalability they are using, as well as of the header part of the NALUs is needed. Specifically, the header consists of the specific parameters dependency_id, temporal_id and quality_id, which correspond to Coarse Grain Scalability, Temporal Scalability and Medium Grain Scalability, respectively. NALUs associated with larger id numbers are of the least importance in the dependency hierarchy and hence may be removed for reducing the bit-rate at the cost of reduced video quality.

Finally, multiview video coding is another type of layered video, compression technique which carries out efficient compression of sequences captured simultaneously from multiple cameras into a single video stream. Multi-view video coding exploits the inter-view redundancy and employs inter-view prediction. The different encoded views can also be considered to be different layers. Since the first encoded view is used as the reference for predicting other views, it is relatively more important and should hence be treated as the BL, while the remaining views may be treated as the ELs.

Multiview video coding has been ratified by several video coding standards. To start with, MPEG provided the so-called multiview profile [31] for MPEG-2. Further more, JVT standardized multi-view video coding as an amendment of the H.264/AVC standard [10]. Finally, multi-view extensions have been included in the second version of HEVC in October 2014.

1.1.4 Preliminaries for the Evaluation of Video Transmission

This section introduces some preliminaries associated with our video simulations, including the video quality metric, the format of the H.264 bitstream and the evaluation framework of scalable video coding.

1.1.4.1 Video Quality Metric

Video quality metrics can be categorised into two main types, namely the objective video metric and subjective quality assessment. The former one makes uses of mathematical models to characterize the quality of video sequences while the later relies on human observers to rate the quality of a video.

Here we commence by introducing the most common video quality metric, namely the peak signal-to-noise ratio (PSNR). Let us commence by stipulating the following assumptions

- $(m \times n)$: the resolution of the luminance frame;

- $I(i,j)$: the original/reference video pixel value at position (i,j) of a specific video frame;
- $K(i,j)$: the reconstructed pixel value at position (i,j) of a specific video frame;

Then the PSNR is calculated as follows

$$\begin{aligned} MSE &= \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\left[I(i,j)-K(i,j)\right]^2 \\ PSNR_{dB} &= 20\cdot\log_{10}\frac{\max(I)}{\sqrt{MSE}}. \end{aligned} \tag{1.1}$$

However, small mean square error (MSE) values may result in infinite PSNR values. Hence the following modified PSNR calculation is employed

$$\begin{aligned} MSE &= \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\left[I(i,j)-K(i,j)\right]^2 \\ PSNR_{dB} &= 20\cdot\log_{10}\frac{\max(I)}{\sqrt{MSE'}} \\ MSE' &= \max(1, MSE), \end{aligned} \tag{1.2}$$

which results in a maximum PSNR value of 48.13 dB.

1.2 Overview of Iterative Source-Channel Decoding

In this section we take a brief look at the research history of Iterative Source-Channel Decoding (ISCD), with the major contributions listed in Table 1.3.

Shannon's source and channel-coding separation theorem [38] has been the basic philosophy followed by most state-of-the-art video codecs [2, 10] used in wireless transceivers. However, due to its assumption of using idealized Gaussian channels, as well as potentially infinite encoding/decoding delay and complexity, Shannon's source and channel-coding theorem is not strictly applicable in practical applications. As a matter of fact, it is almost impossible to remove all the redundancy of the video source, therefore there exits a certain amount of residual correlation. In the light of this, joint source-channel coding (JSCC) [54] was proposed for wireless scenarios, where the residual redundancy of the video source and the intentional redundancy of the channel codecs was jointly exploited for increasing the error resilience of the system.

An iterative decoding structure has been proposed to tackle the JSCD problem [39, 55], where Bauer and Hagenauer proposed a MAP decoding based soft-input/soft-output module for decoding reversible variable length codes (RVrLC). Although some alphabetical statistics of the source signals have been exploited by

Year	Author(s)	Contribution
1948	Shannon [38]	stated his source-channel separation theorem.
2000	Bauer and Hagenauer [39]	derived a maximum a posteriori (MAP) decoding based soft-input/soft-output module for decoding reversible variable length codes (RVrLC) to be used in an iterative decoding structure.
2001	Görtz [40]	proposed the ISCD as an iterative approximation for optimal Joint Source-Channel Decoding (JSCD).
	Fingscheidt and Vary [41]	proposed the Softbit Source Decoding (SBSD) for the error-resilient recovery of speech signals.
2004	Kliewer *et al.* [42,43]	proposed soft decoding of images using a Markov Random Field (MRF) model in an ISCD scheme.
2005	Adrat and Vary [44]	proposed new design guidelines and channel coding component for ISCD using Extrinsic information transfer (EXIT) charts.
	Kliewer [45]	proposed an ISCD scheme relying on Variable Length Coding (VrLC) [46] using a symbol-based soft-input *A Posteriori* Probability (APP) decoder.
2009	Thobaben and Kliewer [47]	proposed a novel VrLC construction method for ISCD, providing both robustness and good source compression properties.
	Nasruminallah and Hanzo [48]	proposed an ISCD scheme with relying on so-called short block code [49] and SBSD for transmitting H.264 [10] video.
2011	Schmalen *et al.* [50]	improved the computation of EXIT charts related to the ISCD process, which overcomes the so-called overshooting effect and provides improved guidelines for ISCD system designs.
	Schmalen *et al.* [51]	proposed a new mapping method to map the source codec parameters to certain bit patterns for achieving a reduced error floor after ISCD decoding.
2012	Huo *et al.* [52]	modelled the correlation of video signals by a first-order Markov process and proposed a spatio-temporal ISCD system for exploiting both the intra-frame and the inter-frame correlations.
2014	Wu *et al.* [53]	employed ISCD for decoding distributed channel coded cooperative transmissions.

Table 1.3: Major contributions on iterative source-channel decoding (ISCD).

the RVrLC soft decoder for error concealment, the intrinsic correlation among the source signals has not been considered. The ISCD philosophy was later proposed by Görtz [56, 40] as an iterative approximation for optimal JSCD, which exploited the correlation among the source signals modelled by a first-order Markov process. Extrinsic information was exchanged between the source encoder and decoder in a turbo-like maner. Fingscheidt and Vary also used the first-order Markov process to model speech signals and derived the Softbit source decoding (SBSD) concept [41], which extracts softbit information from the intrinsically correlated speech signals. Adrat and Vary [44] took a further step by carefully designing the bit mappings in order to map symbols of the source signals to bit patterns, so that the soft-bit based ISCD became capable of achieving an improved performance. With the aid of EXIT charts, they determined maximum number of beneficial iterations for ISCD. Taking source compression coding into consideration, Kliewer [45] proposed an ISCD scheme combined with VrLC [46] using a symbol-based soft-input A Posterior Probability (APP) decoder, where the characteristics of the source were incorporated into the VrLC soft-decoder. The VrLC construction method to be used for ISCD was proposed in [47], which provide robustness and good source compression properties. Schmalen *et al.* observed in their work the existence of error floors in the traditional ISCD systems, therefore they improved the computation of EXIT charts for the sake of overcoming the so-called overshooting effect and proposed a new mapping method to further reduce the error floor after ISCD decoding [51, 50]. Wu *et al.* [53] employed ISCD for decoding distributed channel coded cooperative transmissions for different transmission channels, where the residual redundancy in the source codes as well as the artificial redundant information introduced by the distributed channel encoder was jointly exploited for the sake of reducing the effects of decoding errors.

When applied to image or video transmission, the ISCD scheme requires sophisticated modelling of the source. Kliewer, Görtz and Mertins [42, 43] exploited the *a priori* information inherent in images and invoked both MAP image estimation using Markov random field theory as well as the Markov–Gibbs correspondence relationship discovered in [57]. More explicitly, they modelled images using a Markov Random Field (MRF) for the sake of generating the APPs of the pixels by exploiting the correlation of the adjacent ones and they designed a soft-in-soft-out (SISO) module for exchanging extrinsic information with the adjacent modules in order to perform ISCD. In [48], an ISCD system relying on variable code-rate short block codes (SBC) was proposed for H.264 [10] video telephony, where the partitioned H.264 bitstream contained the partitions A, B and C [10], which were then further encoded by SBC. The artificial redundancy imposed by the SBC on the H.264-encoded bitstreams can be exploited by the softbit source decoder, which performs iterative decoding by exchanging extrinsic information with the channel decoder. Finally, the authors of [52] used a first-order Markov process to model the correlation inherent in the video signals. A spatio-temporal ISCD relying on forward error correction (FEC) codes was proposed. For the decoding process, both the horizontal and vertical intra-frame correlations, as well as the inter-frame correlations, were exploited by the receiver. The corresponding 3-D information exchange was monitored by using 3-D EXIT chart, in order to achieve an improved system performance.

1.3 Overview of Unequal Error Protection

Under the framework of Shannon's source and channel-coding separation theorem [38], all source signals are assumed to have the same importance and are protected equally by the channel encoder. However, real-life applications are often required to transmit source signals having unequal importance, such as layered video streams. Therefore unequal error protection (UEP) techniques have been developed to offer more effective protection. The major contributions to UEP are listed in Table 1.4.

Year	Author(s)	Contribution
1967	Masnick and Wolf [58]	proposed the concept of UEP, which protects more important data with the aid of stronger FEC.
1988	Hagenauer [59]	investigated a UEP system using rate-compatible punctured convolutional (RCPC) codes.
1993	Calderbank and Seshadri [60]	proposed multilevel codes for UEP.
1996	Albanese *et al.* [61]	proposed priority encoded transmission (PET) scheme for sending messages over lossy packet-based networks by exploiting their specific priorities.
2000	Wu *et al.* [62]	proposed a packet-level UEP scheme employing hard decoded FEC codes for mitigating the packet loss events at the application layer.
2001	Chou *et al.* [63]	proposed a layered multicast system of audio and video using PET.
2003	Stockhammer *et al.* [64]	proposed a bit-level UEP for transmitting H.264/AVC coded video in wireless environments.
2004	Zhang *et al.* [65]	designed a HARQ scheme to provide UEP for video transmission by appropriately sharing the bitrate budget between the source and channel encoders.
2005	Brüggen and Vary [66]	conceived an UEP scheme relying on specific power-allocation according to the bit error sensitivity of different transmitted data.

	Barmada *et al.* [67]	proposed a video transmission scheme for partitioned H.264/AVC coded video using HM-QAM to provide UEP.
2006	Chang *et al.* [68]	invoked an UEP scheme, where the intra/inter-coded frame of the H.264 coded video are protected using different levels of protection provided by adaptive hierarchical QAM.
2007	Rahnavard *et al.* [69]	proposed an UEP scheme provided by low-density parity-check (LDPC) codes.
2009	Park *et al.* [70]	conceived a UEP scheme for transmission SVC over an orthogonal frequency-division multiplexing (OFDM) channel.
2010	Li *et al.* [71]	designed an UEP scheme to transmit the H.264 coded video over frequency selective fading channels, combining hierarchical QAM with sub-carrier mapping to provide UEP.
2011	Arslan *et al.* [72]	optimised a coded progressive image transmission system.
	Hellge *et al.* [73]	proposed a UEP scheme for transmission over mobile channels by exploiting the hierarchy of video layers.
2012	Khalek *et al.* [74]	introduced an APP/MAC/PHY cross-layer UEP architecture for improving the perceptual quality of delay-constrained scalable video transmission.
	Chang *et al.* [75]	proposed a Hierarchical Modulation (HM) based system providing multiple levels of UEP.
	Alajel *et al.* [76]	conceived an UEP system for color plus depth-map aided stereoscopic video using hierarchical 16-QAM.
2014	Wu *et al.* [77]	proposed cross-layer-operation aided of UEP schemes for video transmission by using UEP Raptor codes at the application layer, and UEP RCPC codes at the physical layer.
	Nguyen *et al.* [78]	investigated the performance of UEP in wireless relay networks.

Table 1.4: Major contributions on unequal error protection for video communications.

The pioneer work on UEP by Masnick [79] used linear FEC codes for assigning different priority to the encoded signals. The digits of a codeword were differentiated according to their “protection level” and sophisticated methods were proposed for generating linear codes exhibiting beneficial FEC properties. Power-allocation [66] can also be readily used for UEP. As a further advance, Hagenauer designed rate-compatible convolutional codes (RCPC) for an UEP system [59] where different coding-rates were generated by variable rate puncturing. By contrast, in [80] convolutional codes (CC) having configurable free distances were used. Low-density parity-check (LDPC) codes [81,82,83,69] were also used for designing UEP schemes.

The aforementioned contributions laid the theoretical foundations of UEP, often using artificially generated signal models of different error sensitivities. Further contributions tackled the problem of transmitting real source signals, like realistic video signals, which is more application-specific and more challenging in terms of meeting the system requirements. Again, it is intuitively appealing to consider FEC for UEP system designs. In [77], the cross-layer design of FEC schemes was investigated by using UEP Raptor codes at the application layer (AL), and UEP RCPC codes at the physical layer (PHY) for the prioritized video packets, which were prioritized based on their contribution to the received video quality. The authors of [74] introduced an APP/MAC/PHY cross-layer architecture that improved the perceptual quality of delay-constrained scalable video transmission. Furthermore, an online QoS-to-QoE mapping technique was proposed in [74] for quantifying the QoE reduction imposed by each video layer using both the ACK history and a variety of perceptual video-quality metrics. The authors of [70] studied the channel-dependent adaptation capability of SVC by conceiving a solution for transmission over an orthogonal frequency division multiplexing (OFDM) based broadband network relying on cross-layer optimization. The FEC protected UEP schemes may be classified into two categories, namely the packet-level schemes [84, 85, 86, 87, 88, 89, 90, 91] and bit-level schemes [92,93,94,95,96,73,97,98]. The packet-level contributions [84,85,86,87,88,89,90,91] usually employ hard decoded FEC codes for mitigating the packet loss events at the application layer [62], while the bit-level ones operate at the physical layer and rely on soft decoded FEC codes for correcting bit-errors in wireless scenarios [64]. Traditional UEP schemes designed for layered video transmission only handle the different importance of separate video layers by assigning different-rate FEC codes to them. By contrast, the recent contributions [87, 90, 73, 91, 98] explored the dependencies amongst the layers and conceived UEP schemes by exploiting this sophisticated feature. Specifically, the unsuccessful decoding of the BL instructed the video decoder to discard all the ELs depending on it, regardless whether they had or had not been successfully decoded. Naturally, this course of action wasted the transmit power assigned to the dependent layers. Thus we proposed in our previous work [98] a bit-level inter-layer coded FEC (IL-FEC) scheme that embeds the BL into the FEC coded ELs, so that the reception of the BL can be improved with the aid of the ELs using soft decoding. In our subsequent work [99] we conceived a sophisticated on-line real-time video distortion estimation technique, which is suitable for diverse channel conditions and system configurations. More explicitly, in [99] we proposed an on-line code rate optimization method for minimizing the video distortion.

In different circumstances, UEP is achieved by designing modulation and coded modulation schemes [100,101,60,102,103,104,94,105,75,106,78,71,107,67,68,108,76]. Typically, two types of modulation schemes attracted substantial research interests, namely multilevel coded modulation (MCM) [60,102,103,104,94,105] and hierarchical modulation (HM) [75,106,78,71,107,67,68,108,76]. The latter one can be considered as a specific subset of the former one. Multilevel coded modulation was originally proposed for the purpose of UEP in [60], where the 2-dimensional signal constellation was partitioned into disjoint subsets and the bit streams of different importance were mapped to different constellation subsets. Later the theoretical upper bounds and computer simulation based verification of the performance of multilevel block coded modulation schemes designed for UEP and for multistage decoding were presented in [102,103]. The author of [104] and [94] improved the set partitioning of QAM used by TCM for achieving better UEP performance. Finally, a multilevel space-shift-keying scheme was designed by the authors of [105] to provide UEP for transmission of Rician fading channels.

On the other hand, HM is another popular technique of achieving UEP in communication systems, which has already been widely accepted by industry and hence become an integral part of the DVB-T/-H standard [109,110]. Similar to MCM, the symbols of HM schemes are carefully partitioned, but rather than uniformly distributing the signal space, HM maps the bits to symbols by carefully creating protection groups. The attainable Bit Error Ratio (BER) of HM has been investigated in [111], and in cooperative communications systems [112,113].

Numerous systems were developed to achieve UEP by exploiting the properties of HM schemes. An UEP system was proposed in [75] for creating multiple levels of UEP, instead of the traditional two-level designs. The authors of [106,107] proposed HM-QAM like MIMO systems and analysed the attainable UEP properties. The aforementioned contributions carried out their investigations without actually considering multimedia sources that are used in real-life, while the following works explicitly considered real image or video transmission systems. More specifically, the authors of [72] jointly optimised the channel coding parameters and the HM parameters in a coded-HM-aided progressive image transmission system, in order to achieve minimum graphical signal distortion. The partitioning mode of H.264/AVC coded video was considered in [67], and UEP was provided by HM-QAM for the video partitions of different importance. Similarly, H.264 coded video was considered [68], where the intra-coded frame (I-frame) and predictive coded frame (P-frame) were protected using different levels of protection provided by the adaptive HM-QAM. The authors of [71] conceived an UEP scheme for transmitting the H.264 coded video over frequency selective fading channels. HM-QAM was jointly considered in conjunction with mapping the source bits the sub-carriers for providing the video layers of high importance with better error protection. Finally, colour-map plus depth-map based stereoscopic video was considered in [76], using HM 16-QAM. Since the color component had a more significant impact on the reconstructed video quality, the color sequence was mapped to the layers of higher protection in the HM-QAM constellation , so as to achieve high-quality 3-D video perception.

A range of UEP schemes have been conceived for HARQ [63,114,65,115,116,117, 118, 119, 120] in order to improve the video quality of layered video systems. The authors of [65] proposed UEP by appropriately sharing the bitrate budget between the source and channel encoders based on either the minimum-distortion or on the minimum-power consumption criterion. In [118,119], UEP was achieved by assigning each video layer a different ARQ-based retransmission limit. Another stream of contributions [63,65,115,116,117] adopted the so-called limited-retransmission based priority encoding transmission (PET) scheme [61], where UEP is achieved by varying the source block-length across the different source layers, while keeping the FEC-coding block-length fixed. This allows the PET to have a packetization scheme ensuring that the source layers of an FEC-coded block are dropped according to their significance, commencing by dropping the least significant one first.

1.4 Overview of Inter-Layer Forward Error Correction Codes

As reviewed in Section 1.1.3, layered video compression techniques, such as scalable video coding (SVC) [11], are capable of encoding video into several layers associated with different importance. The more important layers should be given higher priority during transmission, since their corresponding dependent layers rely on them for successful decoding. Therefore, unequal error protection (UEP) [58] can be utilized for providing the different-sensitivity layers with different levels of protection in order to achieve an improved overall video quality.

Unequal error protection [121] was reviewed in Section 1.3. Specifically, traditional UEP schemes designed for layered video transmission simply handle the different importance of the separate video layers by assigning different-rate FEC codes to them. By contrast, the more sophisticated recent contributions [87, 90, 73, 91, 98] explored the dependencies amongst the layers and conceived UEP schemes by exploiting this inherent feature. More specifically, the unsuccessful decoding of the BL will instruct the video decoder to discard all the ELs depending on it, regardless whether they have or have not been successfully decoded. Naturally, this course of action wasted the transmit power assigned to the dependent layers.

We will introduce the IL-FEC coding scheme of [122, 98] and its principles in this section. According to the IL-FEC encoding scheme, the systematic bits of the base layer will be embedded into the systematic bits of its dependent enhancement

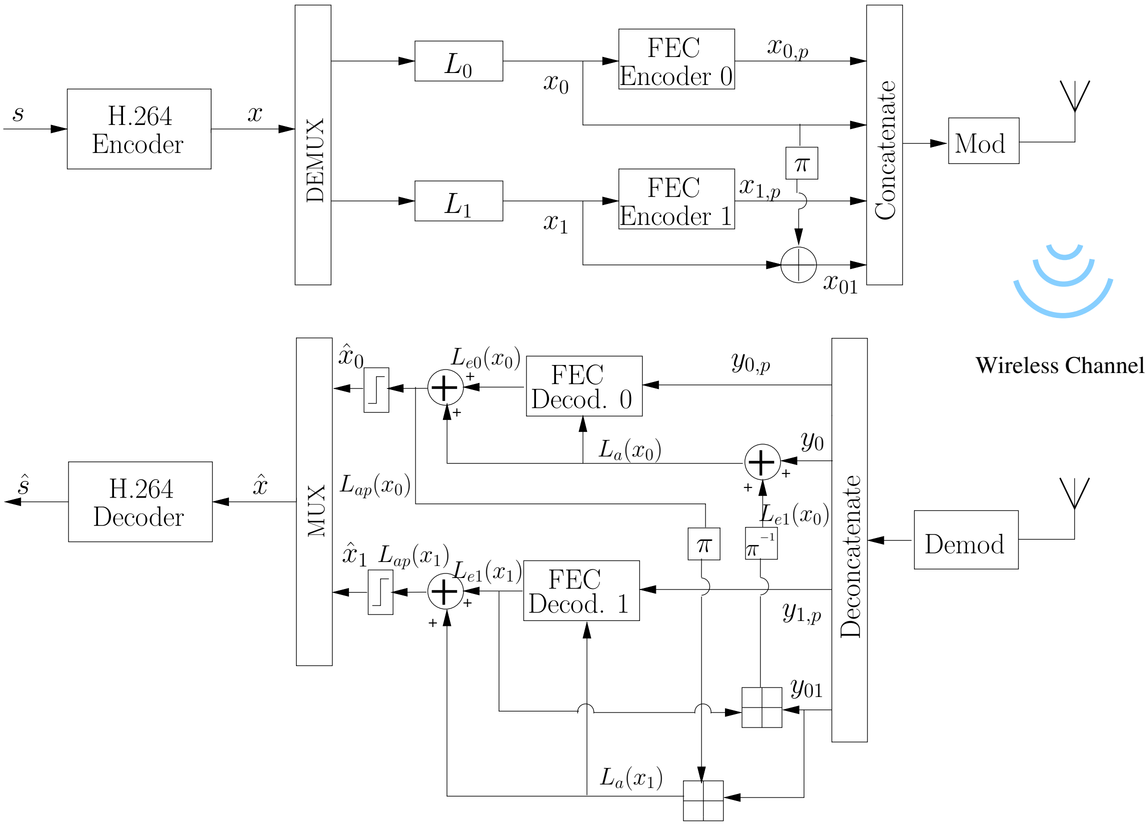


Figure 1.3: IL-FEC encoding architecture of layered video for two layers. [122, 98]

layers. When the base layer is corrupted during transmission, the inter-layer decoding process is invoked for using the information of the enhancement layers with the goal of recovering the BL with the aid of the BL information embedded into the EL. When the base layer is successfully decoded, the soft information of the enhancement layers can be recovered without any loss of any information. This inter-layer technique can be readily combined with arbitrary UEP techniques for further improving the achievable performance.

The architecture of the IL-FEC codec is depicted in Figure 1.3. Specifically, two video layers are used for transmission in order to illustrate the IL-FEC regime, namely the BL L_0 and the EL L_1. Again, this regime can be readily applied to multiple layered video bit-streams in the context of different architectures. As a component of the IL-FEC scheme, the FEC can be any popular systematic LDPC [123], turbo or convolutional code.

In the rest of this section, we will provide a description of the encoder architecture in Section 1.4.1 and the decoder will be discussed in Section 1.4.2. Consequently, the PSNR performance of the IL-FEC coded layered video transmission techniques can be estimated “on-the-fly” by an optimization algorithm at the transmitter so as to achieve the best system performance using the method detailed in [99].

1.4.1 Inter-Layer FEC Encoding

As shown in Figure 1.3, the video signal s is encoded by the layered video encoder into the bitstream x, which consists of the BL L_0 and the EL L_1. The latter one depends on the former one for successful decoding, while L_0 and L_1 are separated by the *DEMUX* module of Figure 1.3. Then each layer is encoded using its individual FEC code. For layer L_0 the input bit sequence x_0 is encoded by its FEC encoder in order to produce the parity bits x_{0p}, and for layer L_1 the parity bits x_{1p}.

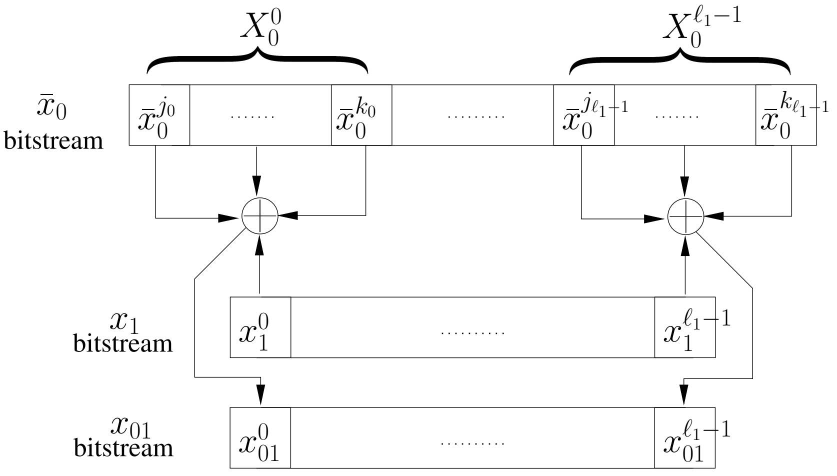
According to the IL-FEC mechanism of Figure 1.3, x_0, as the systematic part of the encoded layer L_0, is interleaved and then embedded into the systematic part x_1 of L_1 in order to produce the bit sequence x_{01}, using the bit-wise XOR operation. When L_0 and L_1 are different in length, the solution proposed in [98] may be invoked, which will be detailed later. Then the systematic bits x_0 and the parity bits $x_{0,p}$ of the BL are concatenated, along with the embedded systematic bits x_{01} of the EL and the parity bits $x_{1,p}$ of the original EL. Finally, the concatenated bitstream is modulated and transmitted over the wireless channel, which can be modelled as an uncorrelated Rayleigh-faded channel as in [98] or [124] .

Now let us elaborate a little further on the *XOR* operation that is invoked to embed x_0 into the bitstream of x_1. Let us assume that the length of x_0 and x_1 is ℓ_0 and ℓ_1, respectively. Furthermore, x_1 and x_{01} are of the same length. In order to generate each bit x_{01}^i, the corresponding x_1^i carries out the conventional binary *XOR* operation with a subset of bits from the interleaved version $\bar{x}_0$ of x_0, namely X_0^i, $0 \leq i < \ell_1$. Therefore, the *XOR* operation performed on X_0^i and x_1^i in order to generate x_{01}^i can be expressed as follows [98]:

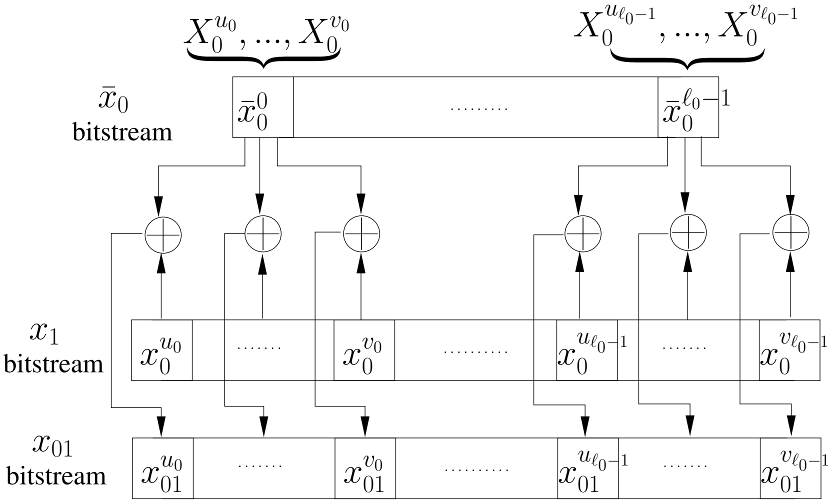
$$x_{01}^i = x_1^i \oplus \sum_{\bar{x}_0^r \in X_0^i} \oplus \bar{x}_0^r, 0 \leq i < \ell_1, \tag{1.3}$$

where $\oplus$ denotes the binary *XOR* operation.

Now let us discuss the *XOR* operation and the definition of X_0^i on a case by case basis. Firstly, if the bit sequences x_0 and x_1 are equal in length, then X_0^i consists only of $\bar{x}_0^i$. According to Eq. (1.3), the *XOR* operation is performed on a bit by bit basis. Now let us discuss the scenarios where the bit sequences x_0 and x_1 have different lengths, which is likely to be the case in reality. When x_0 have more bits than x_1, the *XOR* operation is carried out as shown in Figure 1.4a. Observe from Figure 1.4a that each bit set X_0^i consists of several bits that start from the index j_i and span to the index of k_i, according to $\bar{x}_0^{j_i}, \cdots, \bar{x}_0^{k_i}$. Those bits are then combined using the binary *XOR* operation, together with x_1^i, in order to obtain x_{01}^i. On the other hand, when x_0 consists of less bits than x_1, X_0^i may contain a single bit from $\bar{x}_0$, while multiple X_0^is, namely $X_0^{u_i}, ..., X_0^{v_i}$, share the same bit from $\bar{x}_0$. The *XOR*



(a) Definition of XOR between x_0 and x_1 when $\ell_0 > \ell_1$



(b) Definition of XOR between x_0 and x_1 when $\ell_0 < \ell_1$

Figure 1.4: Schematics of the bit-wise XOR operation used in IL-FEC, where ℓ_0 and ℓ_1 indicate length of bit sequences x_0 and x_1, respectively.

operation for this situation is depicted in Figure 1.4b. According to Eq. (1.3), the same bit from $\bar{x}_0$ performs the binary *XOR* operation with different x_1^is, in order to generate the corresponding x_{01}^is.

1.4.2 Inter-Layer FEC Decoding

After transmission through the wireless channel, the received signals are subject both to noise and fading, therefore we benefit from IL-FEC decoding, as shown in Figure 1.3. Before IL-FEC decoding is carried out, the demodulated signals are partitioned into separate bitstreams, namely the systematic information of the BL y_0, the parity information of the BL $y_{0,p}$, the base-layer-related systematic information of the EL y_{01} and the parity information of the EL $y_{1,p}$.

Since the EL L_1 depends on the BL L_0 for successful video decoding, the FEC decoder 0 of L_0 is invoked first. As shown in Figure 1.3, both the LLR $L_a(x_0)$ of the *a priori* information about x_0 (which is only the LLR of the systematic information y_0 at the beginning of the decoding process) and the parity information $y_{0,p}$ are exploited for providing extrinsic LLRs for x_0, namely $L_{e0}(x_0)$. When combined with the *a priori* LLR $L_a(x_0)$, $L_{e0}(x_0)$ will produce the *a posteriori* LLR $L_{ap}(x_0)$. Then $L_{ap}(x_0)$ is hard-decoded on a bit-by-bit basis for producing $\hat{x}_0$,

FEC decoder 1 of Figure 1.3 carries on to decode L_1 from the received signal. Since there is no direct *a priori* LLR $L_a(x_1)$ for the systematic information of L_1, we have to produce $L_a(x_1)$ using the following procedure. As shown in Figure 1.3, $L_a(x_1)$ is retrieved using y_{01} and the extrinsic LLR $L_{ap}(\bar{x}_0)$ of the interleaved x_0 by carrying out the frame-wise boxplus operation. Before we carry on explaining, let us first define the bitwise boxplus operation $\boxplus$. Let u_1, u_2 be a pair of arbitrary binary variables, then we have $L(u_1) \boxplus L(u_2) = L(u_1 \oplus u_2)$ [125], which may be further formulated as [126]

$$\begin{aligned} L(u_1) \boxplus L(u_2) = & \log \frac{1 + e^{L(u_1)} e^{L(u_2)}}{e^{L(u_1)} + e^{L(u_2)}} \\ = & \text{sign}\left(L(u_1)\right) \cdot \text{sign}\left(L(u_2)\right) \cdot \min\left(|L(u_1)|, |L(u_2)|\right) \\ & + \left[\log\left(1 + e^{-|L(u_1)+L(u_2)|}\right) - \log\left(1 + e^{-|L(u_1)-L(u_2)|}\right)\right]. \end{aligned} \tag{1.4}$$

Therefore, the previously mentioned frame-wise boxplus operation between the two streams of signals is carried out as follows. For each element $L_a(x_1^i)$ in $L_a(x_1)$, we have [98]

$$\begin{aligned} L_a(x_1^i) &= L\left(x_{01}^i \oplus \sum_{\bar{x}_0^r \in X_0^i} \oplus \bar{x}_0^r\right) \\ &= L\left(x_{01}^i\right) \boxplus \sum_{\bar{x}_0^r \in X_0^i} \boxplus L_{ap}(\bar{x}_0^r), \end{aligned} \tag{1.5}$$

where the $\boxplus$ operator represents the bitwise boxplus operation, and $L\left(x_{01}\right)$ is the LLR representing y_{01}. Using $L_a(x_1)$ and the LLR of $y_{1,p}$, FEC decoder 1 generates the extrinsic LLR for x_1, namely $L_{e1}(x_1)$. When combined with the *a priori* LLR $L_a(x_1)$, $L_{e1}(x_1)$ will produce the *a posteriori* LLR $L_{ap}(x_1)$. Then $L_{ap}(x_1)$ is hard-decoded on a bit-by-bit basis to produce $\hat{x}_1$.

On the other hand, in order to exploit the information of x_0 embedded into x_{01} for facilitating the decoding process of FEC decoder 0, $L_a(x_0)$ should be updated after the decoding action of FEC decoder 1 of Figure 1.3. The updated $L_a(x_0)$ consists of two parts, namely the LLR of y_0 received, as well as of the extrinsic information $L_{e1}(\bar{x}_0^p)$ provided by FEC decoder 2 of Figure 1.3. The corresponding extrinsic information $L^i_{e1}(\bar{x}^p_0)$, where $0 \le p < \ell_0$, for any $\bar{x}^p_0 \in X^i_0$ may be derived as follows [98]

$$\begin{aligned} L^i_{e1}(\bar{x}^p_0) &= L\left(x^i_1 \oplus x^i_{01} \oplus \sum_{\bar{x}^r_0 \in X^i_0, r \neq p} \oplus \ \bar{x}^r_0 \right) \\ &= L_{e1}\left(x^i_1\right) \boxplus L\left(x^i_{01}\right) \boxplus \sum_{\bar{x}^r_0 \in X^i_0, r \neq p} \boxplus \ L_{ap}(\bar{x}^r_0). \end{aligned} \tag{1.6}$$

Therefore, the extrinsic information of $\bar{x}^p_0$, namely $L_{e1}(\bar{x}^p_0)$, depends on all the X^i_0 sets that $\bar{x}^i_0$ belongs to. The final form of $L_{e1}(\bar{x}^p_0)$ can be expressed as [98]

$$L_{e1}(\bar{x}^p_0) = \sum_{\left\{ i | x^p_0 \in X^i_0 \right\}} L^i_{e1}(\bar{x}^p_0). \tag{1.7}$$

Finally, $L_a(x_0)$ that is used by FEC decoder 0 of Figure 1.3 can be updated by adding the deinterleaved version of $L_{e1}(\bar{x}^p_0)$ together with the the LLR of y_0, which can be expressed as

$$L_a(x^p_0) = L_{e1}(x^p_0) + LLR\left(y^p_0\right), \tag{1.8}$$

where $LLR\left(\right)$ is the LLR conversion from the demodulated signals. Then the two FEC decoders of Figure 1.3 carry on with the iterative decoding process, until the affordable number of decoding iterations is reached.

Since all the transmitted singals are assumed to be protected by a Cyclic Redundancy Check (CRC) code, $\hat{x}_0$ could be checked for correctness. Once the BL L_0 is successfully decoded at any stage of the IL-FEC decoding process, decoder 0 of Figure 1.3 will provide decoder 1 with the perfectly decoded bit sequence x_0. As a result, Eq. (1.5) will be simplified as [122]

$$\left.\begin{aligned} L_a(x^i_1) &= \prod_{x^r_0 \in X^i_0} \text{sign}\left(\tilde{x}^r_0\right) \cdot L\left(x^i_{01}\right) \\ &= \prod_{x^r_0 \in X^i_0} \text{sign}\left(\tilde{x}^r_0\right) \cdot LLR\left(y^i_{01}\right) \end{aligned}\right., 0 \le i < \ell_1, \tag{1.9}$$

where $\tilde{x}^r_0$ is the modulated version of x^r_0. Actually this shows a beneficial property of the IL-FEC, namely that if the BL has been sucessfully decoded without any errors, only sign flipping is required for recovering the LLR $L_a(x^i_1)$ of the systematic information for decoding x_1, leaving the protection of x_1 unaffected.

Again, the base layer does not rely on the EL for decoding its information and the systematic information of the enhancement layer can be extracted from the perfectly decoded base layer. However, when the base layer cannot be decoded independently, the iterative decoding philosophy exchanging extrinsic information between the base layer and the enhancement layer will be activated, where the check information y_{01}

will be utilized to correct the errors inflicted upon y_0^i. Note that when the base layer cannot be decoded correctly, the enhancement layer must be dropped. Furthermore, since the transmitted signal x_{01} has the same length as the length of the bit sequence x_1, we do not require any extra parity bits. Furthermore, the iterative decoding philosophy is activated only when the base layer cannot be decoded correctly, hence we impose an acceptable complexity and delay. If the decoded bit sequence $\hat{x}_0$ of the base layer is corrupted after the IL-FEC decoding stage seen in Figure 1.3, it will be dropped together with the enhancement layer $\hat{x}_1$. Otherwise they will be forwarded to the H.264 decoder for generating the reconstructed video signal $\hat{s}$.

1.5 Overview of Indoor Visible Light Communications

Visible Light Communication (VLC) [127] belongs to the family of Optical Wireless Communication (OWC) techniques relying on the visible light frequency band between 400 and 800 THz (1 THz = 10^{12} Hz) corresponding to the wavelengths between 375 and 750 nm. VLC has mainly been designed for indoor applications, while the so-called Free Space Optical (FSO) systems operating at near Infrared (IR) frequencies have been conceived for outdoor scenarios. VLC systems employ light emitting diodes (LEDs) which can be pulsed at high frequencies beyond the human eye's fusion frequency, where information is transmitted by modulating the signal in form of optical intensities, which can be received by photo-detectors.

The compelling virtues offered by VLC have been driving the related research interests as well as commercial investments. Firstly, the bandwidth offered by VLC in the optical electro-magnetic spectral band is huge and may conveniently complement the bandwidth resources of radio frequency (RF) communication systems. Secondly, VLC systems are free from electro-magnetic interference (EMI) with the existing radio systems, hence requiring no frequency-planing. Furthermore, VLC combines the illumination and communication functions. Finally, since VLC systems rely on light, their coverage is confined to a room. Hence a high grade of privacy can be guaranteed.

1.5.1 Brief History of VLC Development

The earliest research of employing LEDs for VLC systems originated from Japan. In 1999, Pang *et al.* proposed the idea of fast switching of LEDs and the modulation of the visible light for communications [128]. The LED traffic lights are used as communications devices, which emits modulated and encoded visible lights to broadcast audio or road information. In the following year, researchers at Keio University begin their pioneering research on utilising white LEDs designed both for illumination as well as communication [129, 130], where both the influence of interference and of reflection of visible light were considered. Further research on VLC was combined with Power Line Communication (PLC). In [131], the authors designed a PLC/VLC system prototype, which achieved a data rate of 100 kbps and suggested that VLC could indeed be deployed as an attractive last-mile solution for indoor applications.

In 2003, the Visible Light Communications Consortium (VLCC) [132] was found in Japan, owing to the rapid research and fast development on VLC systems. As a successor of the VLCC, the Visible Light Communications Associations (VLCA) was founded later in 2014. In 2008, the European Union sponsored the Home Gigabit AccessOMEGA project, aimed at providing heterogeneous communication through OWC, PLC, RF communications in order to build "wireless homes" [133, 134]. Aiming at transmitting at Gigabit rates for domestic users, they have demonstrated the capability of the system by allowing the user to download several HD video streams in parallel [135]. Finally, the Li-Fi (Light-Fidelity) Consortium was formed in Oslo, Norway in 2011, as a complement to Wi-Fi. Additionally, some other research groups played vital roles in the development of VLC systems, namely the Center for Ubiquitous Communication by Light (UC-Light) [136], the Center on Optical Wireless Applications (COWA) [137], the Smart Lighting Engineering Research Center (SLERC) [138] *etc.*

As a result, some of the techniques have been standardized successively. In Japan, the VLCC proposed two standards in 2007, namely the Visible Light Communication System Standard (JEITA CP-1221) and the Visible Light ID System Standard (JEITA CP-1222) [139]. The IEEE 802.15.7 VLC standard was published in 2011 for employment in Wireless Personal Area Networks (WPAN) [127]. From a commercial perspective, in 2007 a company called LVX specializing in VLC systems signed a contract with the city of St. Cloud, Minnesota, which is the first city to commercially deploy the VLC technology [140]. The major events in the history of VLC development are listed in Figure 1.5.

1.5.2 LED Employment In VLC Systems

Since the development of the first visible-spectrum LED emitting red lights in 1962 [144], we have seen gradually proliferating commercial developments. A great breakthrough took place in 1994, when the high-brightness blue LED was invented by Shuji Nakamura [145], hence leading to the invention of high-brightness white LEDs that are suitable for illumination. Since then we have seen rapid improvements in the brightness of the white LEDs, leading to more and more commercial lighting products. It is expected that the majority of new energy-efficient lighting installations will be LED-based by the year 2018 [146]. As it turns out, LEDs excel not only in energy-efficient illumination, but they are also capable of providing extra spectrum for VLC, with no negative influence on their illumination functionality. These promising applications of LEDs are all facilitated by the impressive properties of LEDs, which are listed below:

Year	Authors/ Institutions	Contributions
1999	Pang *et al.* [128]	proposed high-switching-speed LEDs and their modulation for visible light communications.
2000	Tanakat *et al.* [141]	proposed white LED for both communications and illumination.
2003	Komine and Nakagawa [131]	designed a Power-Line Communication (PLC)/VLC system prototype.
	Visible Light Communications Consortium (VLCC) [132]	VLCC was established in Japan (http://www.vlcc.net/).
2007	VLCC [139]	proposed two standards: Visible Light Communication System Standard and Visible Light ID System Standard.
2008	European Union [134]	sponsored the Home Gigabit Access (OMEGA) project.
2010	LVX, Minnesota [140]	signed a contract with the city of St. Cloud, Minnesota, which is the first city to commercially deploy the VLC technology.
2011	IEEE 802.15 Working Group [127]	published IEEE 802.15.7 VLC standard.
	Li-Fi (Light-Fidelity) Consortium [142]	was formed in Oslo, Norway, as a complement to Wi-Fi.
2014	VLCA (Visible Light Communications Associations) [143]	is established as a successor of VLCC in Japan for further standardization of VLC.

Table 1.5: Major milestones in the history of VLC systems

- High luminous efficiency: the luminous efficacy of InGaN based white LEDs achieved 303 [lm/W] in 2014 [147, 148], while the average luminous efficacy level is as low as 5-35 [lm/W] for incandescent light bulbs and 60-100 [lm/W] for fluorescent lamps, because they generate less heat [149].
- Long lifetime: LEDs exhibit a longer quality service time of up to 25,000-50,000 hours, which is much longer than the 6,000 to 15,000 hours provided by conventional incandescent bulbs [150]. Furthermore, LEDs seldom encounter abrupt failure but gradually fail by dimming over time.
- Very High Inherent Bandwidth: When used as a VLC transmitter, over 100 [Mbps] of bandwidth can be achieved by LEDs, which enables high-speed communication [133].
- Color and Dimming: Unlike traditional lighting methods, LEDs can emit light of different colours without using any color filters. Furthermore, the dimming of LEDs is easy, either with the aid of pulse-width modulation or by reducing the forward current [151, 152].

Due to their impressive characteristics and with the increased demand for indoor energy efficient light sources, LEDs are rapidly replacing traditional lighting devices and are eminently suitable for VLC systems.

Several types of LEDs suitable for VLC are currently available on the market, namely Phosphor Converted LEDs (PC-LEDs) [153], Multi-Chip LEDs (MC-LEDs) [154], Organic LEDs (OLEDs) [155] and Micro-LEDs [156]. The former two types of LEDs are most commonly adopted. More specifically, PC-LEDs are single-chip LEDs that provide only a single channel for signal transmission in VLC, while MC-LEDs offer several channels for transmitting different signals. PC-LEDs use phosphor to filter the light generated by a blue InGaN LED chip in order to produce light of different colours in order to synthesize white light both for illumination and transmission. The bandwidth of the PC-LEDs is relatively low because of the limited efficiency of the phosphor-based conversion. On the other hand, white light is produced by the MC-LEDs upon combining the coloured light generated by separate chips, which can be controlled separately in terms of their intensities. Therefore, different colours are obtainable and high-throughput VLC modulation methods exploit this multi-channel property. Furthermore, MC-LEDs offer a higher bandwidth than PC-LEDs.

The suitability of different LEDs for VLC deployment was investigated in [157, 158, 159, 160, 161]. The authors of [158] designed a hardware prototype for short-range broadcasting using a white LED. In [162] the non-linear distortions introduced by LEDs in the transmission chain were studied by examining the performance of OFDM. Further research disseminated in [163] built up a model that incorporates amplitude distortion in order to describe the LED's transfer function. The non-linear distortions imposed by the LEDs on VLC systems using OFDM were also studied in [159], where different techniques were proposed for compensating the non-linear distortions. The optical noise faced by VLC was tackled by the authors in [164], where the effects of optical background noise on the performance was characterized and Manchester coding was proposed for mitigating the effect of optical noise. In [165], a VLC system based on a single 50-μm gallium nitride LED was presented, which had a significantly higher bandwidth than commercially available white lighting LEDs, while in [166] a high-performance blue filter was shown to increase the modulation bandwidth, whilst dramatically reducing the bit error ratio of the VLC system. Finally, a dual-purpose offline LED driver combined with illumination control and VLC capabilities was presented in [167], where pulsewidth modulation (PWM) based dimming was invoked at frequencies above the audible range, while maintaining a high data transfer rate. The major advances in LED design are listed in Table 1.6.

Year	Authors	Contributions
1962	Holonyak	developed the first visible-spectrum LED emitting red light.
1994	Nakamura [145]	invented a high-brightness blue LED.
2009	Elgala *et al.* [158]	proposed a novel hardware prototype using white LEDs for short-range VLC broadcasting.
	Elgala *et al.* [162]	studied the non-linear distortions introduced by LEDs in the transmission chain by examining the performance of OFDM.
2010	Elgala *et al.* [163]	developed a model that incorporates amplitude distortion in order to describe the non-linear LED transfer function.
2012	Mesleh *et al.* [159]	proposed different techniques in order to eliminate the non-linear distortion in a coded VLC system.
2013	Chow *et al.* [164]	characterized the effect of optical background noise on the performance and proposed Manchester coding in order to mitigate it.
2014	Tsonev *et al.* [165]	presented a VLC system based on an single 50-μm gallium nitride LED, which has a significantly improved bandwidth.
2015	Wang *et al.* [166]	proposed a high-performance blue filter for VLC systems to increase the modulation bandwidth and to reduce the bit error ratio.
	Modepalli *et al.* [167]	proposed a dual-purpose offline LED driver with illumination control and VLC capabilities.

Table 1.6: Major research contributions on the development of LEDs for VLC systems

1.5.3 Modulation for VLC Systems

In VLC, the modulated signals are transmitted by switching the LEDs at high frequencies without affecting the brightness perception of the human eye. The emitted light carrying the information can then be picked up by the photodetectors at the receiver and demodulated. There are many modulation techniques available for transmitting signals in VLC systems, but typically three modulation methods were standardize by the IEEE 802.15.7 recommendation [127], namely On-Off Keying (OOK), Pulse Position Modulation (PPM) and Colour-Shift Keying (CSK). More specifically, OOK and PPM were included as the modulation scheme of Physical layer I (PHY I) and Physical layer II (PHY II), while CSK represents the modulation scheme of Physical layer III (PHY III) [127]. The first two, i.e. OOK and PPM can be categorized as single carrier modulations, while CSK is transmitted with the aid of MC-LED using multiple carriers. We will introduce the research contributions on the modulation schemes designed for VLC systems using these three examples in the rest of this section.

The OOK scheme is a commonly used low-complexity modulation scheme, which treats each bit as a symbol and maps each bit 1 (“on”) to high amplitude and bit 0 (“off”) to zero amplitude. OOK is often combined with Non-return-to-zero (NRZ) codes, where bit 0 is mapped to the negative counterpart of the bit 1 signal. According to [168], a low-cost high-speed VLC system using OOK-NRZ-based modulation was shown to achieve a 456 Mbit/s data rate. In [169], a bitrate of 340 Mb/s was achieved within the bandwidth of 151 MHz by invoking a post-equalization circuit using OOK-NRZ modulation. Although not standardized in IEEE 802.15.7, Pulse Amplitude Modulation (PAM) [170, 171] has been employed in many VLC systems [172]. PAM is more bandwidth efficient than OOK due to the fact that it uses symbols of multiple intensity levels, which, however, makes PAM vulnerable to the nonlinearity of the LEDs.

In contrast to OOK and PAM, PPM [173] maps the information to the pulse positions rather than to the amplitude. More specifically, in L-PPM, a symbol is associated with a time slot for the transmission of a logical 1 bit out of L time slots within a symbol period. PPM has a lower average power requirement than OOK [174], but it is less bandwidth efficient [175], since it uses multiple time slots to transmit a single symbol. Moreover, the synchronization requirements are stricter, when receiving PPM symbols [176]. Moreover, the channel capacity can be more closely approached by using a PPM scheme [177, 178] and at a given SNR its BER performance is better than that of OOK [179]. Several variants of PPM have been adopted by VLC systems. For instance, Variable PPM (VPPM), which is the PPM variant that is specified in the IEEE VLC standard, controls the pulse width in order to support dimming. Multiple PPM (MPPM), on the other hand, relies on several light pulses instead of a single one and was shown to be twice more bandwidth efficient than PPM at the same power efficiency [180, 181, 182, 183]. Furthermore, the pulse positions in Overlapping PPM (OPPM) can be overlapped in order to achieve an increased data rates in a given bandwidth [184]. Finally, Differential PPM (DPPM) further improves the transmission capacity of PPM [185, 186]. The single carrier modulation related VLC studies are listed in Table 1.7.

The aforementioned modulation schemes are suitable both for FSO and for indoor VLC. On the other hand, CSK is most suitable for indoor VLC applications due to the more benign indoor channel conditions. The PC-LEDs that are suitable for single carrier transmissions cannot be used for CSK modulation, which relies on MC-LEDs. The CSK related studies are listed at a glance in Table 1.8, whilst their deeper discussions will be be provided in Chapter 4.

1.6 Motivation and Methodology

Shannon’s source and channel-coding separation theorem is not strictly applicable to realistic video encoders associated with finite complexity and delay, which fail

Year	Authors	Contributions
1983	Proakis *et al.* [170]	described On-Off Keying (OOK) and pulse amplitude modulation (PAM).
1989	Sugiyama *et al.* [181]	proposed multiple pulse position modulation (MPPM) to improve the bandwidth efficiency of PPM.
1993	Shalaby [184]	conceived overlapping PPM (OPPM).
1998	Ghassemlooy *et al.* [187]	proposed digital pulse interval modulation (DPIM) for optical wireless communication (OWC).
1999	Shiu *et al.* [185]	designed differential PPM (DPPM) for OWC.
2003	Ohtsuki [188]	proposed subcarrier-index modulation (SIM) for OWC.
2007	Fan *et al.* [189]	compared of PPM and pulse width modulation (PWM) for application in OWC.
2010	Nguyen *et al.* [180]	compared the bandwidth efficiency of MPPM with classic PPM.
2011	Rufo *et al.* [174]	compared to PPM with OOK at a given average power requirement in VLC systems.
2012	Arnon [176]	studied the effect of clock jitter using PPM in VLC applications.
2013	Fujimoto *et al.* [168]	demonstrated a low-cost high-speed VLC system using OOK-Non-return-to-zero (NRZ) signalling at 456 Mbit/s of data rate.
2014	Li *et al.* [169]	demonstrated the benefits of a VLC system using OOK-NRZ modulation for achieving a bitrate of 340 Mb/s in a bandwidth of 151 MHz.
2015	Biagi *et al.* [190]	evaluated PPM relying on space time block coding in VLC systems.

Table 1.7: Major contributions on single carrier modulations designed for VLC systems

to remove all the correlation of the source signal and hence result in a residual redundancy. In order to improve the attainable system performance, we set out to exploit this redundancy for improving the channel decoding performance.

We first investigate the amount of redundancy that resides within the transmitted uncompressed video, which exhibits both spatial (adjacent-pixel) correlation and temporal (inter-frame) correlation. Inspired by the existing lossless video codecs introduced in Section 1.1.2, we set out to reduce the redundancy by compressing the video for the convenience of transmission and storage. More explicitly, in Chapter 2, we used low-complexity frame-differencing (FD) based compression for removing the temporal redundancy and a variable-length code (VrLC) for removing the residual spatial redundancy of the FD signal. Furthermore, ISCD techniques will be conceived for exploiting the residual redundancy of the compressed video. More specifically, a new pixel-correlation model will be used in Chapter 2 to conceive a SISO module for

Year	Authors	Contributions
2010	Drost *et al.* [191]	proposed a specific CSK constellation design utilizing Billiards algorithm.
2012	Monterio *et al.* [192]	advocated a particular configurable colour constellation design.
	Bai *et al.* [193]	investigated various properties of CSK modulation and the trade-offs amongst them.
2014	Monterio *et al.* [194]	proposed a design framework for high-order CSK constellations.
	Drost *et al.* [195]	proposed optimization techniques for VLC systems having an arbitrary number of LEDs, each with an arbitrary number of emitters, as well as pre-compensation methods for mitigating the channel distortions.
	Singh *et al.* [196]	designed a CSK modulation format based on four colours.
	Luna-Rivera *et al.* [197]	proposed a time-division multiple access based multi-user CSK system.
	Pergoloni *et al.* [198]	combined PPM and CSK in a hybrid VLC system.
2015	Jiang *et al.* [199]	proposed a three-stage concatenated CSK system and designed its optimal signal labelling.

Table 1.8: Major contributions on indoor CSK techniques

extracting soft information from the source. Finally, in Chapter 2 we will optimize the ISCD process.

In the traditional UEP schemes conceived for layered video communication, variable-rate FEC schemes were invoked for the different layers [92, 93, 94, 95, 96, 73, 97, 98]. By contrast, the recent contributions of [87, 90, 73, 91, 98] explored the dependencies amongst the layers and conceived UEP schemes by exploiting this sophisticated feature. Specifically, the unsuccessful decoding of the BL resulted in instructing the video decoder to discard all the ELs depending on it, regardless whether they have or have not been successfully decoded. Naturally, this course of action wasted the transmit power assigned to the dependent layers. Thus a bit-level inter-layer coded FEC (IL-FEC) scheme [98] that embeds the BL into the FEC coded ELs has been introduced in Section 1.4, so that the reception of the BL can be improved with the aid of the ELs using soft decoding.

Additionally, in Chapter 3 we will conceive an adaptive THARQ (ATHARQ) transmission scheme in support of IL-FEC coded layered video for minimizing the video distortion under the constraint of a given total number of transmission time slots. In the original treatise on IL-FEC [98], the transmission environment of THARQ was not considered. Furthermore, the merits of IL-FEC schemes have not been investigated in the context of THARQ transmission schemes. However, the packet scheduling schemes should be carefully designed by ensuring that instead of the sequential packet transmissions assumed in [98], we exploit the specific characteristics of each IL-FEC coded packet. Furthermore, a method of on-line optimization will be developed for our IL-ATHARQ transmission scheme, in order to find the most appropriate FEC code rate distribution among the video layers for the sake of reducing the video distortion. Our technique proposed in Chapter 3 is significantly different from the existing contributions, such as the packet-level framework of [61]. More specifically, most contributions on packet-level UEP [61] operate at the packet-level and consider the Binary Erasure Channel (BEC). Furthermore, packet-level UEP employs hard-decoded codes, such as the family of (N, K) block codes. By contrast, IL-FEC typically relies on bit-level FEC decoders using soft decoding, such as a Recursive Systematic Convolutional (RSC) code. Furthermore, our focus is on wireless channels.

Despite the plethora of studies on optimizing the CSK constellation for reducing its error rate, no efforts have been invested in investigating the feasibility of CSK-aided UEP schemes. Hence we conceive a new CSK modulation scheme, which is capable of providing various levels of error probability for the different layers of signals. Hence we conceive a hierarchical CSK (HCSK) modulation scheme capable of generating inter-dependent layers of signals associated with different error probabilities. Finally, we will demonstrate the benefits of HCSK by transmitting scalable video sources.

1.7 Outline of the Thesis

Having briefly reviewed the literature of video coding, as well as of unequal error protection [200] and of the iterative source-channel decoding philosophies, let us now outline the organization of the thesis, which is displayed in Figure 1.5. The rest of the chapters are organized as follows:

- **Chapter 2: Markov Random Field Model Assisted Joint Source-channel Decoding**

  As outlined in Figure 1.5, in this chapter we will introduce a joint source-channel decoding system based on the Markov random field model. In Section 2.2 we will introduce our video codec and the structure of the transmitter. Section 2.3 presents the algorithms conceived for all the decoding modules and then the corresponding JSCC algorithm is detailed in the context of loss-less video compression, followed by our 3D EXIT chart results in Section 2.4.2. The performance of our system operating both with and without FD is compared to our benchmarkers in Section 4.5. Finally, we offer our conclusions in Section 2.6.
- **Chapter 3: Adaptive Truncated HARQ Aided Layered Video Streaming Relying On Inter-Layer FEC Coding**

  As illustrated in Figure 1.5, in Chapter 3 we conceive an adaptive THARQ

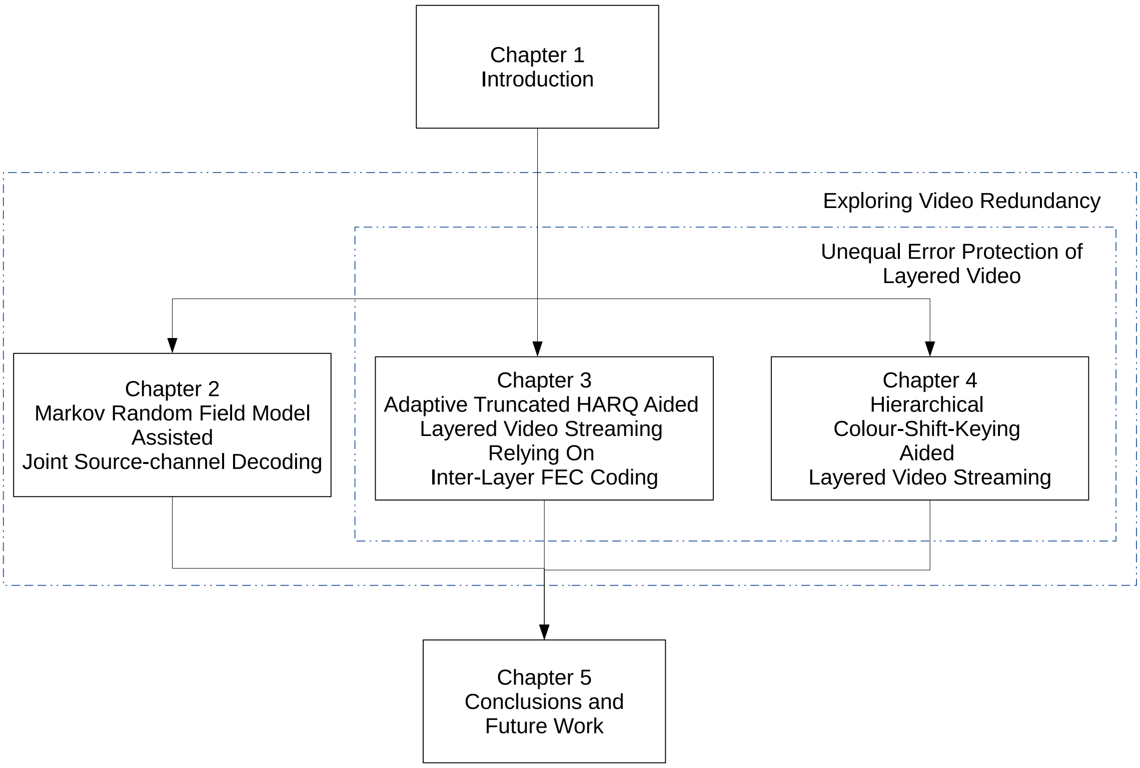


Figure 1.5: Outline of the thesis

(ATHARQ) algorithm for IL-FEC coded layered video streaming for the sake of minimizing the video distortion under the constraint of a given total number of transmission TSs. Section 3.2 details the IL-FEC transmitter and receiver model, as well as the proposed ATHARQ protocol along with the benchmark schemes we used in this treatise. The algorithm of our IL-ATHARQ retransmission controller is described in Section 3.3, followed by the details of coding-rate optimization conceived for our IL-ATHARQ system in Section 3.4. The performance of our IL-ATHARQ scheme as well as of the rate-optimized IL-ATHARQ scheme using a RSC codec are compared to the benchmarkers in Section 3.5 using different video sequences, followed by characterizing both the effects of the delay as well as well of the channel quality prediction errors on the attainable system performance. Finally, we conclude in Section 3.6.

- **Chapter 4: Hierarchical Colour-Shift-Keying Aided Layered Video Streaming**

  As shown in Figure 1.5, in Chapter 4 we conceive a HCSK modulation constellation that can be readily configured to provide different error probabilities for the different video layers. Section 4.2 introduces the CSK transmitter and receiver model, including the signal constellation and coding, the optical-domain channel model as well as the soft demodulation method. We then introduce the construction of our novel M-ary hierarchical colour-shift keying (M-HCSK) modulation and its layered demodulation method in Section 4.3. We continue by optimizing our UEP system using M-HCSK-RSC transmission to convey scalable video sequences, as detailed in Section 4.4. The performance of our optimised M-HCSK-RSC video transmission system using different HCSK constellation sizes is compared to the relevant benchmarker schemes in Section 4.5 using different video sequences. Finally, we conclude in Section 4.6.

- **Chapter 5: Conclusions and Future Work**
  As outlined in Figure 1.5, in this concluding chapter, the major findings of our work are summarized along with our suggestions for future research.

1.8 Novel Contributions

The thesis is based on the publications and manuscript submissions of [201, 52, 124]. The novel contribution of this thesis includes the following:

- We proposed a video codec suitable for lossless video compression and ISCD. We further developed the soft source decoder of [43] for the transmission of video sequences rather than still video pictures. More explicitly, we used low complexity frame-differencing (FD) based compression for removing the temporal redundancy and a VrLC for removing the residual spatial redundancy of the FD signal, while exploiting the correlations amongst the FD pixels within the current frame with the aid of our MRF model based SISO module. A novel three-stage ISCD structure is proposed, which outperforms the two-stage architecture. Furthermore, we examined the convergence of the three-stage iterative decoding process using 3D extrinsic information transfer (EXIT) charts [201, 52].
- We intrinsically amalgamated IL-FEC coding with the THARQ-aided transmission of layered video. We conceived an ATHARQ transmission scheme for adaptively scheduling the IL-FEC coded video layer packets for minimizing the video distortion under the constraint of a certain total number of transmission time slots. We developed a method of on-line optimization for our IL-ATHARQ transmission scheme, in order to find the optimal FEC code rate distribution, appropriately sharing the redundancy among the video layers that results in a reduced video distortion. Quantitatively, about 2.5 dB of PSNR video quality improvement may be observed at an E_b/N_0 of 15 dB over the traditional THARQ benchmarker. Alternatively, we will demonstrate that an E_b/N_0 reduction of about 5.3 dB can be achieved at a PSNR of 38.5 dB [124].
- We conceived a HCSK modulation constellation that can be readily configured to provide different error probabilities for the different video layers. We were the first to introduce HM into the conventional CSK constellations [193, 202, 195, 196, 203] and to design an UEP scheme based on it. We provided a HCSK design example for transmitting scalable video sources relying on RSC codes as our FEC scheme. Our optimisation metric was the quality of the received

video. We demonstrated that the proposed system outperforms the traditional EEP scheme by about 1.7 dB of optical SNR at a video peak signal-to-noise ratio (PSNR) of 37 dB and by up to 6.5 dB at a PSNR of 36 dB.

Chapter 2

Markov Random Field Model Assisted Joint Source-Channel Decoding

2.1 Introduction

As reviewed in Section 1.1, the state-of-the-art video codecs [2, 10] used in wireless tranceivers tend to rely on Shannon's source and channel-coding separation theorem [38]. However, they often fail to remove all the redundancy of the video source, hence they typically exhibit a residual correlation. Furthermore, Shannon's source-and-channel-coding theorem is only valid for idealized Gaussian channels and for a potentially infinite encoding/decoding delay as well as complexity, which is unaffordable in practical applications. As reviewed in Section 1.2, joint source-channel coding (JSCC) [54] was proposed for wireless scenarios, where the residual redundancy of the video source and the intentional redundancy of the channel codecs was jointly exploited for increasing the error resilience of the system. More specifically, soft bit source decoding (SBSD) was proposed in [41] , while iterative source and channel decoding (ISCD) was advocated in [44]. Both of these techniques were designed for exploiting the *a posteriori* probabilities (APPs) of the source parameters using a Markov process model. Furthermore, ISCD exchanged extrinsic information between the source decoder and channel decoder in a turbo-like iterative decoding process.

Variable Length Codes (VrLCs) constitute a family of lossless source compression schemes and ISCD has been adopted in a number of VrLC-aided systems as well. The authors of [55] proposed an ISCD scheme and the bit-based trellis structure aided VrLC decoder of [204]. By contrast, a reversible VrLC was invoked as the outer code and a convolutional code was utilized as the inner code in [205, 206]. In [45], Kliewer and Thobaben exploited the residual redundancy after source encoding using a symbol-based soft-input APP decoder for packetized variable-length encoded correlated source signals and demonstrated a substantial error protection improvement with the aid of this iterative decoding technique. The authors of [207] proposed a novel Irregular VrLC (IrVrLC) scheme for near-capacity JSCC and achieved an improved performance in their serially concatenated system upon combining IrVrLC with Trellis Coded Modulation (TCM) by invoking iterative decoding. The authors of [208] also adopted the IrVrLC scheme, but this time the IrVrLC was concatenated with a rate-1 precoded fast frequency-hopping (FFH) M-ary frequency-shift keying (MFSK) scheme. Three-stage iterative decoding was adopted at the decoder, where extrinsic information was exchanged among the demodulator, the inner decoder and the outer decoder.

When applied to image or video transmission, the ISCD scheme requires sophisticated modelling of the source. In [209], a source decoder and maximum *a posteriori* probability (MAP) channel decoder was applied for jointly decoding the motion vectors of H.264 [10] coded video streams. Kliewer, Görtz and Mertins [43] exploited the *a priori* information inherent in images and invoked both MAP image estimation using Markov random field theory as well as the Markov–Gibbs correspondence relationship discovered in [57]. More explicitly, they modelled images using a Markov Random Field (MRF) for the sake of generating the APPs of the pixels by exploiting the correlation of the adjacent ones and they designed a soft-in-soft-out (SISO) module for exchanging extrinsic information with the adjacent modules in order to perform ISCD.

Against this background, in this chapter we propose a video codec suitable for lossless video compression and ISCD. Specifically, we further developed the soft source decoder of [43] for the transmission of video sequences rather than still video pictures. More explicitly, we used low complexity frame-differencing (FD) for removing the temporal-domain inter-frame redundancy. The low-complexity FD technique does not rely on power-thirsty motion-vector-based motion-compensation and as its other benefit, it does not require the transmission of motion vectors. However, this results in more residual redundancy than the more complex motion-vector based motion-compensation. Furthermore, VrLC is used for removing the residual spatial redundancy of the FD signal, while exploiting the correlations amongst the FD pixels within the current frame with the aid of our MRF model based SISO module. By adopting the MRF for modelling of the video pixels, we can obtain maximum *a posteriori* (MAP) image estimates from the 2-D spatial redundancy between the video pixels, while simpler models like the Markov chain can only explore the 1-D spatial redundancy. Although the estimation of the MRF parameters is difficult compared to the 1-D Markov models, we adopted the estimation methods in [43], which is shown to be effective. Moreover, a novel three-stage ISCD structure is proposed, which outperforms the two-stage architecture. Furthermore, we examined the convergence of the three-stage iterative decoding process using 3D extrinsic information transfer (EXIT) charts.

The rest of this chapter is organized as follows. In Section 2.2, we introduce our video codec and the structure of the transmitter. Section 2.3 presents the algorithms conceived for all the decoding modules and then the corresponding JSCC algorithm is detailed in the context of loss-less video compression, followed by our 3D EXIT chart results. The performance of our system operating both with and without FD is compared to our benchmarks in Section 2.5. Finally, we offer our conclusions in Section 2.6.

2.2 System Structure

The structure of the proposed system is shown in Fig. 2.1. First we consider an uncompressed video sequence consisting of N frames, represented in color YUV or RGB video format. Each sub-sequence is constituted by a single color component of the video (e.g. the Y component in a YUV format video), which is then processed by the system of Fig. 2.1 having color-specific parameter configurations. The n^{th} frame $\boldsymbol{F}_n$ of the single-color video sequence is considered here as an example, assuming $1 \leq n < N$, where $\boldsymbol{F}_n$ contains $(R \times C)$ pixels. Each frame is divided into a series of r-by-c sized macroblocks (MBs) (r and c divide R and C respectively), and each macroblock is scanned into a one-dimensional vector $\boldsymbol{U}_l^n = \left(u_{l,1}^n, ..., u_{l,M}^n\right)$, which contains $M = (r \times c)$ pixels. The vector index l obeys $0 \leq l < L$, where L is the total number of sub-blocks in $\boldsymbol{F}_n$, i.e. we have $L = (R \times C)/M$. Here $u_{l,m}^n$ is the random variable that represents a single source symbol / pixel value and its realization i_U belongs to the legitimate symbol / pixel set $\mathbf{I}_U = \{0, 1, \cdots, 2^K - 1\}$. Moreover, $u_{l,m}^n$ may also be represented by its corresponding bit-vector $\boldsymbol{u}_{l,m}^n$ consisting of K bits.

Most of the inter-frame redundancy may be eliminated by subtracting $\hat{\boldsymbol{F}}_n$, namely the prediction of $\boldsymbol{F}_n$, from $\boldsymbol{F}_n$. Then the frame-difference has to be quantized, VrLC-encoded and transmitted, as seen in Fig. 2.1. Firstly, the stored vectors $\tilde{\boldsymbol{U}}_{l=\{0,1,\cdots,L\}}^{n-1}$ generated from the previously reconstructed $(n-1)^{st}$ frame $\tilde{\boldsymbol{F}}_{n-1}$ are used for predicting $\boldsymbol{F}_n$. Although here we used simple frame-differencing, motion-vector-based motion-compensation may also be used, which is the norm in most standardized video codecs [2]. Then the corresponding vector $\hat{\boldsymbol{U}}_l^n = \left(\hat{u}_{l,1}^n, ..., \hat{u}_{l,M}^n\right)$ is extracted from the predicted frame $\hat{\boldsymbol{F}}_n$ and subtracted from the current vector $\boldsymbol{U}_l^n$, as seen in Fig. 2.1, in order to generate the FD vector $\boldsymbol{E}_l^n = \left(e_{l,1}^n, \cdots, e_{l,M}^n\right)$. The FD element $e_{l,m}^n$ can also be directly mapped into the $(K+1)$ bit vector $\boldsymbol{e}_{l,m}^n$. The specific realization i_E of each $e_{l,m}^n = u_{l,m}^n - \hat{u}_{l,m}^n$ value belongs to the set $\mathbf{I}_E = \{1 - 2^K, \cdots, -1, 0, 1, \cdots, 2^K - 1\}$. The FD $\boldsymbol{E}_l^n$ is then quantized by the quantizer Q of Fig. 2.1 and its quantized version $\boldsymbol{W}_l^n$ is added to $\hat{\boldsymbol{U}}_l^n$ for reconstructing $\boldsymbol{U}_l^n$. The reconstructed version $\tilde{\boldsymbol{U}}_l^n$ is then stored and will be used for the prediction of

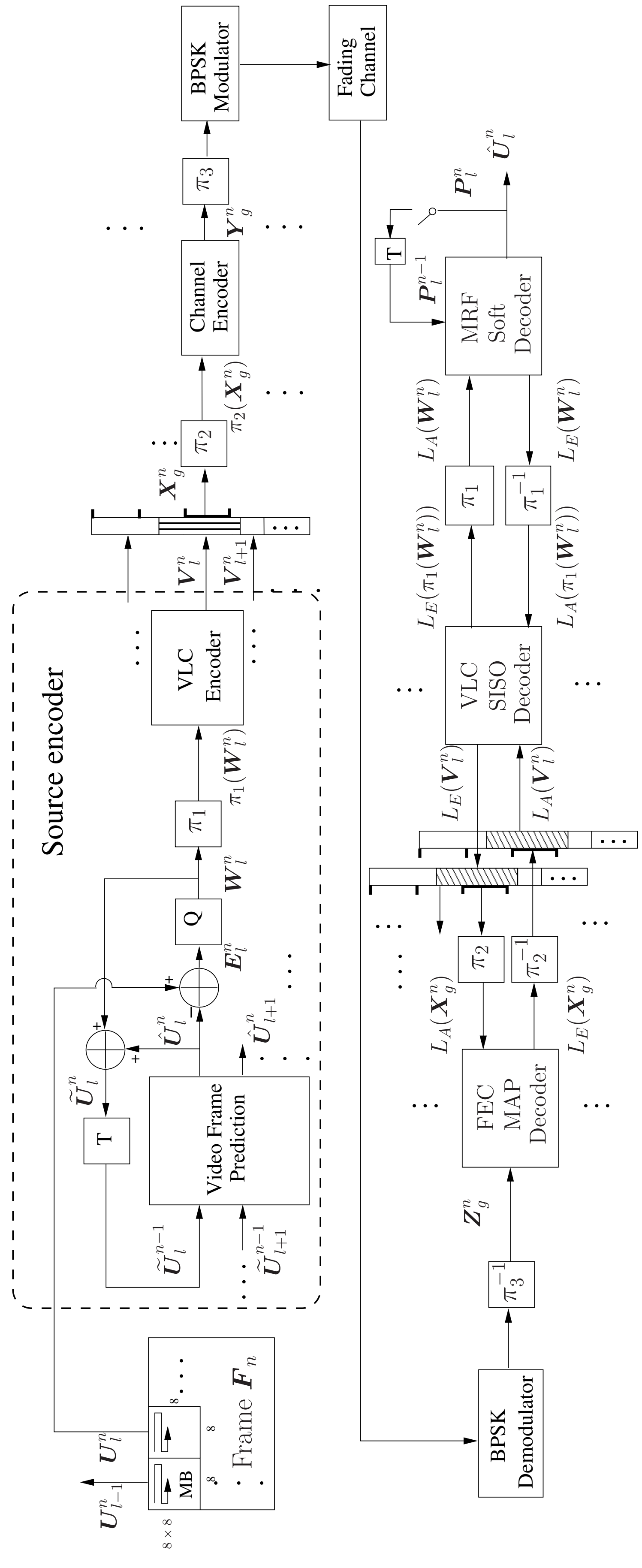


Figure 2.1: The proposed video encoder and the three-stage joint source-channel decoding receiver

the next video frame. For the sake of avoiding severe error propagation due to the differential structure of the video encoder, $\tilde{\boldsymbol{U}}^{n-1}_{l=\{0,1,\cdots,L\}}$ are periodically reset in order to allow the insertion of intra-frame-coded frames, where the number of consecutive inter-frame-coded frames in a specific group of pictures is denoted by N_{ip}. In the intra-coded scenario, where each frame is encoded separately, the elements of the vectors $\tilde{\boldsymbol{U}}^{n-1}_{l=\{0,1,\cdots,L\}}$ are set to a constant.

The output of the quantizer Q is denoted by $\boldsymbol{W}^n_l = \left(w^n_{l,1}, ..., w^n_{l,M}\right)$, where each vector element results from the quantization of the m^{th} FD pixel, and $\boldsymbol{W}^n_l$ is interleaved on a symbol-by-symbol basis [210] before being encoded by the VrLC encoder. The output bit vector of the VrLC encoder is denoted by $\boldsymbol{V}^n_l$, which by definition has a variable length of J^n_l. We assume that the vector lengths J^n_l and the VrLC codebooks are 'known' at the receiver. Before FEC coding, we first store the vectors $\boldsymbol{B}^n_l$ in a buffer. Then the interleaver π_2 of Fig. 2.1 having an input length of L_{π_2} is applied. Each time the interleaver π_2 reads L_{π_2} bits from the VrLC encoder's output buffer, which are represented by the vector $\boldsymbol{X}^n_g$, $1 \leq g \leq G = \left\lceil \left(\sum_{l=1}^{L} J^n_l\right) / L_{\pi_2} \right\rceil$. The scrambled version of its input is passed to the FEC channel encoder of Fig. 2.1, yielding the encoded vector $\boldsymbol{Y}^n_g$ of Fig. 2.1, which has a length of L_{π_3}. Finally this vector is then interleaved by the channel interleaver π_3 before transmission over the channel. We listed the parameters to be used in this chapter in Table 2.1.

2.3 MRF-Based SISO Source Decoding

In this section we will introduce our soft source decoder architecture invoked for exploiting the spatial correlations amongst the neighboring video or FD pixels and detail its decoding algorithm.

2.3.1 MRF-Based Soft Image Decoder

In this chapter we consider lossless VrLC source encoding, which implies that the quantizer Q of Fig. 2.1 imposes no quantization distortion. Moreover, the following derivation may also be used for lossy encoding with minor amendments. We commence by formulating the decoding algorithm for frame-differencing based video source encoding, noting that the decoding method of its intra-frame-encoded counterpart may be readily obtained as a special case. For our low-complexity frame-differencing technique we have $\hat{\boldsymbol{U}}^n_l = \tilde{\boldsymbol{U}}^{n-1}_l$.

Given $e^n_{l,m} = u^n_{l,m} - \hat{u}^n_{l,m} = u^n_{l,m} - \tilde{u}^{n-1}_{l,m}$, we have

$$P_{apri}\left(u^n_{l,m} = i_U\right) = \sum_{\substack{i_U \in \mathbf{I}_U \\ i_E \in \mathbf{I}_E}} P\left(\tilde{u}^{n-1}_{l,m} = i_U - i_E\right) P_{apri}\left(e^n_{l,m} = i_E\right) \tag{2.1}$$

	Description
N	The number of frames consisted in the uncompressed video sequence
n	The video frame index
N_{ip}	The number of consecutive inter-frame-coded frames in a specific group of pictures
$\boldsymbol{F}_n$	The n^{th} frame in the uncompressed video sequence
R, C	The vertical and horizontal pixel-size of a video frame
r, c	The vertical and horizontal pixel-size of a macroblock
L	The total number of sub-blocks in $\boldsymbol{F}_n$
$\boldsymbol{U}_l^n$	The one-dimensional vector scanned from the l th macroblock from the n th frame
$u_{l,m}^n$	The random variable that represents the m th source symbol / pixel value from $\boldsymbol{U}_l^n$
i_U	One realization of $u_{l,m}^n$, which belongs to the legitimate symbol / pixel set $\mathbf{I}_U = \{0, 1, \cdots, 2^K - 1\}$
K	The number of bits used to represent i_U
$\hat{\boldsymbol{F}}_n$	The prediction of $\boldsymbol{F}_n$
$\hat{\boldsymbol{U}}_l^n$	The vector extracted from the l th macroblock of the predicted frame $\hat{\boldsymbol{F}}_n$, where $\hat{\boldsymbol{U}}_l^n = \left(\hat{u}_{l,1}^n, ..., \hat{u}_{l,M}^n\right)$
$\tilde{\boldsymbol{U}}_l^{n-1}$	The vector generated from the previously reconstructed $(n-1)^{st}$ frame $\tilde{\boldsymbol{F}}_{n-1}$
$\boldsymbol{E}_l^n$	The FD vector
$e_{l,m}^n$	The m th element of $\boldsymbol{E}_l^n$
$\boldsymbol{W}_l^n$	The quantized version of $\boldsymbol{E}_l^n$
$\boldsymbol{V}_l^n$	The output bit vector of the VrLC encoder
J_l^n	The vector length of $\boldsymbol{V}_l^n$
$\boldsymbol{X}_g^n$	The vector of L_{π_2} bits from the VrLC encoder's output buffer
$\boldsymbol{Y}_g^n$	The encoded vector of interleaved $\boldsymbol{X}_g^n$, which has a length of L_{π_3}
$T_{MRF}\left(\cdot\right)$	The transfer function of $I_{E,MRF}$ vs $I_{A,MRF}$
T	The 'temperature' MRF parameter
δ	The exponent in the potential function of Eq. (2.7)
N_{MRF}	The number of inner iteration loops used by the MRF soft decoder
I_{inner}	The number of inner iteration loops between the FEC decoder and the VLC decoder
I_{outer}	The number of inner iteration loops between the MRF decoder and the VLC decoder
L_{c_av}	The average VrLC codeword length
R_s	The coding rate of the source encoder
C_R	The compression ratio of the source encoder, where $C_R = 1/R_s$
R_n	The 'natural' code rate (NCR) of the uncompressed video
R_c	The channel coding rate
R_{total}	The total coding rate in the system depicted in Fig. 2.1

Table 2.1: Parameters used in Chapter 2.

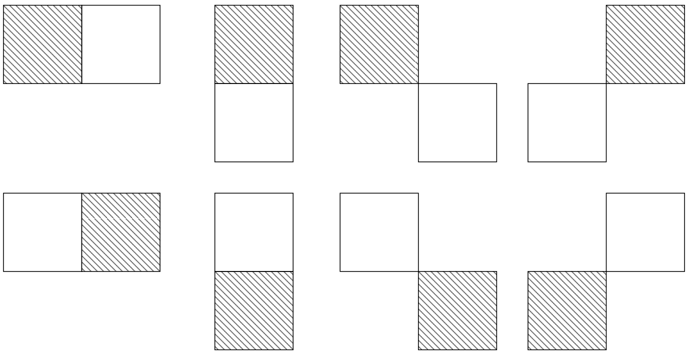


Figure 2.2: The cliques used in the MRF soft source coding of the proposed system, the shaded boxes represent the current pixel and the other one represents the neighboring pixels in the cliques

and

$$P_{apos}\left(e^{n}_{l,m}=i_E\right)=\sum_{\substack{i_U\in\mathbf{I}_U\\ i_E\in\mathbf{I}_E}} P\left(\tilde{u}^{n-1}_{l,m}=i_U-i_E\right)P_{apos}\left(u^{n}_{l,m}=i_U\right). \tag{2.2}$$

This SISO source decoding module may either exchange the symbol probabilities or the log-likelihood ratios (LLRs) of each bits with the inner modules in the concatenated system of Fig. 2.1. Let us denote the k^{th} bit of the FD signal's bit-representation $\boldsymbol{e}^{n}_{l,m}$ by $\boldsymbol{e}(k)$. Similarly, we can define the vector $\boldsymbol{i}_E$ as the realization of $\boldsymbol{e}^{n}_{l,m}$ and its k^{th} bit as $\boldsymbol{i}_E(k)$. If provided with the *a priori* LLR of $\boldsymbol{E}^{n}_{l}$ denoted by $L_{apri}\left(\boldsymbol{E}^{n}_{l}\right)$, we obtain the *a priori* probability (APrP) of $P_{apri}\left(e^{n}_{l,m}=i_E\right)$ in the form of:

$$P_{apri}\left(e^{n}_{l,m}=i_E\right)=\bar{c}^{n}_{l,m}\cdot\exp\sum_{k=0}^{K}\frac{\boldsymbol{i}_E(k)}{2}\cdot L_{apri}\left[\boldsymbol{e}(k)\right]. \tag{2.3}$$

On the other hand, if the *a posteriori* probability (APP) of $e^{n}_{l,m}$ is provided, the APP LLRs of $\boldsymbol{E}^{n}_{l}$ may be readily obtained as:

$$L_{apos}\left(\boldsymbol{e}(k)\right)=\ln\left[\frac{\sum_{\forall i_E\in\mathbf{I}_E:\boldsymbol{e}(k)=0}P_{apos}\left(e^{n}_{l,m}=i_E\right)}{\sum_{\forall i_E\in\mathbf{I}_E:\boldsymbol{e}(k)=1}P_{apos}\left(e^{n}_{l,m}=i_E\right)}\right]. \tag{2.4}$$

Here we briefly describe the soft decoding process of [43] using the MRF model, where the APP of each video pixel value may be calculated by combining the soft information gleaned from the channel with the intrinsic correlations between the pixel itself and the surrounding pixels. Let us simply express the event $u^{n}_{l,m}=i_U$ as i_U.

Then the APP of each source pixel may be expressed as

$$\begin{aligned} P_{apos}\left(i_U\right) &= P\left[i_U \left|\mathcal{N}_{\tilde{i_U}}, P_{apri}\left(i_U\right)\right.\right] \\ &= \bar{c} \cdot P\left(i_U \left|\mathcal{N}_{\tilde{i_U}}\right.\right) \cdot P_{apri}\left(i_U\right), \end{aligned} \tag{2.5}$$

where $\mathcal{N}_{\tilde{i_U}}$ represents the 8-pixel neighbourhood set of the pixel $u^n_{l,m}$. However, in order to avoid an excessive complexity, we also defined a smaller subgroup of mutually dependent symbols, which we referred to as 'cliques' [43]. Then we proceed by examining the system using different cliques sizes for the sake of achieving the best possible performance. Based on our related investigations not included here for the sake of space economy, we adopted the size-2 cliques shown in Fig. 2.2, where each clique relates $u^n_{l,m}$ to a single adjacent pixel. The variable $P\left(i_U \left|\mathcal{N}_{\tilde{i_U}}\right.\right)$ represents the APP of $u^n_{l,m}$ modeled using the so-called Markov-Gibbs correspondence [57], which is generated from the adjacent pixels, and it is expressed as

$$P\left(i_U \left|\mathcal{N}_{i_U}\right.\right) = \frac{1}{Z} \exp\left[-\frac{1}{T}\sum_{c\in C} V_c\left(i_U, i'_U\right)\right], \tag{2.6}$$

where Z is a normalization factor, T is defined as the 'temperature' parameter, and $V_c\left(i_U, i'_U\right)$ is the potential function, which was expressed as

$$V_c\left(i_U, i'_U\right) = \left|i_U - i'_U\right|^{\delta} \tag{2.7}$$

for our specific case, where i'_U represents the value of a particular pixel other than $u^n_{l,m}$ in the given clique c. When summing up the potential functions formulated in Eq. (2.7) over the entire collection of cliques C, we arrive at the summation term seen in the exponent of Eq. (2.6), which is referred to as the energy function. The choice of the parameters T in Eq. (2.6) and δ of Eq. (2.7) is video-content-dependent and the method provided in [43] may be invoked for finding the optimal value, which will be detailed in Section 2.3.3.

2.3.2 Decoding Procedure of MRF-Based Decoder

Finally, we describe the decoding procedure of the MRF soft decoder, which may contain N_{MRF} number of inner iteration loops:

1. Use Eq. (2.1) and Eq. (2.3) for generating the APrPs of $\boldsymbol{U}^n_l$, and carry out the MAP estimation of the pixels.
2. Use Eq. (2.5) in conjunction with Eq. (2.6) and Eq. (2.7) to generate the APPs of $\boldsymbol{U}^n_l$.
3. Re-estimate the pixel values in $\boldsymbol{U}^n_l$ with the aid of the APPs using MAP estimation.

4. If the number of affordable iterations N_{MRF} is not exceeded, go to Step 2), otherwise proceed to 5.
5. Store the APPs of $\boldsymbol{U}_l^n$ as $P\left(\tilde{u}_{l,m}^{n-1}\right)$ to be used in the decoding of the next frame.
   (a) If the system-level iteration is completed, then we output the current estimate of $\boldsymbol{U}_l^n$s.
   (b) If the extrinsic probability of $\boldsymbol{E}_l^n$ is required, then we use Eq. (2.2) for generating the APPs of $\boldsymbol{E}_l^n$. As a result, we can generate the extrinsic probabilities of each element by using $P_{ext}\left(e_{l,m}^n = i_E\right) = \frac{P_{apos}\left(e_{l,m}^n = i_E\right)}{P_{apri}\left(e_{l,m}^n = i_E\right)}$
   (c) If the extrinsic LLRs of $\boldsymbol{e}_{l,m}^n$ are required, then Eq. (2.2) can be used for generating the APPs of $\boldsymbol{E}_l^n$. As a next step, the *a posteriori* LLRs of $\boldsymbol{e}_{l,m}^n$ can be obtained using Eq. (2.4). Finally, the extrinsic LLRs of each element can be obtained by using $L_{ext}\left[\boldsymbol{e}\left(k\right)\right] = L_{apos}\left[\boldsymbol{e}\left(k\right)\right] - L_{apri}\left[\boldsymbol{e}\left(k\right)\right]$.

2.3.3 Choice of Decoder Parameters

In order to find the optimal parameters for the MRF decoder, namely parameter T in Eq. (2.6) and δ of Eq. (2.7), we followed the regime described in [43]. Firstly, the mutual information (MI) $I_{A,MRF}$ between the input *a priori* LLR $L_{apri}\left(\boldsymbol{E}_l^n\right)$ and $\boldsymbol{E}_l^n$, as well as the MI $I_{E,MRF}$ between the extrinsic LLR $L_{ext}\left(\boldsymbol{E}_l^n\right)$ and $\boldsymbol{E}_l^n$ are measured according to different settings of Ts and δs. Therefore EXIT charts can be generated showing the value of $I_{E,MRF}$ versus $I_{A,MRF}$, as shown in Fig. 2.3, where the EXIT curves of two cases are provided, namely for $T = 2$, $\delta = 0.3$ and for $T = 2$, $\delta = 0.9$. We can observe from Fig. 2.3 that the EXIT curve starts from the point of 0 and can reach the top-right point of $I_{A,MRF} = I_{E,MRF} = 1$.

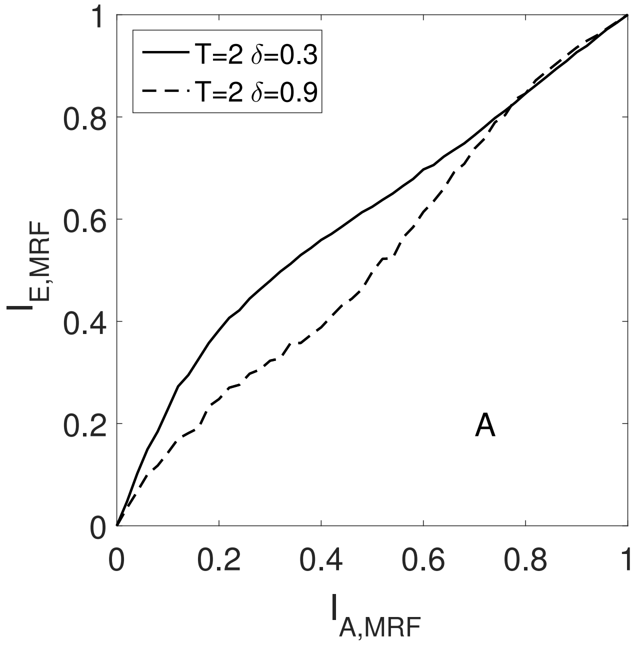


Figure 2.3: EXIT curves showing the value of $I_{E,MRF}$ versus $I_{A,MRF}$ of the MRF soft decoder of two cases, namely when $T = 2, \delta = 0.3$ and $T = 2, \delta = 0.9$.

According to [43], the error-correction capability of the MRF soft decoder coincides with the area underneath the EXIT curve showing the MI characteristic of the decoder. As shown in Fig. 2.3, different settings of T and δ values result in different curve shapes within the EXIT chart, thus enclose different areas, which can be formulated as

$$A\left(T,\delta\right) = \int_0^1 T_{MRF}\left(\varepsilon, T, \delta\right) d\varepsilon,$$

where $T_{MRF}\left(\cdot\right)$ denotes the $I_{E,MRF}$ vs $I_{A,MRF}$ function.

Therefore we scan the values of T and δ over the ranges of $(0, 20]$ using the step size of 1 and $(0, 1.5]$ using step size of 0.1, respectively. The corresponding value of $A\left(T,\delta\right)$ obtained from the output of the MRF soft decoder using the Y, U or V frames of the '*Akiyo*' sequence can be plotted in 3D-format as shown in Fig. 2.4 to Fig. 2.6. We can observe that $A\left(T,\delta\right)$ associated with $T = 2$ and $\delta = 0.3$ has the maximum value for the Y frames of the '*Akiyo*' sequence, while $T = 2$ and $\delta = 0.4$ is the case for the U and V frames. Therefore we chose the above settings for our simulations for optimized performance according to [43].

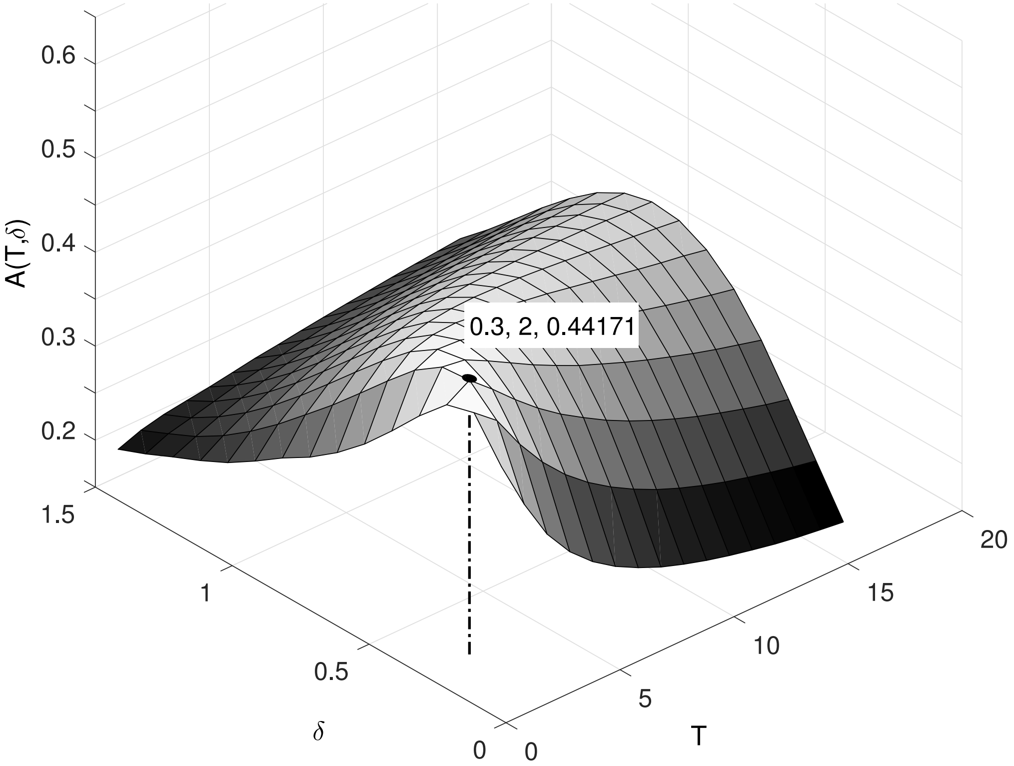


Figure 2.4: The value of $A(T,\delta)$ versus T and δ as the parameters of the MRF soft decoder. The Y frames of the '*Akiyo*' sequence were used.

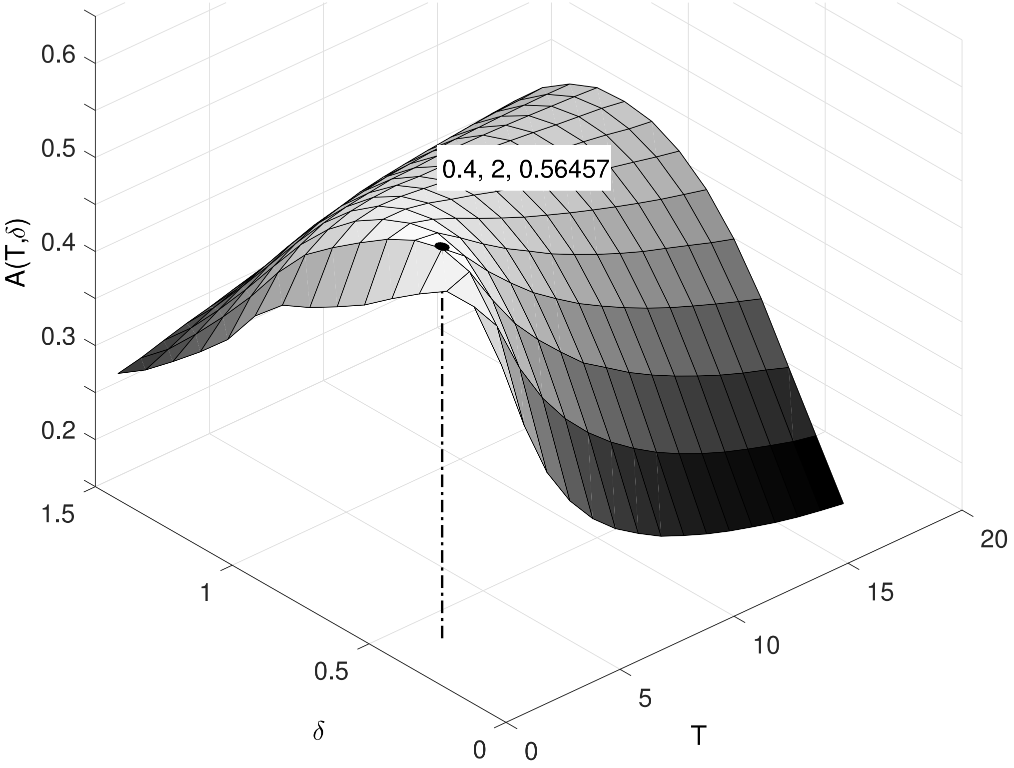


Figure 2.5: The value of $A(T,\delta)$ versus T and δ as the parameters of the MRF soft decoder. The U frames of the '*Akiyo*' sequence were used.

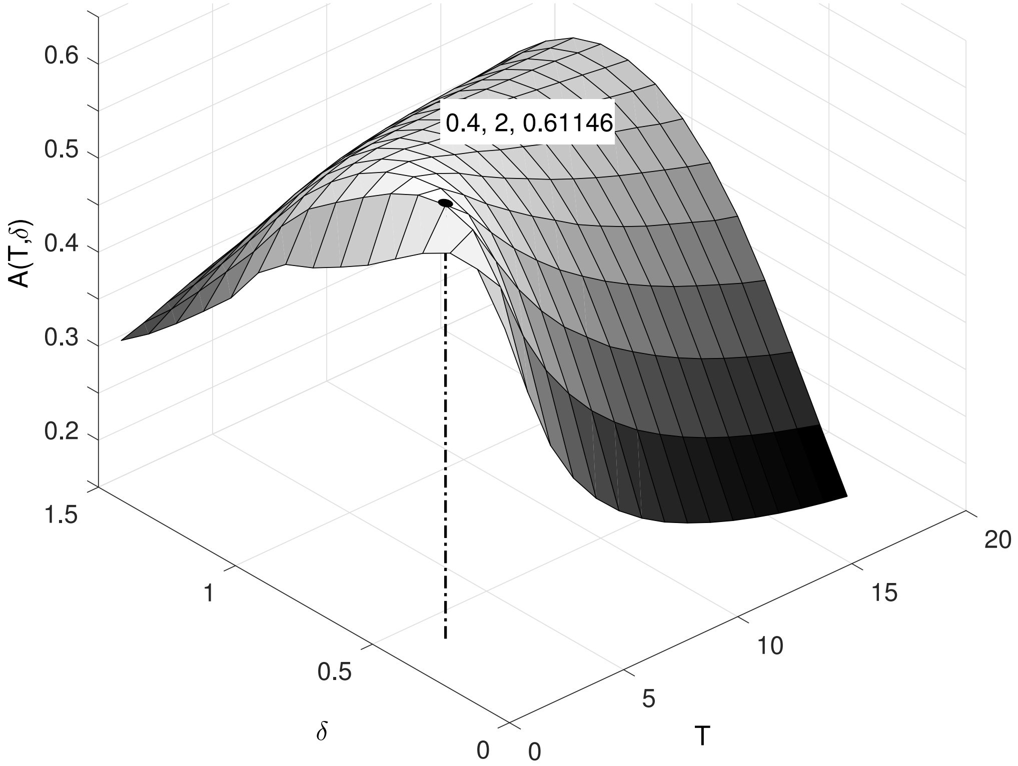


Figure 2.6: The value of $A\left(T,\delta\right)$ versus T and δ as the parameters of the MRF soft decoder. The V frames of the '*Akiyo*' sequence were used.

2.4 VrLC-compressed Iterative Joint Source-Channel Decoding Aided Video Transmission

2.4.1 VrLC Encoder

We use two classic VrLC codes, namely Huffman codes [211] and the reversible variable length code (RVrLC) of [205, 206] in our design examples. With the advent of sophisticated VrLC soft decoding algorithms [212], they constitute excellent candidates for characterizing our iterative decoding process, as it will be demonstrated in Section 2.4.2.

As described in Section 2.2 with reference to Fig. 2.1, the VrLC encoder takes its input symbols from the interleaved version of $\boldsymbol{W}_l^n$, which is the quantized version of $\boldsymbol{E}_l^n$. Naturally, the VrLC codec has to be trained for the sake of generating the required codebooks. Since each color component of the video stream is processed separately, the VrLC codec also has to be trained separately for each color component for the sake of determining the probability of occurrence for each legitimate symbol by observing $\boldsymbol{W}_l^n$, in order to generate the codebook as detailed in [2].

At this stage the redundancy exhibited by the correlation of the neighboring symbols is not removed, which however may be efficiently exploited for error-protection using the soft source decoder. More specifically, the VrLC encoder ignores the spatial correlation amongst the adjacent symbols and considers the symbol sequence as being i.i.d. The average VrLC codeword length L_{c_av} is lower-bounded by the 'entropy' of $\boldsymbol{W}_l^n$s, which is calculated as

$$H = -\sum_{w \in \mathbf{I}_E} P(w) \cdot \log_2 P(w), \tag{2.8}$$

where $P(\cdot)$ stands for the probability of occurrence for a specific symbol. First we define the coding rate R_s of the source encoder as the length of the bit-representation of the encoder's input divided by that of its output, which yields $R_s = 8/L_{c\_av}$ in our case. Then we may define the compression ratio of the source encoder as $C_R = 1/R_s$. Let us consider for example the Y component of the 300-frame, (176 × 144)-pixel quarter common intermediate format (QCIF) and 4:2:0 YUV represented 'Akiyo' video sequence as an example and evaluate the average VrLC-length as a function of the intra-frame-period N_{ip} in the top trace of Fig. 2.7. The intra-frame-encoding scheme constitutes a special case of the inter-frame-coded scheme which is associated with $N_{ip} = 1$. Since each source symbol is represented by $K = 8$ bits, its realization I_U is limited to the range of $\{0, 1, \ldots, 255\}$. As the default value stored in $\tilde{\boldsymbol{U}}_l^{n-1}$ is 128 when encoding the intra-coded frame, the system of Fig. 2.7 yields a differencing output within the range of $\{-128, \ldots, 0, \ldots, 127\}$. Accordingly, the legitimate range of the FD symbols' realization is spread over $\{-255, \ldots, 0, \ldots, 255\}$. The histogram portrayed in the lower half of Fig. 2.7 for comparing the $N_{ip} = 1$ and $N_{ip} = 20$ scenarios shows a highly peaked probability distribution for inter-frame-coding, also suggesting that more efficient compression is expected than for intra-frame coding, because the FD signal is close to zero with a high probability and the VrLC tends to be more efficient for high-peak PDF, rather than for near-uniform PDFs. It is shown at the top of Fig. 2.7 that L_{c_av} decreases rapidly, when N_{ip} is increased from 1 to 5 and continues to drop as N_{ip} further decreases, which indicates a lower compression ratio C_R because of the increased temporal distance of the intra-frame coded frames. According to the upper trace of Fig. 2.7, for the Y component we may only achieve $L_{c\_av} = 7.1$ and $R_s = 1.13$ for intra-frame encoding, while we have $L_{c\_av} = 2$ and $R_s = 4$ for $N_{ip} = 20$ in the inter-frame encoding case.

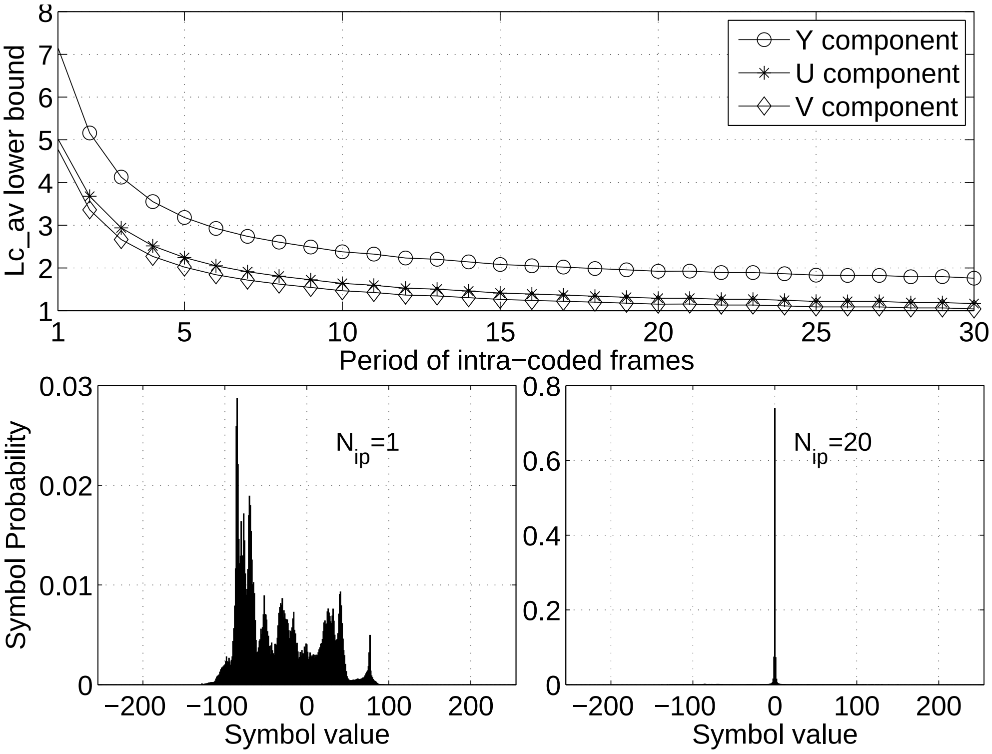


Figure 2.7: Lower-bound of the average codeword length of the YUV VrLC coded 'Akiyo' video sequences and the histogram characterizing the symbol occurrence probabilities of the Y color component, for $N_{ip} = 1$ and $N_{ip} = 20$ of intra-coded frame periods. For the sake of generating our treatment, we handled the intra-frame-coded scenario as part of the inter-frame-coded framework and differentiated it with the aid of $N_p = 1$. This is why the intra-coded pixels are spread over $[-128, \ldots, 127]$, as detailed in the text.

2.4.2 Three Stage Iterative Decoding Process

As shown in Fig. 2.1 and Fig. 2.8, the three-stage iterative decoder consists of three main SISO decoding modules, namely the soft channel decoder, the VrLC SISO

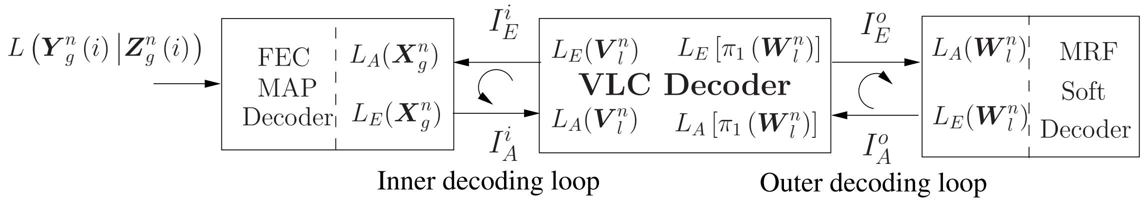


Figure 2.8: Decoding modules that constitute the three-stage iterative decoder, with the signal buffer and interleavers omitted for visual clarity.

decoder and the MRF-based soft source decoder. The decoding process includes two iterative decoding loops, which are the inner loop where the soft channel decoder exchanges its soft bit information estimates with the VrLC SISO decoder, and the outer loop where the VrLC SISO decoder exchanges the soft-information with the MRF-based source decoder in a iterative manner. Furthermore, the decoder operates on a frame-by-frame basis, which implies that both the channel decoder and the VrLC decoder have to complete processing the current video frame, before commencing the processing of the next frame. Finally, if the current frame is the FD frame, its decoding will rely on the APP result of the previous frame.

In the three-stage iterative decoding process, the channel decoder invokes the BCJR algorithm and forwards the extrinsic bit-LLR information $L_E(\boldsymbol{X}_g^n)$ to the VrLC SISO soft decoder, which is evaluated from the LLR $L\left(\boldsymbol{Y}_g^n(i)\,|\boldsymbol{Z}_g^n(i)\right)$, $0 < i < L_{\pi_3}$ of the de-interleaved channel output, while the *a priori* bit LLR $L_A(\boldsymbol{X}_g^n)$ is gleaned from the VrLC decoder, as seen in Fig. 2.1. The interleaver π_2 de-interleaves $L_E(\boldsymbol{X}_g^n)$ of Fig. 2.1 received from the channel decoder, and again, stores the results in a buffer for the VrLC decoder, which uses it as *a priori* information in the next phase after all the signal blocks $\boldsymbol{X}_{g=\{1,\cdots,G\}}^n$ of Fig. 2.1 have been processed. At the same time the interleaver π_2 reads its input from another buffer, which was previously loaded with the extrinsic output of the VrLC decoder and generates the interleaved version as $L_A(\boldsymbol{X}_g^n)$.

Furthermore, as shown in Fig. 2.8, the four-port VrLC SISO soft decoding module reads the *a priori* LLR of $\boldsymbol{V}_l^n$ from the frame buffer, and provides extrinsic LLRs for the FEC decoder module seen in Fig. 2.8 in order to participate in the inner iteration loop. As for the outer iterative loop, the VrLC SISO decoder of Fig. 2.8 exchanges its extrinsic information of $\boldsymbol{W}_l^n$s with the MRF SISO decoder. As for the decoding algorithm of the VrLC SISO decoder, a symbol-based trellis representation of the VrLC decoding process as well as the corresponding symbol-by-symbol MAP decoding method were detailed in [212]. However, Bauer and Hagenauer only used a VrLC as their outer code in the serially concatenated system, and the corresponding decoding module only exchanged soft information with the channel decoder. By contrast we adopted the concept of a SISO APP module [213] for further developing the method of [212] in order to arrange for the VrLC decoder to act as an intermediate code, which exchanges extrinsic symbol probabilities with the soft source decoder as well. Moreover, for the convenience of observing the convergence behavior of the system using EXIT charts, we also allow the exchange of extrinsic bit LLRs between the VrLC decoder and the MRF decoder. Hence, we use the natural mapping of the symbols in $\boldsymbol{W}_l^n$s to bits, and use an expression similar to Eq. (2.4) for directly generating the extrinsic bit LLRs from the extrinsic symbol probabilities.

Finally, we characterize the convergence behavior of the proposed three-stage iterative decoding process using the Akiyo sequence and Huffman VrLCs, as mentioned in Section 2.4.1. Let us denote the MI between $L_A(\boldsymbol{V}_l^n)$ and $\boldsymbol{V}_l^n$ by I_A^i, where the superscript i indicates that the information is exchanged in the inner decoding loop (o indicates the outer loop), and the subscript A indicates that the MI term represents the *a priori* information (E indicates the extrinsic information). Similarly, we can define I_E^i as the MI between $L_E\left(\boldsymbol{V}_l^n\right)$ and $\boldsymbol{V}_l^n$, I_A^o as the MI between $L_A\left[\pi_1\left(\boldsymbol{W}_l^n\right)\right]$ and $\pi_1\left(\boldsymbol{W}_l^n\right)$, while I_E^o as the MI between $L_E\left[\pi_1\left(\boldsymbol{W}_l^n\right)\right]$ and $\pi_1\left(\boldsymbol{W}_l^n\right)$. Since the FEC MAP decoder exploits the channel information $\boldsymbol{Z}_g^n$, its EXIT function inherently depends on E_b/N_0 and hence may be written as $I_A^i = f_{FEC}\left[I_E^i, E_b/N_0\right]$. By contrast, the inner VrLC SISO decoder's EXIT function may be formulated as $I_E^i = f_{VLC}^i\left[I_A^i, I_A^o\right]$, while the outer VrLC SISO decoder's EXIT function can be expressed as $I_E^o = f_{VLC}^o\left[I_A^i, I_A^o\right]$. Finally, the EXIT function of the soft source decoder may be formulated as $I_A^o = f_s^o\left[I_E^o\right]$.

The resultant 3D EXIT charts recorded at the E_b/N_0 value of 12.4 dB are presented in Fig. 2.9 and Fig. 2.10, while a Y frame of the 'Akiyo' sequence is used as the source for our example. The EXIT functions of the receiver components involved in the inner decoding loop are shown in Fig. 2.9, while Fig. 2.10 portrays the EXIT functions of the components involved in the outer decoding loop. We may observe from Fig. 2.10 that the EXIT curve of the MRF-based soft source decoder may not reach the (1,1) point in the I_A^o-I_E^o plane, thus the decoding trajectory is prevented from reaching the (1,1,1) point of perfect convergence in the 3D EXIT chart. However, the trajectory may still closely approach the (1,1) point in the I_A^i-I_E^o plane. It is shown in Fig. 2.10 that when we have $I_A^o = 0$, the EXIT curve of the FEC decoder and that of the Huffman VrLC decoder intersect at the point of (0.93,0.42) in the I_A^i-I_E^i plane, which means that the two-stage (VrLC-RSC) iterative decoder operating without the soft-source decoder may exhibit a residual error floor due to the poor codeword distances of the Huffman code. However, after four iterations using the three-stage decoding process of Fig. 2.9, the decoding trajectory reached the (0.98,0.91) point when $I_A^o = 0.68$, which suggests a substantial improvement in the BER performance.

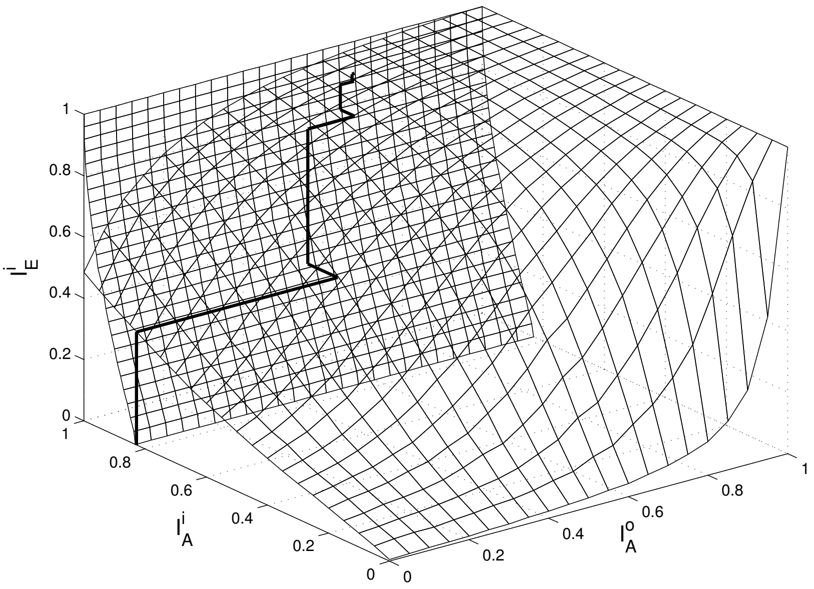


Figure 2.9: Inner decoding functions and decoding trajectory for an E_b/N_0 of 12.4 dB. The Y frame of the '*Akiyo*' sequence was used as our source, and a Huffman code was employed as the VrLC code. The schematic of Fig. 2.1 and the parameters of Table 2.1 were used.

2.5 Performance Results

2.5.1 Parameter Configurations

This section details the parameters used in our following simulations. For the video source, the first 50 frames of the (176 × 144)-pixel quarter common intermediate format (QCIF) 4:2:0 YUV representation based 'Akiyo' sequence is employed. Thus the frame size is given by $R = 176$ and $C = 144$ for the QCIF luminance-frame and $R = 88$ and $C = 72$ for the chroma frame. The macroblocks size used is $(r \times c) = (8 \times 8) = 64$, thus each luminance frame is divided into $22 \times 18 = 396$ macroblocks, while each QCIF chroma-frame is divided into $(11 \times 9) = 99$ macroblocks.

As described in Section 2.3, the search method provided in [43] is invoked for generating the optimal values for the MRF parameters T and δ of Eq. (2.6) and Eq. (2.7) according to the video sequence, but the details of the search process are omitted here for conciseness. As a result, $T = 2$ and $\delta = 0.3$ were chosen for the decoding of the QCIF luminance-frame, while $T = 2$ and $\delta = 0.4$ for the chroma-frames. The number of inner-loop iterations was $I_{MRF} = 1$ for the MRF soft source decoder throughout this Chapter.

As for the FEC, a 2/3-rate RSC encoder using the generator polynomials of $g_1 = 1011$, $g_2 = 1101$ is employed, which can be represented as $\mathrm{G} = [1, g_2/g_1]$, where g_1 denotes the feedback polynomial and g_2 is the feedforward polynomial. The

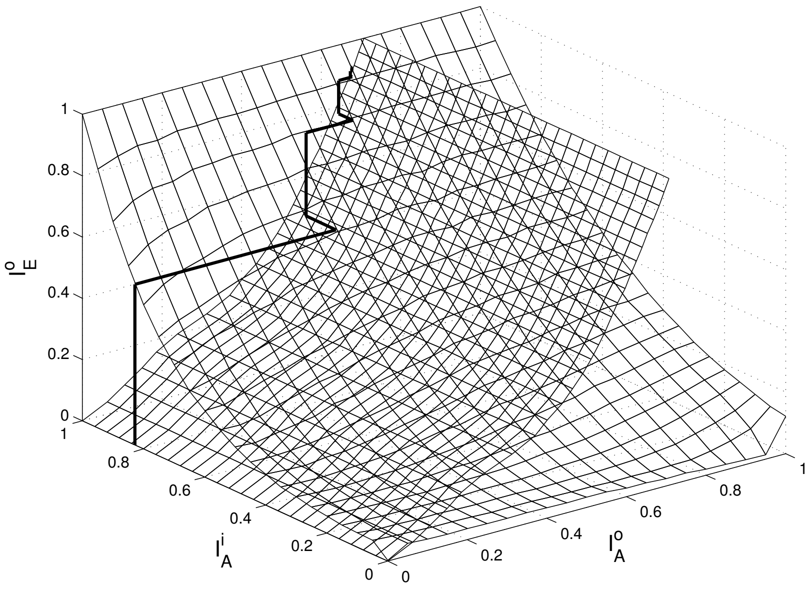


Figure 2.10: Outer decoding functions and decoding trajectory for an E_b/N_0 of 12.4 dB. The Y frame of the '*Akiyo*' sequence was used as our source, and a Huffman code was employed as the VrLC code. The schematic of Fig. 2.1 and the parameters of Table 2.1 were used.

corresponding puncturing pattern is $[1\,1; 1\,0]$, where the two rows correspond to the output of the systematic bit and that of the g_2, respectively. The interleaver length is $L_{\pi_2} = 512$ and random interleavers are used.

The BPSK modulated bits are transmitted over an uncorrelated non-dispersive Rayleigh channel. To specify the E_b/N_0 rario used in our simulations, first we have to define the coding rate of our system. Firstly, the concept of the 'natural' code rate (NCR) R_n [52] should be introduced. According to Shannon's channel capacity theorem [38] proposed for the transmission of i.i.d source bits, the true entropy of the video sequence should be considered when calculating the energy efficiency per bit. Let us denote the true entropy of the video source file by S_e, while the uncompressed video file has a size of S_r bits. Then the NCR becomes $R_n = S_e/S_r$, which may be physically interpreted as the code rate of a natural inherent channel code's redundancy. More explicitly, this redundancy is equivalent to an identical-rate external channel code's redundancy. This NCR will be exploited by our receiver. However, we have a predicament, since the true entropy of the video source cannot be readily evaluated. Therefore we generated the encoded version of the source sequence by the near-lossless coding mode of the H.264 codec [2, 10], which was utilized for approximating its entropy S_e. Since the compression ratio of the Akiyo clip described above using the H.264 codec is 8.7, the NCR R_n of the Akiyo clip becomes 1/8.7. Finally, in the simulations the E_b/N_0 (dB) value is calculated as $E_b/N_0 = 10\log_{10}\frac{E_s}{N_0 R_{\text{total}}}$, where R_{total} is defined as

$$R_{\text{total}} = R_n R_s R_c. \tag{2.9}$$

Thus the BER versus E_b/N_0 curves should be shifted to the right according to the redundancy inherent in the encoded sequence, as well as according to the artificial redundancy imposed by both channel coding and VrLC coding.

2.5.2 System Performance Results

In this section we present the simulation results characterizing our proposed scheme, which was introduced in Section 2.4.2 alongside a couple of benchmarkers, while using the parameter configurations described in Section 2.5.1 and $I_{inner} = I_{outer} = 4$ iterations for both the outer and inner iteration loops. Here two types of measurements are utilized for characterizing the attainable video quality, namely the PSNR versus E_b/N_0 curves and the bit error ratio (BER) versus E_b/N_0 curves. When measuring the PSNR values of the reconstructed video, we employ the technique used in the H.264 reference software JM and set the minimum value of the total averaged mean squared error (MSE) value between the reconstructed and the original frame to 1, so that encountering infinite PSNR values resulting from a perfectly reconstructed video frame can be avoided. Hence the maximum unimpaired video PSNR that may be obtained at the receiver is about 48.1 dB.

Here we introduce the two benchmarkers that are used for comparison with our proposed systems, namely the Lossless-H.264-RSC system and the MRF-RSC two-stage iterative decoding system of [43]. For our first benchmarker, the same RSC codec is used for transmitting and receiving the video compressed in the near-lossless mode of the JM software, as in our proposed system. For the latter one, we used the MRF-based soft decoder introduced in [43] in conjunction with $I = 4$ iterations, where simple natural mapping of the source symbols to bits was used, along with the optimal MRF parameters.

The BER versus E_b/N_0 performance of our scheme is presented in Fig. 2.11, while the PSNR versus E_b/N_0 performance is shown in Fig. 2.12. In order to choose the best system configurations, we carried out comparisons between the systems having different intra-coded frame periods, different VrLC codes and different methods of information exchange between the soft source decoder and the VrLC SISO decoder, which were described in Section 2.4.2. As seen in Fig. 2.11 and Fig. 2.12 the dashed lines represent the system relying on Huffman codes, while the solid lines indicate the ones using the RVrLC codes of [205, 206]. Since a RVrLC has a symmetric construction, which supports both forward and backwards decoding, it is typically less susceptible to error-propagation than Huffman VrLCs. Observe in Fig. 2.11 that the RVrLC codes result in a steeper BER curve and better BER as well as PSNR performance. For example, at a BER of 1×10^{-4}, the RVrLC aided system

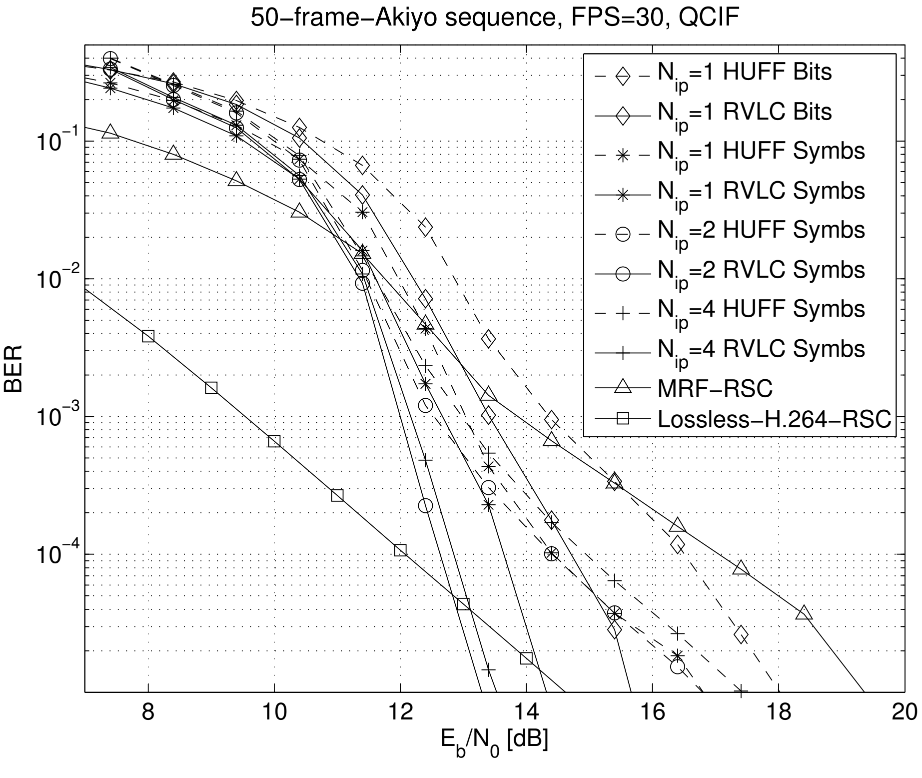


Figure 2.11: BER comparison of the MRF-VrLC-RSC system and of the Lossless-H.264-RSC and MRF-RSC [43] benchmarkers using the *Akiyo* sequence for transmission over uncorrelated Rayleigh channels. The schematic of Fig. 2.1 and the parameters of Table 2.1 were used.

outperforms the Huffman coded system by about 1 dB, 2 dB and 2.2 dB in terms of the E_b/N_0 required for N_{ip} = 1, 2, 4, respectively. Furthermore, the system that exchanges extrinsic symbol probabilities rather than extrinsic LLRs has a better performance when $N_{ip} = 1$, while the former one outperforms the latter one by about 1 dB and 2.1 dB when using RVrLC and Huffman codes, respectively. Finally, our comparisons between the systems having different N_{ip} values was also carried out, and the immediate conclusion emerged that $N_{ip} = 2$ provided a better performance than $N_{ip} = 1$ and $N_{ip} = 4$. This is because the low NCR of the uncompressed video ($N_{ip} = 1$) radically and excessively reduces the bit-energy. By contrast, the error-propagation effects become severe for $N_{ip} = 4$. Thus the system associated with $N_{ip} = 2$ and relying on the RVrLC code as well as on symbol-based information exchange within the inner decoding loop is preferred. It can also be seen in Fig. 2.11 and Fig. 2.12 that a BER of about 9×10^{-3} and a PSNR of about 44 dB can be achieved by our scheme at an E_b/N_0 of 11.4 dB in the best case.

By observing Fig. 2.11 we may see that our proposed scheme outperforms the MRF-RSC system by about 3.4 dB at a BER of 1×10^{-4}. Although the Lossless-H.264-RSC scheme achieves a better BER performance at low E_b/N_0 values, its bits are sensitive to transmission errors. It can be seen from Fig. 2.12 that at a PSNR of

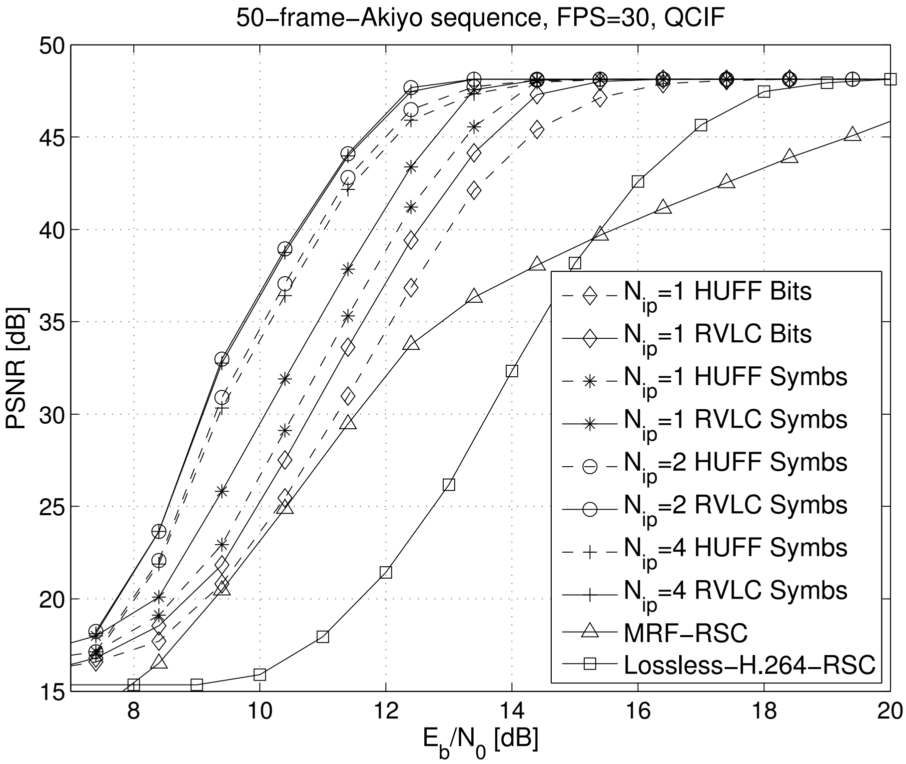


Figure 2.12: Reconstructed video quality of the MRF-VrLC-RSC system and of the Lossless-H.264-RSC and MRF-RSC [43] benchmarkers using the *Akiyo* sequence for transmission over uncorrelated Rayleigh channels. The schematic of Fig. 2.1 and the parameters of Table 2.1 were used.

Figure 2.13: A frame comparison of the decoded *Akiyo* sequence at $E_b/N_0 = 11.4$ dB. The frames are reconstructed by Lossless-H.264-RSC, MRF-RSC [43], MRF-VrLC-RSC($N_{ip} = 1$), MRF-VrLC-RSC($N_{ip} = 2$), respectively. The schematic of Fig. 2.1 and the parameters of Table 2.1 were used.

45 dB our scheme outperforms the MRF-RSC and the Lossless-H.264-RSC scheme by about 8 dB and 5.2 dB, respectively.

Finally, our subjective comparison of the decoded Akiyo sequence at $E_b/N_0 = 11.4$ dB is portrayed in Fig. 2.13, demonstrating that the proposed MRF-VrLC-RSC scheme is capable of substantially better recovering the error-infested video than the benchmarks.

Schemes	E_b/N_0 Gain
MRF-RSC	-2.8 dB
$N_{ip} = 1$ HUFF Bits	2.6 dB
$N_{ip} = 1$ RVrLC Bits	3.1 dB
$N_{ip} = 1$ HUFF Symbs	3.5 dB
$N_{ip} = 1$ RVrLC Symbs	4 dB

Schemes	E_b/N_0 Gain
$N_{ip} = 2$ HUFF Symbs	4.8 dB
$N_{ip} = 2$ RVrLC Symbs	5.2 dB
$N_{ip} = 4$ HUFF Symbs	4.7 dB
$N_{ip} = 4$ RVrLC Symbs	5.1 dB

Table 2.2: E_b/N_0 gains obtained by the different video transmission systems investigated in Fig. 2.12 compared to the Lossless-H.264-RSC scheme used as our benchmark at a PNSR of 45 dB.

2.6 Summary and Conclusions

In this chapter we proposed a video codec suitable for lossless video compression and ISCD.

Firstly, in Section 2.2 our video codec and the structure of the transmitter was introduced. In Section 2.3 we further developed the soft source decoder of [43] for the transmission of video sequences, rather than considering still video pictures. More explicitly, we used low complexity frame-differencing (FD) for removing the temporal redundancy and a VrLC for removing the residual spatial redundancy of the resultant FD signal, while exploiting the correlations amongst the FD pixels within the current frame with the aid of our MRF model based SISO module. In Section 2.4 we presented the algorithms conceived for all the decoding modules and then the corresponding JSCC algorithm was detailed in the context of loss-less video compression. In Section 2.4.2, the novel three-stage ISCD structure of Fig. 2.8 is proposed, which outperformed the two-stage architecture. Furthermore, in Section 2.5 we examined the convergence of the three-stage iterative decoding process using 3D extrinsic information transfer charts.

Finally, the performance of our system operating both with and without FD was compared to our benchmarkers in Fig. 2.11 and Fig. 2.12. Specifically, we demonstrated that our scheme outperformed the MRF-RSC and the Lossless-H.264-RSC schemes of Fig. 2.12 by about 8 dB and 5.2 dB at a PSNR of 45 dB, respectively. The detailed E_b/N_0 gains obtained by the different video transmission systems of Fig. 2.12 are listed in Table 2.2, compared to the Lossless-H.264-RSC scheme used as the benchmarker at a PNSR of 45 dB.

Chapter 3

Adaptive Truncated HARQ Aided Layered Video Streaming Relying On Inter-Layer FEC Coding

3.1 Introduction

The wireless channel is subject to impairments imposed both by the noise and fading. Thus the task of transmitting video contents to users equipped with various terminals is a challenging one both in terms of source coding and transmission techniques [2]. As reviewed in Section 1.1.3, layered video coding [28] is a widely used scheme conceived for handling this heterogeneous networking problem. By providing multiple layers of different importance, layered video coding is capable of supporting progressive reception of video steams, depending both on the prevalent channel conditions and on the hardware requirements of the individual users. More specifically, the most important layer is referred to as the BL, while the ELs are capable of providing additional video quality refinements during instances of higher channel qualities. Hence the popular video standards [31,11,10,32,33] are capable of supporting layered video coding, as reviewed individually in Section 1.1.3.

Further to the introduction carried out in Section 1.4 of Chapter 1, we further consider hybrid automatic retransmission request (HARQ) aided systems, which rely on the combination of two error correction mechanisms that are capable of improving the reliability of transmissions: automatic retransmission request (ARQ) and FEC, where the original signals are retransmitted upon requests, when the signals cannot be flawlessly decoded by the FEC decoders. In Type-I HARQ, the transmitter retransmits the original packet upon reception of a negative acknowledgement (NACK) feedback. In order to provide a more reliable decision concerning the original packet and to achieve a diversity gain, the best approach at the receiver is to combine the various corrupted retransmitted signals according to the maximal ratio combining (MRC) principle, which is carried out by adding the LLRs of several packet replicas. This approach is also referred to as Type-I HARQ relying on Chase Combining (CC) [214]. In Type_II HARQ, incremental redundancy (IR) generated from the original packet in form of additional parity bits is transmitted instead of the original packet upon receiving a NACK feedback. Finally, all the information is appropriately combined at the receiver. This scheme is often referred to as Type_II HARQ with IR.

Due to the delay-constraints of near-real-time video transmission systems, only the employment of truncated HARQ (THARQ), relying on a limited number of retransmissions is realistic. The energy efficiency of THARQ protocols designed for a single-user link or assisted by relay stations was considered in [215]. The closed-form analytical expressions of the achievable throughput, of the average packet delay and of the packet loss rate was provided in [216], where the maximization of the system throughput was also carried out. The performance analysis of a wireless network using adaptive modulation and coding combined with THARQ-CC at the data link was presented in [217]. The transmission of control messages using adaptive modulation and coding was considered in [218] in the scenario of voice over Internet Protocol (VoIP) services supported by THARQ. However, the associated video characteristics had not been addressed in the aforementioned THARQ schemes. As a further advance, a video transmission system was proposed and analysed in [219], which relied both on THARQ and selective combining, as well as on rate-compatible punctured convolutional (RCPC) codes for transmission over fading channels. A finite-state Markov model was used for representing the Rayleigh fading channels. An improved video quality was achieved by the proposed scheme at a limited delay. Layered video has been considered for transmission using HARQ schemes in either unicast or multicast scenarios [220,221,222,223,224]. The authors of [220] presented a theoretical analysis as well as rich experimental results for characterizing both unicast and multicast scenarios for transmission over packet-erasure channels, while the authors of [221, 222, 223, 224] provided solutions for multicast systems transmitting layered video using various HARQ schemes.

The transmission of layered video can be protected by UEP [79], as reviewed in Section 1.3. Amongst numerous recent works by other researchers on UEP, a bit-level inter-layer coded FEC (IL-FEC) scheme proposed in [98] embeds the BL into the FEC coded ELs, so that the reception of the BL can be improved with the aid of the ELs using soft decoding. Subsequently, a sophisticated on-line real-time video distortion estimation technique was conceived in [99], which is suitable for diverse channel conditions and system configurations. More explicitly, in [99] we proposed an on-line code rate optimization method for minimizing the video distortion.

A range of UEP schemes have been conceived for HARQ [65, 114, 63, 115, 116, 117, 118, 119, 120] in order to improve the video quality of layered videos. The authors of [65] proposed UEP by appropriately sharing the bitrate budget between the source and channel encoders based on either the minimum-distortion or on the minimum-power consumption criterion. In [118, 119], UEP was achieved by assigning each video layer a different retransmission limit. Another stream of contributions [65, 63, 115, 116, 117] adopted the so-called limited-retransmission based priority encoding transmission (PET) scheme [61], where UEP is achieved by varying the source block-length across the different source layers, while keeping the FEC-decoding block-length fixed. This allows the PET to have a packetization scheme that ensures that the source layers of an FEC-coded block are dropped according to their significance, commencing by dropping the least significant one first.

Against this background, in this chapter, we conceive an adaptive THARQ (ATHARQ) transmission scheme in support of IL-FEC coded layered video for minimizing the video distortion under the constraint of a given total number of transmission time slots. In our previous work [98], the transmission environment of THARQ was not considered. Furthermore, the merits of IL-FEC schemes have not been investigated in the context of THARQ transmission schemes. However, the packet scheduling schemes should be carefully designed by ensuring that instead of the sequential packet transmissions assumed in [98], we have to exploit the specific characteristics of each IL-FEC coded packet. Furthermore, we develop a method of on-line optimization for our IL-ATHARQ transmission scheme, in order to find the most appropriate FEC code rate distribution among the video layers that reduces the video distortion. Type-I HARQ relying on Convolutional Codes (CC) is used for simplicity, because our focus is on the design of the scheduling schemes. Our proposed technique is significantly different from the existing contributions, such as the PET framework of [61], as detailed below. Firstly, our transmission scheme is proposed for wireless channels, while most contributions on PET [61] operate at the packet-level and consider the BEC. Secondly, IL-FEC typically relies on bit-level FEC decoders using soft decoding, such as a RSC code, while PET employs hard-decoded codes, such as the family of (N, K) block codes.

Against this background, the rationale and novelty of this chapter is summarized as follows.

1. We intrinsically amalgamated IL-FEC coding with the THARQ-aided transmission of layered video. We conceived an ATHARQ transmission scheme for adaptively scheduling the IL-FEC coded video layer packets for minimizing the video distortion under the constraint of a certain total number of transmission time slots.
2. We develop a method of on-line optimization for our IL-ATHARQ transmission

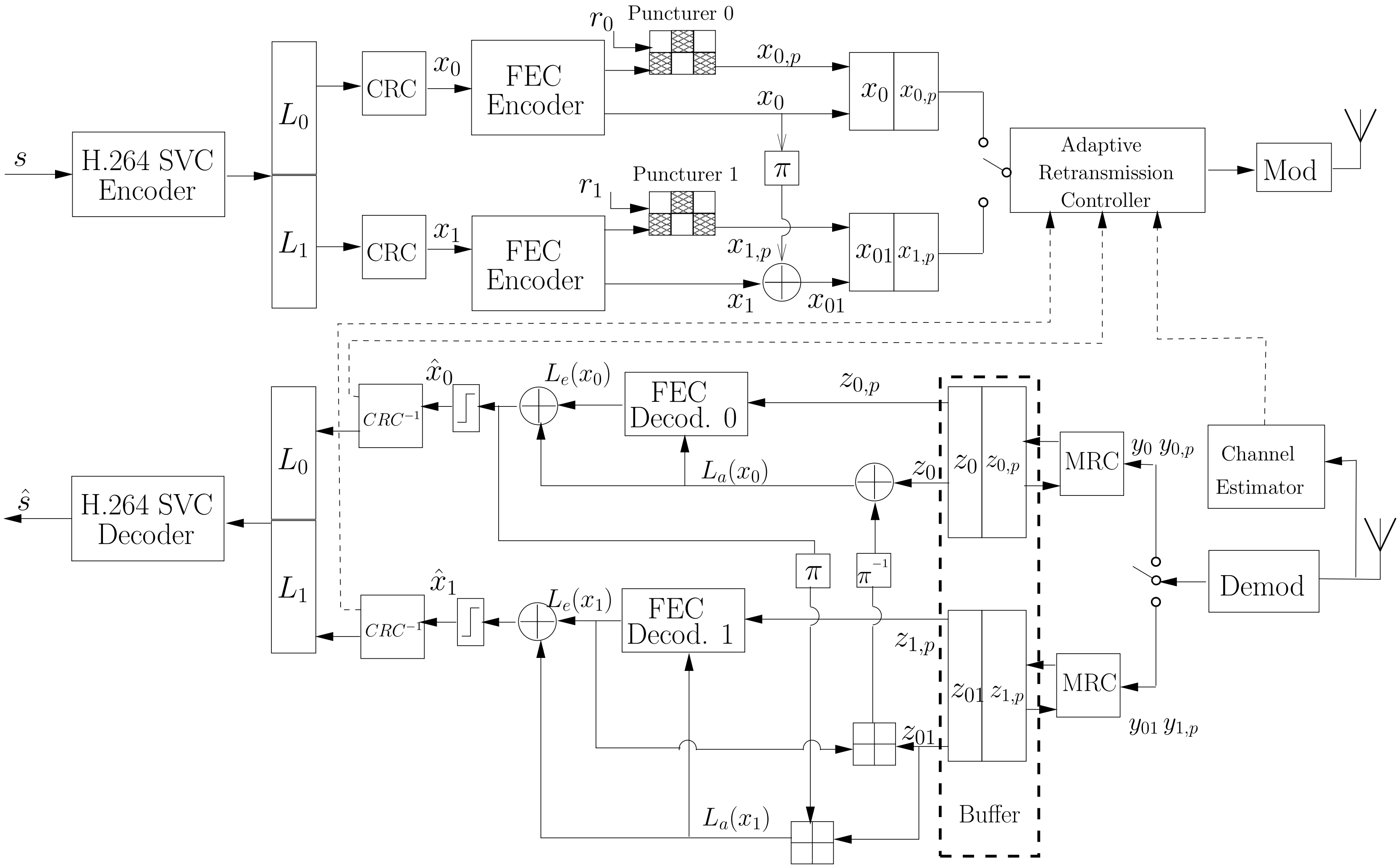


Figure 3.1: Block diagram of the proposed ATHARQ-IL-FEC coded SVC H.264/AVC coded video system, where r_0 and r_1 represent the code rates for FEC encoder 0 and 1, respectively.

scheme, in order to find the optimal FEC code rate distribution, sharing among the video layers that results in a reduced video distortion. Quantitatively, about 2.5 dB of PSNR video quality improvement may be observed at an E_b/N_0 of 15 dB, over the traditional THARQ benchmaker. Alternatively, we will demonstrate that an E_b/N_0 reduction of about 5.3 *dB* at a PSNR of 38.5 *dB* can be achieved.

The rest of this chapter is organized as follows. Section 3.2 details the IL-FEC transmitter and receiver model, as well as the proposed ATHARQ protocol along with the benchmark schemes we used in this chapter. The algorithm of our IL-ATHARQ retransmission controller is described in Section 3.3, followed by the details of the coding-rate optimization of the IL-ATHARQ system in Section 3.4. The performance of our IL-ATHARQ scheme as well as the rate-optimized IL-ATHARQ scheme using a RSC codec are compared to the benchmarks in Section 3.5 using different video sequences, followed by characterizing both the effects of the delay as well as well of the channel quality prediction errors on the attainable system performance. Finally, we conclude in Section 3.6.

3.2 System Overview

Here we introduce the IL-ATHARQ-aided inter-layer video transceiver shown in Fig. 3.1. The system consists of two major parts: the inter-layer FEC (IL-FEC) protected video codec and the retransmission control protocol. Firstly, the former one is introduced based on [98], where the inter-layer FEC architecture is described in detail. We will briefly describe both the IL-FEC transceiver architecture, as well as the IL-ATHARQ protocol in this section.

3.2.1 Transmitter Model

The original video sequence is firstly encoded into a scalable video stream by invoking the SVC extension of H.264 [10]. The compressed video stream consists of the layers L_0, L_1,... L_n with the dependency of $L_0 \Leftarrow L_1 \Leftarrow \ldots \Leftarrow L_n$, where each item on the right of the $\Leftarrow$ symbol depends on all the items to the left of it. To utilize the n-th layer for successful decoding, the decoder has to invoke the information from all the previous $(n-1)$ layers. For simplicity of illustration, only the pair of layers L_0 and L_1 are used in our description of IL-FEC, where L_0 is the BL and L_1 is the EL.

As shown in Fig. 3.1, each layer of the SVC encoded video is protected by the Cyclic Redundancy Check (CRC) encoder. Then each layer is encoded using their individual FEC code, typically an RSC code. Since each layer is allowed to have its own specific code-rate, the FEC encoded layers are passed through their individual puncturer, which may have different puncturing rates. We assume that the punctured layers have FEC code rates of r_0 and r_1, respectively. For layer L_0 the input bit sequence x_0 is encoded and punctured in order to produce the parity bits x_{0p}, and for layer L_1 the parity bits x_{1p}.

As part of the IL-FEC mechanism, the systematic part of the encoded layer L_0, namely x_0, is interleaved and then embedded into the systematic part x_1 of L_1, using the bit-wise XOR operation, producing the bit sequence x_{01}. In the case that L_0 and L_1 are different in length, the solution detailed in [98] may be invoked. Then the systematic bits and the parity bits of the BL are concatenated. Similarly, the EL, which contains the systematic bits and parity bits of the original EL are also concatenated. For each time slot, the adaptive retransmission controller picks the packets from one of the two layers and transmits them using BPSK over the wireless channel, which is modeled as an uncorrelated Rayleigh-faded channel.

3.2.2 Receiver Model

At the receiver, the likelihood of the demodulated bits is identified. If L_0 is received, the demodulated sequence consist of y_0 and y_{0p}, which represent the likelihood of the systematic information x_0 and that of the parity information x_{0p} for L_0. If L_1 is received, the demodulated sequence consist of y and y_{1p}, corresponding to x_{01} and x_{1p}. Then the identified likelihood information is combined with that of the information already stored in the corresponding buffer, using maximum ratio combining (MRC). Let z' be the likelihood before combining, and z afterwords. Then we have $z_0 = z'_0 + y_0$, $z_{0p} = z'_{0p} + y_{0p}$, $z_{01} = z'_{01} + y_{01}$ and $z_{1p} = z'_{1p} + y_{1p}$.

After updating the buffers, the decoder carries out the IL-FEC decoding process. The pair of FEC decoders shown in Fig. 3.1 invokes the BCJR algorithm [225] to produce the extrinsic information for x_0 and x_1, given the *a priori* information of their systematic bits and parity bits.

At the beginning of the decoding process, the FEC decoder 0 of Fig. 3.1 generates the extrinsic information $L_e(x_0)$ using the accumulated parity-bit-related information $z_{0,p}$ and the systematic-bit-related information $L_{apr}(x_0)$. Since the FEC decoder 1 has no information to contribute initially, decoder 0 uses z_0 directly from the buffer as $L_{apr}(x_0)$. Given the extrinsic information, we can obtain the *a posteriori* information by $L_{aps}(x_0) = L_e(x_0) + L_{apr}(x_0)$. The temporary decoding result $\hat{x}_0$ is obtained by making a hard decision concerning $L_{aps}(x_0)$. The subsequent CRC checker will check, whether we have $\hat{x}_0 = x_0$ and if so, then $L_{aps}(x_0)$ will be replaced by the perfect LLR of x_0. Then the interleaved $L_{aps}(x_0)$ and z_{01} together will provide the a priori information of $L_{apr}(x_1) = L_{aps}\left[\pi\left(x_0\right)\right] \boxplus z_{01}$ for the FEC decoder 1, where $\pi\left(\centerdot\right)$ represents the interleaving-based permutation, while $\pi^{-1}\left(\centerdot\right)$ the corresponding deinterleaving function. Furthermore, given the bits u_1 and u_2, the 'boxplus' operation $\boxplus$ is defined as follows:

$$\begin{aligned} L\left(u_1 \oplus u_2\right) &= L(u_1) \boxplus L(u_2) \\ &= \log \frac{1 + e^{L(u_1)} e^{L(u_2)}}{e^{L(u_1)} + e^{L(u_2)}}. \end{aligned} \tag{3.1}$$

For the second decoding phase, given the *a priori* information $L_{apr}\left(x_1\right)$ of x_1 and the *a priori* information $L_{apr}\left(z_{1,p}\right)$ of its parity bits, the FEC decoder 1 of Fig. 3.1 generates the extrinsic information $L_e(x_1)$. In turn the function $\pi^{-1}\left(L_e(x_1) \boxplus z_{01}\right)$ will provide part of the *a priori* information for x_0, so that the FEC decoder 0 is supplied with the improved *a priori* information $\pi^{-1}\left(L_e(x_1) \boxplus z_{01}\right) + z_0$ for the systematic bits. Again, the *a posteriori* information is generated by $L_{aps}(x_1) = L_e(x_1) + L_{apr}(x_1)$, upon which the hard decision yielding $\hat{x}_1$ will be carried out and the CRC checker of Fig. 3.1 will be invoked to check its correctness.

By iteratively repeating the above two decoding phases, the decoder exploits the information embedded in the EL L_1 for the sake of assisting the decoding of the BL L_0, without affecting the performance of the L_1 transmission, as long as L_0 is successfully decoded. The iterations are terminated, when either the CRC of all the layers indicates success, or the affordable maximum number of iterations has been reached. In this chapter we set the maximum number of iterations to $T = 2$.

3.2.3 Major Assumptions and Transmission Protocol

Again, for the sake of limiting the delay imposed, we consider limited-delay THARQ as our transmission technique. In our scenario we map each layer to a single packet, which also correspond to a single network abstraction layer unit (NALU), since we adopted the SVC profile of the H.264 video codec. The packets corresponding to the different layers are likely to have different lengths of bits, depending on the lengths of the NALUs generated by the SVC codec.

The traditional THARQ transmission protocol conceived for the FEC coded video layers is shown in Fig. 3.2(a). The BL is transmitted first, followed by the ELs. Each layer is transmitted a maximum number of n times, regardless, whether or not it is correctly received. However, according to the dependency between the video layers, there is no need to transmit the ELs, if the BL is lost. Therefore it is sub-optimum to assign the same retransmission limit to each layer. Thus we adapt the traditional THARQ by defining a total maximum retransmission limit for a specific video slice. Explicitly, for a total of n transmissions, the BL is allowed to have a higher number of transmissions than the EL. This plausible prioritization principle may be readily extended to an arbitrary number of layers, where the less dependent layers are granted more transmission opportunities than the more dependent layers.

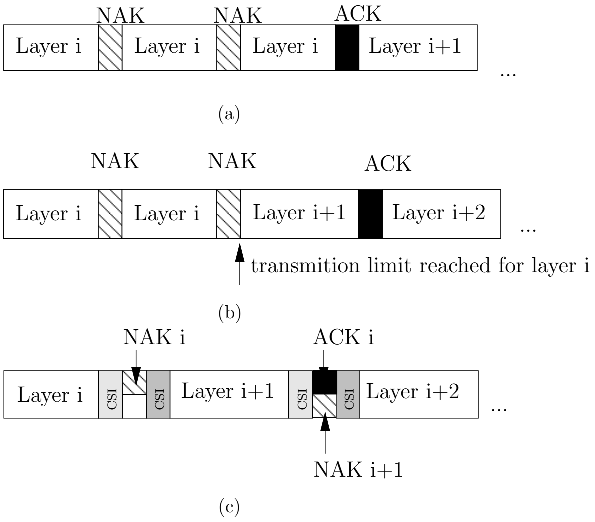


Figure 3.2: THARQ schemes conceived for scalable video transmission: (a) pure THARQ (b) IL-FEC aided THARQ where each video layer has a different retransmission limit (c) proposed IL-FEC aided adaptive THARQ

With the introduction of IL-FEC coding into the THARQ aided SVC coded video steam, the constraints imposed may be relaxed, because the IL-FEC coded layers of higher dependencies may have sufficient information concerning the layers of lower dependencies and hence they may be capable of recovering them, even they were incorrectly recovered during the previous transmissions. As a result, it may in fact become wasteful to complete the recovery of the BL before transmitting the ELs. Hence we have to carefully consider the choice of transmission limits for each layer. The total number of transmissions dedicated to a specific video slice remains the same as defined previously for fair comparison. The philosophy of this scheme is illustrated by a specific example in Fig. 3.2(b).

We set out to improve the THARQ regime introduced above, which relies on the CRC check result of the decoded layers. In the traditional regime, the transmitter only knows whether the current layer has or has not been successfully recovered. However, it has no quantitative knowledge about the specific grade of degradation imposed on the unsuccessfully decoded layers in the buffer, if any. Similarly, the transmitter has no knowledge of the next transmission's contribution towards the successful decoding of the video slice. With the goal of improving the performance, we set out to estimate both and hence to make better-informed decisions. Therefore in our new regime, the receiver has to provide a feedback for the transmitter concerning the CRC result, as well as additionally has to feed back the channel state information (CSI) of both the most recent transmission and of the next transmission.

In recent transmission systems such as the Worldwide Interoperability for Microwave Access (WiMAX) [226] arrangement, the data subframe is divided into a number of mini-slots using Time Division Duplexing (TDD). In reciprocal channels encountered in TDD systems, the CSI of the next transmission can be acquired by incorporating pilot symbols into the reverse-direction data frames, which allows us to estimate the CSI at the receiver and assuming that the coherence-time is longer than the frame length [227], use it for predicting the channel to be encountered during the next transmission. Owing to using this low-complexity zero-order channel prediction, a prediction error may be introduced at this stage. At the reception of a packet, the receiver sends a feedback message to the transmitter, which includes the CRC results of the layers of interest, plus the estimated CSI of this specific transmission. The latter one assists the transmitter in rectifying the previous CSI prediction error at the transmitter, and as a benefit, this measure prevents error propagation in the subsequent prediction process. As shown in Fig. 3.2(c), the transmitter becomes capable of estimating the contribution of each possible transmission at the receiver, hence intelligently controlling the re-transmission process by maximising the video quality after the next transmission attempt.

Indeed, the new scheme introduces overheads in terms of requiring extra bandwidth for accommodating the feedback channel. However, the HARQ feedback only requires a few bits for conveying the CRC flag of the layers queuing in the buffer as candidates for transmission, as it will be detailed in Section 3.3. Hence we may readily assume that this does not impose a heavy burden on the feedback channel and assume furthermore that it is transmitted without any errors. By contrast, the channel estimate feedback is more error-sensitive, because it is a floating-point number and thus it requires more bandwidth. Consequently it may not be justifiable to assume perfect feedback reception, since the CSI feedback may be subject to channel impairments. Since the CSI feedback is used for predicting the channel of the next transmission, we will take into account this factor by considering the CSI impairments to be modelled by extra additive noise and as being part of the additive prediction error, which will be detailed in Section 3.3.

As mentioned, each video layer is packaged into a single NALU and can be transmitted over a single channel instance. By contrast, when the video layers of high-resolution sequences are represented by more bits, each video layer may be packaged into several NALUs and transmitted over different channel instances. In that case, advanced channel estimation techniques [228, 229] may be adopted for acquiring the channel estimates for the sake of predicting the video qualities, which is beyond the scope of this chapter.

3.3 Adaptive Truncated HARQ Transmission

In this section, we will describe our adaptive truncated HARQ aided IL-FEC coded video streaming scheme, which is used in the “Adaptive Retransmission Controller” block of Fig. 3.1.

As described at the end of Section 3.2.3, our adaptive transmission algorithm aims for minimizing the reconstructed video distortion at the receiver by carefully choosing the sequential order of transmitting the different video layers, given the total number of transmissions. Again, the wireless channel is assumed to impose uncorrelated block-fading between different time slots, which remains constant for a time-slot and then it is independently faded for the next time-slot. However, the above-mentioned TDD-related reciprocity still allows us to exploit the correlation of the forward and reverse links for typical packet-lengths that are shorter than the coherence-time. Nonetheless it is impossible to predict all the channel information for all the time slots, let alone to find a globally optimal transmit schedule depending on the predicted CSI information. Instead, we conceive an adaptive algorithm, which is sub-optimal but practical and seeks to achieve the minimization of the reconstructed video distortion for the next single transmission only, given the prediction of the forthcoming channel condition obtained by using the protocol described in Section 3.2.3.

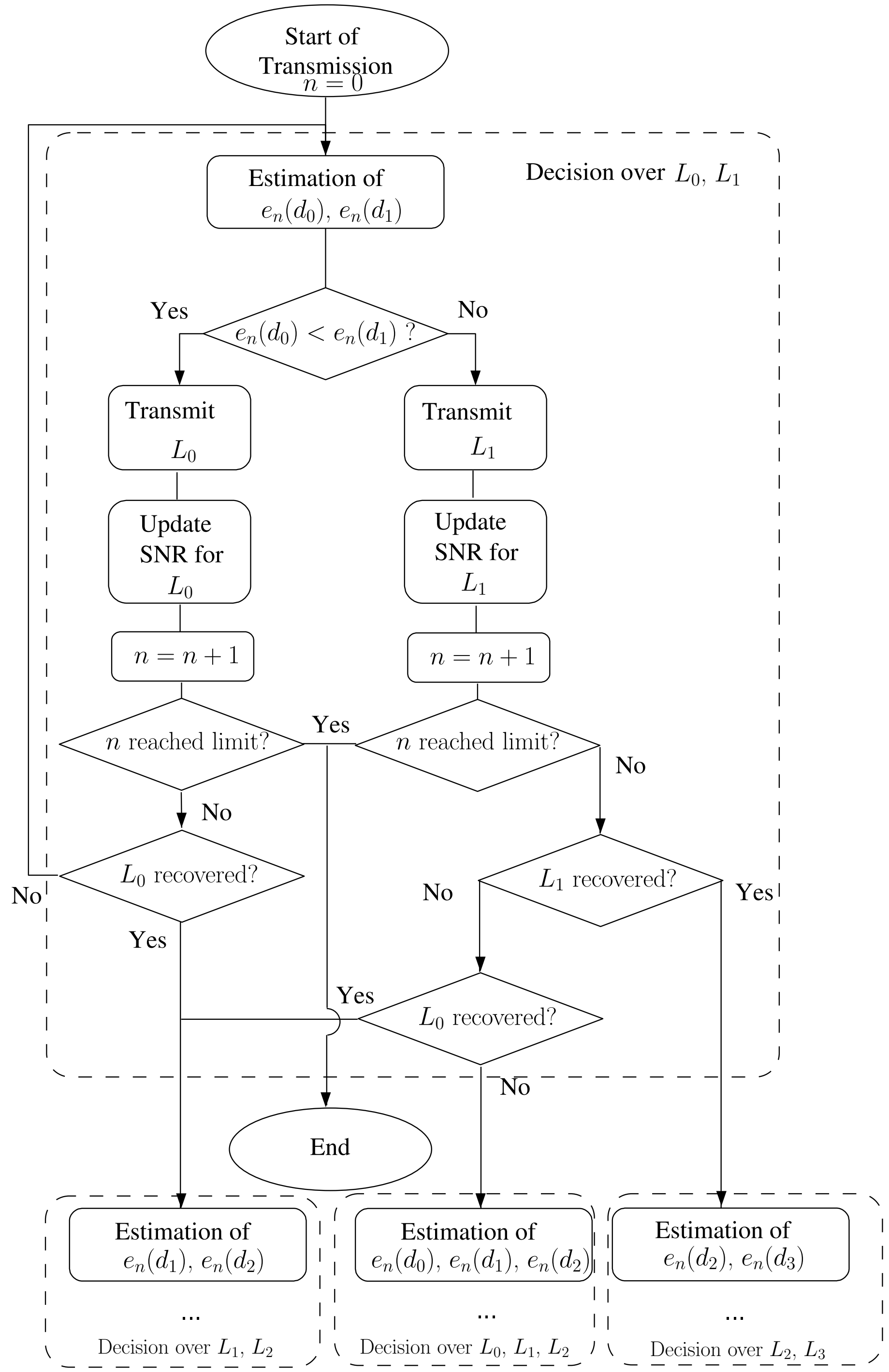


Figure 3.3: Adaptive THARQ algorithm for the Adaptive Retransmission Controller

In order to characterize the behavior of the receiver seen in Fig. 3.1 relying on the proposed algorithm, a classic RSC codec is used as the FEC code. However, the employment of our proposed techniques is not limited to the RSC codec.

Before introducing the adaptive retransmission control algorithm, let us define the symbols to be used in our discussion in Table 3.1:

The scheduling procedure of the adaptive retransmission controller is shown in Fig. 3.3. Each time the algorithm considers a number of video layers, for making

	Description
N_T	The limit of the total number of transmission time slots (TS)
N_L	The total number of video layers
L_n	The n th layer of the compressed video stream
$\lvert h_n\rvert$	The amplitude of the channel at TS n, $1 \leq n \leq N_T$
A_n	$A_n = \lvert h_n\rvert^2$
$\left\lvert \widetilde{h}_n \right\rvert$	The prediction of $\lvert h_n\rvert$ as described in Section 3.2.3
ϵ_i	The video distortion due to the corruption or absence of layer i, which is measured using the peak signal-to-noise ratio (PSNR), $0 \leq i < N_L$
SNR_n	The N_L-element vector, which represents the SNR values of the signals in the receiver buffers after the n-th transmission. In other words, $SNR_{n,i}$, which is the i-th element of SNR_n, represents the SNR of the signals in the receiver buffer as defined in Section 3.2.2, $1 \leq n \leq N_T$, $0 \leq i < N_L$
$SNR_n\left(\cdot\right)$	The N_L-element vector, which represents the predicted value of SNR_n, depending on both SNR_{n-1}, as well as on the channel conditions and on the scheduling decisions, etc.
ℓ_i	The length of the bitstream of layer i, $0 \leq i < N_L$
R	The overall coding rate of the system
$\boldsymbol{r}$	The vector including the FEC coding rates of all the layers, where r_i is the coding rate of layer i, $0 \leq i < N_L$
$\mathcal{D}$	The decision-set including all the possible choices concerning which particular layer to transmit, where $\mathcal{D} = \{d_1, d_2, \ldots, d_{N_L}\}$. The decision vector d_k has N_L binary elements, where $d_{k,i}$ is defined as $d_{k,i} = \begin{cases} 1 & k = i \\ 0 & k \neq i \end{cases}$, $0 \leq i < N_L$, $0 \leq k < N_L$, which means that the k-th layer is chosen for transmission
D_n	The actual decisions adopted for transmitting at TS n, where we have $D_n \in \mathcal{D}$
$p_{n,i}\left(d_k\right)$	The packet error ratio (PER) of layer L_i at TS n, using decision d_k, when layer L_{i-1} is correctly decoded
δ	The estimation error tolerance threshold defined in Eq. (3.16)
T_p	The lookup-table (LUT) used for the estimation of the PER

Table 3.1: Parameters used in Chapter 3.

decisions, as shown in the dashed boxes of Fig. 3.3.

In order to generalize the scheduling process, the concept of a decision window can be introduced, which contains the layers to be chosen by our algorithm, as shown in Fig. 3.4. This decision window always contains one more layer than the layers transmitted during the most recent history, which is layer $(i+j+1)$ in the example shown in Fig. 3.4. Upon reception of an ACK for a successfully recovered layer, the layer is removed from the window. At the beginning of the entire transmission, the decision window only contains L_0 and L_1.

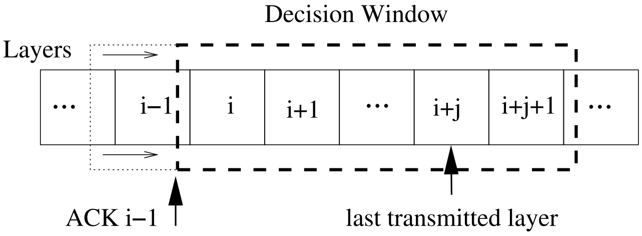


Figure 3.4: Decision Window of the Adaptive THARQ algorithm

Now we set out to describe the algorithm in each of the dashed boxes of Fig. 3.3. As described in Section 3.2.3, the prediction of the forthcoming channel amplitude is given by $\left|\widetilde{h}_n\right|$. Furthermore, when using a Maximum Ratio Combining (MRC) receiver, given the E_b/N_0 value, the combined signal's SNR in the receiver's buffer can be predicted as:

$$\begin{aligned} SNR_n\left(d_k\right) &\doteq SNR_n\left(d_k, \left|\widetilde{h}_n\right|, SNR_{n-1}\right) \\ &= \frac{E_b}{N_0} R \left|\widetilde{h}_n\right|^2 \cdot d_k + SNR_{n-1} \end{aligned}, \tag{3.2}$$

which means that the SNR value of the buffered signal corresponding to the k-th layer will increase by the value of $\frac{E_b}{N_0} R \left|\widetilde{h}_n\right|^2$, if the decision d_k is made. The channel prediction $\widetilde{h}_n$ can be expressed as [230]

$$\widetilde{h}_n = h_n + \zeta_n, \tag{3.3}$$

where ζ_n is the prediction error or imperfect feedback error with a variance of σ_e^2.

To obtain the estimate of the distortion, each layer's PER $p_{n,i}\left(d_k\right)$ and its distortion ϵ_i should be acquired. The latter can be obtained by the so-called offline removal-decoding test, which was advocated in [231]. Explicitly, ϵ_i is measured by comparing the PSNRs of the reconstructed video with the bit stream of the i-th video layer removed, and the one relying on the intact bit stream for the i-th video layer. Again, this measurement is carried out offline, before the transmissions begin.

Let us now consider the “conditional” PER of layer i, which refers to the PER of the i-th video layer at TS n corresponding to decision d_k, given that layer $(i-1)$ has been successfully recovered, which is denoted by $p_{n,i}(d_k)$. In order to formulate the PER $p_{n,i}(d_k)$, here we introduce the function $f_i(\centerdot)$ which was defined by Eq. (17), (18) in [99]:

$$f_i\left(SNR_{n,i}, \ell_i, r_i\right) = 1 - \left[1 - T_p\left[SNR_{n,i}, I_a, r_i\right]\right]^{\ell_i/\ell}. \tag{3.4}$$

Given the $SNR_{n,i}$ of the buffered signal corresponding to layer i, the systematic bit-length ℓ_i, and the FEC coding rate r_i, the $f_i(\centerdot)$ function gives the PER estimate $p_{n,i}$ of layer i. At the right side of Eq. (3.4), the pre-generated look-up table (LUT) T_p is used for obtaining the PER assuming a fixed systematic bit-length of ℓ. Furthermore, T_p is a 4-dimensional LUT that has three input parameters to index the specific PER needed. Apart from the aforementioned $SNR_{n,i}$ and ℓ_i, the mutual information (MI) I_a gleaned from the estimation of the decoding output of layer $i+1$ is needed. Further details concerning the estimate of I_a can be found in [99]. Therefore the PER $p_{n,i}(d_k)$, which is the estimate of layer i, can be readily formulated as

$$\begin{aligned} p_{n,i}\left(d_k\right) &\doteq p_{n,i}\left(d_k, \left|\widetilde{h}_n\right|, SNR_{n-1}\right) \\ &= f_i(SNR_n\left(d_k\right), \ell_i, \ldots, \ell_{N_L-1}, \\ &\quad r_i, \ldots, r_{N_L-1}). \end{aligned} \tag{3.5}$$

Firstly, the $SNR_n(d_k)$, which is required for the estimation of the PER $p_{n,i}$ can be obtained from Eq. (3.2). As for the decoding process, the IL decoder commences its operation from the specific layer having the highest grade of dependency, which is layer (N_L-1). Then it exchanges information between the decoding of two consecutive video layers during each iteration, as illustrated in Section 3.2.2. Given the intact layer $(i-1)$, the successful decoding of layer i depends on the assistance of layer $(i+1)$, which in turn depends on layer $(i+2)$, etc. Therefore the estimation of the “conditional” PER in Eq. (3.5) depends on the properties of all the layers spanning from i to (N_L-1), which includes both the lengths of their coded blocks and their coding rates. Specifically, the layer (N_L-1) associated with the highest grade of dependency but receiving no extra protection from the other layers has the “conditional” PER that only depends on the layer (N_L-1) itself, which is formulated as

$$p_{n,N_L-1}\left(d_k\right) = f_{N_L-1}\left[SNR_n\left(d_k\right), \ell_{N_L-1}, r_{N_L-1}\right]. \tag{3.6}$$

Given the PER expression of $p_{n,i}(d_k)$, the expected distortion of the decoded video at the receiver during TS n can be formulated as:

$$\begin{aligned} e_n(d_k) &\doteq e_n\left(d_k, SNR_{n-1}, |h_n|^2\right) \\ &= p_{n,0}(d_k)\cdot\epsilon_0 + \\ &[1-p_{n,0}(d_k)]\cdot p_{n,1}(d_k)\cdot\epsilon_1+\ldots \\ &= \sum_{i=0}^{N_L-1}\epsilon_i\cdot p_{n,i}(d_k)\cdot\prod_{j=0}^{i-1}[1-p_{n,j}(d_k)] \end{aligned}, \quad (3.7)$$

where $p_{n,i}(d_k)\cdot\prod_{j=0}^{i-1}(1-p_{n,j}(d_k))$ represents the PER of layer i, when the layers spanning from 0 to layer $(i-1)$ have already been successfully received.

The retransmission controller of Fig. 3.3 opts for transmitting the specific video layer that ends up with the minimum distortion of the decoded video. Hence the final decision carried out by the controller is

$$D_n\left(SNR_{n-1}, |h_n|^2\right) = \underset{d_k\in\mathcal{D}}{\arg\min}\left\{e_n(d_k)\right\}, \quad (3.8)$$

which may be compactly expressed as D_n. At the commencement of transmissions, the module estimates the distortions that two different scheduling decisions would impose, namely when transmitting L_0 or transmitting L_1, which may be denoted by d_0/d_1. As shown in Fig. 3.3, the receiver buffer is empty at the beginning of a transmission session, and both d_0 as well as d_1 are compared by the retransmission controller, as the potentially available choices. If we have $D_1 = d_0$, which means that the controller of Fig. 3.3 decides to transmit L_0, and L_0 is successfully recovered, then the controller will consider both d_1 and d_2, provided that more transmission TSs are available. Otherwise, if the receiver failed to decode L_0, then both d_0 and d_1 will be reconsidered as retransmission candidates for the next retransmission attempt. On the other hand, if we have $D_1 = d_1$, which implies that L_1 is selected for transmission, and L_1 is successfully recovered, the controller of Fig. 3.3 will naturally move on to consider both d_2 and d_3, if possible. However, if L_1 was unsuccessfully decoded, the coded packet of L_1 will be stored in the receiver's buffer, and hence we have to further consider the decoding outcome of L_0. If the decoding of L_0 turned out to be successful, the transmitter will consider retransmitting L_1 or transmitting the new L_2 packet. But if L_0 also failed, with d_0 and d_1 on the table, d_2 should also be considered. Since the IL-encoded L_2 includes the redundancy protecting L_1, it may be capable of improving the decoding of L_1. In turn, the improved decoding of L_1 may become capable of providing beneficial information for the BL L_0.

After each transmission, the transmitter will receive the updated version of the channel's amplitude $|h_n|$, in order to replace the predicted version $\left|\widetilde{h}_n\right|$, as described by our protocol in Section 3.2.3. Here the updated version may be considered accurate, since it exactly equals $|h_n|$. The retransmission controller of Fig. 3.3 will then update the estimation of the SNRs of the signals in the receiver buffer, using both the E_b/N_0 value, as well as the past transmission decision records and channel amplitudes, as formulated in Eq. (3.2). The SNR estimate of SNR_{n-1} can be expressed as

$$SNR_{n-1} = \frac{E_b}{N_0} R \cdot \sum_{t=1}^{n-1} D_t \left| h_t \right|^2, \tag{3.9}$$

and upon substituting it into Eq. (3.2), we arrive at:

$$SNR_n\left(d_k\right) = \frac{E_b}{N_0} R \cdot \left(d_k \left| \widetilde{h}_n \right|^2 + \sum_{t=1}^{n-1} D_t \left| h_t \right|^2 \right). \tag{3.10}$$

Finally, if the TS limit N_T is reached, the transmissions are concluded.

3.4 FEC Coding Rate Optimization

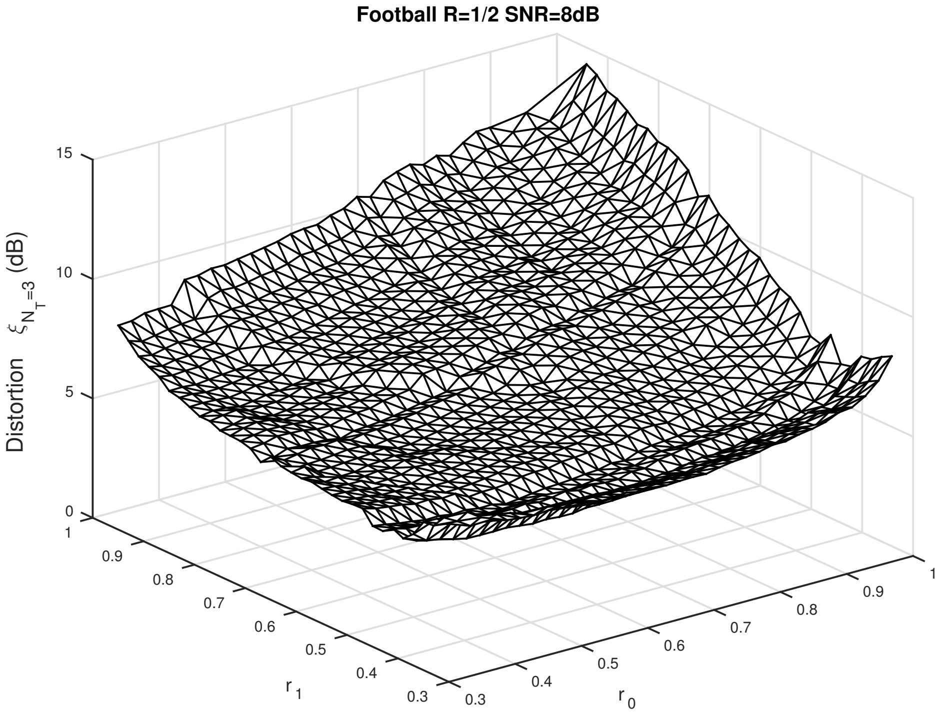
We have described our IL-ATHARQ algorithm in Section 3.3, which aims for beneficial layer-scheduling, whilst relying on a fixed FEC coding rate. However, the FEC coding rate itself has yet to be optimized, for the sake of improving the achievable system performance. Specifically, with the total coding rate being R, the best distribution sharing of the coding rates among the different layers has to be found for minimizing the video distortion. Therefore in this section, we focus our attention on finding the most appropriate FEC coding rate for our IL-ATHARQ algorithm of Section 3.3.

According to Eq. (3.8), the distortion of the reconstructed video frame after the n-th transmission is given by:

$$\mathcal{E}_n\left(SNR_{n-1}, \left|h_n\right|^2\right) = \min_{d_k \in \mathcal{D}} \left\{ e_n\left(d_k\right) \right\}. \tag{3.11}$$

Given the video distortion definition in Eq. (3.7), by substituting SNR_{n-1} from Eq. (3.9) into Eq. (3.11), we get:

$$\begin{aligned} &\mathcal{E}_n\left(\left|h_1\right|^2, \ldots, \left|h_n\right|^2\right) \\ &\quad = \mathcal{E}_n\left(A_1, \ldots, A_n\right) \\ &\quad = \min_{d_k \in \mathcal{D}} \left\{ e_n\left(d_k, \frac{E_b}{N_0} R \cdot \sum_{t=1}^{n-1} D_t A_t, A_n \right) \right\}. \end{aligned} \tag{3.12}$$



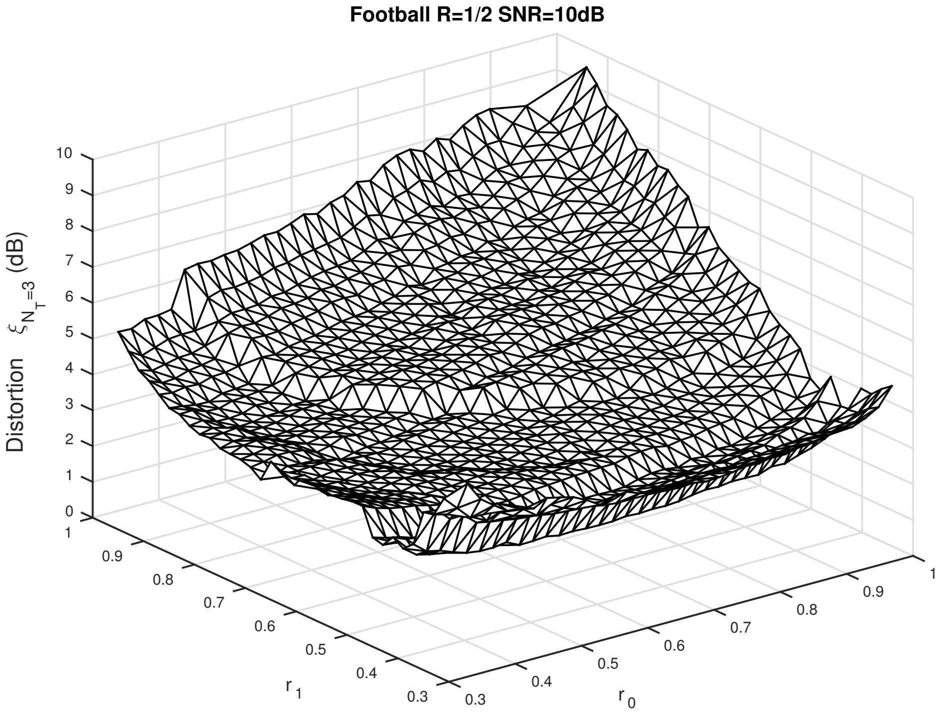


Figure 3.5: Video distortion versus FEC coding rate performance according to Eq. (3.12)

Finally, since in our scenario an uncorrelated block-faded channel is considered, the expected value of the video distortion after the n-th transmission can be expressed as

$$\begin{array}{c} E\left(\mathcal{E}_n\right) = \int f\left(A_1\right) dA_1 \int f\left(A_2\right) dA_2 \cdots \\ \int \mathcal{E}_n\left(A_1, \ldots, A_n\right) f\left(A_n\right) dA_n \end{array}, \tag{3.13}$$

where $f\left(\cdot\right)$ is the probability density function (PDF) of the fading channel. For Rayleigh-faded channels $f\left(\cdot\right)$ is given by the PDF of the Gamma distribution. Since Eq. (3.13) is difficult to evaluate in a closed form, we carried out Monte-Carlo simulations using Eq. (3.12). It is worth noting that each experiment is based on low-complexity table-look-up operations without any actual encoding or decoding operations, therefore imposing an affordable complexity. A numerical example is provided in Fig. 3.5 for the Football sequence in terms of the video distortion vs coding rates.

For a given E_b/N_0 value, we aim for minimizing the video distortion, when the maximum transmission limit N_T is reached. The corresponding objective function (OF) of our optimization problem can be expressed as

$$\underset{\boldsymbol{r}\in\Gamma}{\arg\min} \left\{E\left(\mathcal{E}_{N_T}\right)\right\}, \tag{3.14}$$

where the combination of the coding rates $\boldsymbol{r}$ has to satisfy the total coding rate constraint. In other words, $\boldsymbol{r}$ belongs to the set Γ of all the possible coding rate combinations, which can be expressed as

$$\Gamma = \left\{ \boldsymbol{r} \middle| \sum_{i=0}^{N_L-1} \frac{\ell_i}{r_i} = \frac{\sum_{i=0}^{N_L-1} \ell_i}{R} \right\}. \tag{3.15}$$

Naturally, the system performance formulated in Eq. (3.14) may be affected by excessive video distortion estimation errors at lower E_b/N_0 values. Therefore an amended version of Eq. (3.14) can be formulated as

$$\boldsymbol{r}_{op} = \begin{cases} \underset{\boldsymbol{r}\in\Gamma}{\arg\min} \left\{E\left(\mathcal{E}_{N_T}\right)\right\} & \begin{array}{c} \min \left\{E\left(\mathcal{E}_{N_T}\right)\right\} < \\ \underset{\boldsymbol{r}=\boldsymbol{r}_{as}}{E} \left(\mathcal{E}_{N_T}\right) + \delta \end{array} \\ \boldsymbol{r}_{as} & otherwise \end{cases}, \tag{3.16}$$

where $\boldsymbol{r}_{as} \in \Gamma$ represents the code rates in ascending order, and the BL has the lowest FEC coding rate, i.e. the highest protection. Still referring to Eq. (3.16), δ is the

	Football	Soccer	Crew
Representation	YUV 4:2:0		
Format	QCIF	4CIF	4CIF
Bits Per Pixel	8		
FPS	15	60	60
Number of Frames	30	60	60
Video Codec	SVC-H.264		
GOP	15		
Scalability	MGS		
Bitrate	2297 kbps	15.36 mbps	11.77 mbps
Error-Free PSNR	40.46 dB	42.62 dB	42.82 dB
Error Concealment	Frame-Copy		

Table 3.2: The video parameters used in our investigations

estimation error tolerance threshold, which is found experimentally.

To evaluate the effect of the value of δ on the final PSNR performance of the system, we carried out simulations for various settings of δ, and the corresponding PSNR results are shown in Fig. 3.6. We can observe from Fig. 3.6a for $N_T = 3$ using the Football sequence that the PSNR performance is not very sensitive to the δ values at high E_b/N_0 values, say for $13 \sim 16$ dB. However, a slight improvement of PSNR can be observed for lower E_b/N_0 values around the δ values of 0.5 dB. Similar trends can also be observed for $N_T = 4$ in Fig. 3.6b, except that a marginal PSNR reduction is encountered upon increasing δ at E_b/N_0 values above 10 dB. Since the estimation error $\mathcal{E}_{N_T}$ is difficult to model analytically, we found the optimum value of δ experimentally. In order to improve the PSNR performance at lower E_b/N_0 values, we set $\delta = 0.6$ for our simulations, which was also found to be beneficial for the other video sequences investigated in Section. 3.5.

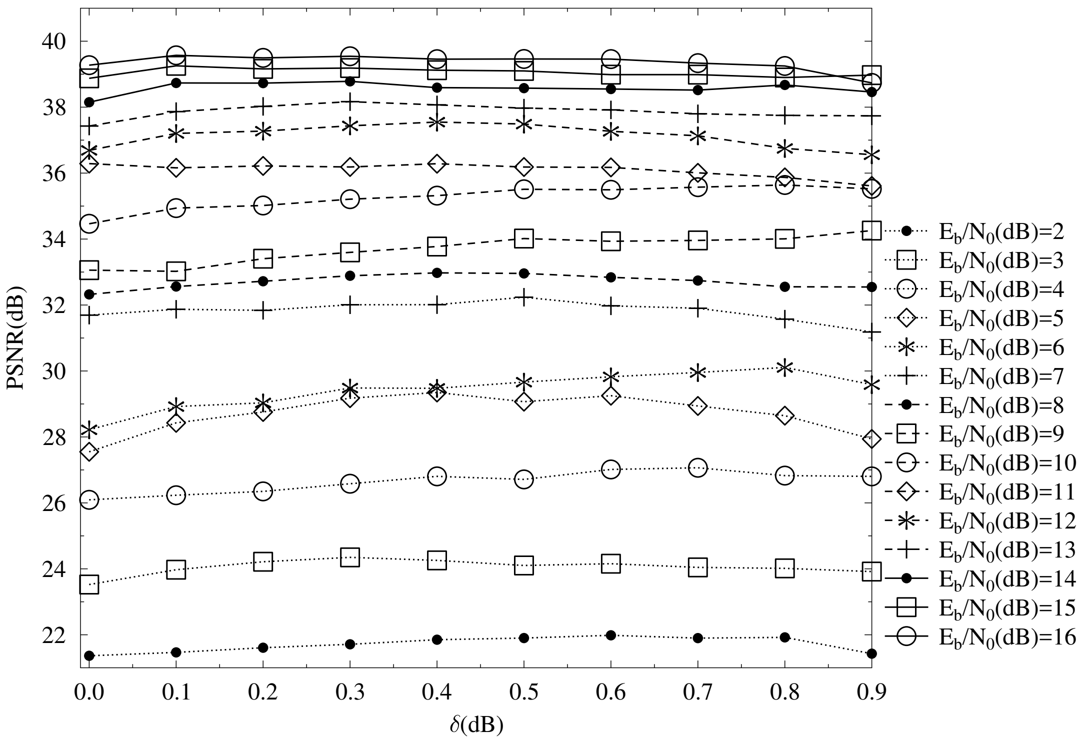
3.5 System Performance

In this section, we will quantify the attainable performance gain of our proposed ATHARQ-IL transmission scheme, as well as the additional performance gain of our rate-optimized ATHARQ-IL scheme. Furthermore, we will characterize both the delay performance and the robustness of the aforementioned systems against channel prediction errors. The main system parameters are listed in Table 3.2. Three 4:2:0

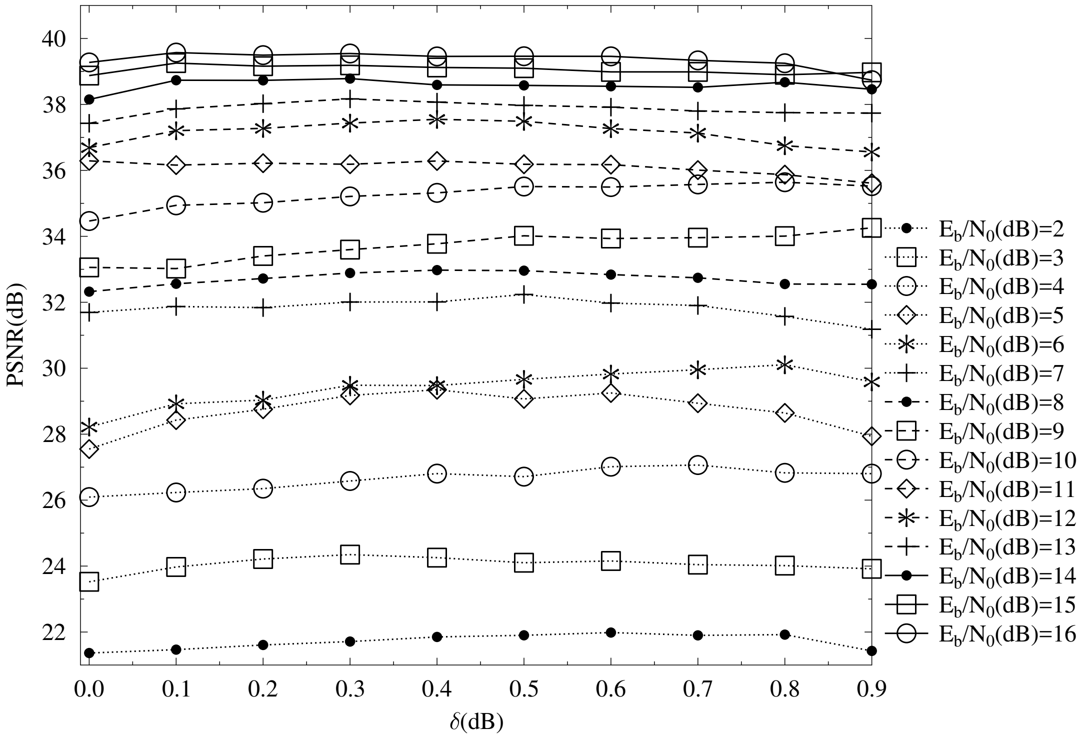
YUV format video sequences were chosen for transmissions, namely the Football, the Soccer and the Crew video clips. The 15-frame Football sequence is in the (176×144)-pixel quarter common intermediate format (QCIF) and has a frame rate of 15 frames/second (FPS). The other two 60-frame sequences are in the (704×576)-pixel 4CIF format, and were recorded at 60 FPS.

We use the JSVM H.264/AVC reference video codec as the SVC codec. The video encoder relies on a group of pictures (GOP) duration of 15 frames and the bi-directionally predicted (B) frames are disabled. We enabled the Medium Grain Scalability (MGS) [11,232] feature for encoding the video sequences into three layers with the aid of the standardized quantization parameters (QP) of 40, 32 and 24, respectively. The average PSNRs achieved by the decoder for different sequences are 40.46 dB, 42.62 dB and 42.82 dB, respectively.

Based on our configuration of the SVC encoder, each slice is encoded into three layers and each layer is encapsulated into a network abstraction layer unit (NALU) [10]. The NALUs are transmitted sequentially using our proposed system. Should the CRC check of a certain NALU indicate a decoding failure, these NALUs are discarded. The SVC decoder uses the low-complexity error concealment method of frame-copying in order to compensate for the lost frames.



(a) PSNR vs δ with RO-IL-ATHARQ(δ) for ***Football***, $\mathbf{N_T = 3}$.



(b) PSNR vs δ with RO-IL-ATHARQ(δ) for ***Football***, $\mathbf{N_T = 4}$

Figure 3.6: PSNR versus δ performance with RO-IL-ATHARQ(δ) systems. The *Football* sequence was transmitted over block-fading non-dispersive uncorrelated Rayleigh channels. The schematic of Fig. 3.1 and the parameters of Table 3.1 were used.

The RSC code having a code-rate of 1/3 and the generator polynomials of $[1011, 1101, 1111]$

is employed as the FEC code in our system. The reconfigurable puncturers employed are capable of adjusting the FEC code rate on a fine scale, ranging from its original 1/3 to 1, thus providing a wide range of design options. The FEC encoded signals are BPSK modulated *and transmitted through a block-fading non-dispersive uncorrelated Rayleigh channel.* The total coding rate of the system is assumed to be 1/2. The channel is static for each FEC encoded NALU, but it is faded independently between NALUs. As we are considering delay-constrained systems, we characterize the attainable performance of the proposed scheme using two scenarios, where either $N_T = 3$ or $N_T = 4$ transmissions are allowed in total, respectively. The regimes and their settings characterized in this section are listed in Table 3.3.

Regime	Coding Rate (r_0, r_1)	Other Parameters
THARQ	$({}^1/_2, {}^1/_2)$, $R = {}^1/_2$	
IL-THARQ	$({}^1/_2, {}^1/_2)$, $R = {}^1/_2$	(n_0, n_1): limit of transmission times for the video layers
IL-ATHARQ	$({}^1/_2, {}^1/_2)$, $R = {}^1/_2$	
	$({}^1/_3, {}^1/_2)$, $R = {}^1/_2$	
	$({}^1/_3, {}^1/_3)$, $R = {}^1/_2$	
RO-IL-ATHARQ	$R = {}^1/_2$	non-modified
	$R = {}^1/_2$	$\delta = 0.6$ for all the sequences

Table 3.3: The regimes with their settings characterized in Section 3.5, where R stands for the overall coding rate of the system.

3.5.1 Off-line LUTs Generation

As described in Eq. (3.4) of Section 3.3, the estimation of the PER relies on the LUT T_p. Here we describe the implementation of the LUT T_p that is used in our experiments. As mentioned in Section 3.3, the LUT T_p is indexed by three parameters, namely SNR, I_a, r. To generate T_p, we fix the block-length ℓ of the FEC and obtain the outputs, namely the extrinsic information I_e and the PER $p(\ell)$ of the component FEC by scanning the practical coding parameter ranges of SNR, I_a, r at certain intervals. Specifically, the SNR is considered over the range of $[0, 25]$ dB, using a step-size of 0.2 dB, I_a is scanned over the range of $[0, 1]$ at intervals of 0.01, and finally r is scanned across the range of $[0.33, 1]$ at intervals of 0.02. This makes the total number of legitimate settings $n_{Tp} = n_{snr} n_I n_r = 126 \times 101 \times 33$, which is 419,958. All 5 items corresponding to each setting can be individually stored as floats

SNR	I_a	r	I_e	$p(l)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
10.6	0.834	0.93	0.206	1
10.6	0.834	0.94	0.257	1
10.6	0.834	0.95	0.333	0.993
10.6	0.834	0.96	0.409	0.973
10.6	0.834	0.97	0.527	0.88
10.6	0.834	0.98	0.687	0.533
10.6	0.834	0.99	0.851	0.207
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Table 3.4: Example of the LUT $T_p(SNR, I_s, r)$.

in 8 bytes. Thus the total size of the LUT T_p is 16 MB. In Table 3.4, we show an example of the LUT T_p that is used in our simulations.

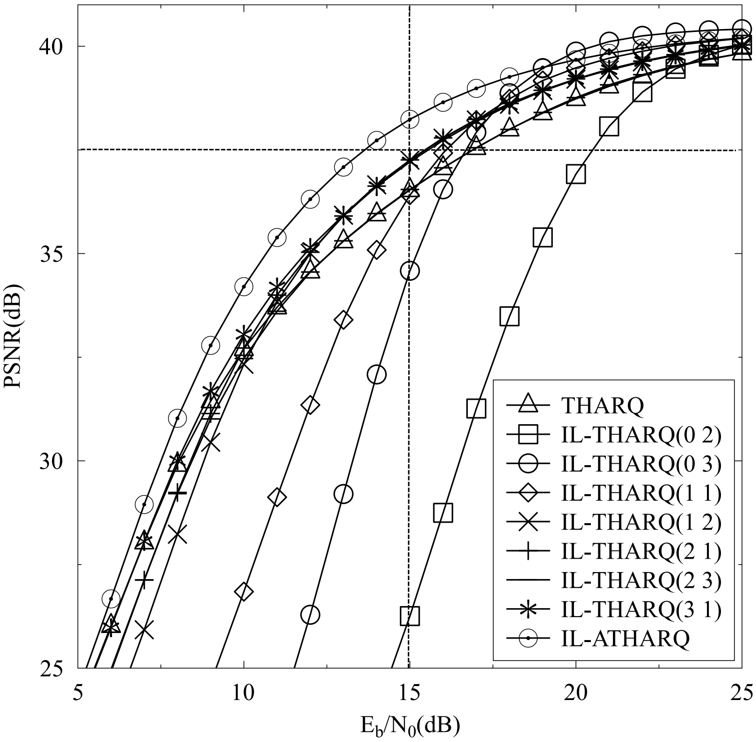
3.5.2 Performance of the Adaptive Rate Controller

In order to demonstrate the attainable performance gain of our IL-ATHARQ algorithm, we compare its PSNR performance to that of the aforementioned traditional THARQ as well as to that of the IL-THARQ scheme relying on fixed transmission limits, using the Football sequence, as listed in Table 3.2. For the IL-THARQ scheme, we use the compact form of IL-THARQ(n_0, n_1) to represent different configurations, where n_0 and n_1 denote the number of transmission times allowed for L_0 and L_1, respectively. Since the total number of transmission is fixed to N_T, L_2 is allowed to transmit as long as L_0 and L_1 have completed their transmission, provided that the total transmission attempts N_T has not been exceeded. At this stage we assume that all the three video layers are encoded using the same FEC coding rate of 1/2, which is the total coding rate of the system. The corresponding results are shown in Fig. 3.7.

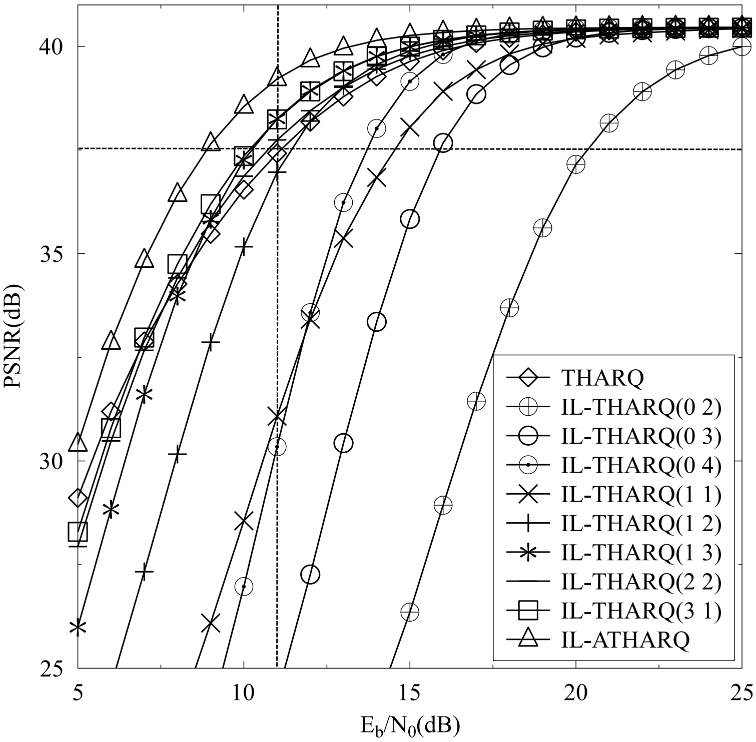
Observe in Fig. 3.7 that the PSNR versus E_b/N_0 performances of our proposed IL-ATHARQ system, relying on $N_T = 3$ or $N_T = 4$ transmissions are portrayed separately in Fig. 3.7a and Fig. 3.7b. Observe in Fig. 3.7a that the IL-THARQ(n_0, n_1) schemes perform differently for the different configurations of n_0 and n_1, given $N_T = 3$. We opted for characterizing the most typical combinations of n_0 and n_1, noting that others have similar results, hence we limited the number of combinations to make the figure more readable. It is clear from Fig. 3.7a that the IL-THARQ(0, n_1) class of systems performs relatively poorly at low E_b/N_0 values, because the BL L_0 is never transmitted, and the recovery of L_0 solely depends on the accumulation of the MI provided by the information embedded in L_1, which is not necessarily beneficial, given the limited number of transmission slots. However, this drawback turns into a benefit, when the E_b/N_0 value reaches higher levels, where L_0 can be readily recovered with the aid of L_1 and the remaining TSs can be saved for transmitting other layers for the sake of improving the video quality. The traditional THARQ scheme performs better than the IL-THARQ(n_0, n_1) schemes at low levels of E_b/N_0 because this scheme prioritizes the transmission of L_0 and indeed, recovers L_0 with a high probabiliry. However, this scheme is not so efficient at high E_b/N_0 values, because each layer is transmitted at least once, which is not always necessary in IL-based schemes, since the skipped layer can be recovered later using the information embedded into the other layers.

Compared to the benchmarks, our proposed IL-ATHARQ scheme results in an improved performance by virtue of its adaptive nature. As observed in Fig. 3.7a, the IL-ATHARQ scheme outperforms the traditional THARQ scheme all the way and achieves an E_b/N_0 reduction of about 3.8 dB at a PSNR of 38.5 dB. Alternatively, 1.8 dB of PSNR video quality improvement may be observed at an E_b/N_0 of 15 dB. Furthermore, IL-ATHARQ also outperforms most IL-THARQ(n_0, n_1) schemes, except for the IL-THARQ(0, 3) scheme, which shows an exceptionally good performance at sufficiently high E_b/N_0 values and slightly outperforms IL-ATHARQ. This may be due to the inaccuracy of the distortion estimation function invoked by IL-ATHARQ. Nonetheless, an approximately 1.9 dB of power reduction is achieved by the IL-ATHARQ arrangement compared to the IL-THARQ(0, 3) scheme at a PSNR of 38.5 dB. Alternatively, about 1.1 dB of PSNR video quality improvement may be observed at an E_b/N_0 of 15 dB, compared to the IL-THARQ(2, 1) scheme, which is the best performer amongst the IL-THARQ(n_0, n_1) schemes at an E_b/N_0 of 15 dB.

Similarly, observe from Fig. 3.7b that given $N_T = 4$, the IL-ATHARQ scheme outperforms the traditional THARQ arrangement and achieves an E_b/N_0 reduction of about 2.7 dB at a PSNR of 38.5 dB. Alternatively, about 1.9 dB of PSNR video quality improvement may be observed at an E_b/N_0 of 11 dB. IL-ATHARQ outperforms all of the IL-THARQ(n_0, n_1) schemes at all the E_b/N_0 values considered. More specifically, about 1.5 dB of power reduction is achieved by the IL-ATHARQ scheme compared to the IL-THARQ(1, 3) scheme at a PSNR of 38.5 dB. Alternatively, about 1.0 dB of PSNR video quality improvement may be observed at an E_b/N_0 of 11 dB compared to the IL-THARQ(1, 3) scheme, which is the best performance amongst the IL-THARQ(n_0, n_1) schemes at the E_b/N_0 of 11 dB. Generally speaking, the video performance gain becomes relatively modest upon increasing $N_T = 3$ to $N_T = 4$. This is because the adaptive scheduling of the layers becomes less important, when there are sufficient TSs.



(a) PSNR vs E_b/N_0 for ***Football*** *with* $\mathbf{N_T} = \mathbf{3}$



(b) PSNR vs E_b/N_0 for ***Football*** *with* $\mathbf{N_T} = \mathbf{4}$

Figure 3.7: PSNR versus E_b/N_0 performance of our proposed IL-ATHARQ system versus both the traditional THARQ transmission and the IL-HARQ scheme in conjunction with fixed transmission limits (IL-THARQ(n_0, n_1)) as benchmarks. The **Football sequence** is used for transmission over block-faded non-dispersive uncorrelated Rayleigh channels. The schematic of Fig. 3.1 and the parameters of Table 3.1 were used.

3.5.3 Optimized Coding Rates

In order to characterize the PSNR versus E_b/N_0 performance both of our proposed RO-IL-ATHARQ system and of the modified RO-IL-ATHARQ scheme, we compare them to two benchmarks, namely to a fixed-rate IL-ATHARQ scheme and to traditional THARQ transmission over block-fading non-dispersive uncorrelated Rayleigh channels, as listed in Table 3.3. These comparisons are shown in Fig. 3.8-3.13, which were carried out using three different video sequences, namely the Football, the Soccer and the Crew sequences, as listed in Table 3.2. Three different IL-ATHARQ transmission schemes denoted by $Rate\,(r_0, r_1)$ were simulated, namely the $Rate\,(1/2, 1/2)$, the $Rate\,(1/3, 1/2)$ and the $Rate\,(1/3, 1/3)$ schemes, where the $Rate\,(1/2, 1/2)$ scheme is the same as the IL-ATHARQ scheme we used in Section 3.5.2.

The results recorded for the Football sequence with the aid of three transmission TSs are shown in Fig. 3.8. It can be observed that all the fixed-rate IL-ATHARQ transmission schemes outperformed the pure THARQ transmission. The RO-IL-ATHARQ scheme outperforms all other schemes at high E_b/N_0 values, but its performance becomes inferior to the $Rate\,(1/3, 1/3)$ IL-ATHARQ scheme below the E_b/N_0 value of 10 dB, owing to the inaccurate distortion estimation. However, when the modified RO-IL-ATHARQ scheme associated with $\delta = 0.6$ is adopted, the system outperforms the other schemes at lower E_b/N_0 values, and achieves an E_b/N_0 reduction of about 5.3 dB at a PSNR of 38.5 dB, over the THARQ benchmark. This represents an additional 1.5 dB of E_b/N_0 reduction compared to the $Rate\,(1/2, 1/2)$ IL-ATHARQ scheme used in Section 3.5.2. Alternatively, about 2.5 dB of PSNR video quality improvement may be observed at an E_b/N_0 of 15 dB, which is an additional 0.7 dB of PSNR improvement compared to the $Rate\,(1/2, 1/2)$ IL-ATHARQ scheme.

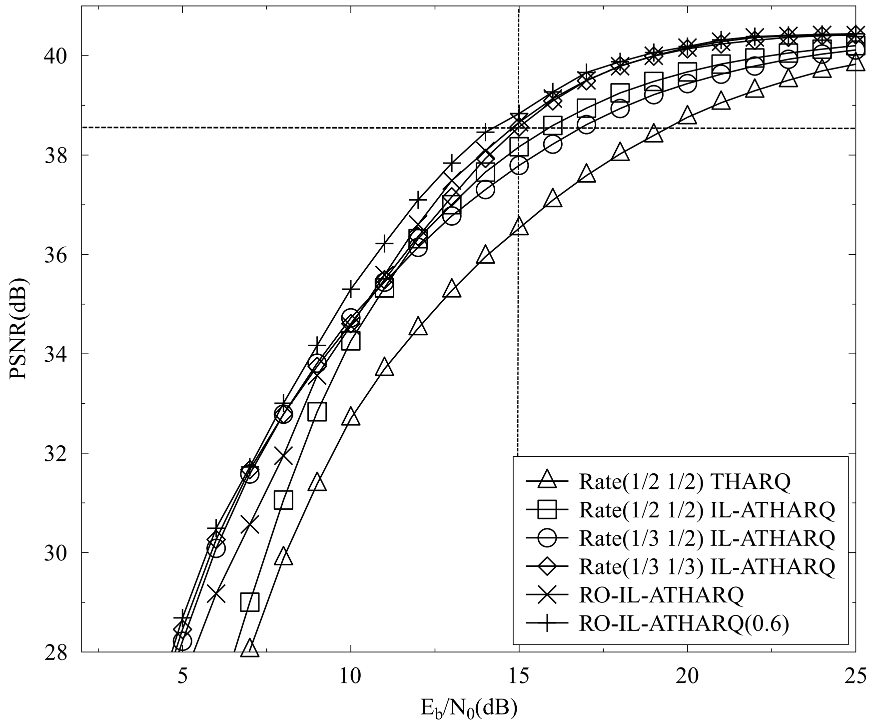


Figure 3.8: PSNR vs E_b/N_0 performance of our proposed RO-IL-ATHARQ system and of the modified RO-IL-ATHARQ scheme in comparison to both the IL-ATHARQ transmission and to the traditional THARQ transmission as benchmarks. The ***Football* sequence** is used for transmission over block-fading non-dispersive uncorrelated Rayleigh channels, and $\mathbf{N_T = 3}$.. The schematic of Fig. 3.1 and the parameters of Table 3.1 were used.

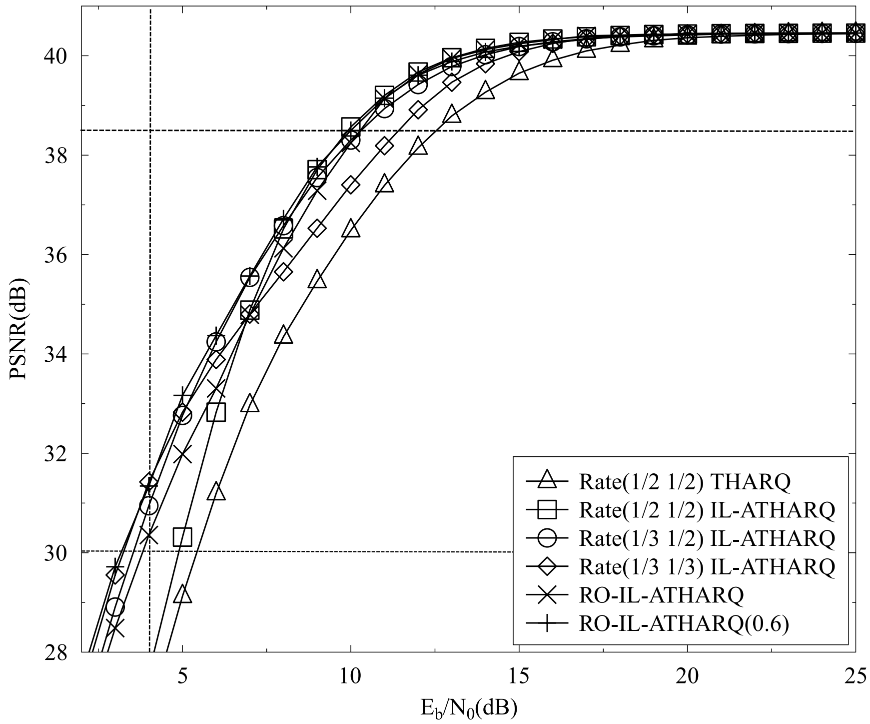


Figure 3.9: PSNR vs E_b/N_0 performance of our proposed RO-IL-ATHARQ system and of the modified RO-IL-ATHARQ scheme in comparison to both the IL-ATHARQ transmission and to the traditional THARQ transmission as benchmarks. The ***Football* sequence** is used for transmission over block-fading non-dispersive uncorrelated Rayleigh channels, and $\mathbf{N_T = 4}$. The schematic of Fig. 3.1 and the parameters of Table 3.1 were used.

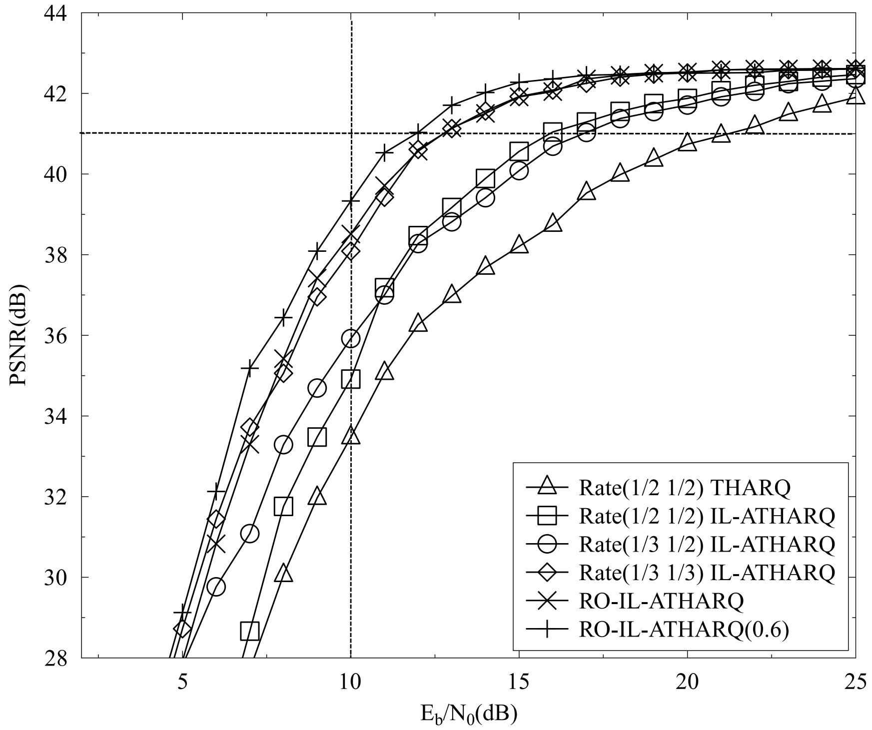


Figure 3.10: PSNR vs E_b/N_0 performance of our proposed RO-IL-ATHARQ system and of the modified RO-IL-ATHARQ scheme in comparison to both the IL-ATHARQ transmission and to the traditional THARQ transmission as benchmarks. The ***Soccer* sequence** is used for transmission over block-fading non-dispersive uncorrelated Rayleigh channels, and $\mathbf{N_T = 3}$. The schematic of Fig. 3.1 and the parameters of Table 3.1 were used.

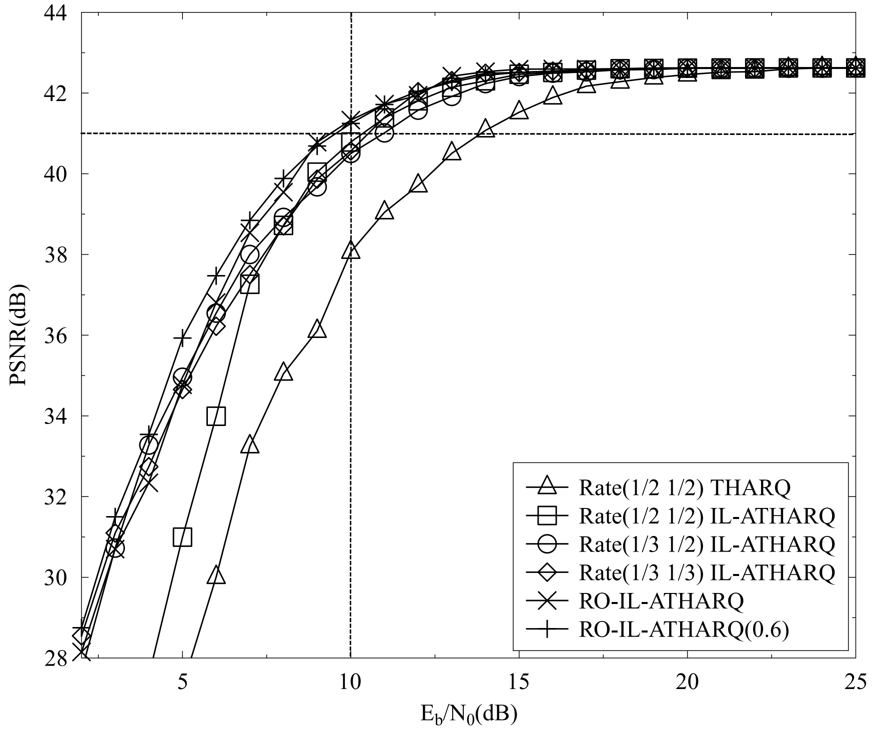


Figure 3.11: PSNR vs E_b/N_0 performance of our proposed RO-IL-ATHARQ system and of the modified RO-IL-ATHARQ scheme in comparison to both the IL-ATHARQ transmission and to the traditional THARQ transmission as benchmarks. The ***Soccer* sequence** is used for transmission over block-fading non-dispersive uncorrelated Rayleigh channels, and $\mathbf{N_T = 4}$. The schematic of Fig. 3.1 and the parameters of Table 3.1 were used.

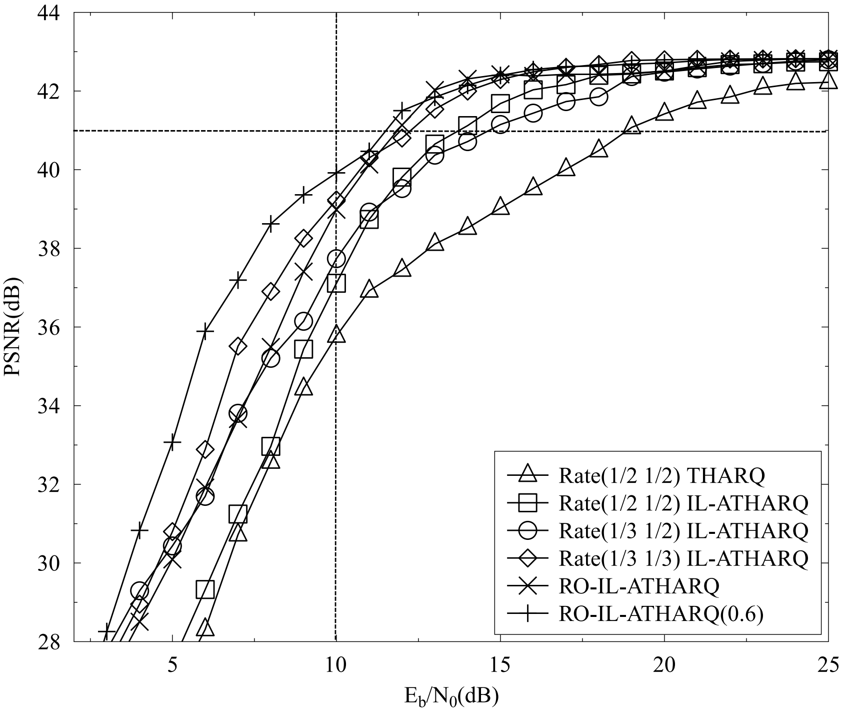


Figure 3.12: PSNR vs E_b/N_0 performance of our proposed RO-IL-ATHARQ system and of the modified RO-IL-ATHARQ scheme in comparison to both the IL-ATHARQ transmission and to the traditional THARQ transmission as benchmarks. The ***Crew* sequence** is used for transmission over block-fading non-dispersive uncorrelated Rayleigh channels, and $\mathbf{N_T = 3}$. The schematic of Fig. 3.1 and the parameters of Table 3.1 were used.

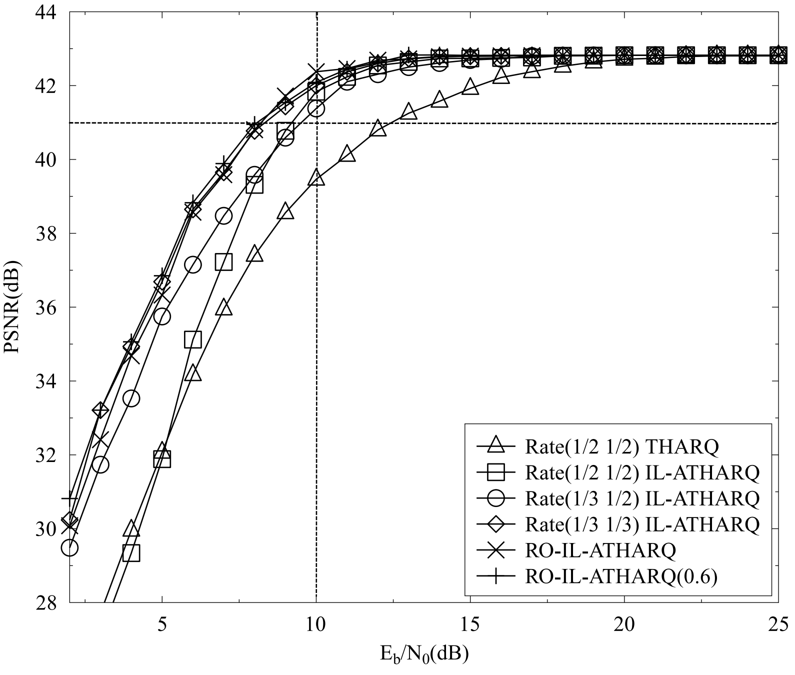


Figure 3.13: PSNR vs E_b/N_0 performance of our proposed RO-IL-ATHARQ system and of the modified RO-IL-ATHARQ scheme in comparison to both the IL-ATHARQ transmission and to the traditional THARQ transmission as benchmarks. The ***Crew* sequence** is used for transmission over block-fading non-dispersive uncorrelated Rayleigh channels, and $\mathbf{N_T = 4}$. The schematic of Fig. 3.1 and the parameters of Table 3.1 were used.

Given $N_T = 4$, Fig. 3.9 shows a slightly different trend, where the different fixed-rate IL-ATHARQ schemes exhibit a better performance over certain E_b/N_0 regions. Specifically, the *Rate* $(1/2, 1/2)$ scheme outperforms the rest for E_b/N_0 values above 8 dB, while the *Rate* $(1/3, 1/3)$ scheme performs better below E_b/N_0 of 5 dB and the *Rate* $(1/3, 1/2)$ regime excels in the region between 5 and 8 dB. Again, we can observe that the modified RO-IL-ATHARQ scheme associated with $\delta = 0.6$ outperforms all the fixed-rate IL-ATHARQ schemes across the entire E_b/N_0 region we are interested in. It achieves a similar E_b/N_0 reduction as the *Rate* $(1/2, 1/2)$ IL-ATHARQ at a PSNR of 38.5 dB. However, a 2.4 dB of E_b/N_0 reduction is observed at a PSNR of 30 dB, while only 0.5 dB of E_b/N_0 reduction is attained by the *Rate* $(1/2, 1/2)$ IL-ATHARQ. Alternatively, about 4.7 dB of video PSNR improvement may be observed at an E_b/N_0 of 4 dB, while only 0.8 dB of PSNR improvement is achieved by the *Rate* $(1/2, 1/2)$ IL-ATHARQ.

Similar trends can be observed, when the Soccer or Crew sequences are used, as shown in Fig. 3.10 to Fig. 3.13. We infer from these results that our RO-IL-ATHARQ scheme is applicable to video sequences of diverse natures, and it is capable of achieving a beneficial performance gain for both $N_T = 3$ and 4. The subjective comparison of the decoded videos associated with our different regimes is discussed in Section 3.5.4.

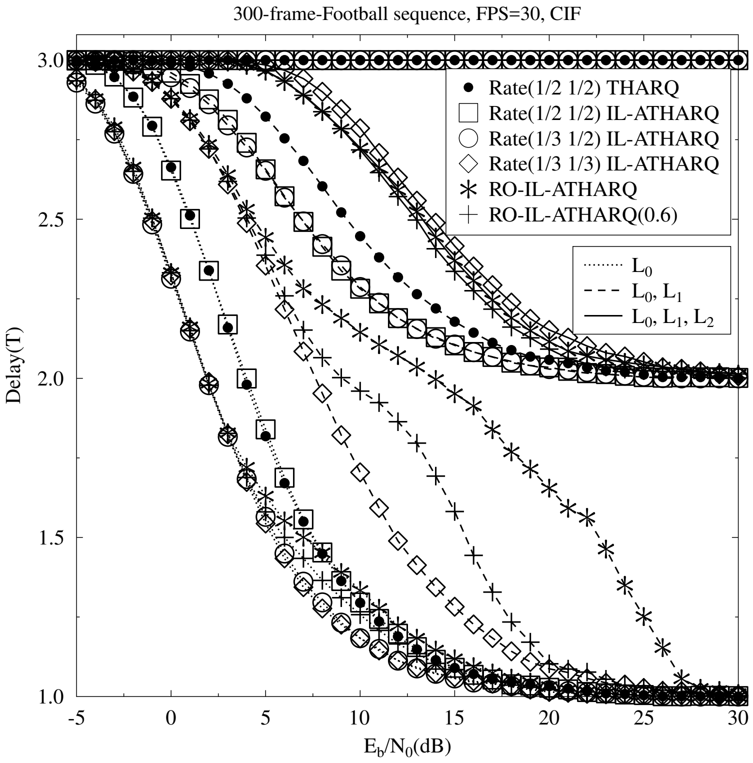
3.5.4 Subjective Comparison

Explicitly, Fig. 3.14 shows the subjective comparison of the decoded video frames associated with our different regimes using the Soccer sequence and $N_T = 3$ at the E_b/N_0 value of 10 dB. The 26-th frame of the recovered videos of some of our schemes are shown in the top row of Fig. 3.14. The rate$(1/2, 1/2)$ IL-THARQ(1 1) scheme is more error-prone according to Section 3.5.2, and in this regime all three layers of this frame failed to be recovered, and so did all their preceding frames. The difference frame, which is obtained by subtracting the recovered frame from the 26-th frame of the original video, has substantial non-zero values. Continuing from left to right, we can observe that the frames corresponding to the rate$(1/2, 1/2)$ THARQ scheme, the rate$(1/2, 1/2)$ IL-ATHARQ scheme and the RO-IL-ATHARQ(0.6) scheme are becoming sharper and containing more intricate video details, while the corresponding difference frames having less and less non-zero values, which indicates the improvement of the video quality.

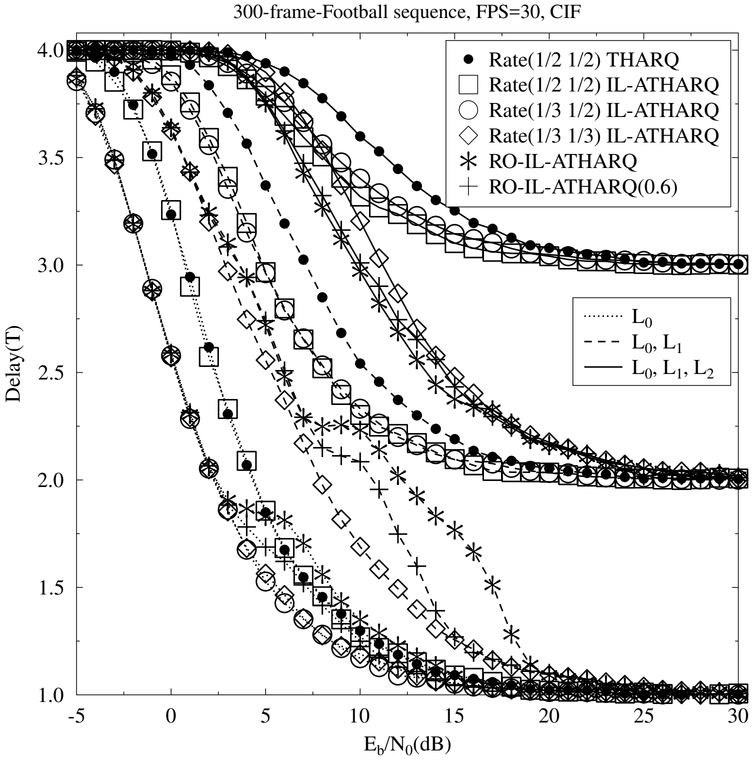
3.5.5 Transmission Delay

Figure 3.14: Comparison of decoded frames of the 26-th frame at E_b/N_0 of 10 dB for the *Soccer* sequences and $N_T = 3$. The upper five columns (from left to right) indicate frames of the original video, the rate($1/2, 1/2$) IL-THARQ(1 1) scheme, rate($1/2, 1/2$) THARQ scheme, the rate($1/2, 1/2$) IL-ATHARQ scheme and the RO-IL-ATHARQ(0.6) scheme, respectively. The lower row correspond to the difference frames between the top ones and the original video frame.

In Fig. 3.15, the average number of TSs required for receiving the i-th layer employing various transmission schemes is displayed. When $N_T = 3$ is used, the average number of TSs versus the E_b/N_0 characteristics are shown in Fig. 3.15a. When the *Rate* ($1/2, 1/2$) THARQ, *Rate* ($1/2, 1/2$) IL-ATHARQ and *Rate* ($1/3, 1/2$) IL-ATHARQ schemes are used along with $N_T = 3$, three TSs are occupied, regardless of the E_b/N_0 value. On the other hand, observe in Fig. 3.15a that the *Rate* ($1/3, 1/3$) IL-ATHARQ, RO-IL-ATHARQ and RO-IL-ATHARQ(0.6) schemes only use two TSs on average, in order to successfully receive all transmissions at high E_b/N_0 values. Furthermore, as seen in Fig. 3.15a, both the RO-IL-ATHARQ(0.6) and RO-IL-ATHARQ



(a) Average time slots used vs E_b/N_0 for *Football*, $N_T = 3$



(b) Average time slots used vs E_b/N_0 for *Football*, $N_T = 4$

Figure 3.15: Average transmission time slots versus E_b/N_0 performance comparison of the THARQ, IL-ATHARQ, RO-IL-ATHARQ, and modified RO-IL-ATHARQ schemes for the *Football* sequence, transmitted over the quasi-static non-dispersive uncorrelated Rayleigh fading wireless channels. The schematic of Fig. 3.1 and the parameters of Table 3.1 were used.

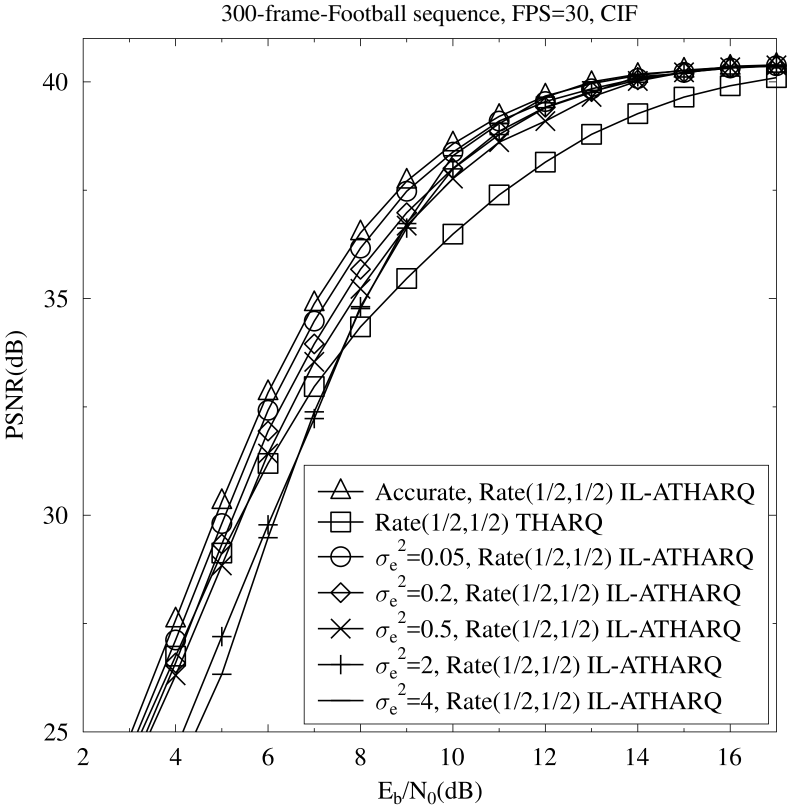
schemes require slightly less transmission TSs than the *Rate* $(1/3, 1/3)$ IL-ATHARQ at the same E_b/N_0 value. It can also be observed in Fig. 3.15a that for the first two layers, namely for L_0 and L_1, the *Rate* $(1/2, 1/2)$ THARQ, *Rate* $(1/2, 1/2)$ IL-ATHARQ and *Rate* $(1/3, 1/2)$ IL-ATHARQ schemes require two TSs on average at sufficiently high E_b/N_0 values, while the *Rate* $(1/3, 1/3)$ IL-ATHARQ, RO-IL-ATHARQ and RO-IL-ATHARQ(0.6) schemes only need one TS. It can also be inferred from Fig. 3.15a that the rate-optimized schemes occupy more TSs by successfully conveying L_0 and L_1. The reason behind this phenomenon is that the optimization algorithm strikes a more balanced compromise instead of assigning all the resources for protecting L_0 and L_1, where a reasonable reduction of the protection of L_1 can be compensated by successfully decoding both L_1 and L_2 in a single reception, if the latter one is well protected and ends up possessing high MI values. Finally, if we consider the transmission of L_0, we find from Fig. 3.15a that the *Rate* $(1/3, 1/2)$ and *Rate* $(1/3, 1/3)$ IL-ATHARQ, as well as the RO-IL-ATHARQ and RO-IL-ATHARQ(0.6) generally necessitates less transmission TSs, than the *Rate* $(1/2, 1/2)$ THARQ and *Rate* $(1/2, 1/2)$ IL-ATHARQ.

In Fig. 3.15b, the average number of TSs used versus the E_b/N_0 is portrayed for $N_T = 4$. Similar trends can be observed to those recorded in Fig. 3.15a for $N_T = 3$. Hence we conclude that the IL-ATHARQ is capable of efficiently reducing the number of TSs required for transmission, and the RO-IL-ATHARQ, although optimized for minimum distortion, additionally occupies less TSs.

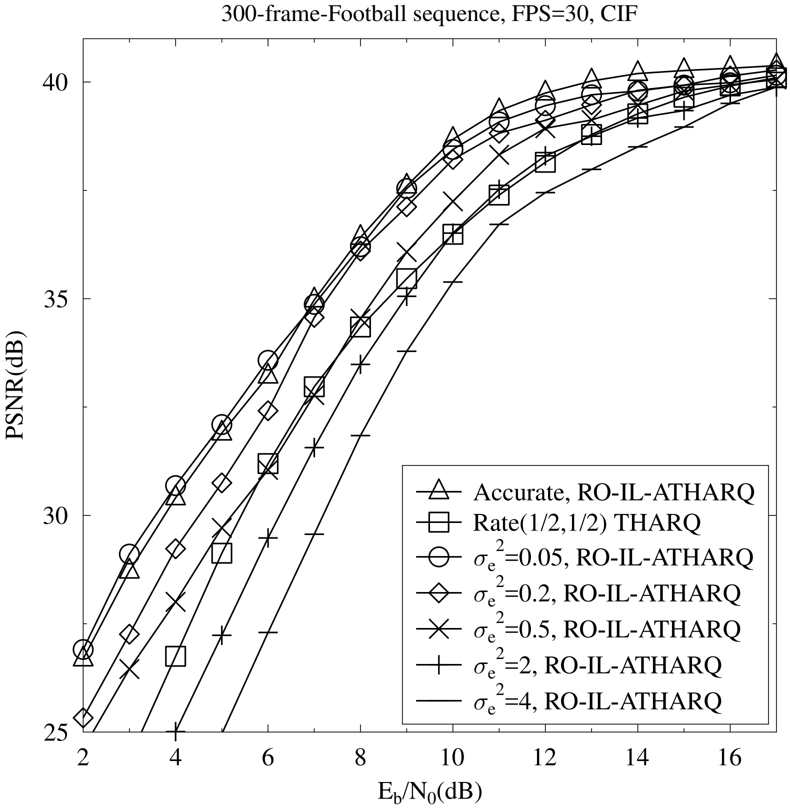
3.5.6 Effect of Channel Prediction Errors

To demonstrate the effect of the channel prediction errors on the performance of our proposed system, we include the simulation results for both the IL-ATHARQ and RO-IL-ATHARQ(0.6) transmission schemes contaminated by channel prediction errors [230] in Fig. 3.16, which obeyed a Gaussian distribution.

As shown in Fig. 3.16a, the performance of the *Rate* $(1/2, 1/2)$ IL-ATHARQ degrades with the increase of σ_e^2 of the channel prediction error. The PSNR performance was affected predominantly in the lower E_b/N_0 range by the channel prediction error. For example, at the E_b/N_0 of 6 dB, the PSNR associated with $\sigma_e^2 = 0.5$ is 1.35 dB lower than the one relying on perfect channel prediction, while the system's performance with $\sigma_e^2 = 4$ is 3.35 dB worse. At the E_b/N_0 of 11 dB, the PSNR performance associated with $\sigma_e^2 = 0.5$ is 1.5 dB worse than the one with perfect channel prediction, while that in conjunction with $\sigma_e^2 = 4$ is 1 dB worse. The IL-ATHARQ schemes still exhibit a performance gain over the THARQ benchmark system for E_b/N_0 values above 8 dB.



(a) PSNR vs E_b/N_0 with IL-ATHARQ for *Football*, $N_T = 3$



(b) PSNR vs E_b/N_0 with RO-IL-ATHARQ for *Football*, $N_T = 4$

Figure 3.16: PSNR versus E_b/N_0 performance in the presence of channel prediction errors, which affected the *Rate* (1/2, 1/2) IL-ATHARQ, RO-IL-ATHARQ system for different σ_e^2 values against that of the THARQ system used as the benchmark. The schematic of Fig. 3.1 and the parameters of Table 3.1 were used.

As for the performance of the RO-IL-ATHARQ system, we can observe in Fig. 3.16b that the channel prediction error affected the systems more severely than for the IL-ATHARQ systems. At the E_b/N_0 of 6 dB, the PSNR performance associated with $\sigma_e^2 = 0.5$ is 2.4 dB worse than that of perfect channel prediction, while the system performance relying on $\sigma_e^2 = 4$ is 6.1 dB worse. At the E_b/N_0 of 11 dB, the PSNR performance of $\sigma_e^2 = 0.5$ is 1 dB lower than the one with perfect channel prediction, while that associated with $\sigma_e^2 = 4$ is 2.6 dB worse. The system's performance recorded for $\sigma_e^2 = 2$ or $\sigma_e^2 = 4$ is even worse than that of the THARQ benchmark system due to the severe error propagation imposed by the channel prediction. We may observe that in Fig. 3.16b the PSNR performance achieved with the aid of perfect channel prediction is not substantially better than the one associated with $\sigma_e^2 = 0.05$, when we have $E_b/N_0 \leq 6$ dB. Recall that the unmodified RO-IL-ATHARQ scheme of Section 3.4 is subject to a certain level of PER estimation errors introduced by the algorithm itself and it is even more so in conjunction with larger N_T values because of the error propagation. The effect of the channel estimation error may be deemed comparable to that of the PER estimation errors, provided that it is not excessive. The interaction of these two types of errors may not be additive.

3.6 Conclusions

We conceived an adaptive THARQ algorithm for IL-FEC coded layered video streaming for the sake of minimizing the video distortion under the constraint of a given total number of transmission TSs. The adaptive retransmission controller predicts the channel conditions and estimates the SNR values at the receiver for the sake of appropriately configuring the transmitter with the aid of a feedback message. The specific video layer, which would most effectively reduce the video distortion at the receiver is chosen for transmission.

Our simulation results of Section 3.5 firstly showed the performance gain of our IL-ATHARQ algorithm, by comparing its PSNR performance to that of the aforementioned traditional THARQ as well as to that of the IL-THARQ scheme relying on fixed transmission limits, using the Football sequence and the parameters listed in Table 3.2. Specifically, our simulation results of Fig. 3.7a show that the IL-ATHARQ scheme outperforms the traditional THARQ scheme all the way and achieves an E_b/N_0 reduction of about 3.8 dB at a PSNR of 38.5 dB in conjunction with $N_T = 3$. Alternatively, a 1.8 dB of PSNR video quality improvement may be observed in Fig. 3.7a at an E_b/N_0 of 15 dB. Given $N_T = 4$, the IL-ATHARQ scheme outperforms the traditional THARQ arrangement and achieves an E_b/N_0 reduction of about 2.7 dB at a PSNR of 38.5 dB. Alternatively, about 1.9 dB of PSNR video quality improvement may be observed in Fig. 3.7b at an E_b/N_0 of 11 dB. The E_b/N_0 gains obtained using the aforementioned schemes compared to traditional THARQ

Transmission Schemes	$N_T = 3$	$N_T = 4$
IL-THARQ(0 2)	-1.8 dB	-8.8 dB
IL-THARQ(0 3)	1.9 dB	-4 dB
IL-THARQ(0 4)	2 dB	-1.8 dB
IL-THARQ(1 1)	2 dB	-3 dB
IL-THARQ(1 2)	1.9 dB	0.2 dB
IL-THARQ(1 3)	na	1.2 dB
IL-THARQ(2 1)	1.9 dB	na
IL-THARQ(2 2)	na	0.5 dB
IL-THARQ(2 3)	0 dB	na
IL-THARQ(3 1)	1.9 dB	1.2 dB
IL-ATHARQ	3.8 dB	2.7 dB

Table 3.5: E_b/N_0 gains obtained by our IL-ATHARQ algorithm as well as the IL-THARQ scheme relying on fixed transmission limits, compared to traditional THARQ at a PNSR of 38.5 dB, using the Football sequence and parameters as listed in Table 3.2.

at a PNSR of 38.5 dB are listed in Tab. 3.5, using the Football sequence and the parameters specified in Table 3.2

Furthermore, in Section 3.4 we developed an on-line optimization technique for our IL-ATHARQ transmission scheme of Fig. 3.1, in order to find the most beneficial FEC code rate for each of the video layers that results in a reduced video distortion. A method of estimating the video distortions related to each code rate assignment was conceived for the IL-ATHARQ transmission.

Our simulation results of Fig. 3.8-3.13 demonstrated that the proposed RO-IL-ATHARQ system and the modified RO-IL-ATHARQ scheme of Fig. 3.1 outperformed the benchmarkers shown in Table 3.3, namely three fixed-rate IL-ATHARQ schemes and the traditional THARQ for transmission over block-fading non-dispersive uncorrelated Rayleigh channels, using the parameters listed in Table 3.2. Specifically, our proposed system of Fig. 3.8 using the modified RO-IL-ATHARQ scheme of Section 3.4 outperforms the traditional THARQ system by an E_b/N_0 value of about 5.3 *dB* at a PSNR of 38.5 *dB*, given $N_T = 3$ and using the Football sequence. Alternatively, an approximately 2.5 dB of PSNR video quality improvement may be observed in Fig. 3.8 at an E_b/N_0 of 15 dB, when employing a RSC code. The E_b/N_0 gains obtained using the aforementioned schemes over traditional THARQ at a PNSR of 38.5 dB are listed in Tab. 3.6, using three video sequences and the parameters specified in Table 3.2.

Finally, we characterized the effects of the delay in Fig. 3.15 and found that the IL-ATHARQ is capable of efficiently reducing the number of TSs required for transmission, and that the RO-IL-ATHARQ additionally occupies a reduced number of TSs. We also characterized the effects of channel quality prediction errors on the

Transmission Schemes	Sequences					
	Football	Soccer	Crew	Football	Soccer	Crew
	$N_T = 3$			$N_T = 4$		
Rate(1/2 1/2) IL-ATHARQ	3.8 dB	3.4 dB	3 dB	2.8 dB	2.5 dB	1.3 dB
Rate(1/3 1/2) IL-ATHARQ	2.5 dB	3 dB	3.2 dB	2 dB	3 dB	1.8 dB
Rate(1/3 1/3) IL-ATHARQ	4.8 dB	5.2 dB	4.5 dB	1.1 dB	2.5 dB	2.9 dB
RO-IL-ATHARQ	4.9 dB	5.7 dB	4.1 dB	2 dB	3.4 dB	2.9 dB
RO-IL-ATHARQ (0.6)	5.3 dB	6.2 dB	6.1 dB	2.8 dB	3.6 dB	3 dB

Table 3.6: E_b/N_0 gains obtained by our proposed RO-IL-ATHARQ system and the modified RO-IL-ATHARQ scheme, as well as the systems specifications shown in Table 3.3, namely three fixed-rate IL-ATHARQ schemes, compared to the traditional THARQ at a PNSR of 38.5 dB. All tested systems assume block-fading non-dispersive uncorrelated Rayleigh channels, using parameters as listed in Table 3.2.

attainable system performance in Fig. 3.16 to demonstrate the robustness of our system and quantified the affordable range of prediction errors.

Chapter 4

Hierarchical Colour-Shift-Keying Aided Layered Video Streaming

4.1 Introduction

Given the growing popularity of indoor multimedia applications, such as Voice over IP (VoIP), smart TV and online video games, video streaming substantially contributes to the exponentially increased amount of mobile data traffic. As a counter-measure, Heterogeneous Networks (HetNet) relying on smaller cells are relied upon, in order to achieve a higher system capacity for coping with the growing demands. However, the denser reuse of spectral resources will potentially impose an increased amount of interference [233]. In the light of these problems, Visible Light Communications (VLC) [130,234,235,127,236] has emerged as an appealing access method to provide extra license-free bandwidth, as a compliment to WiFi, femtocells, ultra wideband (UWB) and mm-wave communications in order to support indoor communications. Specifically, VLC supports high-rate downlink communications, in addition to its primary function of illumination in indoor environments. The recent IEEE standard 802.15.7 [127] on VLC also paves the way for wider commercial exploitation of VLC.

As reviewed in Chapter 1, colour-shift keying (CSK) relies on multiple light sources and photo-detectors corresponding to different colours, as introduced in the IEEE 802.15.7 standard, along with on-off Keying (OOK) and variable pulse position modulation (VPPM). Although the basic CSK modulation scheme was conceived in [127], recent research optimized it further for specific scenarios [193,202,195,196,203]. Drost and Sadler [202] examined the design of CSK signaling constellations conceived for an additive white Gaussian noise channel (AWGN) and optimized the location of the constellation points. In their follow-on work [195], VLC systems employing an arbitrary number of LEDs were investigated and they pre-compensated the channel effects at the transmitter. A new CSK modulation format based on four colours was presented in [196] by Singh *et al.*, while Monteiro and Hranilovic [194] set out to optimize high-order CSK constellations for specific communication channels. Finally, different signal labelling strategies were designed in [237] for diverse color constellations and detection schemes in order to improve the attainable Bit Error Ratio (BER) performance.

Again, in order to satisfy the escalating mobile data demands, efficient video streaming should be considered for the sake of conceiving sophisticated transmission schemes. Layered video coding [28] is a widely used scheme designed for employment in heterogeneous networking problems. The concept of layered video coding relies on the provision of multiple layers of different importance, namely the base layer (BL), which conveys the most important video bits required for representing a low-resolution video, as well as the enhancement layers (ELs), which provide additional video quality refinements, when a higher channel quality allows their transmission. Therefore, layered video coding is capable of supporting different-resolution terminals having different quality-requirements. It is also suitable for progressive reception of video steams according to the diverse network and channel conditions of different users. Therefore layered video coding is widely supported by the popular video standards [238, 31, 11, 10, 32]. Specifically, partitioned video coding [10] was provided by H.264 in order to generate multiple layers of different error-sensitivity. The Moving picture expert group (MPEG) provided the so-called multiview profile (MVP) [31], which generates different-angle encoded camera-views as different layers. Scalable video coding (glsSVC) [11, 10] constitutes an extension of the H.264/AVC standard [10], and it generates an encoded stream containing multiple inter-dependent layers. Finally, an extension referred to as scalable high-efficiency video coding (SHVC) [239, 240] is being developed for the state-of-the-art high efficiency video coding (HEVC)/H.265 scheme [238], in order to support flexible scalability.

Due to the fact that the different video layers typically exhibit a distinct level of importance in terms of their contributions to the quality of the recovered video, unequal error protection (UEP) [79], as reviewed in Chapter 1, constitutes a promising technique of conveying the video streams through noisy channels. More specifically, UEP can be achieved by using specific forward error correction (FEC) schemes carefully combined with particular modulation or coded modulation schemes. In the context of UEP schemes, either the packet-level regimes [86,88,89,90,91,241,73,242] or the bit-level schemes [92,93,94,95,96,73,97,98,243] may be adopted. The packet-level contributions of [84, 85, 86, 87, 88, 89, 90, 91] refer to the schemes, where hard decoded FEC codes, e.g. Reed-Solomon (RS) codes or Luby transform (LT) codes are invoked for mitigating the packet loss events at the application layer [62]. By contrast, the bit-level arrangements operate at the physical layer and rely on soft-decoded FEC codes for correcting bit-errors in wireless scenarios [64,98,124]. Specifically, the authors of [98] proposed a bit-level inter-layer coded FEC (IL-FEC) scheme that embeds the BL into the FEC coded ELs, so that the reception of the BL can be improved with the aid of the ELs using soft decoding.

To elaborate a little further, UEP can also be achieved by appropriately designing modulation and coded modulation schemes [100,101,60,102,103,104,94,105,75,106, 78,71,244,245,107,67,68,108,76]. Both multilevel coded (MLC) modulation [60,102, 103,104,94,105] and hierarchical modulation (HM) [75,106,78,71,244,245,107,67, 68,108,76] have attracted substantial research attention, albeit the latter one can be considered as a special subset of the former one. Multilevel coded modulation was originally proposed by Calderbank and Seshadri for providing UEP [60], where the 2-dimensional signal constellation is partitioned into disjoint subsets and the bit streams of different importance are carefully mapped to the different-integrity bit-positions of the constellation subsets. Later the theoretical upper bounds and the computer simulation-based performance of multilevel block coded modulations designed for UEP and multistage decoding were presented in [102,103]. Aydinlik and Salehi [94], as well as Kim *et al.* [104] improved the set partitioning of QAM used by Trellis Coded Modulation (TCM) to achieve an improved UEP performance. Finally, a multilevel space-shift-keying scheme was designed by Zamkotsian *et al.* [105] to provide UEP for transmission over Rician fading channels.

Again, HM is another popular UEP technique that has already been widely accepted by industry and hence became an integral part of the DVB-T/-H standard [109, 110]. Similar to MLC modulation, the symbols of the HM schemes are carefully partitioned into different-integrity groups. Since the different bits in a symbol are used for selecting specific sub-groups of the constellation points, they have different error probabilities. HM has also been investigated in the context of cooperative communications, where the different-priority layers can be routed via different paths [248,112,111,249].

Numerous systems have been developed for UEP by exploiting the properties of HM schemes. To elaborate, an UEP system was proposed by Chang *et al.* [75] to support multiple protection levels, instead of the traditional twin-level designs. HM based on QAM was investigated as the UEP modulation method [250]. The authors of [245,106,107] proposed HM-QAM-style MIMO systems and analysed the attainable UEP properties. However, the authors of [245, 106, 107] carried out their investigations without actually considering multimedia sources. By contrast, the authors of [244] jointly optimised the channel coding parameters and the HM parameters in a coded-HM-aided progressive image transmission system, in order to minimize the video distortion. The streaming of partition-mode H.264/AVC [10] coded video was considered in [67] and UEP was provided by the HM-QAM scheme for protecting the video partitions of different importance. Similarly, H.264 coded video was also considered by Chang *et al.* [68], where the intra-coded frames (I-frame) and predictive

Year	Authors	Contributions
1993	Calderbank *et al.* [60]	first proposed multilevel coded modulation for providing UEP.
1995	Morimoto *et al.* [246]	designed a hierarchical modulation scheme for satellite communication.
1996	Morimoto *et al.* [247]	proposed hierarchical image transmission for multimedia mobile communications.
2000	Morelos *et al.* [103]	presented the theoretical upper bounds of multilevel block coded modulation designed for UEP and multistage decoding.
	Isaka *et al.* [102]	
2001	Kim and Pottie [104]	improved the set partitioning of QAM used by TCM to improve UEP performance.
2005	Barmada *et al.* [67]	proposed HM-QAM UEP for the streaming of partitioned H.264/AVC [10] coded video.
2006	Hossain *et al.* [111]	proposed an UEP scheme for voice and data transmission over fading channels using adaptive hierarchical modulation.
	Chang *et al.* [68]	designed an UEP scheme for H.264 coded video streaming, where I-frame and P-frame protected using different levels of protection provided by adaptive HM-QAM.
2007	Noh *et al.* [245]	proposed UEP-aided MIMO-OFDM systems relying on hierarchical signal constellations.
2008	Aydinlik and Salehi [94]	improved the set partitioning of QAM to achieve an improved UEP performance.
2009	Chang *et al.* [248]	derived the explicit closed-form expressions of the BER for cooperative communication systems combined with HM.
2010	Li *et al.* [71]	conceived an UEP scheme for transmitting H.264 coded video over frequency selective fading channels, using OFDM-based HM-QAM.
2011	Arslan *et al.* [244]	optimised the channel coding parameters and the HM parameters in a coded HM-aided progressive image transmission system.
2012	Chang *et al.* [106]	proposed HM-based space shift keying (SSK)-type modulation for UEP.
	Alajel *et al.* [76]	proposed HM-QAM for depth-map based stereoscopic video streaming.
2014	Zamkotsian *et al.* [105]	designed a multilevel space-shift-keying scheme to provide UEP for transmission over Rician fading channels.
	Quazi and Xu [107]	proposed HM-QAM-style MIMO systems and analysed the attainable UEP properties.
2015	Sun *et al.* [112]	investigated triple-layer turbo-trellis-coded HM in the context of cooperative communications.

Table 4.1: Major contributions on UEP schemes using MLC or HM

coded frames (P-frame) were protected using different levels of protection provided by adaptive HM-QAM. Li *et al.* [71] conceived an UEP scheme for transmitting H.264 coded video over frequency selective fading channels, using OFDM-based HM-QAM for carefully mapping the video bits to the OFDM-subcarriers to provide the video layers of high importance with better error protection. Finally, depth-map based stereoscopic video streaming was considered using HM 16-QAM in [76]. Since the color component of the video sequence has a more significant impact on the reconstructed video quality than the depth map, the color-component of the 3D video was mapped to the layers of better protection in the HM-QAM scheme. A brief review of the major contributions on UEP schemes using MLC and HM can be found in Table 4.1.

Although VLC offers a large bandwidth for data transmission [251], and the current standard supports a data rate of 48 Mb/s (or 384 Mbps) when CSK is combined with channel coding [127], there is a surging demand for higher data transmission rates for applications like the streaming of 4K or higher-resolution videos, which may have data-rates of hundreds of Mbps [252]. Additionally, multiple-user and multiple-source scenarios [253, 254, 255] also require higher data rates. Finally, the received SNR of the indoor VLC systems dramatically drops, when the receiver does not have a clear line of sight to the lighting system [256,251]. Therefore, it is necessary to protect the system against any video performance degradation due to dropping some of the video-layers during indoor VLC transmission. In order to maintain the maximum possible video quality in the presence of error-sensitive video data, UEP is adopted by using hierarchical modulation. Furthermore, in this chapter the philosophy of our enabling technique can also be adopted by millimetre wave based transmission systems.

Despite the rich research efforts invested in optimizing the CSK constellation for reducing the BERs, no efforts have been dedicated to investigating the feasibility of UEP CSK schemes. Furthermore, no CSK-aided video systems have been disseminated in the open literature. Hence our inspiration is to close this open problem.

Explicitly, the rationale and novelty of this chapter is summarized as follows.

1. We conceive a hierarchical CSK (HCSK) modulation constellation that can be readily configured to provide different error probabilities for the different video layers. We are the first to introduce HM into the conventional CSK constellations [193, 202, 195, 196, 203] and to design an UEP scheme based on it.
2. We provide a HCSK design example for transmitting scalable video sources relying on the RSC code as the FEC.
3. Our optimisation metric is the quality of the received video. We will demonstrate that the proposed system outperforms the traditional EEP scheme by

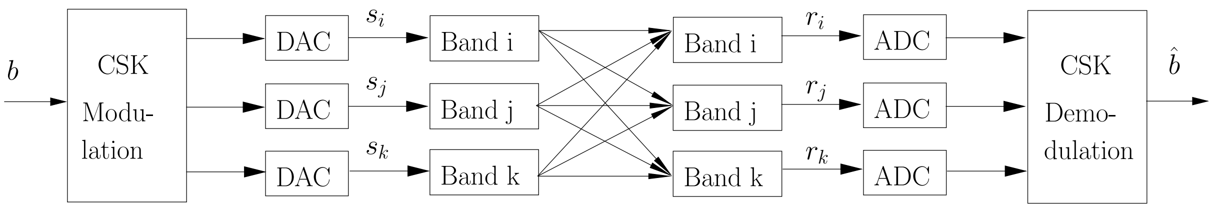


Figure 4.1: Transceiver architecture of a basic M-CSK transceiver

about 1.7 dB of optical SNR at a video peak signal-to-noise ratio (PSNR) of 37 dB and up to 6.5 dB at a PSNR of 36 dB.

The rest of this chapter is organized as follows. Section 4.2 introduces the CSK transmitter and receiver model, including the signal constellation and coding, the optical domain channel model as well as the soft demodulation method. We then introduce the construction of our novel M-HCSK modulation and its layered demodulation method in Section 4.3. We continue by optimizing a UEP system using M-HCSK-RSC transmission to convey scalable video sequences, as detailed in Section 4.4. The performance of our optimised M-HCSK-RSC video transmission system using different HCSK constellation sizes is compared to the relevant benchmarker schemes in Section 4.5 using different video sequences. Finally, we conclude in Section 4.6.

4.2 CSK Modulation

In this section we will briefly describe the basics of CSK modulation that are associated with the transceiver architecture of Fig. 4.1. Specifically, we will describe the constellation of the CSK modulation, the channel model of the optical-domain propagation and the corresponding demodulation methods. The symbols to be used in this chapter is listed in Table 4.2.

4.2.1 CSK Constellation and Coding

As shown in Fig. 4.1, the bit sequence $\mathbf{b}$ is modulated on to the CSK symbols. For the M-ary CSK modulation, every group of $N_b = \log_2 M$ incoming bits, denoted by $\mathbf{b}_s$, is mapped to an M-ary CSK symbol. According to the IEEE 802.15.7 standard [127], the CSK symbols are modulated using the combination of three distinct sources of visible lights, having different wavelength/frequencies. Therefore the CSK symbols are represented by the three-dimensional vectors containing the intensities of the three components of the light. Explicitly, given a CSK symbol $\mathbf{s}$, it can be represented by $\mathbf{s} = [s_i, s_j, s_k]$, $\mathbf{s} \in \mathcal{S}$, where each element represents the power of the corresponding light source and $\mathcal{S} = \{\mathbf{s}_0, \mathbf{s}_1, \ldots, \mathbf{s}_{M-1}\}$ is the collection of all the M legitimate symbols in M-ary CSK. According to the standard [127], the output light intensity

	Description
M	Used to define M-ary CSK modulation
$\mathbf{b}$	The bit sequence to be transmitted
$\mathbf{b}_s$	The bit sequence representing a symbol
N_b	The number of bits in a symbol, which is defined by $N_b = \log_2 M$
$\mathbf{s}$	The CSK symbol $\mathbf{s}$, which can be represented by $\mathbf{s} = [s_i, s_j, s_k]$, $\mathbf{s} \in \mathcal{S}$
$\hat{\mathbf{s}}$	The estimate of $\mathbf{s}$
$\mathcal{S}$	$\mathcal{S} = \{\mathbf{s}_0, \mathbf{s}_1, \ldots, \mathbf{s}_{M-1}\}$ is the collection of all the M legitimate symbols in M-ary CSK
$\mathbf{s}_m$	The m th legitimate symbol in M-ary CSK , $\mathbf{s}_m \in \mathcal{S}$
$\mathbf{q}$	The 2-dimensional vector representing s, which can be expressed as $\mathbf{q} = [x, y]$
$\mathbf{v}_i$, $\mathbf{v}_j$, $\mathbf{v}_k$	The three vertices of the triangular area of the legitimate CSK symbols mapped into the CIE1931 coordinates of $\mathbf{v}_i = [x_i, y_i]$, $\mathbf{v}_j = [x_j, y_j]$ and $\mathbf{v}_k = [x_k, y_k]$
$\mathbf{c}$	The centroid of the triangular area of the legitimate CSK symbols
$\mathbf{H}$	The (3×3)-element matrix, where each element represents the channel gain between a specific light source and a PD
$\mathbf{n}$	The AWGN noise vector represented by $\mathbf{n} = [n_i, n_j, n_k]$
σ^2	The total power of $\mathbf{n}$
γ_o	The optical SNR defined as $\gamma_o = E_s/\sigma^2$
E_s	The average of the power of all possible M-ary CSK symbols, which is defined as $E_s = \mathbb{E}\left\{\|\mathbf{s}\|^2\right\}$
N_L	The total number of M-HCSK layers in each symbol defined as $N_L = N_b/2$
l	Index of the l-th M-HCSK layer
$\mathbf{b}_{sL_l}$	The 2 bits that correspond to the l-th M-HCSK layer, where $l \in \{0, 1, \ldots, N_L - 1\}$
$\mathbf{d}_p$	The basic vectors (BVs) defined in Eq. (4.18), where $p \in \{i, j, k\}$
h	The index of the M-HCSK sub-areas, $h \in \{0, 1, 2, 3\}$
$\mathbf{v}_{p,h}$	The vertices of each of the 4 areas of layer $l = 1$, where $p \in \{i, j, k\}$, $h \in \{0, 1, 2, 3\}$
$\mathbf{d}_{p,h}$	The BVs of each of the 4 areas of layer $l = 1$, where $p \in \{i, j, k\}$, $h \in \{0, 1, 2, 3\}$
$\mathbf{c}_h$	The centroids of each of the 4 areas of layer $l = 1$, where $h \in \{0, 1, 2, 3\}$
α	The scaling factor defined in Eq. (4.21)
δ_0	The parameter used for describing the scaling factor α, which is defined in Eq. (4.21)
δ_l	The value of δ_0 of Eq. (4.21), when generating the constellation of Layer l.
N_v	The number of layers consisted in the compressed video stream
Lv_n	The n th layer of the compressed video stream
ℓv_n	The length of the n-th layer Lv_n
Bf_l	The l th buffer used in Fig. 4.11b, $l \in \{0, \ldots, N_L - 1\}$
ℓ_{Bf}	The bit size of the buffer Bf_l
N_T	The total number of M-HCSK frame types
p_n	The conditional FER associated with each video layer's FER
ϵ_n	The distortion associated with each video layer
$p_{t,l}$	The FER of the l-th layer of the t-th M-HCSK frame
$\delta_{t,l}$	The δ_l used for the l-th layer of the t-th M-HCSK frame, where $l \in \{0, 1, \ldots, N_L - 2\}$

Table 4.2: Parameters used in Chapter 4.

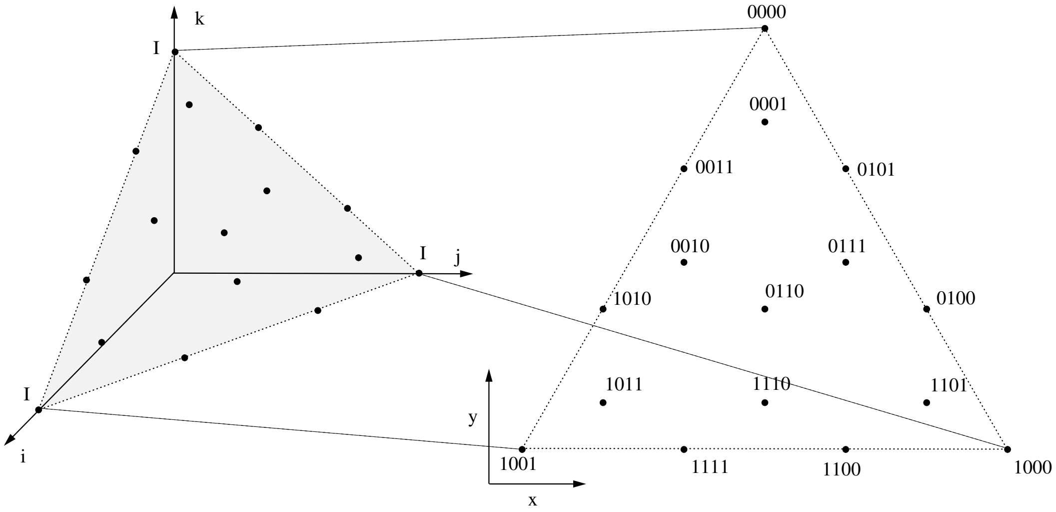


Figure 4.2: Colour constellations of 16-CSK without considering white balance, where the operating intensity is normalised to 1

of each of the light sources should be limited, i.e. we have

$$0 < s_p < I \tag{4.1}$$

for $p = i, j, k$. Furthermore, in most applications the total intensity of the CSK symbol must be constant, which is expressed as:

$$\sum_{p=i,j,k} s_p = I \quad . \tag{4.2}$$

Therefore, in a three dimensional space, the legitimate CSK symbols are confined to a triangular area on the $\sum_p s_p = I$ plane, as defined by the borderlines $\begin{cases} s_i + s_j & = I \\ s_k & = 0 \end{cases}$, $\begin{cases} s_i + s_k & = I \\ s_j & = 0 \end{cases}$ and $\begin{cases} s_j + s_k & = I \\ s_i & = 0 \end{cases}$, as shown in Fig. 4.2. Usually the symbol intensities s_p are normalized by setting the intensity boundary I to unity.

Wavelength Band [nm]	Band	Centre [nm]	(x, y)	$\mathbf{b}_s$	$[s_\mathrm{i}, s_\mathrm{j}, s_\mathrm{k}]$
380 - 478	Band k (s_0)	429	(0.169,0.007)	1 0	[0 0 1]
540 - 588	Band j (s_1)	564	(0.402,0.597)	0 0	[0 1 0]
726 - 780	Band i (s_2)	753	(0.734,0.265)	1 1	[1 0 0]
-	s_3	-	(0.435,0.290)	0 1	[0.333 0.333 0.333]

Table 4.3: The information-carrying colours of 4-CSK modulation specified by the IEEE standard, where the constellation point s_3 is constituted by a weighted mixture of the red, yellow and blue colours ©IEEE [127].

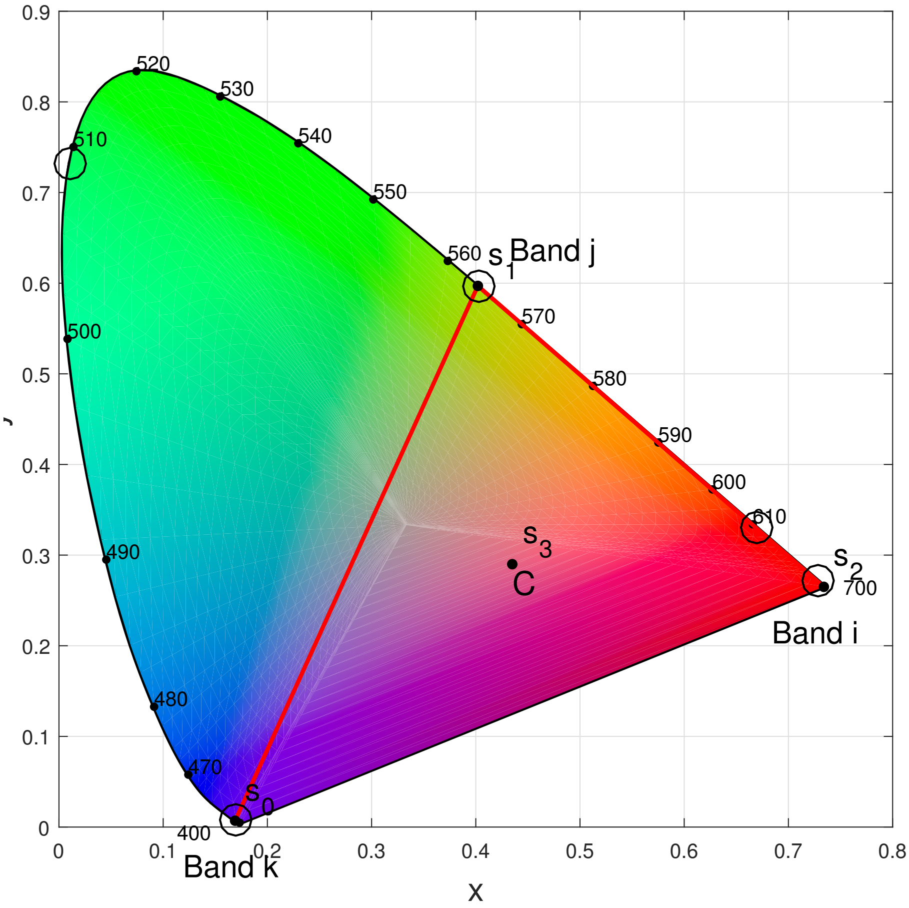


Figure 4.3: The CIE1931 colour plane for CSK modulation IEEE, where point (C) in the centroid of the triangle. The point s_3 and the three corner points represent four wavelength conveying 2 bits/symbol in 4-ary CSK ©IEEE [127].

Observe that due to the additional constraint formulated in Eq. (4.2), the above-mentioned three dimensional representation of the CSK symbols only has two degrees of freedom. In fact the CSK symbols may also be represented by the CIE 1931 colour space [257], which uses the 2-dimensional vector of $\mathbf{q} = [x, y]$ to represent s. In the 2-dimensional representation, the three light components we used, i.e$[1, 0, 0]$, $[0, 1, 0]$ and $[0, 0, 1]$, which correspond to the three vertices of the triangular area of the legitimate CSK symbols, are mapped into the new coordinates of $\mathbf{v}_i = [x_i, y_i]$, $\mathbf{v}_j = [x_j, y_j]$ and $\mathbf{v}_k = [x_k, y_k]$, respectively. Therefore the following equations may be used for defining the new 2-dimensional symbols:

$$\begin{aligned} x &= s_i x_i + s_j x_j + s_k x_k \\ y &= s_i y_i + s_j y_j + s_k y_k \end{aligned}, \tag{4.3}$$

which can also be formulated as:

$$\mathbf{q} = s_i \mathbf{v}_i + s_j \mathbf{v}_j + s_k \mathbf{v}_k, \tag{4.4}$$

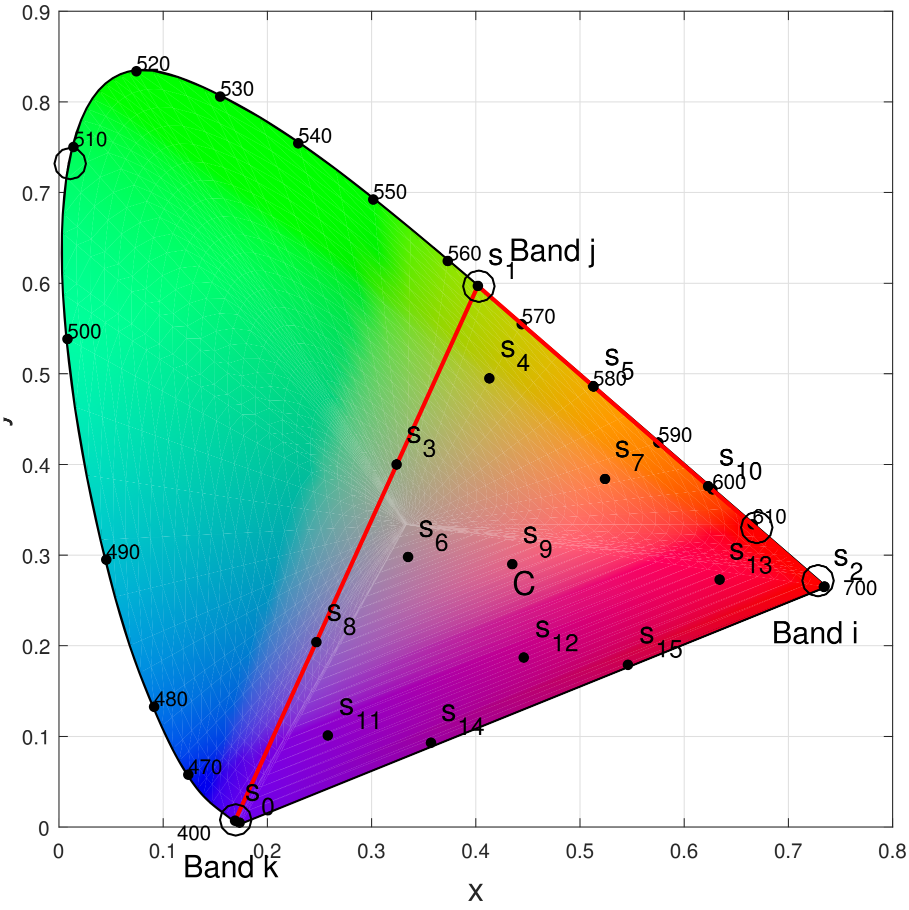
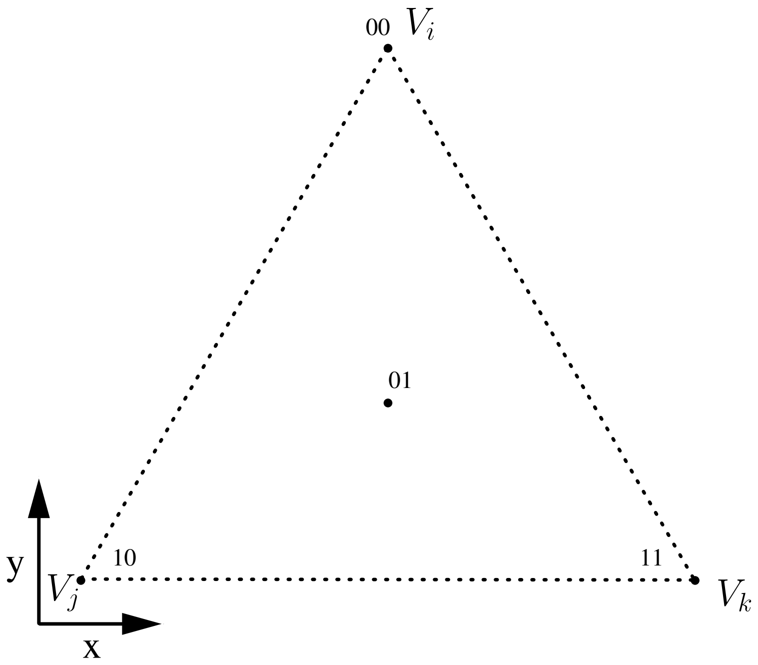
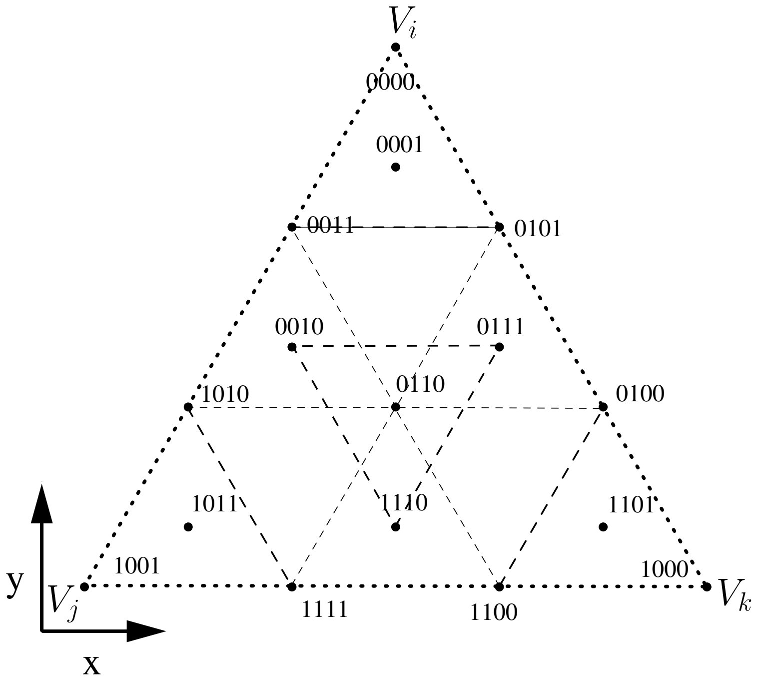


Figure 4.4: The CIE1931 colour plane for CSK modulation IEEE, where point (C) in the centroid of the triangle. The point s_3 - s_{15} and the three corner points represent sixteen wavelength conveying 4 bits/symbol in 16-ary CSK ©IEEE [127].

and each vector **s** can be mapped to a unique vector **q**, as shown in Fig. 4.2. The coordinates $\mathbf{v}_i$, $\mathbf{v}_j$ and $\mathbf{v}_k$ of the 2-dimensional vertices are also referred to as the centre of colour bands, and a total of 7 predefined settings were specified by the standard [127]. As the terminology suggests, each of them corresponds to a certain band of frequencies' colours. According to [127], there are only 9 valid colour band combinations consist a legitimate triangles. The points on the colour plane represent the CSK symbols generated by modulating the intensities of different colours of the legitimate colour band combinations on a RGB chip. For instance, the components are mapped to the Red, Yellow and Blue (RYB) LED transmitters according to the mapping scheme of Table 4.3. Explicitly, for the corner points of the triangle seen in Fig. 4.3, only one of the RYB LEDs is activated, as also seen in Table 4.3, but for the point s_3, all of the RYB LEDs are activated according to the weights seen at the bottom right corner of Table 4.3. As expected, these weights add up to unity. In analogy to classic Frequency Shift Keying (FSK), which detects the energy at the output of the detection filters, here the PDs detect the different visible-light colours falling into the 400 - 800 [THz] band. This justifies the terminology of CSK.



(a) Constellation for 4-CSK



(b) Constellation for 16-CSK

Figure 4.5: Constellation for 4-CSK and 16-CSK in IEEE 802.15.7 [127]

Wavelength Band [nm]	Band	Centre [nm]	(x, y)	$\mathbf{b}_s$	$[s_\mathrm{i}, s_\mathrm{j}, s_\mathrm{k}]$
380 - 478	Band k (s_0)	429	(0.169,0.007)	1 0 0 1	[0 0 1]
540 - 588	Band j (s_1)	564	(0.402,0.597)	0 0 0 0	[0 1 0]
726 - 780	Band i (s_2)	753	(0.734,0.265)	1 0 0 0	[1 0 0]
-	s_3	-	(0.324,0.400)	0 1 1 0	[0.333 0.334 0.332]
-	s_4	-	(0.413,0.495)	0 0 0 1	[0.111 0.779 0.115]
-	s_5	-	(0.513,0.486)	0 1 0 1	[0.334 0.666 0]
-	s_6	-	(0.335,0.298)	0 0 1 0	[0.110 0.445 0.445]
-	s_7	-	(0.524,0.384)	0 1 1 1	[0.445 0.444 0.111]
-	s_8	-	(0.247,0.204)	1 0 1 0	[0.0004 0.334 0.6656]
-	s_9	-	(0.435,0.290)	0 1 1 0	[0.333 0.334 0.333]
-	s_{10}	-	(0.623,0.376)	0 1 0 0	[0.666 0.334 0]
-	s_{11}	-	(0.258,0.101)	1 0 1 1	[0.112 0.110 0.777]
-	s_{12}	-	(0.446,0.187)	1 1 1 0	[0.445 0.111 0.445]
-	s_{13}	-	(0.634,0.273)	1 1 0 1	[0.777 0.111 0.112]
-	s_{14}	-	(0.357,0.093)	1 1 1 1	[0.3327 0.0003 0.667]
-	s_{15}	-	(0.546,0.179)	1 1 0 0	[0.667 0 0.333]

Table 4.4: The information-carrying colours of 16-CSK modulation specified by the IEEE standard, where the constellation points s_3 - s_{15} are constituted by the weighted mixture of the red, yellow and blue colours ©IEEE [127].

In IEEE 802.15.7 [127], the constellation settings are provided for 4-CSK, 8-CSK and 16-CSK. We can find the definition of symbol points for 4-CSK and 16-CSK in Fig. 4.5. For 4-CSK, we can observe in Fig. 4.5a that three symbol points are exactly on location of the vertices $\mathbf{v}_i$, $\mathbf{v}_j$ and $\mathbf{v}_k$ of the triangle, while the fourth symbol is at the centroid of the triangle. This design maximizes the Euclidean distances of the 4-CSK symbol points. For the 16-CSK constellation design of Fig. 4.5b, we can see that the entire triangular area is divided into 9 smaller triangles, which are congruent with each other and have 1/3 of the edge-length of the original triangle. As many as 10 of the 16-CSK symbol points are on the vertices of these smaller triangles, while the remaining 6 of them are the centroids of 6 of the the smaller triangles that share at least one side with the entire triangle. Furthermore, if we join the constellation points “0010”, “0111” and “1110” together, we will obtain a new small triangle that has the centroid of “0110” and it is congruent with the other small triangles. Therefore the constellation points of 16-CSK can simply be divided into the vertices and centroids of four congruent small triangles, where each setting of the 4 symbol points in each small triangle is similar to the 4-CSK symbol settings. We will further explore this interesting property later in this chapter.

The two-dimensional colour coordinates $\mathbf{q} = (x, y)$ of the symbols are generated by the intensity of the three light sources $\mathbf{s} = [s_\mathrm{i}, s_\mathrm{j}, s_\mathrm{k}]$, which represents the power of the light source. Therefore we can readily obtain the intensity vector $\mathbf{s}$ transmitted from the two-dimensional colour coordinates $\mathbf{q}$, under the settings of Table 4.3 and Table 4.4. Here we illustrate the conversion using 2 examples:

Example 1: For the 4-CSK modulation, given Band i, j and k of the CIE1931 colour plane of Fig. 4.3, the basic coordinates of the source colours RYB are $\mathbf{q}_R = \mathbf{v}_i$ = [0.734,0.265], $\mathbf{q}_Y = \mathbf{v}_j$ = [0.402,0.597] and $\mathbf{q}_B = \mathbf{v}_k$ = [0.169,0.007]. Then, to generate the specific colour at the point s_3 represented by $(x, y) = (0.435, 0.290)$ in Fig. 4.3, the intensity vector $\mathbf{s} = [s_{\mathrm{i}}, s_{\mathrm{j}}, s_{\mathrm{k}}]$ is obtained by solving the following equation:

$$\begin{cases} 0.435 = 0.734 s_{\mathrm{i}} + 0.402 s_{\mathrm{j}} + 0.169 s_{\mathrm{k}} \\ 0.290 = 0.265 s_{\mathrm{i}} + 0.597 s_{\mathrm{j}} + 0.007 s_{\mathrm{k}} \\ s_{\mathrm{i}} + s_{\mathrm{j}} + s_{\mathrm{k}} = 1, \end{cases} \tag{4.5}$$

where the resultant intensity vector becomes $\mathbf{s} = [s_{\mathrm{i}}, s_{\mathrm{j}}, s_{\mathrm{k}}]$ = [0.333049 0.334023 0.332928]. The intensity vector of other colour constellation points may be obtained in the same way, which has also been listed in Table 4.3.

Example 2: For the 16-CSK modulation, the basic coordinates of the source colours RYB for Band i, j and k are still $\mathbf{q}_R = \mathbf{v}_i$ = [0.734,0.265], $\mathbf{q}_Y = \mathbf{v}_j$ = [0.402,0.597] and $\mathbf{q}_B = \mathbf{v}_k$ = [0.169,0.007]. The constellation points of 16-CSK modulation on the CIE 1931 colour plane are shown in Fig. 4.4, which are represented by the two-component colour coordinates and three-component intensity vectors listed in Table 4.4. For example, the two-component coordinates of point s_{12} are represented by $(x, y) = (0.446, 0.187)$ in Fig. 4.4, which the intensity vector $\mathbf{s} = [s_{\mathrm{i}}, s_{\mathrm{j}}, s_{\mathrm{k}}]$ is obtained by

$$\begin{cases} 0.446 = 0.734 s_{\mathrm{i}} + 0.402 s_{\mathrm{j}} + 0.169 s_{\mathrm{k}} \\ 0.187 = 0.265 s_{\mathrm{i}} + 0.597 s_{\mathrm{j}} + 0.007 s_{\mathrm{k}} \\ s_{\mathrm{i}} + s_{\mathrm{j}} + s_{\mathrm{k}} = 1. \end{cases} \tag{4.6}$$

Hence, the resultant intensity vector becomes $\mathbf{s} = [s_{\mathrm{i}}, s_{\mathrm{j}}, s_{\mathrm{k}}]$ = [0.445, 0.444, 0.111], as shown in Table 4.4.

Hence to summarize, the conversion of two-dimensional colour coordinates $\mathbf{q} = (x, y)$ of the symbol to the three-dimensional optical signal intensity $\mathbf{s} = [s_{\mathrm{i}}, s_{\mathrm{j}}, s_{\mathrm{k}}]$ is achieved by simultaneously solving a simple system of three equations, which is given by Eq. (4.2) and Eq. (4.4). Here we define the coordinate-vector of the three light sources as $\mathbf{v}_i$, $\mathbf{v}_j$ and $\mathbf{v}_k$. Thus the entire CSK modulation process may be interpreted as a one-to-one mapping function.

4.2.2 Optical Domain Channel Model

As shown in Fig. 4.1, the modulated CSK symbol $\mathbf{s}$ is transmitted with the aid of 3 different light sources, and received by three separate Photon-Detectors (PD), which convert the intensities of light into electronic signals. The optical domain channel of Fig. 4.1 is represented with a noisy multipath model, which characterizes the channel-induced degradations, such as the multi-color imbalance, multi-color interference, ambient light-induced noise, as well as device noise. The contaminated received signal $\hat{\mathbf{s}} = [\hat{s}_i, \hat{s}_j, \hat{s}_k]$ can be represented as

$$\hat{\mathbf{s}} = \mathbf{H}\mathbf{s}^T + \mathbf{n}^T, \tag{4.7}$$

where $\mathbf{H}$ is a (3×3)-element matrix, which is represented as

$$\mathbf{H} = \begin{bmatrix} h_{ii} & h_{ji} & h_{ki} \\ h_{ij} & h_{jj} & h_{kj} \\ h_{ik} & h_{jk} & h_{kk} \end{bmatrix}. \tag{4.8}$$

Each element represents the channel gain between a specific light source and a PD, h_{ij} for example is the channel gain between the light source i and PD j, representing the interference imposed.

The noise $\mathbf{n}$ in Eq. (4.8) models three different sources of noise encountered by the indoor visible light optical link [258], namely the ambient light noise, shot noise and thermal noise introduced by the photodetector. The ambient noise, which is usually caused by solar radiation and artificial illumination sources can be modelled as a DC interference and can be mitigated by using an electrical high pass filter at the receiver. The shot noise, which is usually introduced by other illumination sources, can be modelled either using Poisson distribution [258] or Gaussian distribution [193, 202, 195, 196, 203]. In this chapter we adopted the model of a non-dispersive AWGN channel, where $\mathbf{H}$ is simply an identity matrix, while $\mathbf{n} = [n_i, n_j, n_k]$ is the AWGN noise vector modelling both the shot noise and the thermal noise, where $\mathbf{n}$ has a total power of σ^2, with each element having the power of $\sigma_0^2 = \sigma^2/3$. The variance σ_0^2 of the AWGN at each of the PDs, can be modelled as

$$\sigma_0^2 = \sigma_{\mathrm{S}}^2 + \sigma_{\mathrm{T}}^2, \tag{4.9}$$

according to [196], where the shot noise variance σ_{S}^2 is defined by

$$\sigma_{\mathrm{S}}^2 = 2qR\left(P_{Signal}\left(t\right) + P_{Daylight}\right)W \tag{4.10}$$

and the thermal noise variance σ_{T}^2 is given by

$$\sigma_{\mathrm{T}}^2 = \frac{4k_B TW}{r}, \tag{4.11}$$

where W is the bandwidth, q is the charge of an electron, R is the responsivity of PD, $P_{Signal}(t)$ is the instantaneous received power, $P_{Daylight}$ is the mean power received from the diffuse sunlight in indoor environments, k_B is Boltzmann's constant and T is the temperature of the noise equivalent input resistance r in Kelvin. By substituting σ_{S}^2 and σ_{T}^2 in Eq. (4.9) we arrive at

$$\sigma_0^2 = 2\left(qRP_{Signal}(t) + qRP_{Daylight} + \frac{2k_b T}{r}\right)W. \tag{4.12}$$

In the scenario of an indoor office environment [196], the lighting provides approximately 400 lux of luminance, while the daylight is about 50 lux at the centre of the office. Therefore $P_{Daylight}$ is assumed to be approximately 11% of the mean value of $P_{Signal}(t)$ according to [196].

Hereby we define the optical SNR as $\gamma_o = E_s/\sigma^2$, and $E_s = \mathbb{E}\{||\mathbf{s}||^2\}$, which is the average of the power of all possible M-ary CSK symbols.

4.2.3 Soft Demodulator for CSK

Basically, a simple estimate of $\mathbf{s}$, denoted by $\hat{\mathbf{s}}$, can be obtained by compensating the propagation-induced impairment reflected by $\mathbf{H}$ [127]:

$$\hat{\mathbf{s}} = \mathbf{H}^{-1}\mathbf{r}. \tag{4.13}$$

However, to fully exploit the information for post-demodulation decoding, we have to obtain the *a posteriori* probability of each symbol $\mathbf{s}_m$, $\mathbf{s}_m \in \mathcal{S}$, which can be readily expressed as:

$$\begin{aligned}\Pr(\mathbf{s}_m|\mathbf{r}) &= \mathrm{p}(\mathbf{r}|\mathbf{s}_m)\frac{\Pr(\mathbf{s}_m)}{\mathrm{p}(\mathbf{r})} \\ &= \frac{\Pr(\mathbf{s}_m)}{2\pi\sigma\mathrm{p}(\mathbf{r})}\exp\left(-\frac{\|\mathbf{r}-\mathbf{H}\mathbf{s}_m\|^2}{2\sigma^2}\right),\end{aligned} \tag{4.14}$$

where $\Pr(\mathbf{s}_m)$ is the *a priori* probability of the symbol $\mathbf{s}_m$. For simplicity, Eq. (4.14) can be expressed in the log domain as

$$\mathrm{lr}\left(\mathbf{s}_m|\mathbf{r}\right) = A - \left(\|\mathbf{r} - \mathbf{H}\mathbf{s}_m\|^2 / 2\sigma^2\right), \tag{4.15}$$

where $A = \ln\left[\Pr\left(\mathbf{s}_m\right)\right] - \ln\left[2\pi\sigma \mathrm{p}\left(\mathbf{r}\right)\right]$ and A is the same for all $\mathbf{s}_m$ when no *a priori* information is provided, while $\mathrm{lr}\left(\mathbf{s}_m|\mathbf{r}\right)$ is defined as $\ln\left[\Pr\left(\mathbf{s}_m|\mathbf{r}\right)\right]$.

At this stage the estimation of the symbol $\mathbf{s}_m$ can be carried out by using Eq. (4.15) as the maximum-likelihood function, and the MAP estimate can be expressed as

$$\begin{aligned}\hat{\mathbf{s}} &= \arg\max_{\mathbf{s}_m \in \mathcal{S}} \left[\mathrm{lr}\left(\mathbf{s}_m|\mathbf{r}\right)\right] \\ &= \arg\min_{\mathbf{s}_m \in \mathcal{S}} \|\mathbf{r} - \mathbf{H}\mathbf{s}_m\|^2,\end{aligned} \tag{4.16}$$

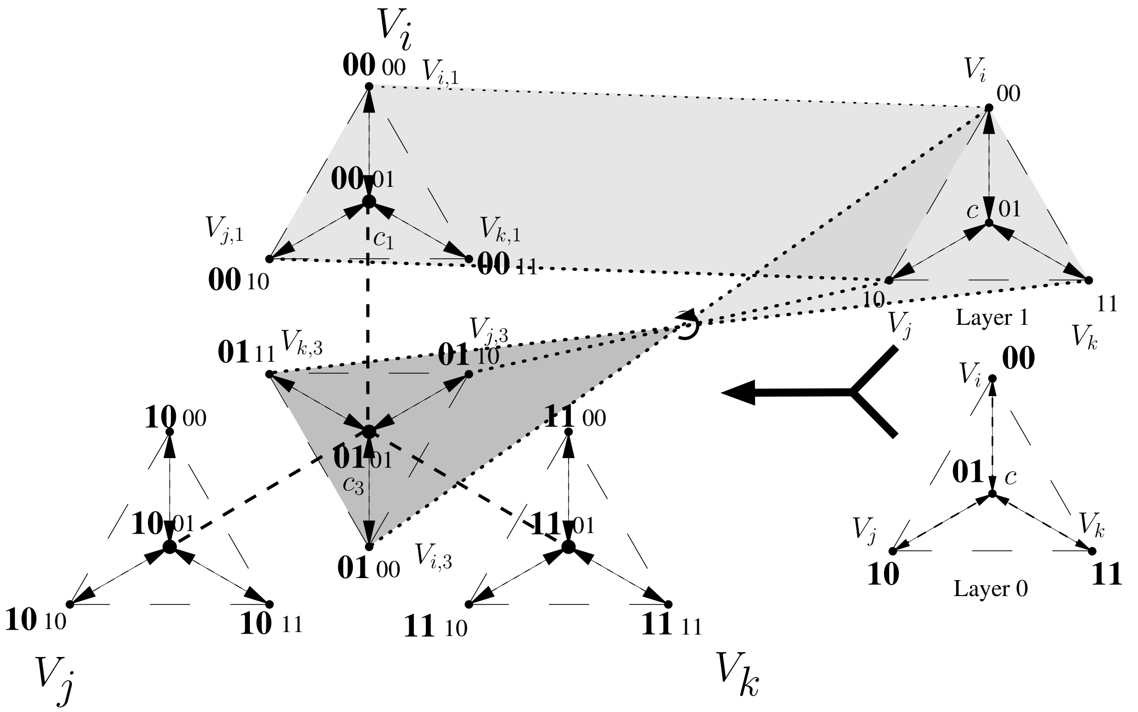
by substituting Eq. (4.15) into the first equation of Eq. (4.16). The legitimate constellation symbol $\mathbf{s}_m$ having the smallest distance from the received symbol $\mathbf{r}$ is chosen as the MAP detection result $\hat{\mathbf{s}}$. Finally, the decoded symbol $\hat{\mathbf{s}}$ is mapped to the bits $\hat{\mathbf{b}}$.

4.3 Hierarchical Colour-Shift Keying

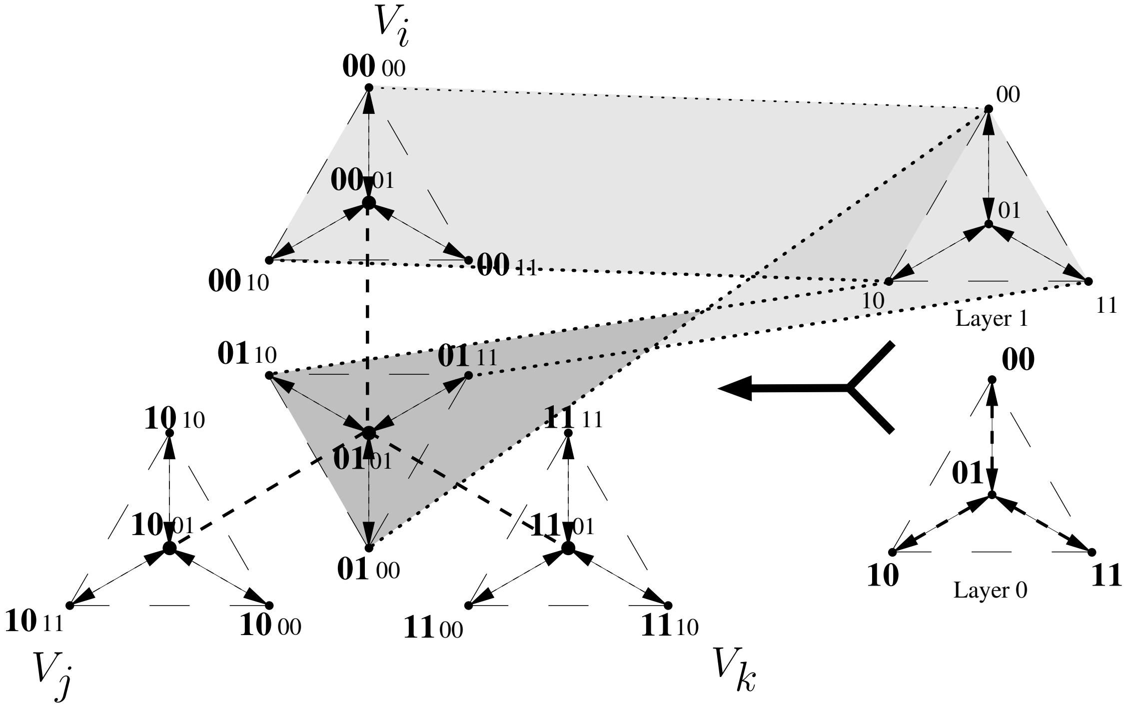
Based on the standard M-CSK regime of [127], we created a variant of the traditional scheme. More specifically, in this section, we will describe our hierarchical CSK (HCSK) modulation schemes as well as their detection methods. Similar to the traditional M-CSK modulation scheme, we will rely on the acronym M-HCSK using M as the total number of signal points in the symbol set S.

4.3.1 The M-HCSK Constellation

As stated in Section 4.2.1, there are $N_b = \log_2 M$ bits in each M-HCSK symbol. Let us denote these bits by $\mathbf{b}_s$. Since the symbol points in the basic 4-CSK constellation of the can be represented by 2 bits, we partition $\mathbf{b}_s$ by ensuring that every 2 bits correspond to a new layer. Therefore we have $\mathbf{b}_s = \left\{\mathbf{b}_{sL_0}, \mathbf{b}_{sL_1}, \ldots, \mathbf{b}_{sL_{N_L}}\right\}$, where $N_L = N_b/2$ is the total number of layers in each symbol and $\mathbf{b}_{sL_l}$, $l \in \{0, 1, \ldots, N_L - 1\}$ represents the 2 bits that correspond to the l-th layer. For example, a 16-HCSK symbol $\mathbf{b}_s = b_0b_1b_2b_3$ contains 4 bits per symbol in $N_L = 2$ layers, where b_0b_1 are assigned to L_0 and b_2b_3 are assigned to L_1.



(a) Type I



(b) Type II

Figure 4.6: Generation of the 16-HCSK modulation constellation using the constellation of a basic 4-CSK, following the Type I or Type II design rules.

Given the symbol constellation of N_L-layer M-HCSK and the symbol constellation of the simple 4-CSK, we can readily obtain the symbol constellation of a (N_L+1)-layer 4M-HCSK. Let us now describe the details of the generation method below.

Firstly, given the three vertices $\mathbf{v}_i$, $\mathbf{v}_j$ and $\mathbf{v}_k$ of a triangular CSK constellation plane, the centroid $\mathbf{c}$ of the triangle can be readily obtained according to:

$$\mathbf{c} = \frac{(\mathbf{v}_i + \mathbf{v}_j + \mathbf{v}_k)}{3}. \tag{4.17}$$

Furthermore, we define the basic vectors (BVs) as the three vectors that have the centroid as the initial point and the three vertices as their individual terminal points, which is formulated as

$$\mathbf{d}_p = \mathbf{v}_p - \mathbf{c}, \tag{4.18}$$

where we have $p \in \{i, j, k\}$.

Let us now study the (N_L+1)-layer (4M)-HCSK constellation of Fig. 4.6. The construction process of (4M)-HCSK can be viewed as replacing the points in a standard 4-CSK constellation with the scaled-down versions of the N_L-layer M-HCSK constellation, which may also be rotated or flipped. The 2-bit 4-CSK symbols $\mathbf{b}_{s-(4-CSK)}$, as the new $\mathbf{b}_{sL_0}$, are then concatenated as a prefix of their corresponding scaled down M-HCSK symbols $\mathbf{b}_{s-(M-HCSK)}$, as the new $\mathbf{b}_{sL_1 \sim L_{N_L-1}}$, in order to form the new symbol $\mathbf{b}_{s-\{(4M)-HCSK\}}$.

Let us now focus our attention on the (4M)-HCSK constellation, where the three vertices of the triangular constellation area are $\mathbf{v}_p$, $p \in \{i, j, k\}$. Since the layers above $l = 0$ are contained in each of the 4 separate sub-areas, we index these sub-areas by h, $h \in \{0, 1, 2, 3\}$. Therefore we will use $\mathbf{v}_{p,h}$, $p \in \{i, j, k\}$ to represent the vertices of each of the 4 areas of layer $l = 1$, where the corresponding BVs are denoted by $\mathbf{d}_{p,h}$ and the centroids by $\mathbf{c}_h$. Since the constellation of each sub-area is transformed from the constellation of the N_L-layer M-HCSK, their vertices $\mathbf{v}_{p,h}$ will coincide with the vertices $\mathbf{v}_p$ of the M-HCSK constellation having the same p, so that the specific sub-area will inherit both the relative location of the constellation points and the specific bit-to-symbol mapping of M-HCSK, once the vertices have been located. Consider the mapping in Fig. 4.6a as an example. When $M = 4$, two 4-CSK constellations are used to form the 16-HCSK constellation, where the symbols of a 4-CSK constellation are used as Layer 0, while the symbols of the other one are used as Layer 1. We proceed by first replacing V_i in the constellation representing Layer 0 at the bottom-right corner of Fig. 4.6a by the constellation representing Layer 1 at the top-right of Fig. 4.6a, forming sub-area $h = 1$. In the resultant framework of the 16-HCSK constellation seen at the left of Fig. 4.6a, the vertices of sub-area $h = 1$ have been determined as $\mathbf{v}_{i,1}$, $\mathbf{v}_{j,1}$ and $\mathbf{v}_{k,1}$. Therefore we obtain the location of the constellation points within the sub-area associated with $h = 1$ by appropriately overlapping $\mathbf{v}_{i,1}$, $\mathbf{v}_{j,1}$ and $\mathbf{v}_{k,1}$ with $\mathbf{v}_i$, $\mathbf{v}_j$ and $\mathbf{v}_k$ of the scaled-down version of the constellation at the top-right of Fig. 4.6a. Since sub-area $h = 1$ is generated by replacing $\mathbf{v}_i$ of the constellation at the bottom-right of Fig. 4.6a, the newly formed constellation points in sub-area $h = 1$ of the 16-HCSK constellation will obey the new bit-mapping by prefixing '00' to their original bit-mapping.

However, to locate $\mathbf{v}_{p,h}$, we first have to decide its relative location with respect to the vertex $\mathbf{v}_p$ of (4M)-HCSK. Below, we will proceed by considering this problem in two different ways.

Although having different arrangements for the 4 sub-areas is possible, we can set up some basic rules:

- In order to replace the 4 constellation points of 4-CSK, the position of the h-th symbol of 4-CSK must be covered by the h-th sub-area of (4M)-HCSK in the replaced constellation. For instance, in Fig. 4.6a the first symbol $\mathbf{v}_i$ of 4-CSK is covered by sub-area $h = 1$ of the 16-HCSK constellation generated.
- Each vertex $\mathbf{v}_p$ of (4M)-HCSK also constitutes one of the vertices in a certain sub-area. For example, in Fig. 4.6a the symbol $\mathbf{v}_i$ of the 16-HCSK constellation generated is also the vertex $\mathbf{v}_{i,1}$ of the sub-area $h = 1$.
- The centroid $\mathbf{c}$ of (4M)-HCSK also constitutes the centroid $\mathbf{c}_3$ of the sub-area associated with $h = 3$.
- Each side of the triangular sub-areas must be parallel to the sides of the entire area of (4M)-HCSK constellation, so that the sub-areas are also proportional to the entire area, having a scaling factor of α. For example, the sides of the 16-HCSK constellation generated in Fig. 4.6a are parallel to or partially coincide with the corresponding sides of sub-area $h = 1$.

According to the above rules, we implemented a pair of designs, namely the Type I and II designs of Table. 4.5. More specifically, the mapping of the sub-area vertices and BVs are specified at certain key constellation points, according to the specifications of Table 4.5a, while the rest of the points obey

$$\mathbf{v}_{p,h} = \mathbf{c}_h + \mathbf{d}_{p,h} \tag{4.19}$$

for $h = 0, 1, 2$ and

$$\mathbf{v}_{p,h} = \mathbf{c}_h - \mathbf{d}_{p,h} \tag{4.20}$$

Vertices	Type I	Type II
$\mathbf{v}_i$	$\mathbf{v}_{i,0}$	$\mathbf{v}_{i,0}$
$\mathbf{v}_j$	$\mathbf{v}_{j,1}$	$\mathbf{v}_{i,1}$
$\mathbf{v}_k$	$\mathbf{v}_{k,2}$	$\mathbf{v}_{i,2}$
$\mathbf{c}_0$	$\mathbf{c}_3$	$\mathbf{c}_3$

(a) Mapping of the sub-area vertices and centroids

BVs of (4M)-HCSK	Type I	Type II
$\alpha\mathbf{d}_i$	$\mathbf{d}_{i,0\sim3}$	$\mathbf{d}_{i,0}$, $\mathbf{d}_{j,1}$, $\mathbf{d}_{k,2}$, $\mathbf{d}_{i,3}$
$\alpha\mathbf{d}_j$	$\mathbf{d}_{j,0\sim3}$	$\mathbf{d}_{j,0}$, $\mathbf{d}_{k,1}$, $\mathbf{d}_{i,2}$, $\mathbf{d}_{k,3}$
$\alpha\mathbf{d}_k$	$\mathbf{d}_{k,0\sim3}$	$\mathbf{d}_{k,0}$, $\mathbf{d}_{i,1}$, $\mathbf{d}_{j,2}$, $\mathbf{d}_{j,3}$

(b) Mapping of the sub-area BVs

Table 4.5: Parameters of the Type I and II constellation designs showing the specific mapping of the sub-area vertices and BVs to those of (4M)-HCSK's.

specially for $h = 3$. Hence the relationship of $\mathbf{d}_{p,h}$ to the BVs of (4M)-HCSK has now been defined. Finally, we will define the scaling factor α by

$$\begin{aligned}\alpha &= \|\mathbf{d}_{p,h}\| \,/\, \|\mathbf{d}_p\| \\ &= \frac{\delta_0\left(\sqrt{M}-2\right)}{2\left(\sqrt{M}-1\right)},\end{aligned} \tag{4.21}$$

so that when $\delta_0 = 1$ in Eq. (4.21), the minimum Euclidean distance amongst the vertices $\mathbf{v}_{p,h}$ of different sub-areas also equals $\|\mathbf{d}_{p,h}\|$. In fact, the locations of the 16-HCSK constellation points associated with $\delta_0 = 1$ coincide with those of the standard 16-CSK of [127]. The scaling factor obeys $0 < \alpha < 1/2$, therefore we have $0 < \delta_0 < 1 + 1/\left(\sqrt{M}-2\right)$. Let us now define δ_l, which is the value of δ_0 used for describing the scaling factor α of Eq. (4.21), when generating the constellation of Layer l.

Let us now investigate the Euclidean distance between the symbol points. Considering the M-HCSK constellation for example, let us partition the constellation points into 4 groups using the definition of the sub-areas. Given $\delta_l = 1$ for $l > 0$, the minimum Euclidean distance among the symbols within each group can be expressed as

$$dist_{ig} = \frac{\delta_0}{\sqrt{M}-1} \min \|\mathbf{d}_p\| \,. \tag{4.22}$$

By contrast, the minimum Euclidean distance between the different symbol groups is

$$dist_{ag} = \left(1 - \frac{\sqrt{M}-2}{\sqrt{M}-1}\delta_0\right) \min \|\mathbf{d}_p\| \,. \tag{4.23}$$

For example, when assuming $\delta_0 = 1$ for the case of the 16-HCSK constellation shown in Fig. 4.6a, we have $dist_{ig} = dist_{ag} = \|\mathbf{d}_i\|/3$, which implies that the minimum Euclidean distances are the same both within and between group(s). From Eq. (4.22) and Eq. (4.23) we find that $dist_{ag}$ decreases as δ_0 increases, which results in an increased symbol error rate (SER) for layer $l = 0$. At the same time, $dist_{ig}$ increases as δ_0 increases, which suggests having a reduced SER for the higher layers associated with $l > 0$, despite the presence of cross-layer interferences. For example, when assuming $\delta_0 = 1.2$ in Fig. 4.6a, we have $dist_{ig} > \|\mathbf{d}_i\|/3$, which means that the symbol '0010' is less likely to be confused with '0001', when contaminated by channel distortions. In this case the last two bits of the symbol belonging to Layer 1 may appear to be better protected. However, at the same time we have $dist_{ag} < \|\mathbf{d}_i\|/3$, which means that symbol '0010' in sub-area $h = 0$ can easily be confused with '0111' in sub-area $h = 3$. In this case both the first and last two bits corresponding to 2 layers are more error-prone. However, looking at the differences between the type I and II symbol mappings in Table 4.5, we find that the neighbouring symbols between each two groups of symbols have the same bit-mapping for $l > 0$, which is aimed at reducing the SER degradation of the higher layers owing to the increased cross-layer interferences encountered upon increasing δ_0. For example, when assuming $\delta_0 = 1.2$ in Fig. 4.6b, we still have $dist_{ag} < \|\mathbf{d}_i\|/3$, which means that symbol '0010' in sub-area $h = 0$ can easily be confused with '0110' in sub-area $h = 3$. In this case only the first two bits corresponding to Layer 0 are more error-prone, while the last two bits corresponding to Layer 1 are less prone to errors, even if the symbol is misjudged, because these two nearby symbols are the same for the last two bits.

4.3.2 Layer-by-layer Demodulation of M-HCSK

As stated in Section 4.3.1, each M-HCSK symbol $\mathbf{s}_m$ can be represented by $\mathbf{b}_s = \left\{\mathbf{b}_{sL_0}, \mathbf{b}_{sL_1}, \ldots, \mathbf{b}_{sL_{N_L-1}}\right\}$, where $\mathbf{b}_{sL_l}$, $l \in \{0, 1, \ldots N_L - 1\}$, corresponds to the l-th layer. Since $\mathbf{b}_{sL_l}$ is a 2-bit symbol, it has 4 possible values, namely '00', '01', '10' and '11'. Let us denote these 4 values by $\mathbf{b}_{h_l}$, where $h_l \in \{0, 1, 2, 3\}$. Therefore, the layer-by-layer demodulation of M-HCSK can be carried out by firstly obtaining the *a posteriori* probability of $\mathbf{b}_{h_l}$s, formulated as:

$$\Pr\left(\mathbf{b}_{h_l}\middle|\,\mathbf{r}\right) = \sum_{\mathbf{b}_s \in \left\{\mathbf{b}_{s'}\middle|\mathbf{b}_{sL_l} = \mathbf{b}_{h_l}\right\}} \Pr\left(\mathbf{b}_s\middle|\,\mathbf{r}\right), \tag{4.24}$$

where $\Pr\left(\mathbf{b}_s\middle|\,\mathbf{r}\right)$ was expressed in Eq. (4.14) and Eq. (4.15). Therefore the decision on $\mathbf{b}_{sL_l}$ can be formulated as:

$$\hat{h}_l = \underset{h \in \{0,1,2,3\}}{\arg\max} \Pr\left(\mathbf{b}_{h_l}\middle|\,\mathbf{r}\right) \tag{4.25}$$

and

$$\hat{\mathbf{b}}_{sL_l} = \mathbf{b}_{\hat{h}_l}. \tag{4.26}$$

However, the procedure represented by Eq. (4.24) imposes a high computational complexity, which requires the calculation and summing of 4^{N_L-1} different $\Pr\left(\mathbf{b}_s\middle|\,\mathbf{r}\right)$s. In order to reduce the complexity imposed, we may simplify Eq. (4.24) by replacing the *a posteriori* probabilities of the $\mathbf{b}_{sL_l}$ represented symbol sub-groups with that of the centroid of the corresponding sub-area of the constellation, yielding:

$$\Pr\left(\mathbf{b}_{h_l}\middle|\,\mathbf{r}\right) = \sum_{h_l = h} \Pr\left(\mathbf{c}_{\underbrace{h_0, h_1, \ldots h_l}_{l+1}}\middle|\,\mathbf{r}\right), \tag{4.27}$$

where $\mathbf{c}_{h_0,h_1,\ldots h_l}$ stands for the centroid of the triangular sub-area constituted by the constellation points of the symbols $\mathbf{b}_s \in \left\{\mathbf{b}_{s'}\middle|\mathbf{b}_{sL_0} = \mathbf{b}_{h_0}, \mathbf{b}_{sL_1} = \mathbf{b}_{h_1}, \ldots, \mathbf{b}_{sL_l} = \mathbf{b}_{h_l}\right\}$. Therefore in Eq. (4.27) the *a posteriori* probabilities $\Pr\left(\mathbf{c}_{\underbrace{h_0, h_1, \ldots h_l}_{l+1}}\middle|\,\mathbf{r}\right)$ are summed over different values of $h_0, h_1, \ldots h_{l-1}$ with a fixed h_l, to get the result for $\Pr\left(\mathbf{b}_{h_l}\middle|\,\mathbf{r}\right)$. By replacing the 4^{N_L-l-1} symbols in a sub-area with their associated centroid, Eq. (4.27) reduces the complexity to 4^l instead of 4^{N_L-1}.

If no FEC is used and hence only detection decisions are required for generating the modulation output, the previous process of Eq. (4.24) and Eq. (4.25) can be further simplified by making the decisions in a layer-by-layer manner, yielding:

$$\hat{h}_l = \underset{h \in \{0,1,2,3\}}{\arg\min} \left\| \mathbf{r} - \mathbf{H}\mathbf{c}_{\underbrace{\hat{h}_0, \hat{h}_1, \ldots, \hat{h}_{l-1}, h}_{l+1}} \right\|^2 . \tag{4.28}$$

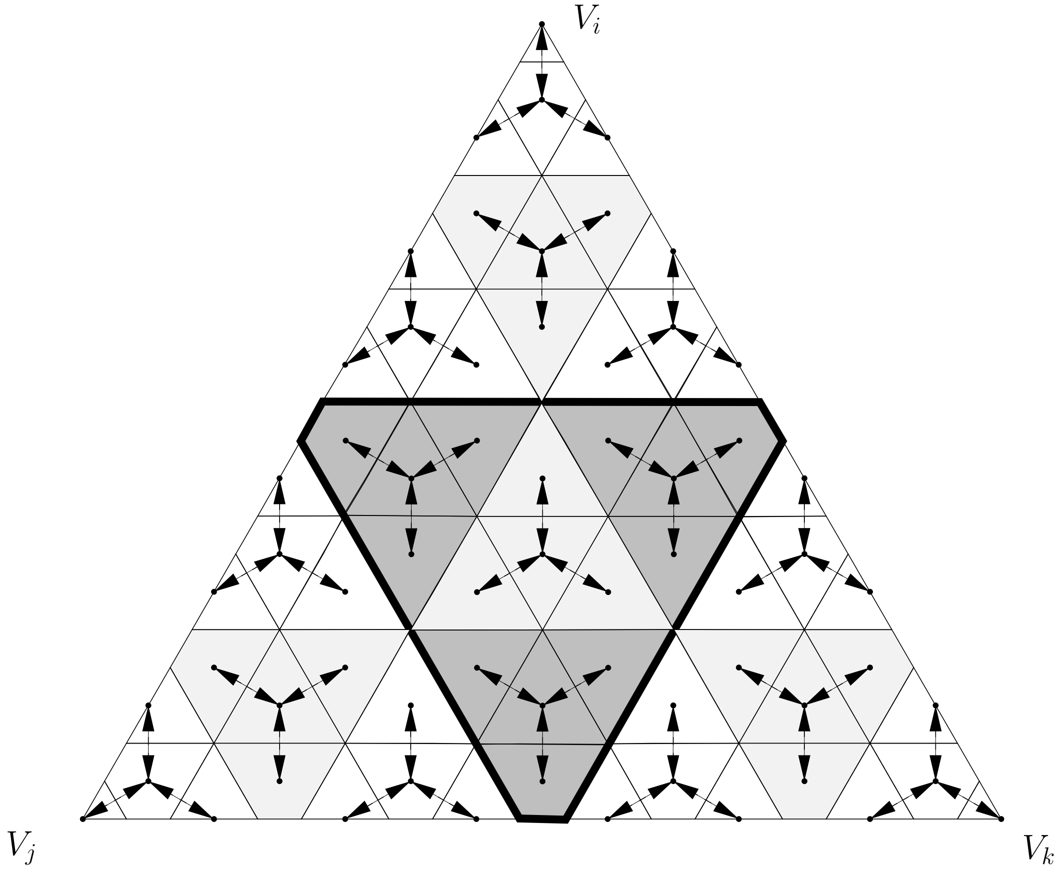


Figure 4.7: Colour constellation of 64-HCSK

Namely, by using Eq. (4.28), we first make decisions of $\hat{h}_0$ associated with layer $l = 0$ and work our way up based on the decisions made for the previous layers. Therefore the estimation $\hat{h}_0, \hat{h}_1, \ldots, \hat{h}_{l-1}$ of $h_0, h_1, \ldots h_{l-1}$ will be attained in the above order, and the next decision for h_l can be readily obtained by applying Eq. (4.28). In this way only four Euclidean distance computations are required for a decision.

By using the reduced complexity demodulation method of Eq. (4.28), we obtained the frame error ratio (FER) versus the values of δ_l using either the 16-HCSK or 64-HCSK for a transmission frame length of 1200 bits, as shown in Fig. 4.8 and Fig. 4.9. The construction of 64-HCSK is carried out using the method of Section 4.3.1, which results in the constellation shown in Fig. 4.7. By observing Fig. 4.8 and Fig. 4.9, we can clearly see the differences between the Type I and Type II constellation designs of Table 4.5 in terms of the FERs of each decoded M-HCSK layer. For example, by comparing Fig. 4.8a and 4.8b, we can see that as the parameter δ_0 defined in Eq. (4.21) for 16-HCSK increases, the FER of Layer 0 increases for both the Type I and II regimes of Table 4.5. However, when the FER of Layer 1 is considered, for Type I it first decreases until around $\delta_0 = 1$, then it is increased again, while it is reduced all along as δ_0 increases for Type I. The reason behind this is that as δ_0 increases, the cross-layer interferences become more significant, as described in Section 4.3.1. While the cross-layer interferences result directly in an increase of the FER of Layer 1 in the

Type I constellation, the mapping of the Type II constellation prevents this, because the adjacent Layer 1 symbols belong to different Layer 0 symbol groups having the same bit mapping in Layer 1, as seen in the constellation of Fig. 4.6b. Similar properties can be observed for the FERs of 64-HCSK in Fig. 4.9a and Fig. 4.9b. Explicitly, observe that the Type II constellation provides a wider range of FERs for the higher layers upon varying δ_l according to Fig. 4.8 and Fig. 4.9, therefore it is more promising in terms of reducing the effect of cross-layer interferences, hence providing a high grade of flexibility in terms of system configuration and optimization.

4.4 Video Transmission System

In this section, we will introduce a video streaming system relying on our M-HCSK modulation scheme and a RSC FEC code, as shown Fig. 4.10.

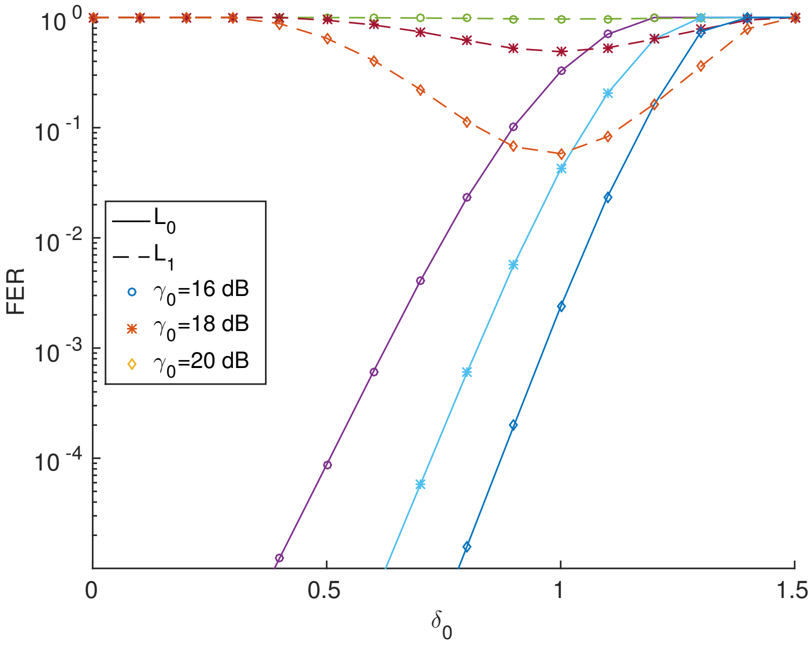
4.4.1 Source Coding

As shown in Fig. 4.10, the original video sequence is firstly encoded into a scalable video stream by invoking the SVC extension of H.264 [10]. The compressed video stream consists of N_v layers, namely of Lv_0, Lv_1,... Lv_n, $n \in \{0, 1, \ldots, N_v - 1\}$, with the dependency of $Lv_0 \Leftarrow Lv_1 \Leftarrow \ldots \Leftarrow Lv_n$, where each item on the right of the $\Leftarrow$ symbol depends on all the items to the left of it. To utilize the n-th layer for successful decoding, the decoder has to invoke the information from all the previous n layers ranging from Lv_0 to Lv_{n-1}. The different layers Lv_n of the compressed video stream have different number of bits. Let us denote the length of the n-th layer by ℓv_n. In order to make the video layers robust to transmission errors, their multiplexing has to be carefully designed.

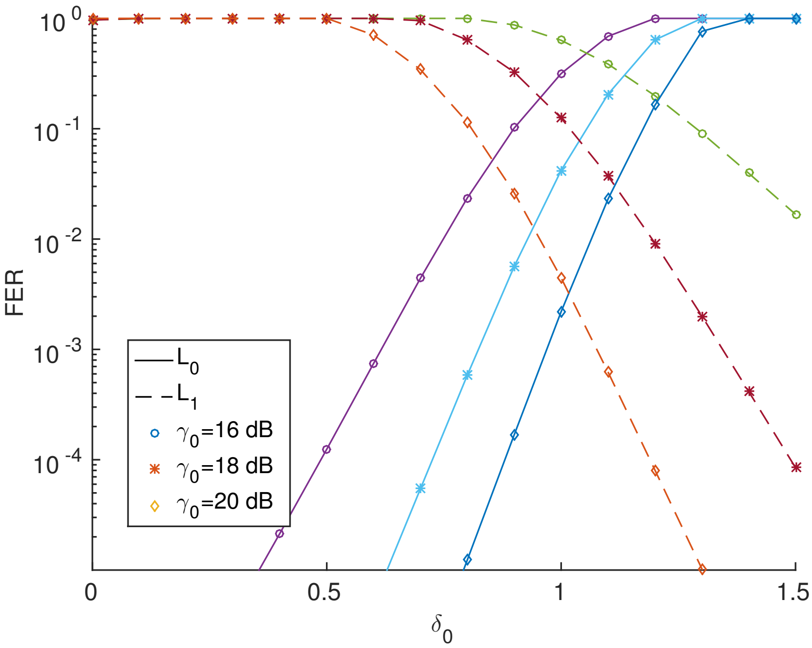
4.4.2 Channel Coding and Video-to-M-HCSK Mapping

Two specific types of FEC codes have been adopted in the IEEE 802.15 standard [127] for VLC, namely a RS code and a convolutional code (CC). In this chapter we employed a RSC code, which is similar to the CC used in the standard. As shown in Fig. 4.11, the classic multiplexing would feed the compressed video stream to the FEC encoder and modulator sequentially. However, the system will not benefit from the layered structure of the video stream and of the M-HCSK modulator for this conventional equal error protection (EEP) scheme. In order to exploit the above-mentioned layered structure, we conceived the beneficial video bit to M-HCSK mapping scheme of Fig. 4.11b.

Hence observe in Fig. 4.11b that according to the number of M-HCSK layers, N_L buffers are used for storing the compressed output of the SVC video encoder. These buffers are denoted as Bf_l, $l \in \{0, \ldots, N_L - 1\}$, ranging from top to bottom. Each

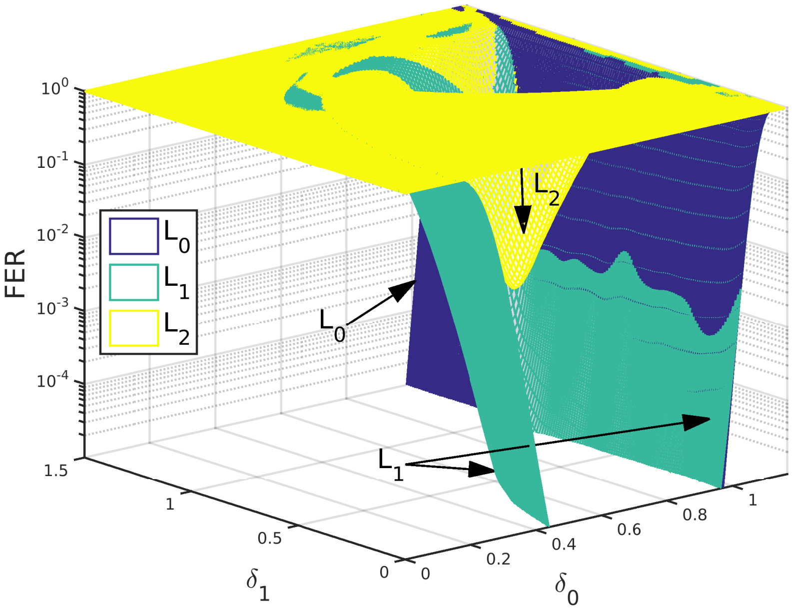


(a) FER vs. δ_0 using Type I 16-HCSK, for various optical SNRs

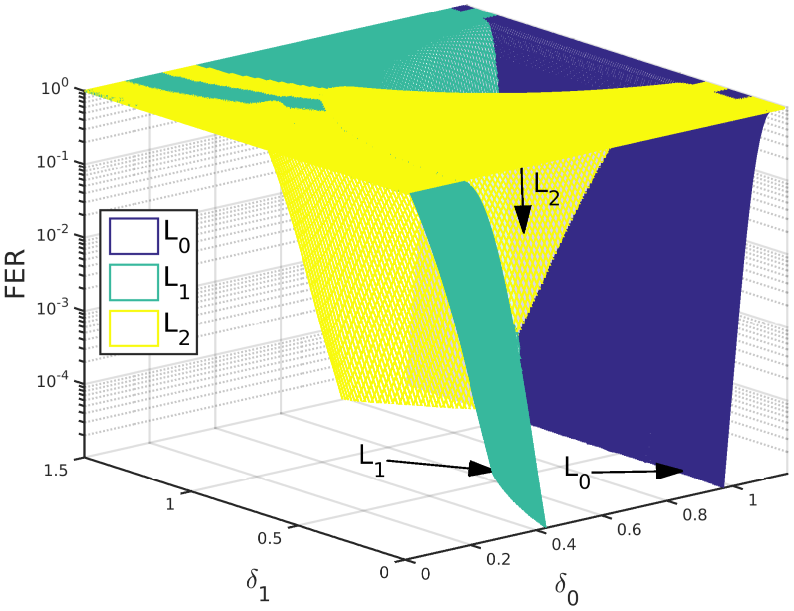


(b) FER vs. δ_0 using Type II 16-HCSK, for various optical SNRs

Figure 4.8: FER vs. δ_l of Type I and II 16-HCSK modulation, using a frame length of 1200 bits. The schematic of Fig. 4.1 and the parameters of Table 4.2 were used.



(a) FER vs. δ_0 and δ_1 using Type I 64-HCSK, for an optical SNR of $\gamma_0 = 28\,dB$



(b) FER vs. δ_0 and δ_1 using Type II 64-HCSK, for an optical SNR of $\gamma_0 = 28\,dB$

Figure 4.9: FER vs. δ_l of Type I and II 64-HCSK modulation, using a frame length of 1200 bits. The schematic of Fig. 4.1 and the parameters of Table 4.2 were used.

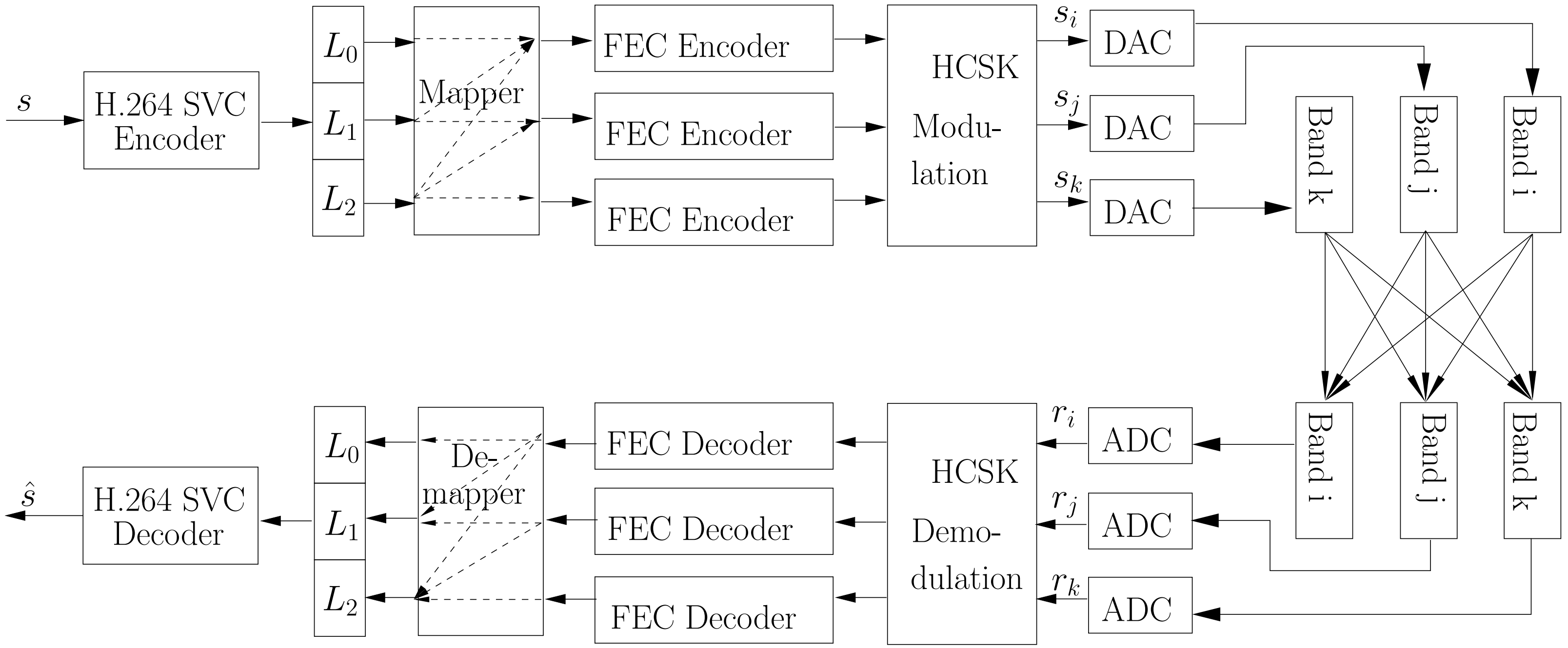
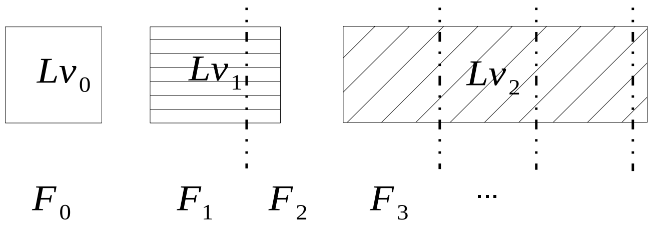
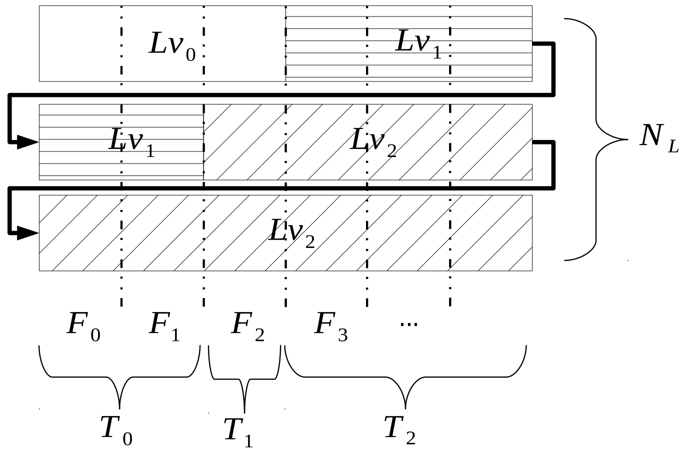


Figure 4.10: Transceiver architecture of the M-HCSK-RSC scheme conceived for scalable video transmission over VLC channels.



(a) EEP



(b) UEP

Figure 4.11: EEP and UEP versions of the video-to-M-HCSK for scalable video transmission.

layer has the same length of

$$\ell_{Bf} = \left\lceil \frac{1}{N_L} \sum_{n \in \{0,1,\ldots,N_v-1\}} \ell v_n \right\rceil \tag{4.29}$$

bits. The N_v video layers are written into those buffers sequentially, commencing from the BL and gradually proceeding with the ELs having higher levels of dependency. When the n-th buffer is filled up, the mapping process will continue from the beginning of the $(n+1)$-st buffer. When the mapping of all the video layers is finished, the remaining storage space left in the buffer will be filled by bits of '0'.

In the channel coding process, N_L FEC encoders read simultaneously from the N_L data buffers, each taking ℓ_F bits at a time. The FEC coded bits of the n-th data buffer will be used as the input of the n-th layer of the M-HCSK modulator. Those N_L FEC encoded data frames are mapped to a single M-HCSK frame. By adopting this multiplexing scheme, we have ensured that the N_v number of video layers having different bit-lengths are mapped to N_L streams having the same lengths. More importantly, we ensure that the video frames corresponding to a single M-HCSK frame are strictly sorted according to their importances, bearing in mind the level of dependencies they have. The FEC frames corresponding to the BL will always occupy the highest-integrity layer of the M-HCSK, while the ELs are mapped to higher layers.

Fig. 4.11b shows an example of the multiplexing results, where we have $N_v = N_L = 3$. Consequently, the M-HCSK frames have three types. Specifically, let us denote the t-th M-HCSK frame by F_t. Then F_0 and F_1 has the type $Lv_0 - Lv_1 - Lv_2$, F_2 has the type $Lv_0 - Lv_2 - Lv_2$, while the rest of the frames has the type $Lv_1 - Lv_2 - Lv_2$. Generally, let us denote the total number of M-HCSK frame types by N_T.

4.4.3 Modulation Optimization

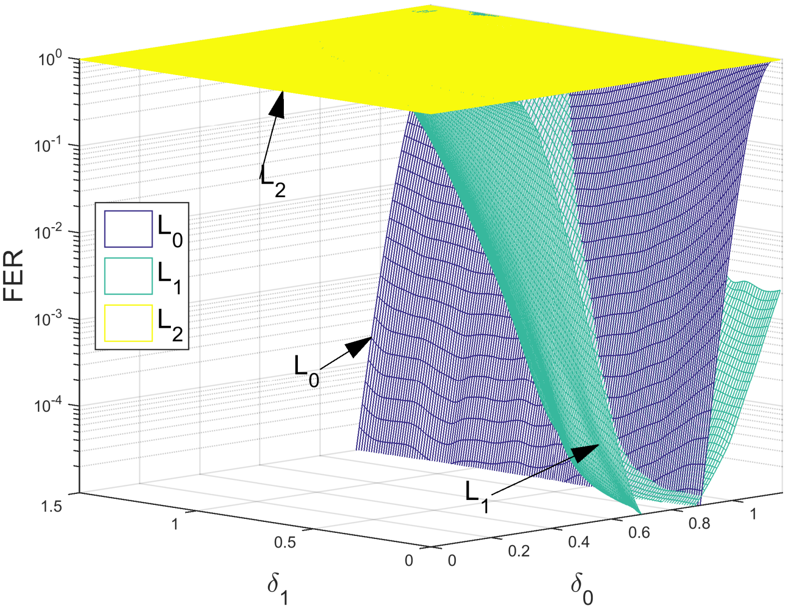
Given the transmission scheme of Fig. 4.10 described in Section 4.4.1 and 4.4.2, we aim for optimizing the modem parameters for to ensuring that the system achieves its best performance. More specifically, the distortion of the video reconstructed at the receiver end should be minimized.

In order to estimate the video distortion, we require the conditional FER p_n associated with each video layer's FER and its distortion ϵ_n. The conditional FER p_n is defined as the FER of the n-th video layer, given that all the bit streams gleaned from the previous layers having an index ranging from 0 to $n-1$ are intact. The distortion ϵ_n imposed by the loss of layer n can be obtained by the so-called off-line removal-decoding test, which was advocated in [99] and [259]. Before the commencement of transmissions, ϵ_n is measured by the process based on the PSNRs of the reconstructed video with the bit stream of the n-th video layer removed, and of the one relying on the intact bit stream of the n-th video layer.

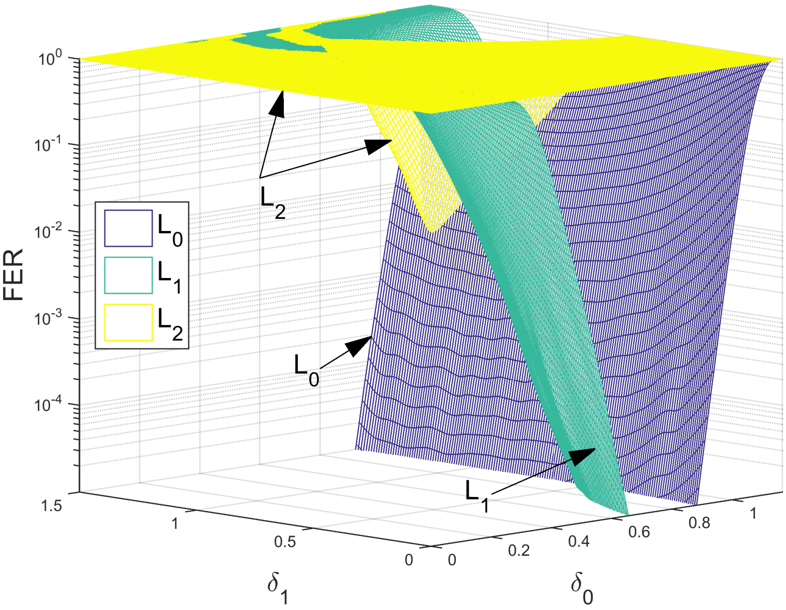
Firstly, let us define another conditional FER $p_{t,l}$, which refers to the FER of the l-th layer of the t-th M-HCSK frame after RSC decoding, provided that the previous l layers in the t-th M-HCSK frame have been successfully recovered. The FER $p_{t,l}$ can be formulated by introducing the function $f_l\left(\centerdot\right)$, which can be expressed as

$$\begin{aligned} p_{t,l} &= f_l\left(\gamma_o, \ell_F, \delta_{t,0}, ..., \delta_{t,N_L-2}\right) \\ &= 1 - \left[1 - \mathcal{T}\left[\gamma_o, \delta_{t,0}, ..., \delta_{t,N_L-2}\right]\right]^{\ell_F/\ell} . \end{aligned} \tag{4.30}$$

As shown in Eq. (4.30), the knowledge of the optical SNR γ_o, the length ℓ_F of each M-HCSK frame and of the M-HCSK parameters $\delta_{t,l}$, $l \in \{0, 1, \ldots, N_L - 2\}$ corresponding to the t-th M-HCSK frame are all required for to estimating $p_{t,l}$. At the right hand side of Eq. (4.30), a pre-stored look-up table (LUT) $\mathcal{T}\left(\cdot\right)$ is used for obtaining the FER assuming a fixed transmission frame-length of ℓ bits. More specifically, the LUT $\mathcal{T}\left(\cdot\right)$ has N_L entries corresponding to the specific FER evaluated as a function of both the SNR and of the M-HCSK parameters.



(a) FER vs. δ_ls using Type I 64-HCSK



(b) FER vs. δ_ls using Type II 64-HCSK

Figure 4.12: FER vs. δ_0 and δ_1 of the M-HCSK-RSC transmission systems of Fig. 4.10 using Type I or II 64-HCSK modulation at the optical SNR of $\gamma_0 = 20\,dB$. A frame length of 1200 bits is used. The schematic of Fig. 4.10 and the parameters of Table 4.2 were used.

Given the expression of $p_{t,l}$ in Eq. (4.30), we can now readily obtain the conditional FER p_n as

$$p_n = 1 - \prod_{\{t,l|\mathfrak{m}(t,l)=n\}} (1 - p_{t,l}), \tag{4.31}$$

which exploits the following assumptions. The mapping function $\mathfrak{m}(t, l)$ of Eq. (4.31) returns the index of the video layer associated with the l-th layer of the t-th M-HCSK frame. Therefore the calculation of p_n in Eq. (4.31) relies on all the parts of the n-th video layer that are distributed across the different M-HCSK frames or layers.

Given the FER expression of p_n, the expected decoded video distortion at the receiver can be formulated as:

$$\begin{aligned} & e(\gamma_o, \delta_{0,0}, \ldots, \delta_{t,l}, \ldots) \\ & = p_0 \cdot \epsilon_0 + [1 - p_0] \cdot p_1 \cdot \epsilon_1 + \ldots \\ & = \sum_{n=0}^{N_V-1} \epsilon_n \cdot p_n \cdot \prod_{j=0}^{n-1} [1 - p_j], \end{aligned} \tag{4.32}$$

where $p_n \cdot \prod_{j=0}^{n-1} [1 - p_j]$ represents the probability of the video layer n being corrupted, while the layers $0 \sim n-1$ have been successfully received. Observe in Eq. (4.32) that the video distortion estimation requires the value of $\delta_{t,l}$, which represents δ_l apoted at time slot t, for $l \in \{0, 1, \ldots, N_L - 2\}$ and $t \in \{0, 1, \ldots, T - 1\}$. Hence there are as many as $(N_L - 1) \cdot T$ $\delta_{t,l}$ values. For instance, 16-HCSK needs T $\delta_{t,l}$ values, while 64-HCSK requires $2T$ $\delta_{t,l}$ values. However, in practice there are only a few different types of M-HCSK frames. Assuming that there are K different M-HCSK frames, we have $K \leq N_v < T$. For instance, if a video stream containing three layers is transmitted over a larger number of time slots, there are at most three different types of M-HCSK frames. Consequently, we can reduce the number of $\delta_{t,l}$ values to $K \cdot T$. Let us furthermore denote these different types of M-HCSK frames by T_k, where we have $k \in \{0, 1, ..., K - 1\}$, as shown in Fig. 4.11b.

Finally, as stated before, we want to optimize the system by minimizing the distortion of the decoded video. Hence the final decision concerning the M-HCSK parameters may be formulated as:

$$\mathbf{D}(\gamma_o) = \underset{0<\delta_{k,l}<\delta_{max,l}}{\arg} \min \{e(\delta_{0,0}, \ldots, \delta_{k,l}, \ldots)\}, \tag{4.33}$$

where $k \in \{0, \ldots, K - 1\}$. The optimization of Eq. (4.33) can be readily handled by optimization tools, such as the function *fmincon* in MATLAB.

	Soccer	Ice	Crew
Representation	YUV 4:2:0		
Format	4CIF		
Bits Per Pixel	8		
FPS	60		
Number of Frames	60		
Video Codec	SVC-H.264		
GOP	15		
Scalability	MGS		
Bitrate	19.30 Mbps	6.01 Mbps	11.50 Mbps
Error-Free PSNR	42.25 dB	44.56 dB	42.89 dB
Error Concealment	Frame-Copy		

Table 4.6: The video parameters

4.5 System Performance

In this section, we will quantify the attainable performance gain of our proposed M-HCSK-RSC transmission scheme. The main video system parameters are listed in Table 4.6. Three 4:2:0 YUV format video sequences were chosen for transmissions, namely the Soccer, the Ice and the Crew video clips. These 60-frame sequences are in the (704 × 576)-pixel 4CIF format, and were recorded at 60 FPS.

We use the JSVM H.264/AVC reference video codec, which relies on a group of pictures (GOP) duration of 15 frames and the bi-directionally predicted (B) frames are disabled. We enabled the Medium Grain Scalability (MGS) [11, 232] feature for encoding the video sequences into three layers with the aid of the standardized video quantization parameters (QP) of 46, 34 and 25, respectively. The average PSNRs achieved by the decoder for the three sequences are 42.25 dB, 44.56 dB and 42.89 dB, respectively.

Based on our configuration of the SVC encoder, each slice is encoded into three layers and each layer is encapsulated into a network abstraction layer unit (NALU) [10]. The NALUs are transmitted sequentially using our proposed system. Should the Cyclic Redundancy Check (CRC) of a certain NALU indicate a decoding failure, these NALUs are discarded. The SVC decoder uses the low-complexity error concealment method of previous frame-copying in order to replace the lost frames.

Category	Regime
Constellation Size	16-HMCSK-RSC with 2 layers
	64-HMCSK-RSC with 3 layers
Constellation Type	Type I
	Type II
UEP/EEP	EEP
	Optimized UEP

Table 4.7: The regimes with their settings characterized in Section 4.5.

Again, an RSC code having a code-rate of 1/2 and the generator polynomials of $[10011, 00110]$ is employed as the FEC code in our system of Fig. 4.10.

Since we are considering delay-constrained systems, we characterize the attainable performance of the proposed scheme using two scenarios, respectively, which are characterized in Table 4.7.

4.5.1 Off-line LUT Generation

As described in Eq. (4.30) of Section 4.4.3, the estimation of the FER relies on the LUT $\mathcal{T}(\cdot)$. Here we describe the generation of the LUT $\mathcal{T}(\cdot)$ used in our experiments. As mentioned in Section 4.4.3, the LUT $\mathcal{T}(\cdot)$ is indexed by two types of parameters, namely the SNR, and the δ_l parameters, where $l \in \{0, 1, \ldots, N_L - 2\}$. To generate the LUT $\mathcal{T}(\cdot)$, we fix the block-length ℓ of the FEC and obtain the FER $p(\ell)$ of the component FEC by scanning the practical coding parameter ranges of SNR and δ_l at certain intervals. Specifically, the SNR is considered over the range of $[8, 26]$ dB, using a step-size of 0.5 dB. The δ_l values are non-linearly scanned over the range of $\left(0, \frac{\sqrt{M}-1}{\sqrt{M}-2}\right)$, using a step-size of 0.01 over the range of $(0, 0.5]$, and using a step-size of 0.005 over the range of $\left(0.5, \frac{\sqrt{M}-1}{\sqrt{M}-2}\right)$. This will constitute a total number of $n_{\mathcal{T}} = n_{snr} n_r^{N_L-1} = 37 \times \left(50 + \left\lfloor \frac{100\sqrt{M}}{\left(\sqrt{M}-2\right)} \right\rfloor \right)^{N_L-1}$ legitimate settings. For the 16-HCSK-RSC system of Table 4.7, we have $N_L = 2$, $M = 16$ and therefore $n_{\mathcal{T}} = 9,250$. On the other hand, we have $N_L = 3$, $M = 64$ for the 16-HCSK-RSC system of Table 4.7 and therefore $n_{\mathcal{T}} = 1,239,093$. For each optimized UEP regime, a LUT is needed. By combining all the LUTs $\mathcal{T}(\cdot)$, for $l \in \{0, 1, \ldots, N_L - 1\}$, we have $2N_L$ items per setting. An example can be found in Tab. 4.8, showing the LUT needed for the Type I 64-HCSK system using our optimized UEP regime. Since each item in the LUTs can be individually stored as floats in 8 bytes, the total size of the LUT for

Optical SNR(dB)	δ_0	δ_1	$p_0(\ell)$	$p_1(\ell)$	$p_2(\ell)$
...	...	...	...	...	...
13	0.95	0.05	0.357741	0.953467	1
13	0.95	0.10	0.307621	0.955615	1
13	0.95	0.15	0.273464	0.957082	1
13	0.95	0.20	0.230841	0.959333	1
13	0.95	0.25	0.196220	0.960815	1
13	0.95	0.30	0.168964	0.969287	1
13	0.95	0.35	0.146773	0.974093	1
...	...	...	...	...	...

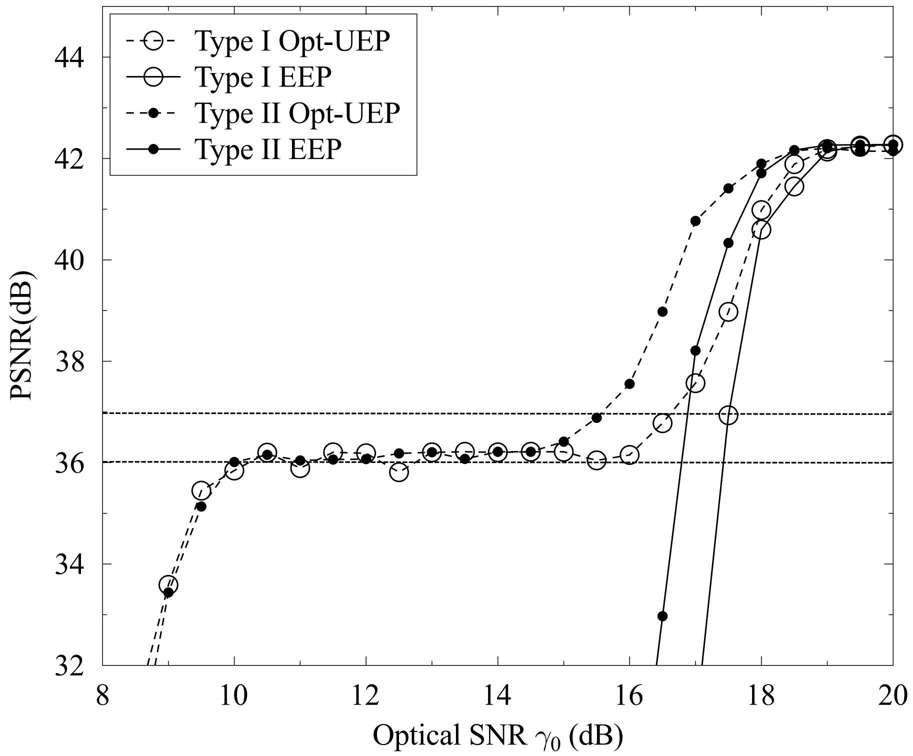
Table 4.8: Example of the LUT of Type I 64-HCSK.

our 16-HCSK-RSC system is 289 KB and that of the 64-HCSK-RSC system is 56.72 MB.

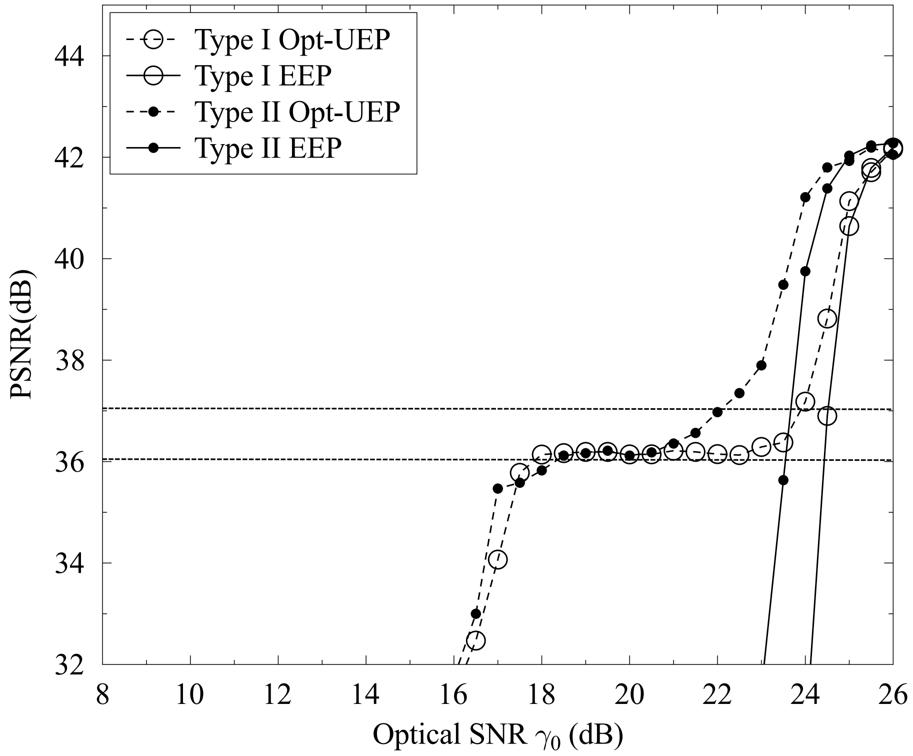
4.5.2 Optimized System Parameters

In order to characterize the PSNR versus optical SNR γ_0 performance of our proposed optimised UEP M-HCSK-RSC system, we compare them to the traditional EEP M-HCSK-RSC scheme for transmission over an AWGN channel, as listed in Table 4.7. These comparisons are shown in Fig. 4.13, 4.14 and 4.15, which were carried out using the above-mentioned three different video sequences, namely the Soccer, the Ice and the Crew sequences, as listed in Table 4.6. The previous Type I and Type II M-HCSK constellation designs were simulated, as described in Section 4.3.1.

The results recorded for the Soccer sequence with the aid of 16-HCSK are shown in Fig. 4.13a. We can observe that both UEP schemes relying on the Type I and Type II constellations outperformed their corresponding EEP counterparts. More specifically, when the Type I constellation design of Table 4.5 is adopted, the optimised UEP system outperforms the EEP benchmark and achieves an optical SNR reduction of about 1.1 dB at a PSNR of 37 dB, which correspond to the scenario of receiving all video layers with a high probability. To elaborate a little further, the staircase-shaped nature of these PSNR curves explicitly indicates the number of video layers received flawlessly. Similarly, for the Type II constellation design of Table 4.5, the optimised UEP system achieves an optical SNR reduction of about 1.4 dB at a PSNR of 37 dB. Additionally, it can be observed that the UEP schemes associated with

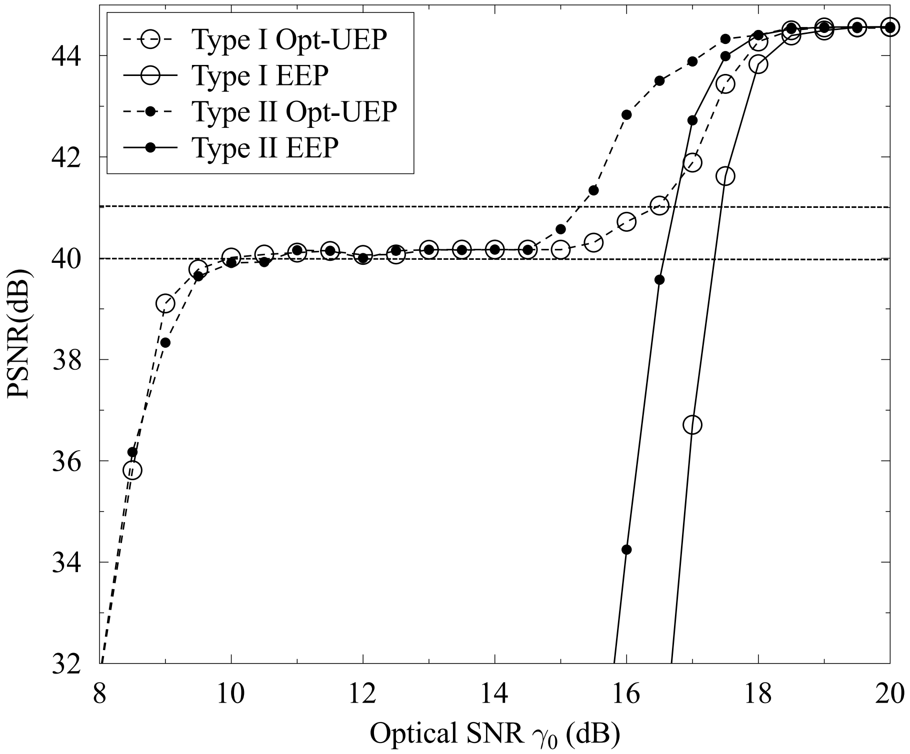


(a) PSNR vs optical SNR γ_0 for *Soccer* with 16-HCSK-RSC

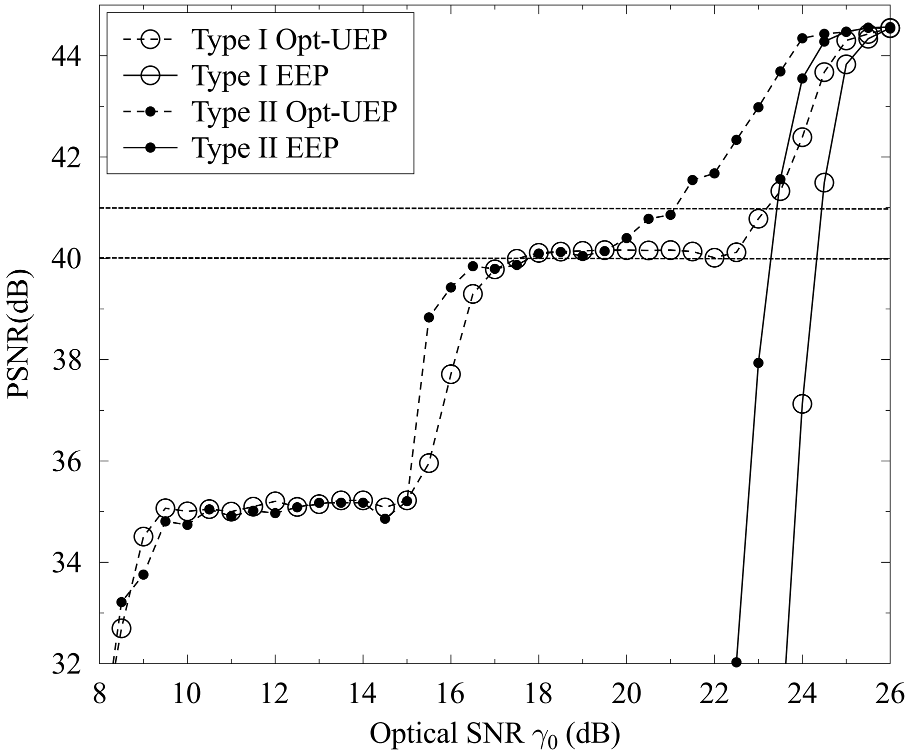


(b) PSNR vs optical SNR γ_0 for *Soccer* with 64-HCSK-RSC

Figure 4.13: PSNR versus optical SNR γ_0 performance of our proposed optimised UEP 16/64-HCSK-RSC system, with either type I or type II HCSK are used as modulation.The EEP HCSK-RSC transmissions are used as benchmarks. The Soccer sequence is used for transmission over AWGN channels. The schematic of Fig. 4.10 and the parameters of Table 4.2 were used.

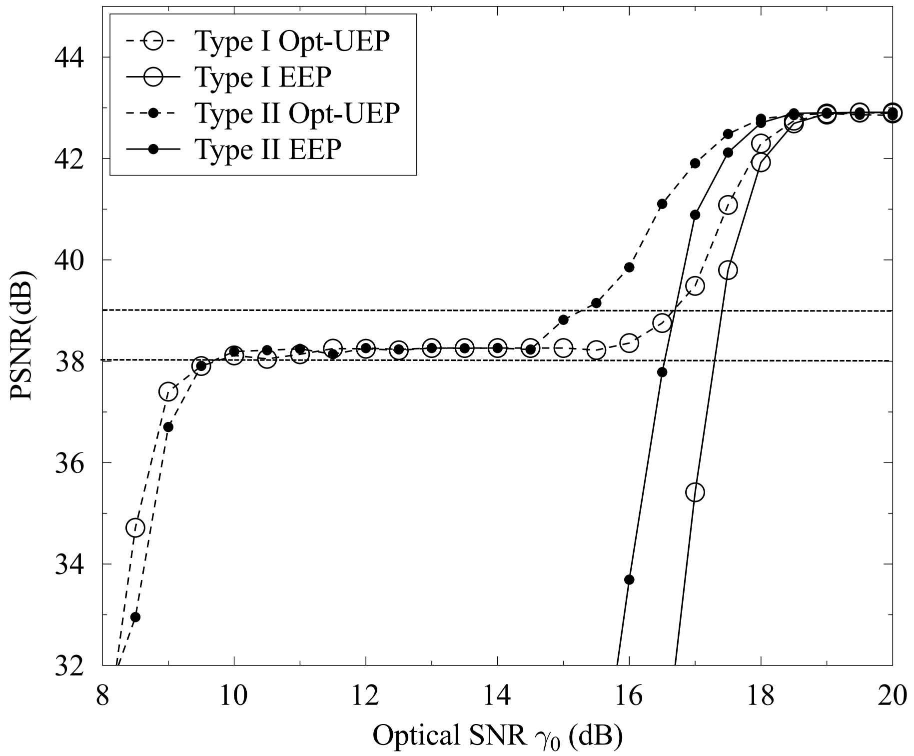


(a) PSNR vs optical SNR γ_0 for *Ice* with 16-HCSK-RSC

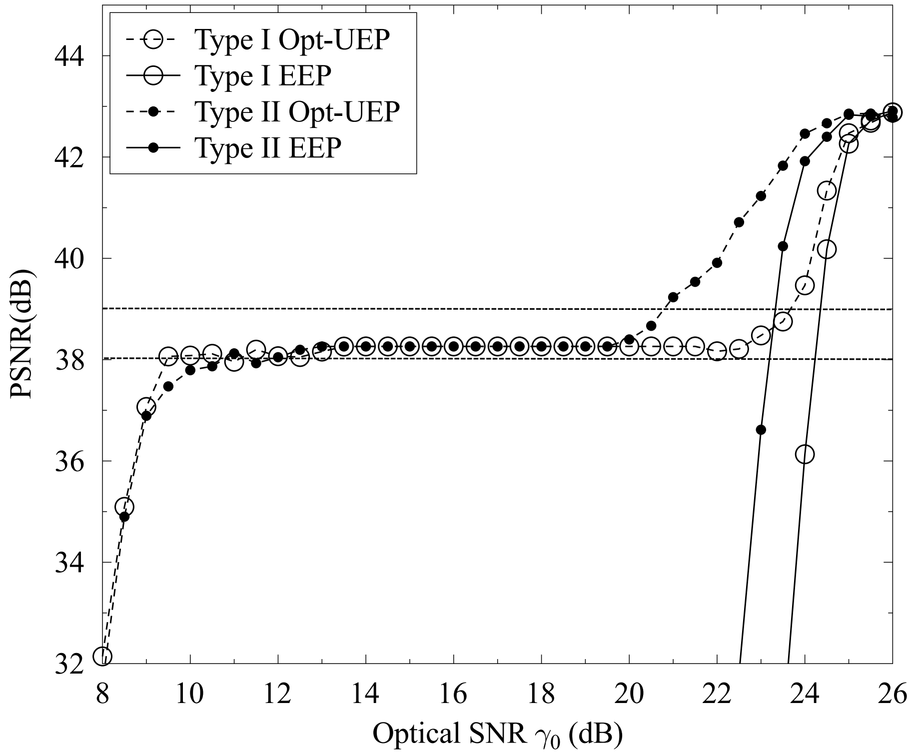


(b) PSNR vs optical SNR γ_0 for *Ice* with 64-HCSK-RSC

Figure 4.14: PSNR versus optical SNR γ_0 performance of our proposed optimised UEP 16/64-HCSK-RSC system, with either type I or type II HCSK are used as modulation.The EEP HCSK-RSC transmissions are used as benchmarks. The Ice sequence is used for transmission over AWGN channels. The schematic of Fig. 4.10 and the parameters of Table 4.2 were used.

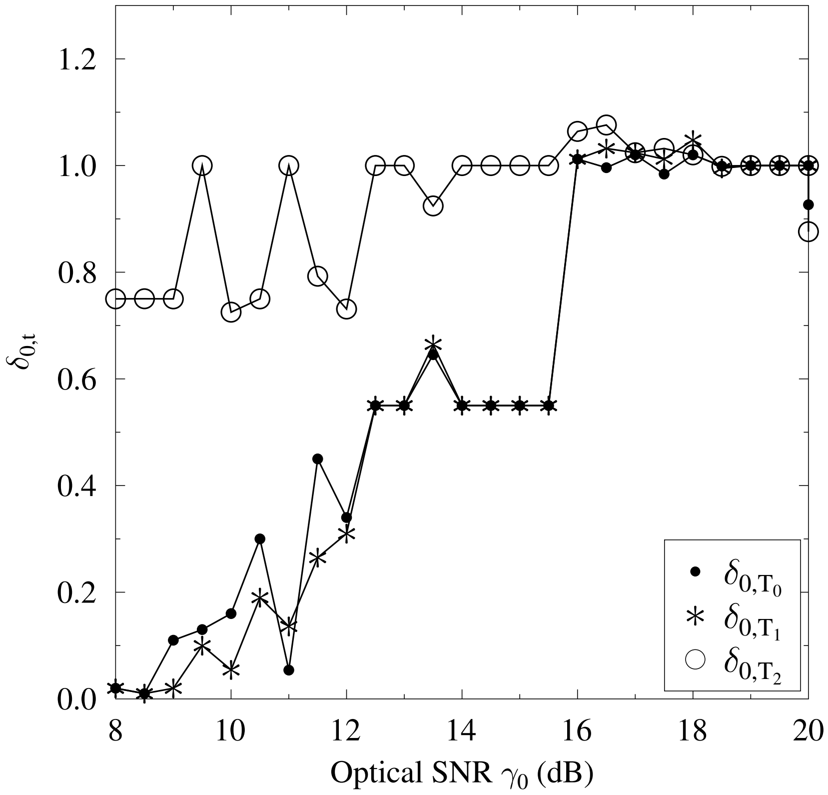


(a) PSNR vs optical SNR γ_0 for *Crew* with 16-HCSK-RSC



(b) PSNR vs optical SNR γ_0 for *Crew* with 64-HCSK-RSC

Figure 4.15: PSNR versus optical SNR γ_0 performance of our proposed optimised UEP 16/64-HCSK-RSC system, with either type I or type II HCSK are used as modulation.The EEP HCSK-RSC transmissions are used as benchmarks. The Crew sequence is used for transmission over AWGN channels. The schematic of Fig. 4.10 and the parameters of Table 4.2 were used.

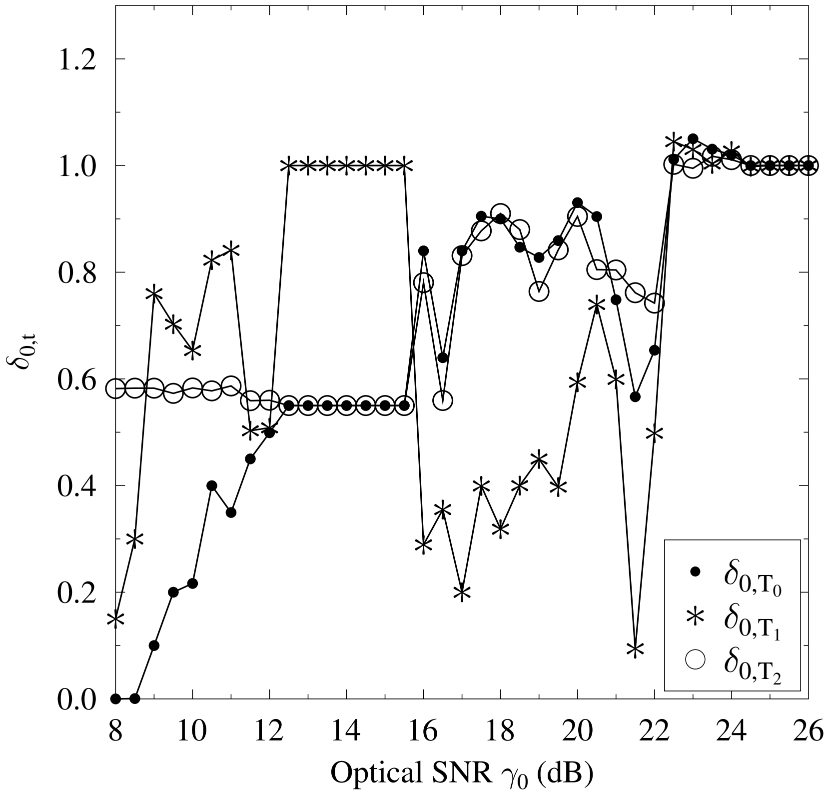


(a) The choice of $\delta_{0,t}$ vs optical SNR γ_0 for 16-HCSK-RSC

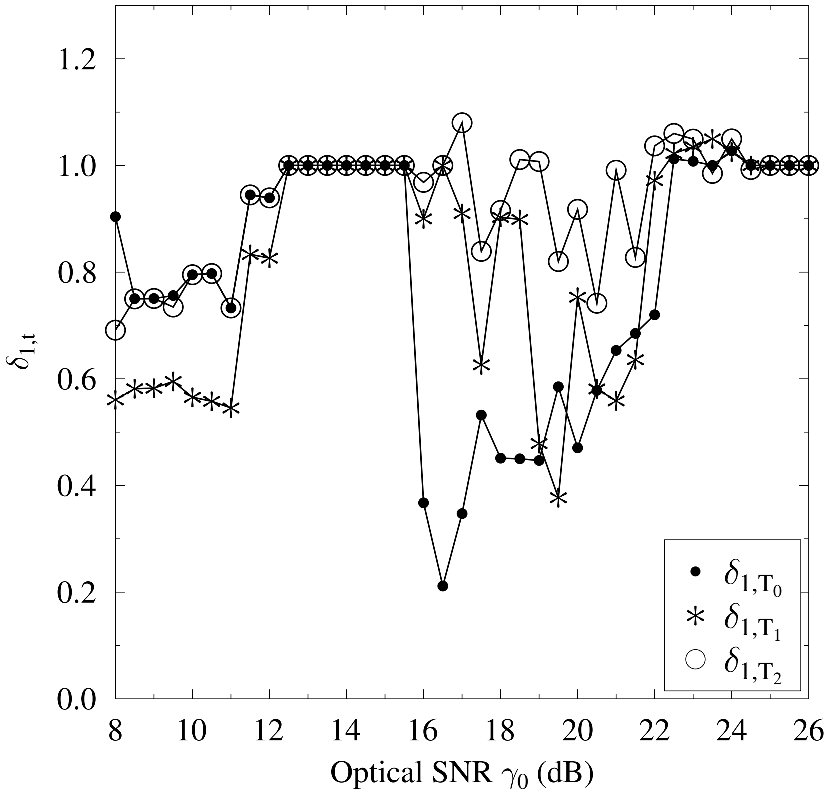
Figure 4.16: The choice of parameters versus optical SNR γ_0 with the optimised UEP 16-HCSK-RSC transmission system for the $1st$ video frame of *Soccer*. The schematic of Fig. 4.10 and the parameters of Table 4.2 were used.

the Type II HCSK constellation perform significantly better than the Type I HCSK, and an optical SNR reduction of about 1.1 *dB* can be achieved at a PSNR of 37 *dB* by using the former over the latter. This suggests that the system associated with the Type II HCSK constellation design is more capable.

Now that we have inspected the performance of our optimised UEP M-HCSK-RSC system at higher optical SNR values, we will also characterise them at lower optical SNRs. A substantial gain can be observed for the UEP schemes in Fig. 4.13a around the PSNR of 36 *dB*. Quantitatively, up to 7.5 *dB* of optical SNR reduction can be obtained by the optimised UEP system using the Type I constellation of Table 4.5, which is increased to 6.5 *dB* for the optimised UEP systems using the Type II constellation. This energy reduction is an explicit benefit of the fact that the optimisation algorithm can decide whether to reduce the energy assigned to the less important layers depending on the channel conditions, owing to the design flexibility of M-HCSK. Once the channel becomes so noisy that it is unrealistic to transmit all the layers successfully, the optimisation will opt for reallocating the energy for transmitting the more important layers, by adjusting the δ_l values to increase the



(a) The choice of $\delta_{0,t}$ vs optical SNR γ_0 for 64-HCSK-RSC



(b) The choice of $\delta_{1,t}$ vs optical SNR γ_0 for 64-HCSK-RSC

Figure 4.17: The choice of parameters versus optical SNR γ_0 with the optimised UEP 64-HCSK-RSC transmission system for the $1st$ video frame of *Soccer.* The schematic of Fig. 4.10 and the parameters of Table 4.2 were used.

Euclidean distance amongst the 16-HCSK constellation subgroups conveying the base layer.

In Fig. 4.16a we can observe the choice of the optimum $\delta_{0,t}$ values versus the optical SNR γ_0 for the optimised UEP 16-HCSK-RSC transmission system for the $1st$ video frame of *Soccer*. We find that for the M-HCSK frame type T_2 δ_{0,T_2} stays close to 1 as γ_0 increases, which is due to the fact the frame type of T_2 happens to be $Lv_2 - Lv_2 - Lv_2$, hence there is no room for optimisation. By contrast, we can see that the values of $\delta_{0,t}$ for T_0 and T_1 are small at a low γ_0, in order to keep a reasonable minimum Euclidean distance for the lower-layer symbols, hence giving little priority to the higher layers. As γ_0 increases, the optimised $\delta_{0,t}$ values gradually settle around 1, as optimisation becomes unnecessary for high γ_0 values.

By contrast, the results recorded for the Soccer sequence with the aid of 64-HCSK are shown in Fig. 4.13b. We can observe that both UEP schemes associated with the Type I and Type II constellation outperformed their corresponding EEP counterparts. More specifically, when the Type I constellation design is adopted, the optimised UEP system outperforms its EEP benchmarker and achieves an optical SNR reduction of about 0.8 dB at a PSNR of 37 dB. Similarly, for the Type II constellation design, the optimised UEP system achieves an optical SNR reduction of about 1.7 dB at a PSNR of 37 dB. Additionally, it can be observed that the UEP schemes associated with the Type II HCSK constellation performs significantly better than that with the Type I HCSK, and an optical SNR reduction of about 1.9 dB can be achieved at a PSNR of 37 dB by using the former over the latter.

Similar trends can be observed, when the Ice or Crew sequences are used, as shown in Fig. 4.14a to Fig. 4.15b. We infer from these results that our optimised UEP M-HCSK-RSC scheme associated with the Type II constellation design is applicable to video sequences of diverse nature, and it is capable of achieving a beneficial performance gain for both $M = 16$ and 64. The subjective comparison of the decoded videos associated with our different regimes is discussed in Section 4.5.3.

4.5.3 Subjective Comparison

Explicitly, Fig. 4.18 shows the subjective comparison of the decoded video frames associated with our different regimes for 16-HCSK as the modulation using the Soccer sequence at an optical SNR value of $\gamma_0 = 10$ dB. The 20-th frame of the recovered videos of some of our schemes are shown in the top row of Fig. 4.18. The EEP 16-HCSK-RSC scheme using the Type I constellation is more error-prone according to Section 4.5.2, and in this regime all three layers of this video frame failed to get recovered, and so did all their preceding frames. The difference frame, which is obtained by subtracting the recovered frame from the 20-th frame of the original video, has substantial non-zero values. Continuing from left to right, we can observe that

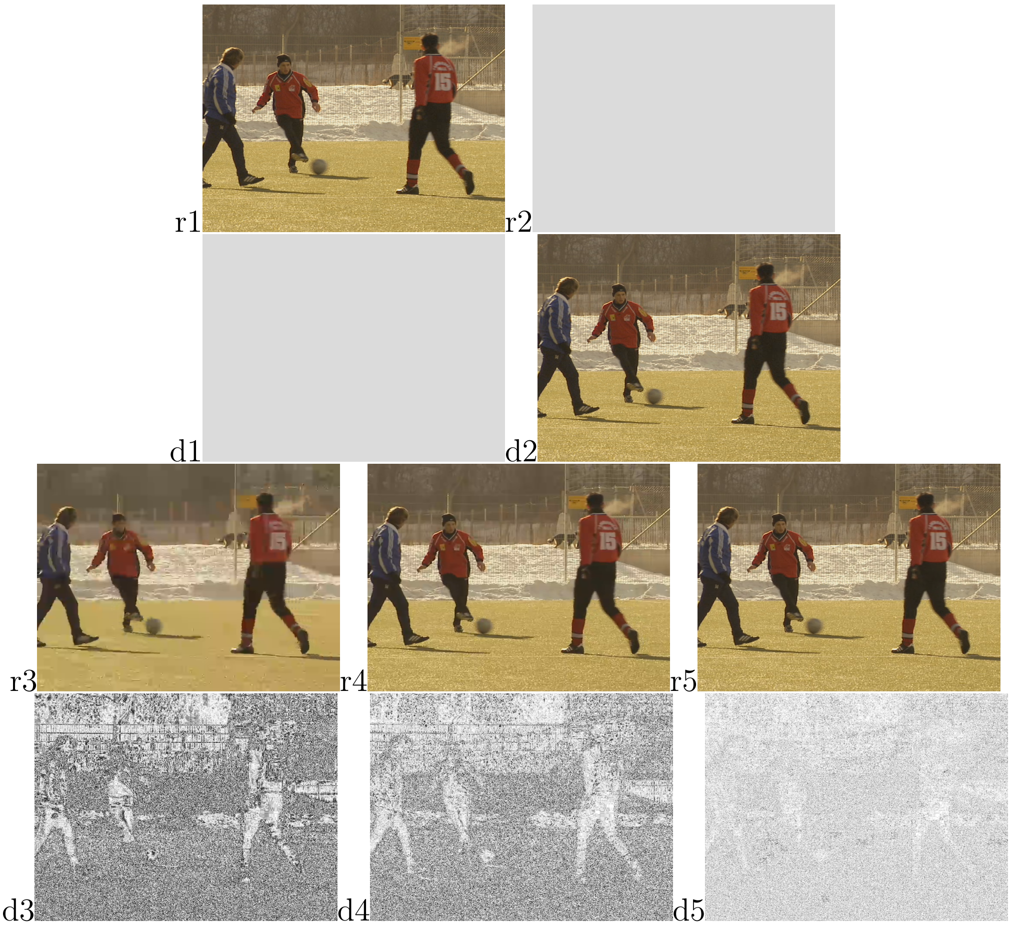


Figure 4.18: Comparison of decoded frames of the 20-th frame at an optical SNR γ_0 of 17 dB for the *Soccer* sequence for a 16-HCSK system. The figures indicated by 'r' indicate the original video frames, the EEP 16-HCSK-RSC scheme using Type I constellation, the optimised UEP scheme using Type I constellation, the EEP scheme using Type II constellation, and the optimised UEP scheme using Type II constellation, respectively. The figures indicated by 'd' correspond to the difference frames between the received and the original video frames. The schematic of Fig. 4.10 and the parameters of Table 4.2 were used.

the frames corresponding to the optimised UEP scheme using the Type I constellation, the EEP scheme using the Type II constellation and the optimised UEP scheme using the Type II constellation become sharper, revealing more intricate video details, while the corresponding difference frames have less and less non-zero values, which indicates the improvement of the video quality.

4.6 Conclusions

In this chapter we conceived hierarchical CSK modulation based on the traditional CSK, which is capable of conveying inter-dependent layers of video signals. In Section 4.3 we proposed the Type I and Type II HCSK, which can be flexibly configured by changing its parameters, as detailed in Section 4.3.1. According to Fig. 4.8 and Fig. 4.9, the Type II constellation allows us to use a wider range of FERs for the higher layers upon varying the constellation-shaping parameter δ_l, therefore it is more promising in terms of reducing the effect of cross-layer interferences. Hence it provides a high grade of flexibility in terms of system configuration and optimization.

Furthermore, we provided a detailed design example for the employment of HCSK transmitting scalable video sources with the aid of a RSC code in Section 4.4. An optimisation method was conceived in Section 4.4.3 in order to enhance the UEP and to improve the quality of the received video.

The results recorded in Section 4.5 characterized the performance of our proposed HCSK-RSC system of Fig. 4.10. We have summarized our results in Table 4.9 and Table 4.10 based on Fig. 4.13-Fig. 4.15.

Observe in Fig. 4.13a for the *Soccer* sequence transmitted with the aid of 16-HCSK that both UEP schemes relying on the Type I and Type II constellations outperformed their corresponding EEP counterparts. More specifically, when the Type I constellation design of Table 4.5 is adopted, the optimised UEP system outperforms the EEP benchmark and achieves an optical SNR reduction of about 1.1 dB at a PSNR of 37 dB, while the Type II constellation design associated with the optimised UEP of Table 4.5 achieves an optical SNR reduction of about 1.4 dB. Additionally, it can be observed that the UEP schemes associated with the Type II HCSK constellation perform significantly better than the Type I HCSK, and an optical SNR reduction of about 1.1 dB can be achieved at a PSNR of 37 dB by using the former over the latter, which suggests that the system associated with the Type II HCSK constellation design is more capable.

More importantly, observe in Fig. 4.13a that quite a substantial gain can be observed for the UEP schemes at the slightly lower optical SNR of 36 dB. Quantitatively, up to 6.5 dB of optical SNR reduction can be obtained by the optimised UEP system using the Type II constellation of Table 4.5, which is increased to 7.5 dB for the optimised UEP systems using the Type I constellation. This energy reduction is an explicit benefit of the fact that the optimisation algorithm can decide whether to reduce the energy assigned to the less important layers depending on the channel conditions, owing to the design flexibility of our M-HCSK of Fig. 4.10.

From the results recorded for the *Soccer* sequence with the aid of 64-HCSK in Fig. 4.13b, we can also observe that both UEP schemes of Section 4.4 associated with the Type I and Type II constellation defined in Section 4.3 outperformed their corresponding EEP counterparts. More specifically, when the Type I constellation design is adopted, the optimised UEP system outperforms its EEP benchmarker and achieves an optical SNR reduction of about 0.8 dB at a PSNR of 37 dB. Similarly, for the Type II constellation design, the optimised UEP system achieves an optical SNR reduction of about 1.7 dB at a PSNR of 37 dB. Additionally, it can be observed that the UEP schemes associated with the Type II HCSK constellation perform

Transmission Schemes		Sequences	Compared at	Difference
16-HCSK-RSC	Type I Opt-UEP over Type I EEP	Soccer	37 dB	1.4 dB
			36 dB	7.5 dB
		Ice	42 dB	0.6 dB
			40 dB	7.3 dB
		Crew	39 dB	0.7 dB
			38 dB	7.4 dB
	Type II Opt-UEP over Type II EEP	Soccer	37 dB	1.1 dB
			36 dB	6.5 dB
		Ice	42 dB	1.2 dB
			40 dB	6.5 dB
		Crew	39 dB	1.3 dB
			38 dB	7.1 dB
	Type II Opt-UEP over Type I Opt-UEP	Soccer	37 dB	1.1 dB
			36 dB	0 dB
		Ice	42 dB	1.3 dB
			40 dB	0 dB
		Crew	39 dB	1.3 dB
			38 dB	0 dB

Table 4.9: Optical SNR γ_0 performance comparisons of different transmission schemes using the 16-HCSK-RSC system of Fig. 4.10. Type I or type II HCSK are used as modulation. Optimised UEP schemes are also compared over EEP schemes. The *Soccer*, *Ice* and *Crew* sequences are used for transmission over AWGN channels. The parameters of Table 4.2 were used.

significantly better than that with the Type I HCSK, and an optical SNR reduction of about 1.9 *dB* can be achieved at a PSNR of 37 *dB* by using the former over the latter.

Similar trends can be observed for the *Ice* and *Crew* sequence, as detailed in Fig. 4.14 and Fig. 4.15, which are also characterized in Table 4.9 and Table 4.10.

Finally, we provided a subjective comparison in Section 4.5.3 for our evaluated transmission schemes using the *Soccer* sequence in Fig. 4.18.

Transmission Schemes		Sequences	Compared at	Difference
64-HCSK-RSC	Type I Opt-UEP over Type I EEP	Soccer	37 dB	0.8 dB
			36 dB	6.7 dB
		Ice	42 dB	0.7 dB
			40 dB	5.4 dB
		Crew	39 dB	0.7 dB
			38 dB	15 dB
	Type II Opt-UEP over Type II EEP	Soccer	37 dB	1.7 dB
			36 dB	5.8 dB
		Ice	42 dB	1.5 dB
			40 dB	6 dB
		Crew	39 dB	2.5 dB
			38 dB	13 dB
	Type II Opt-UEP over Type I Opt-UEP	Soccer	37 dB	1.9 dB
			36 dB	-0.3 dB
		Ice	42 dB	1.6 dB
			40 dB	0 dB
		Crew	39 dB	3 dB
			38 dB	-0.8 dB

Table 4.10: Optical SNR γ_0 performance comparisons of different transmission schemes using the 64-HCSK-RSC system of Fig. 4.10. Type I or type II HCSK are used as modulation. Optimised UEP schemes are also compared over EEP schemes. The *Soccer*, *Ice* and *Crew* sequences are used for transmission over AWGN channels. The parameters of Table 4.2 were used.

Chapter 5

Conclusions and Future Work

In this concluding chapter, a summary of the thesis and the main findings of our investigations will be presented. This will be followed by a range of ideas on future research.

5.1 Summary and Conclusions

This thesis investigated a number of techniques that may be utilized for video compression and communication.

- In Chapter 1, the background of this treatise was provided. Specifically, Section 1.1 gave a historical review of video compression, including both lossy video compression as reviewed in Section 1.1.1 as well as the lossless video compression as portrayed in Section 1.1.2. Additionally, Section 1.1.3 provided the background of layered video coding techniques, while Section 1.1.4 introduced some preliminaries for the evaluation of video communications. Section 1.2 took a brief look at the research history of ISCD, while Section 1.3 reviewed the the subject area of UEP. Section 1.5 was dedicated to indoor VLC systems. Finally, the motivation and methodology of this thesis was provided in Section 1.6 and its outline was detailed in Section 1.7. We described the novel contributions of this treatise in Section 1.8.
- In Chapter 2, we proposed a video codec suitable for lossless video compression and ISCD. Specifically, we further developed the soft source decoder of [43] for the transmission of video sequences rather than still video pictures. More explicitly, we used low complexity frame-differencing (FD) for removing the temporal redundancy and a VrLC for removing the residual spatial redundancy of the FD signal, while exploiting the correlations amongst the FD pixels within the current frame with the aid of our MRF model based SISO module. In Section 2.2, we introduced our video codec and the structure of the transmitter. Moreover, a novel three-stage ISCD structure was proposed, which outperforms

the two-stage architecture. Furthermore, we examined the convergence of the three-stage iterative decoding process using 3D extrinsic information transfer (EXIT) charts. Section 2.3 presented the algorithms conceived for all the decoding modules and then the corresponding JSCC algorithm is detailed in the context of loss-less video compression, followed by our 3D EXIT chart results. The performance of our system operating both with and without FD is compared to our benchmarkers in Section 2.5. Finally, we offered our chapter conclusions in Section 2.6.

- An adaptive THARQ (ATHARQ) algorithm was detailed in Section 3.3 for IL-FEC coded layered video streaming for the sake of minimizing the video distortion under the constraint of a given total number of transmission TSs. The adaptive retransmission controller predicted the channel conditions and estimated the SNR values at the receiver for the sake of appropriately configuring the transmitter. The specific video layer, which would most effectively reduce the video distortion at the receiver were chosen for transmission. Section 3.2 also detailed the IL-FEC transmitter and receiver model, as well as the proposed ATHARQ protocol along with the benchmark schemes we used in this treatise. Furthermore, we developed an on-line optimization technique for our IL-ATHARQ transmission scheme, in order to find the most beneficial FEC code rate for each of the video layers that resulted in a reduced video distortion. A method of estimating the video distortions related to each code rate assignment was conceived for the IL-ATHARQ transmission. The details of the coding-rate optimization of the IL-ATHARQ system was included in Section 3.4. The performance of our IL-ATHARQ scheme as well as the rate-optimized IL-ATHARQ scheme using a RSC codec were compared to the benchmarks in Section 3.5 using different video sequences, followed by characterizing both the effects of the delay as well as well of those channel quality prediction errors on the attainable system performance. Our simulation results demonstrated that the optimized IL-FEC system outperforms the traditional THARQ system by an E_b/N_0 value of about 5.3 dB at a PSNR of 38.5 dB. Alternatively, about 2.5 dB of PSNR video quality improvement may be observed at an E_b/N_0 of 15 dB, when employing a RSC code. Finally, we concluded in Section 3.6.
- In Chapter 4, we introduced a novel video streaming method for transmission over indoor optical domain channel with the aid of HCSK. Section 4.2 introduced the CSK transmitter and receiver model, including the signal constellation and bit-to-symbol mapping, the optical domain channel model as well as the soft demodulation method. We introduced the construction of our novel M-HCSK modulation and its layered demodulation method in Section 4.3, which was capable of generating inter-dependent layers of signals with different

error probability. Its symbol constellation can be readily configured for providing various settings and for redistributing the error probabilities amongst the different layers. Furthermore, we demonstrated the employment of HCSK by transmitting scalable video sources, as detailed in Section 4.4. The system was designed to employ UEP by carefully exploiting both the layered feature of the source and the new HCSK modulator. An optimisation method was conceived in order to enhance the UEP and to improve the quality of the received video. The performance of our optimised M-HCSK-TCM video transmission system using different HCSK constellation sizes was compared to the benchmarkers in Section 4.5 using different video sequences. The simulation results demonstrated that the proposed system outperformed the traditional EEP scheme by about 1.7 dB of optical SNR γ_o at a peak signal-to-noise ratio (PSNR) of 37 dB, while supporting a wide range of system configurations attaining optical SNR reductions of up to 6.5 dB at a PSNR of 36 dB. Finally, we concluded in Section 4.6.

5.2 Design Guidelines

Based on the above-mentioned investigations, below we summarize our general design guidelines conceived for video communication schemes by carefully examining our proposals detailed in Chapters 1-4.

5.2.1 ISCD System Design

In case of interactive lip-synchronized video communications having a maximum of say 100ms encoding/decoding delay as well as a limited complexity, a certain amount of correlation resides within the compressed data-streams. Therefore, the concept of SBSD and ISCD can be envoked for exploiting the APPs of the source parameters. Specifically, an ISCD scheme exchanged extrinsic information between the source decoder and channel decoder in a turbo-like iterative decoding process. For the sake of achieving near-capacity operation, the source-and-channel-codes have to be carefully matched with the aid of EXIT-charts.

5.2.2 Layered Video Transmission Design

1. In our simplified layered video transmission scenarios, we assumed that N video frames each containing L layers are included in a GOP to be transmitted. All video frames are of different importance, and so are the layers included in each frame. The N by L layers are interdependent. Therefore IL techniques such as IL-FECs may be adopted in order to exploit the dependencies amongst the video layers. The transmission resources have to be carefully distributed amongst the different video layers to offer UEP, while taking into consideration

the contribution of each layer to the video reconstruction, in order to achieve the best possible video quality.

2. In delay-tolerant THARQ scenarios, the video layers are allowed to be transmitted or retransmitted over a limited number T timeslots. The quality of the wireless channel during the forthcoming timeslots is usually unpredictable, hence the global optimization of the transmission scheduling may not be feasible.

5.2.3 Hierarchical CSK Modulation Design

Hierarchical CSK (HCSK) modulation constitutes a refined version of the traditional CSK, which is capable of generating interdependent layers of signals associated with different error probability. The low-priority layers cannot be demodulated and decoded until the more important layers have been successfully demodulated and decoded. The symbol constellation can be flexibly configured to provide various settings and error probabilities for the different layers.

5.3 Future Work

In this thesis we have proposed several techniques for video compression and communication, which may be further improved, as detailed below.

5.3.1 Practical Application Scenarios for MRF Soft Decoder

In Chapter 2, we proposed a video codec suitable for lossless video compression and ISCD. Specifically, we further developed the soft source decoder of [43] for the transmission of video sequences rather than still video pictures. However, we only adopted low complexity frame-differencing (FD) for removing the temporal redundancy and a VrLC for removing the residual spatial redundancy of the FD signal. In our future work we will improve our temporal prediction method of frames with the aid of sophisticated motion compensation, and we may consider DCT-based transformation methods for further reducing the intra-frame redundancies, in order to make the codec more practical. Furthermore, we shall replace our VLC with more efficient entropy codecs in order to improve the attainable compression ratio. We may also consider the adoption of the MRF soft decoder in distributed video coding (DVC) [260, 261].

5.3.2 Transmission of Layered Video Using THARQ With IR

In Chapter 3, we conceived an adaptive THARQ transmission scheme in support of IL-FEC coded layered video for minimizing the video distortion under the constraint of a given total number of transmission time slots. Furthermore, we developed a method of on-line optimization for our IL-ATHARQ transmission scheme, in order to find the most appropriate FEC code-rate for the video layers that controls the video distortion. However, Type-I HARQ was used for simplicity, which is not the most bandwidth-efficient technique. Therefore in our future work, we will further develop our THARQ scheme for incremental redundancy aided schemes.

5.3.3 Joint FEC and Modulation Optimization for Indoor Optical Transmission of Layered Video

In Chapter 4, a new CSK modulation scheme was conceived, which is capable of providing different error probabilities for the layers of signals and can be flexibly reconfigured by changing its parameters. We conceived a design example for the HCSK-aided transmission of scalable video sources. The system was designed to provide UEP by exploiting both the layered features of the source and the new HCSK modulator. However, we only configured the parameter settings of the HCSK modulator, which was combined with a fixed coding-rate FEC. In our future work, we will consider systems, where the modulation scheme and the coding rate of the FEC are jointly optimized, in order to achieve the best possible video performance.

5.3.4 Compression and Transmission of Holographic Video

Recently, holographic displays have appeared on the market, which can display scenes from multiple viewing angles without goggles. Since the video sequences captured by cameras positioned at different angles are highly correlated, a large amount of redundancy may be removed with the aid of carefully designed video compression.

Furthermore, we may further investigate the transmission of the holographic information over hostile wireless channels and design UEP schemes for joint FEC and modulation methods for 3D holographic transmission.

5.3.5 Video Streaming in Camera-Phone Networks

Since multiview video communications facilitate the selection of several camera views of a given scene, it may be deemed to be a promising technique also for mobile multiview television. With the popularity of smart-phones equipped with high-resolution cameras, it is possible to capture, share and process multiview video over camera-phone networks. Hence it is necessary to design techniques for multiview video communications in order to combat the packet loss events and to remove the redundancy for the sake of improving the transmission efficiency.

Glossary

A

ACK Acknowledgement.

APP A-Posteriori Probability.

ARQ Automatic Repeat reQuest.

ASP Advanced Simple Profile.

ATHARQ Adaptive THARQ.

AVC Advanced Video Coding.

AVS Advanced Video Standard.

AWGN Additive White Gaussian Noise.

B

BER Bit Error Ratio.

BL Base Layer.

C

CC Convolutional Codes.

CGS Coarse Grain Scalability.

CIF Common Intermediate Format.

CPB Coded Block Pattern.

CRC Cyclic Redundancy Check.

CSI Channel State Information.

CSK Colour-Shift Keying.

D

DCT Discrete Cosine Transform.

DP Data Partitioning.

DPCM Differential Pulse Code Modulation.

DPPM Differential PPM.

DV Digital Video.

DVD Digital Versatile Disc.

E

EEP Equal Error Protection.

EL Enhancement Layer.

EMI Electro-Magnetic Interference.

EXIT EXtrinsic Information Transfer.

F

FEC Forward Error Correction.

FER Frame Error Ratio.

FFH Fast Frequency-Hopping.

FSK Frequency Shift Keying.

G

GOP Group Of Pictures.

H

HARQ Hybrid Automatic Retransmission Request.

HCSK Hierarchical Colour-Shift Keying.

HD High Definition.

HetNet Heterogeneous Networks.

HEVC High-Efficiency Video Coding.

HM Hierarchical Modulation.

I

IEC International Electrotechnical Commission.

IL-FEC Inter-Layer Forward Error Correction.

IR Infrared.

IrVrLC Irregular Variable Length Code.

ISCD Iterative Source-Channel Decoding.

ISDN Integrated Services Digital Networks.

ITU International Telecommunication Union.

ITU-T International Telecommunication Union Telecommunication standardization sector.

J

JPEG Joint Photographic Experts Group.

JSCC Joint Source-Channel Coding.

JSCD Joint Source-Channel Decoding.

K

kbps KBits Per Second.

L

LDPC Low-Density Parity-Check.

LED Light Emitting Diodes.

LLR Log-Likelihood Ratio.

LT Luby Transform.

LUT LookUp Table.

M

MAP Maximum A-Posteriori probability.

MC Motion Compensation.

MC-LEDs Multi-Chip LEDs.

MGS Medium Grain Scalability.

MLC Multilevel Code.

MPEG Moving Picture Experts Group.

MPPM Multiple PPM.

MRC Maximal Ratio Combining.

MRF Markov Random Field.

MSE Mean Square Error.

MVC Multiview Video Coding.

N

NACK Negative Acknowledgement.

NALU Network Abstraction Layer Unit.

NCR Natural Code Rate.

NRZ Non-return-to-zero.

O

OF Objective Function.

OFDM Orthogonal Frequency-division Multiplexing.

OLEDs Organic LEDs.

OMEGA Home Gigabit Access.

OOK On-Off Keying.

OWC Optical Wireless Communication.

P

PAM Pulse Amplitude Modulation.

PC-LEDs Phosphor Converted LEDs.

PD Photon-Detector.

PET Priority Encoding Transmission.

PHY I Physical layer I.

PHY II Physical layer II.

PHY III Physical layer III.

PLC Power Line Communication.

PPM Pulse Position Modulation.

PSNR Peak Signal-to-Noise Ratio.

Q

QCIF Quarter Common Intermediate Format.

QoE Quality of Experience.

QP Quantization Parameter.

R

RCPC Rate-Compatible Convolutional Code.

RD Rate Distortion.

RF Radio Frequency.

RS Reed-Solomon.

RSC Recursive Systematic Convolutional.

RVrLC Reversible Variable-length Code.

S

SBSD SoftBit Source Decoding.

SER Symbol Error Rate.

SHVC Scalability Extension of H.265/HEVC Standard.

SISO Soft In Soft Out.

SMPTE Society of Motion Picture and Television Engineers.

SP Simple Profile.

SVC Scalable Video Coding.

T

TCM Trellis Coded Modulation.

TDD Time Division Duplex.

THARQ Truncated HARQ.

TS Time Slots.

U

UEP Unequal Error Protection.

UHD Ultra High Definition.

UWB Ultra Wideband.

V

VCD Video Compact Disc.

VCEG Video Coding Experts Group.

VLC Visible Light Communication.

VLCC Visible Light Communications Consortium.

VoIP Voice over IP.

VPPM Variable PPM.

VrLC Variable Length Coding.

W

WiMAX Worldwide Interoperability for Microwave Access.

WMV9 Windows Media Video version 9.

WPAN Wireless Personal Area Networks.

Bibliography

[1] S. Bowman and C. Willis, "We media: How audiences are shaping the future of news and information," 2003.

[2] L. Hanzo, P. Cherriman, and J. Streit, *Video Compression and Communications: From Basics to H.261, H.263, H.264, MPEG2, MPEG4 for DVB and HSDPA-Style Adaptive Turbo-Transceivers.* New York: John Wiley, 2007.

[3] G. K. Wallace, "The JPEG still picture compression standard," *Communications of the ACM*, pp. 30–44, 1991.

[4] ITU-T, *H.263 : Video coding for low bit rate communication*, January 2005.

[5] COST 211 and CCITT, *ITU-T Rec. H.120: Codecs for videoconferencing using primary digital group transmission*, November 1988.

[6] ITU-T, *H.261: Video codec for audiovisual services at p x 384 kbit/s*, November 1988.

[7] ISO/IEC 11172, *Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s*, 1993.

[8] IEC 61834-2, *Recording - Helical-scan digital video cassette recording system using 6,35 mm magnetic tape for consumer use*, August 1998.

[9] "Standards facilitate interoperability | society of motion picture & television engineers." https://www.smpte.org/standards. (Visited on 12/09/2015).

[10] Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, *ITU-T Rec. H.264/ISO/IEC 14496-10 AVC: Advanced Video Coding for Generic Audiovisual Services*, March 2010.

[11] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, pp. 1103–1120, September 2007.

[12] Y. Chen, Y.-K. Wang, K. Ugur, M. Hannuksela, J. Lainema, and M. Gabbouj, "The emerging MVC standard for 3D video services," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, pp. 1–13, January 2009.

[13] AVS Workgroup of China, *Audio and Video Coding Standard*, January 2005.

[14] ITU-T, *H.265 : High efficiency video coding*, April 2013.

[15] G. Sullivan, J. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," vol. 22, no. 12, pp. 1649–1668.

[16] T. K. Tan, M. Mrak, V. Baroncini, and N. Ramzan, "Hevc verification test report," *MPEG-H JCT-VC*, May 2014.

[17] S. Takamura, "Lossless video coding," 2007.

[18] "Lossless compression - wikipedia, the free encyclopedia." https://en.wikipedia.org/wiki/Lossless_compression. (Visited on 11/20/2015).

[19] "Huffyuv." http://neuron2.net/www.math.berkeley.edu/benrg/huffyuv.html. (Visited on 11/19/2015).

[20] "Ffv1." http://sourceforge.net/projects/ffdshow/. (Visited on 11/19/2015).

[21] "Arithyuv." http://www.free-codecs.com/download/arithyuv.htm. (Visited on 11/19/2015).

[22] "Msu lossless video codec." http://www.compression.ru/video/ls-codec/index_en.html. (Visited on 11/19/2015).

[23] "Lossless video codec by yuvsoft." http://www.yuvsoft.com/2d-technologies/lossless-video-codec/. (Visited on 11/19/2015).

[24] "Lagarith lossless video codec." http://lags.leetcode.net/codec.html. (Visited on 11/19/2015).

[25] G. J. Sullivan, P. N. Topiwala, and A. Luthra, "The H. 264/AVC advanced video coding standard: Overview and introduction to the fidelity range extensions," in *Optical Science and Technology, the SPIE 49th Annual Meeting*, pp. 454–474, International Society for Optics and Photonics, August 2004.

[26] C. M. G. . M. L. V. Group, "MSU lossless video codecs comparison," March 2007.

[27] "Videolan - x264." http://www.videolan.org/developers/x264.html. (Visited on 11/19/2015).

[28] T. Zhang and Y. Xu, "Unequal packet loss protection for layered video transmission," *IEEE Transactions on Broadcasting*, vol. 45, pp. 243–252, June 1999.

[29] M. Ghanbari, "Two-layer coding of video signals for vbr networks," *IEEE Journal on Selected Areas in Communications*, vol. 7, pp. 771–781, Jun 1989.

[30] S. McCanne, M. Vetterli, and V. Jacobson, "Low-complexity video coding for receiver-driven layered multicast," *IEEE Journal on Selected Areas in Communications*, vol. 15, pp. 983–1001, August 1997.

[31] H. Imaizumi and A. Luthra, *Three-Dimensional Television, Video and Display Technologies*, ch. MPEG-2 Multiview Profile, pp. 169–181. Berlin, Heidelberg, and New York: Springer Verlag, 2002.

[32] 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, pp. 626–642, April 2011.

[33] L. Zhang, G. Tech, K. Wegner, and S. Yea, "Test model 6 of 3D-HEVC and MV-HEVC," vol. N13940, ISO/IEC JTC-1/SC29/WG11, November 2013.

[34] S. Wenger, "H.264/AVC over IP," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, pp. 645–656, July 2003.

[35] J. Boyce, W. Jang, D. Hong, S. Wenger, Y.-K. Wang, and Y. Chen, "High level syntax hooks for future extensions," in *JCT-VC document*, vol. JCTVC-H0388, (San JosÃ©, CA, USA), February 2012.

[36] A. Segall, "BoG report on SHVC," in *JCT-VC document*, vol. JCTVC-K0354, (Shanghai, China), October 2012.

[37] P. Seeling and M. Reisslein, "Video transport evaluation with h.264 video traces," *IEEE Communications Surveys & Tutorials*, vol. 14, no. 4, pp. 1142–1165, 2012.

[38] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423 and 623–656, June and October 1948.

[39] R. Bauer and J. Hagenauer, "Iterative source/channel-decoding using reversible variable length codes," in *Data Compression Conference, 2000. Proceedings. DCC 2000*, pp. 93–102.

[40] N. Görtz, "On the iterative approximation of optimal joint source-channel decoding," *IEEE Journal on Selected Areas in Communications*, vol. 19, pp. 1662–1670, September 2001.

[41] T. Fingscheidt and P. Vary, "Softbit speech decoding: A new approach to error concealment," *IEEE Transaction on Speech and Audio Processing*, vol. 9, pp. 240–251, March 2001.

[42] J. Kliewer, N. Görtz, and A. Mertins, "On iterative source-channel image decoding with Markov random field source models," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, (Montreal, Canada), pp. iv–661–iv–664, May 2004.

[43] J. Kliewer, N. Görtz, and A. Mertins, "Iterative source-channel decoding with Markov random field source models," *IEEE Transactions on Signal Processing*, vol. 54, pp. 3688–3701, October 2006.

[44] M. Adrat and P. Vary, "Iterative source-channel decoding: Improved system design using EXIT charts," *EURASIP Journal on Applied Signal Processing*, vol. 2005, pp. 928–941, May 2005.

[45] J. Kliewer and R. Thobaben, "Iterative joint source-channel decoding of variable-length codes using residual source redundancy," *IEEE Transactions on Wireless Communications*, vol. 4, pp. 919–929, May 2005.

[46] L. Hanzo, R. G. Maunder, J. Wang, and L.-L. Yang, *Near-Capacity Variable-Length Coding: Regular and Exit-Chart Aided Irregular Designs.* John Wiley & Sons Ltd, 2010.

[47] R. Thobaben and J. Kliewer, "An efficient variable-length code construction for iterative source-channel decoding," *IEEE Transactions on Communications*, vol. 57, pp. 2005–2013, Jul 2009.

[48] Nasruminallah and L. Hanzo, "EXIT-chart optimized short block codes for iterative joint source and channel decoding in H.264 video telephony," *IEEE Transactions on Vehicular Technology*, vol. 58, pp. 4306–4315, October 2009.

[49] Nasruminallah and L. Hanzo, "Short block codes for guaranteed convergence in soft-bit assisted iterative joint source and channel decoding," *Electronics Letters*, vol. 44, pp. 1315–1316, October 2008.

[50] L. Schmalen, M. Adrat, T. Clevorn, and P. Vary, "Exit chart based system design for iterative source-channel decoding with fixed-length codes," *IEEE Transactions on Communications*, vol. 59, pp. 2406–2413, Sep 2011.

[51] L. Schmalen and P. Vary, "Iterative source-channel decoding with reduced error floors," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, pp. 1577–1587, Dec 2011.

[52] Y. Huo, C. Zhu, and L. Hanzo, "Spatio-temporal iterative source-channel decoding aided video transmission," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 4, pp. 1597–1609, 2013.

[53] C.-F. Wu and W.-H. Chung, "Iterative symbol decoding of distributed channel encoded fixed-length codes for cooperative communications," *IEEE Transactions on Wireless Communications*, vol. 13, pp. 1528–1539, Mar 2014.

[54] K. Sayood and J. Borkenhagen, "Use of residual redundancy in the design of joint source/channel coders," *IEEE Transactions on Communications*, vol. 39, pp. 838–846, June 1991.

[55] R. Bauer and J. Hagenauer, "On variable length codes for iterative source/channel decoding," in *Data Compression Conference, 2001. Proceedings. DCC 2001.*, pp. 273–282.

[56] N. Görtz, "Joint source channel decoding using bit-reliability information and source statistics," in *International Symposium on Information Theory*, (Cambridge, MA, USA), p. 9, August 1998.

[57] S. Geman and D. Geman, "Stochastic relaxation, gibbs distributions, and the bayesian restoration of images," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. PAMI-6, pp. 721–741, Nov 1984.

[58] B. Masnick and J. Wolf, "On linear unequal error protection codes," *IEEE Transactions on Information Theory*, vol. 13, pp. 600–607, October 1967.

[59] J. Hagenauer, "Rate-compatible puncture convolutional codes (RCPC) and their application," *IEEE Transactions on Communications*, vol. 36, pp. 389–400, April 1988.

[60] A. Calderbank and N. Seshadri, "Multilevel codes for unequal error protection," *IEEE Transactions on Information Theory*, vol. 39, pp. 1234–1248, Jul 1993.

[61] A. Albanese, J. Blomer, J. Edmonds, M. Luby, and M. Sudan, "Priority encoding transmission," *IEEE Transactions on Information Theory*, vol. 42, pp. 1737–1744, November 1996.

[62] D. Wu, Y. T. Hou, and Y.-Q. Zhang, "Transporting real-time video over the Internet: challenges and approaches," *Proceedings of the IEEE*, vol. 88, pp. 1855–1877, December 2000.

[63] P. Chou, A. Mohr, A. Wang, and S. Mehrotra, "Error control for receiver-driven layered multicast of audio and video," *IEEE Transactions on Multimedia*, vol. 3, pp. 108–122, March 2001.

[64] T. Stockhammer, M. Hannuksela, and T. Wiegand, "H.264/AVC in wireless environments," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, pp. 657–673, July 2003.

[65] Q. Zhang, W. Zhu, and Y.-Q. Zhang, "Channel-adaptive resource allocation for scalable video transmission over 3G wireless network," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, pp. 1049–1063, August 2004.

[66] T. Brüggen and P. Vary, "Unequal error protection by modulation with unequal power allocation," *IEEE Communications Letters*, vol. 9, pp. 484–486, June 2005.

[67] B. Barmada, M. Ghandi, E. Jones, and M. Ghanbari, "Prioritized transmission of data partitioned H.264 video with hierarchical QAM," *IEEE Signal Processing Letters*, vol. 12, pp. 577–580, Aug 2005.

[68] Y. C. Chang, S. W. Lee, and R. Komiya, "A low-complexity unequal error protection of H.264/AVC video using adaptive hierarchical QAM," *IEEE Transactions on Consumer Electronics*, vol. 52, pp. 1153–1158, Nov 2006.

[69] N. Rahnavard, H. Pishro-Nik, and F. Fekri, “Unequal error protection using partially regular LDPC codes,” *IEEE Transactions on Communications*, vol. 55, pp. 387–391, March 2007.

[70] J. Park, H. Lee, S. Lee, and A. Bovik, “Optimal channel adaptation of scalable video over a multicarrier-based multicell environment,” *IEEE Transactions on Multimedia*, vol. 11, pp. 1062–1071, October 2009.

[71] P. Li, Y. Chang, N. Feng, and F. Yang, “A novel hierarchical QAM-based unequal error protection scheme for H.264/AVC video over frequency-selective fading channels,” *IEEE Transactions on Consumer Electronics*, vol. 56, pp. 2741–2746, November 2010.

[72] S. S. Arslan, P. C. Cosman, and L. B. Milstein, “Coded hierarchical modulation for wireless progressive image transmission,” *IEEE Transactions on Vehicular Technology*, pp. 4299 – 4313, Nov. 2011.

[73] C. Hellge, D. Gomez-Barquero, T. Schierl, and T. Wiegand, “Layer-aware forward error correction for mobile broadcast of layered media,” *IEEE Transactions on Multimedia*, vol. 13, pp. 551–562, June 2011.

[74] A. A. Khalek, C. Caramanis, and R. W. Heath, “A cross-layer design for perceptual optimization of H.264/SVC with unequal error protection,” *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 7, pp. 1157–1171, 2012.

[75] S.-H. Chang, M. Rim, P. Cosman, and L. Milstein, “Optimized unequal error protection using multiplexed hierarchical modulation,” *IEEE Transactions on Information Theory*, vol. 58, pp. 5816–5840, Sept 2012.

[76] K. M. Alajel, W. Xiang, and Y. Wang, “Unequal error protection scheme based hierarchical 16-QAM for 3-D video transmission,” *IEEE Transactions on Consumer Electronics*, vol. 58, no. 3, pp. 731–738, 2012.

[77] Y. Wu, S. Kumar, F. Hu, Y. Zhu, and J. Matyjas, “Cross-layer forward error correction scheme using raptor and RCPC codes for prioritized video transmission over wireless channels,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, pp. 1047–1060, June 2014.

[78] T. Nguyen, P. Cosman, and L. Milstein, “Double-layer video transmission over decode-and-forward wireless relay networks using hierarchical modulation,” *IEEE Transactions on Image Processing*, vol. 23, pp. 1791–1804, April 2014.

[79] B. Masnick and J. Wolf, “On linear unequal error protection codes,” *IEEE Transactions on Information Theory*, vol. 13, pp. 600–607, October 1967.

[80] V. Pavlushkov, R. Johannesson, and V. Zyablov, “Unequal error protection for convolutional codes,” *IEEE Transactions on Information Theory*, vol. 52, pp. 700–708, February 2006.

[81] N. Rahnavard and F. Fekri, “New results on unequal error protection using LDPC codes,” *IEEE Communications Letters*, vol. 10, pp. 43–45, January 2006.

[82] V. Kumar and O. Milenkovic, “On unequal error protection LDPC codes based on Plotkin-type constructions,” *IEEE Transactions on Communications*, vol. 54, pp. 994–1005, June 2006.

[83] C. Gong, G. Yue, and X. Wang, “Message-wise unequal error protection based on low-density parity-check codes,” *IEEE Transactions on Communications*, vol. 59, pp. 1019–1030, April 2011.

[84] H. Wang, F. Zhai, Y. Eisenberg, and A. Katsaggelos, “Cost-distortion optimized unequal error protection for object-based video communications,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, pp. 1505–1516, December 2005.

[85] H. Ha and C. Yim, “Layer-weighted unequal error protection for scalable video coding extension of H.264/AVC,” *IEEE Transactions on Consumer Electronics*, vol. 54, pp. 736–744, May 2008.

[86] D. Sejdinović, D. Vukobratović, A. Doufexi, V. Šenk, and R. Piechocki, “Expanding window fountain codes for unequal error protection,” *IEEE Transactions on Communications*, vol. 57, pp. 2510–2516, November 2009.

[87] D. Vukobratović, V. Stanković, D. Sejdinović, L. Stanković, and Z. Xiong, “Scalable video multicast using expanding window fountain codes,” *IEEE Transactions on Multimedia*, vol. 11, no. 6, pp. 1094–1104, 2009.

[88] E. Maani and A. Katsaggelos, “Unequal error protection for robust streaming of scalable video over packet lossy networks,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, pp. 407–416, March 2010.

[89] S. Ahmad, R. Hamzaoui, and M. M. Al-Akaidi, “Unequal error protection using fountain codes with applications to video communication,” *IEEE Transactions on Multimedia*, vol. 13, pp. 92–101, February 2011.

[90] K. Nguyen, T. Nguyen, and S.-C. Cheung, “Video streaming with network coding,” *Journal of Signal Processing Systems*, vol. 59, pp. 319–333, June 2010.

[91] M. Halloush and H. Radha, “Network coding with multi-generation mixing: A generalized framework for practical network coding,” *IEEE Transactions on Wireless Communications*, vol. 10, no. 2, pp. 466–473, 2011.

[92] F. Marx and J. Farah, “A novel approach to achieve unequal error protection for video transmission over 3G wireless networks,” *Signal Processing: Image Communication*, vol. 19, no. 4, pp. 313–323, 2004.

[93] S. X. Ng, J. Y. Chung, and L. Hanzo, “Turbo-detected unequal protection MPEG-4 wireless video telephony using multi-level coding, trellis coded modulation and space-time trellis coding,” *IEE Proceedings Communications*, vol. 152, pp. 1116–1124, December 2005.

[94] M. Aydinlik and M. Salehi, “Turbo coded modulation for unequal error protection,” *IEEE Transactions on Communications*, vol. 56, pp. 555–564, April 2008.

[95] Y. C. Chang, S. W. Lee, and R. Komiya, "A fast forward error correction allocation algorithm for unequal error protection of video transmission over wireless channels," *IEEE Transactions on Consumer Electronics*, vol. 54, pp. 1066–1073, August 2008.

[96] Y. C. Chang, S. W. Lee, and R. Komiya, "A low complexity hierarchical QAM symbol bits allocation algorithm for unequal error protection of wireless video transmission," *IEEE Transactions on Consumer Electronics*, vol. 55, pp. 1089–1097, August 2009.

[97] Nasruminallah and L. Hanzo, "Near-capacity H.264 multimedia communications using iterative joint source-channel decoding," *IEEE Communications Surveys and Tutorials*, vol. 14, pp. 538–564, Second Quarter 2012.

[98] Y. Huo, M. El-Hajjar, and L. Hanzo, "Inter-layer FEC aided unequal error protection for multilayer video transmission in mobile TV," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, pp. 1622–1634, September 2013.

[99] Y. Huo, M. El-Hajjar, R. Maunder, and L. Hanzo, "Layered wireless video relying on minimum-distortion inter-layer FEC coding," *IEEE Transactions on Multimedia*, vol. 16, pp. 697–710, April 2014.

[100] L.-F. Wei, "Coded modulation with unequal error protection," *IEEE Transactions on Communications*, vol. 41, pp. 1439–1449, Oct 1993.

[101] S. Gadkari and K. Rose, "Time-division versus superposition coded modulation schemes for unequal error protection," *IEEE Transactions on Communications*, vol. 47, pp. 370–379, Mar 1999.

[102] M. Isaka, M. Fossorier, R. Morelos-Zaragoza, S. Lin, and H. Imai, "Multilevel coded modulation for unequal error protection and multistage decoding. II. asymmetric constellations," *IEEE Transactions on Communications*, vol. 48, pp. 774–786, May 2000.

[103] R. Morelos-Zaragoza, M. Fossorier, S. Lin, and H. Imai, "Multilevel coded modulation for unequal error protection and multistage decoding .i. symmetric constellations," *IEEE Transactions on Communications*, vol. 48, pp. 204–213, Feb 2000.

[104] J. Kim and G. Pottie, "Unequal error protection tcm codes," *IEE Proceedings-Communications*, vol. 148, pp. 265–272, Oct 2001.

[105] M. Zamkotsian, K. Peppas, F. Lazarakis, and P. Cottis, "Multilevel spatial hierarchical modulation: An efficient scheme for unequal error protection under rician fading," *IEEE Transactions on Vehicular Technology*, vol. PP, no. 99, pp. 1–1, 2014.

[106] R. Chang, S.-J. Lin, and W.-H. Chung, "Hierarchical space shift keying for unequal error protection," *IEEE Communications Letters*, vol. 16, pp. 1341–1344, September 2012.

[107] T. Quazi and H. Xu, "Simple unequal error protection mechanism for multimedia traffic using the Alamouti structure with hierarchical modulation and signal space diversity," *IET Communications*, vol. 8, no. 17, pp. 3128–3135, 2014.

[108] H. Nguyen, H. Nguyen, and T. Le-Ngoc, “Signal transmission with unequal error protection in wireless relay networks,” *IEEE Transactions on Vehicular Technology*, vol. 59, pp. 2166–2178, June 2010.

[109] C. Hellge, S. Mirta, T. Schierl, and T. Wiegand, “Mobile TV with SVC and hierarchical modulation for DVB-H broadcast services,” *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, 2009. BMSB ’09.*, pp. 1–5, May. 2009.

[110] ETSI, *Digital Video Broadcasting (DVB); Framing structure, channel coding and modulation for digital terrestrial television*, August 1997. EN 300 744 V1.1.2.

[111] J. Hossain, P. K. Vitthaladevuni, M. S. Alouini, and V. K. Bhargava, “Adaptive hierarchical modulation for simultaneous voice and multiclass data transmission over fading channels,” *IEEE Transactions on Vehicular Technology*, vol. 55, pp. 1181–1194, Jul. 2006.

[112] H. Sun, S. X. Ng, C. Dong, and L. Hanzo, “Decode-and-forward cooperation-aided triple-layer turbo-trellis-coded hierarchical modulation,” *IEEE Transactions on Communications*, vol. 63, pp. 1136–1148, April 2015.

[113] M. K. Chang and S. Y. Lee, “Performance analysis of cooperative communication system with hierarchical modulation over Rayleigh fading channel,” *IEEE Transactions on Wireless Communications*, vol. 8, pp. 2848–2852, Jun. 2009.

[114] T. Stockhammer, H. Jenkac, and C. Weiss, “Feedback and error protection strategies for wireless progressive video transmission,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, pp. 465–482, June 2002.

[115] D. Taubman and J. Thie, “Optimal erasure protection for scalably compressed video streams with limited retransmission,” *IEEE Transactions on Image Processing*, vol. 14, pp. 1006–1019, August 2005.

[116] R. Xiong, D. Taubman, and V. Sivaraman, “Optimal PET protection for streaming scalably compressed video streams with limited retransmission based on incomplete feedback,” *IEEE Transactions on Image Processing*, vol. 19, no. 9, pp. 2382–2395, 2010.

[117] R. Xiong, D. S. Taubman, and V. Sivaraman, “PET protection optimization for streaming scalable videos with multiple transmissions,” *IEEE Transactions on Image Processing*, vol. 22, pp. 4364–4379, November 2013.

[118] J. Zhang, W. Liang, J. Wu, and D. Shi, “A novel retransmission scheme for video services in hybrid wireline/wireless networks,” in *IEEE Vehicular Technology Conference (VTC-Spring), 71st*, pp. 1–5, May 2010.

[119] B. Tirouvengadam, R. Radhakrishnan, and A. Nayak, “CAAHR: Content aware adaptive HARQ retransmission scheme for 4G/LTE network,” in *Fourth International Conference on Ubiquitous and Future Networks (ICUFN)*, pp. 456–461, July 2012.

[120] A. Le Duc, P. Ciblat, and C. Le Martret, “Analysis of a cross-layer hybrid-ARQ scheme: Application to unequal packet protection,” in *IEEE International Conference on Communications (ICC)*, pp. 1–5, June 2011.

[121] C. Hellge, S. Mirta, T. Schierl, and T. Wiegand, “Mobile TV with SVC and hierarchical modulation for DVB-H broadcast services,” in *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, BMSB*, (Bilbao, Spain), pp. 1–5, May 2009.

[122] Y. Huo, X. Zuo, R. G. Maunder, and L. Hanzo, “Inter-layer FEC decoded multi-layer video streaming,” in *IEEE Global Telecommunications Conference (GLOBECOM)*, (Anaheim, CA, USA), pp. 2113–2118, December 2012.

[123] N. Miladinovic and M. Fossorier, “Systematic recursive construction of LDPC codes,” *IEEE Communications Letters*, vol. 8, pp. 302–304, May 2004.

[124] C. Zhu, Y. Huo, B. Zhang, R. Zhang, M. El-Hajjar, and L. Hanzo, “Adaptive truncated HARQ aided layered video streaming relying on inter-layer FEC coding,” *IEEE Transactions on Vehicular Technology*, vol. PP, no. 99, pp. 1–1, 2015.

[125] J. Hagenauer, E. Offer, and L. Papke, “Iterative decoding of binary block and convolutional codes,” *IEEE Transactions on Information Theory*, vol. 42, pp. 429–445, March 1996.

[126] J. Chen, A. Dholakia, E. Eleftheriou, M. Fossorier, and X.-Y. Hu, “Reduced-complexity decoding of LDPC codes,” *IEEE Transactions on Communications*, vol. 53, pp. 1288–1299, August 2005.

[127] IEEE Computer Society, *IEEE Standard for Local and Metropolitan Area Networks - Part 15.7: Short-Range Wireless Optical Communication Using Visible Light.* 2011.

[128] G. Pang, T. Kwan, C.-H. Chan, and H. Liu, “LED traffic light as a communications device,” in *1999 IEEE/IEEJ/JSAI International Conference on Intelligent Transportation Systems, 1999. Proceedings*, pp. 788–793.

[129] Y. Tanaka, T. Komine, S. Haruyama, and M. Nakagawa, “Wireless optical transmissions with white colored LED for wireless home links,” in *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, pp. 1325–1329, 2000.

[130] T. Komine and M. Nakagawa, “Fundamental analysis for visible-light communication system using LED lights,” *IEEE Transactions on Consumer Electronics*, vol. 50, pp. 100–107, February 2004.

[131] T. Komine and M. Nakagawa, “Integrated system of white LED visible-light communication and power-line communication,” vol. 49, no. 1, pp. 71–79.

[132] “Visible Light Communications Consotium (VLCC).” http://www.vlcc.net/. (Visited on 12/02/2015).

[133] J. Grubor, S. Randel, K.-D. Langer, and J. W. Walewski, “Broadband information broadcasting using LED-based interior light,” *IEEE Journal of Lightwave Technology*, vol. 26, pp. 3883–3892, February 2009.

[134] “Home gigabit access project,” in *[Online]. Available: http://www.ict-omega.eu/*, 2012.

[135] OMEGAresults, “OMEGA optical wireless prototypes.”

[136] Bourns College of Engineering, “Center for ubiquitous communication by light,” in *[Online]. Available: http://www.uclight.ucr.edu/*, 2012.

[137] The Pennsylvania State University, “Center on optical wireless applications,” in *[Online]. Available: http://cowa.psu.edu/*, 2012.

[138] Smart Lighting Engineering Research Center, “Synthesizing light for the benefit of humanity,” in *[Online]. Available: http://www.bu.edu/smartlighting*, 2008.

[139] “JEITA / JEITA Standards / AV&IT Technology Standardization / Visible Light Communications.” http://www.jeita.or.jp/. (Visited on 12/02/2015).

[140] “St. Cloud first to sign on for new technology (Press release).” http://www.tmcnet.com/usubmit/2010/11/19/5148608.htm. (Visited on 12/02/2015).

[141] Y. Tanaka, S. Haruyama, and M. Nakagawa, “Wireless optical transmissions with white colored LED for wireless home links,” in *The 11th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, 2000. PIMRC 2000*, vol. 2, pp. 1325–1329 vol.2.

[142] “Li-Fi Consortium.” http://www.lificonsortium.org/. (Visited on 12/02/2015).

[143] “VLCA Visible Light Communications Association.” http://vlca.net/. (Visited on 01/08/2016).

[144] L. Schmitt, *The Bright Stuff: The LED And Nick Holonyak’s Fantastic Trail Of Innovation.* BookBaby, 2012.

[145] S. Nakamura, T. Mukai, and M. Senoh, “Candela-class high-brightness InGaN/AlGaN double-heterostructure blue-light-emitting diodes,” *Applied Physics Letters*, vol. 64, pp. 1687–1689, Mar. 1994.

[146] T. Baumgartner, F. Wunderlich, A. Jaunich, T. Sato, G. Bundy, N. Grießmann, J. Kowalski, S. Burghardt, and J. Hanebrink, “Lighting the way: Perspectives on the global lighting market,” 2012.

[147] C. Inc., “Cree sets new R & D performance record with 254 lumen-per-watt power LED,” in *[Online]. Available: http://www.cree.com/news-and-events/cree-news/press-releases/2012/april/120412-254-lumen-per-watt*, 2012.

[148] "Cree news: Cree first to break 300 lumens-per-watt barrier." http://www.cree.com/News-and-Events/Cree-News/Press-Releases/2014/March/300LPW-LED-barrier. (Visited on 12/03/2015).

[149] "Solid-state lighting: Comparing leds to traditional light sources." http://web.archive.org/web/20090505080533/http://www1.eere.energy.gov/buildings/ssl/comparing.html. (Visited on 12/03/2015).

[150] H. Hecht, "Changing the lights: Are LEDs ready to become the market standard?," *Optics and Photonics News*, vol. 23, no. 3, pp. 44–50, 2012.

[151] N. Kumar and N. Lourenco, "LED based visible light communication: A brief survey and investigation," *Journal of Engineering and Applied Sciences*, vol. 5, pp. 269–307, 2010.

[152] H.-J. Chiu, Y.-K. Lo, J.-T. Chen, S.-J. Cheng, C.-Y. Lin, and S.-C. Mou, "A high-efficiency dimmable led driver for low-power lighting applications," *IEEE Transactions on Industrial Electronics*, vol. 57, pp. 735–743, Feb 2010.

[153] D. Steigerwald, J. Bhat, D. Collins, R. Fletcher, M. Holcomb, M. Ludowise, P. Martin, and S. Rudaz, "Illumination with solid state lighting technology," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 8, pp. 310–320, August 2002.

[154] Multi-Chip LED, "OSRAM opto semiconductors," in *[Online]. Available: http://www.osram-os.com/osram-os/en/product-catalog/led-light-emitting-diodes/multi-chipled/index.jsp*, 2013.

[155] Konica Minolta, "OLED lighting," in *[Online]. Available: http://www.konicaminolta.com/oled/products/index.htm*, 2014.

[156] A. Kelly, J. McKendry, S. Zhang, D. Massoubre, B. Rae, R. Green, R. Henderson, and M. Dawson, "High-speed GaN micro-LED arrays for data communication," in *14th International Conference on Transparent Optical Networks (ICTON)*, (Coventry, UK), pp. 1–5, July 2012.

[157] M. H. Crawford, "LEDs for solid-state lighting: Performance challenges and recent advances," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 15, pp. 1028–1040, July-August 2009.

[158] H. Elgala, R. Mesleh, and H. Haas, "Indoor broadcasting via white LEDs and OFDM," *IEEE Transactions on Consumer Electronics*, pp. 1127–1134, August 2009.

[159] R. Mesleh, H. Elgala, and H. Haas, "LED nonlinearity mitigation techniques in optical wireless OFDM communications systems," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 4, pp. 865–875, November 2012.

[160] R. Zhang and H. S.-H. Chung, "A TRIAC-dimmable LED lamp driver with wide dimming range," *IEEE Transactions on Power Electronics*, vol. 29, pp. 1434–1446, September 2014.

[161] T.-J. Liang, W.-J. Tseng, J.-F. Chen, and J.-P. Wu, “A novel line frequency multi-stage conduction LED driver with high power factor,” *IEEE Transactions on Power Electronics*, vol. 30, pp. 5103–5115, April 2015.

[162] H. Elgala, R. Mesleh, and H. Haas, “A study of led nonlinearity effects on optical wireless transmission using ofdm,” in *IFIP International Conference on Wireless and Optical Communications Networks*, (Cairo), pp. 1–5, 28-30 April 2009.

[163] H. Elgala, R. Mesleh, and H. Haas, “An led model for intensity-modulated optical communication systems,” *IEEE Photonics Technology Letters*, vol. 22, pp. 835–837, June 2010.

[164] C. Chow, C. Yeh, Y. Liu, and P. Huang, “Mitigation of optical background noise in light-emitting diode (led) optical wireless communication systems,” *Photonics Journal, IEEE*, vol. 5, pp. 7900307–7900307, Feb 2013.

[165] D. Tsonev, H. Chun, S. Rajbhandari, J. McKendry, S. Videv, E. Gu, M. Haji, S. Watson, A. Kelly, G. Faulkner, M. Dawson, H. Haas, and D. O’Brien, “A 3-gb/s single-led ofdm-based wireless vlc link using a gallium nitride μ-led,” *IEEE Photonics Technology Letters*, vol. 26, pp. 637–640, April 2014.

[166] S.-W. Wang, F. Chen, L. Liang, S. He, Y. Wang, X. Chen, and W. Lu, “A high-performance blue filter for a white-led-based visible light communication system,” *Wireless Communications, IEEE*, vol. 22, pp. 61–67, April 2015.

[167] K. Modpalli and L. Parsa, “Dual-purpose offline led driver for illumination and visible light communication,” *IEEE Transactions on Industry Applications*, vol. 51, pp. 406–419, Jan 2015.

[168] N. Fujimoto and H. Mochizuki, “477 Mbit/s visible light transmission based on OOK-NRZ modulation using a single commercially available visible LED and a practical LED driver with a pre-emphasis circuit,” in *Optical Fiber Communication Conference and Exposition and the National Fiber Optic Engineers Conference (OFC/NFOEC), 2013*, pp. 1–3, 2013.

[169] H. Li, X. Chen, B. Huang, D. Tang, and H. Chen, “High bandwidth visible light communications based on a post-equalization circuit,” *IEEE Photonics Technology Letters*, vol. 26, no. 2, pp. 119–122, 2014.

[170] J. G. Proakis and M. Salehi, *Digital Communications, 5th Edition.* McGraw-Hill, 2007.

[171] T. H. Chan and S. Hranilovic, “Capacity-achieving probability measure for conditional Gaussian channels with bounded inputs,” *IEEE Transactions on Information Theory*, vol. 51, pp. 2073–2088, June 2005.

[172] S. H. Lee, K.-I. Ahn, and J. K. Kwon, “Multilevel transmission in dimmable visible light communication systems,” *Journal of Lightwave Technology*, vol. 31, no. 20, pp. 3267–3276, 2013.

[173] S. G. Wilson, M. Brandt-Pearce, Q. L. Gao, and M. Baedke, "Optical repetition MIMO transmission with multipulse PPM," *IEEE Journal on Selected Areas in Communications*, vol. 23, pp. 1901–1910, September 2005.

[174] J. Rufo, C. Quintana, F. Delgado, J. Rabadan, and R. Perez-Jimenez, "Considerations on modulations and protocols suitable for visible light communications (VLC) channels: Low and medium baud rate indoor visible ligth communications links," in *Consumer Communications and Networking Conference (CCNC), 2011 IEEE*, pp. 362–364, 2011.

[175] S. He, G. Ren, Z. Zhong, and Y. Zhao, "M-ary variable period modulation for indoor visible light communication system," *IEEE Communications Letters*, vol. 17, no. 7, pp. 1325–1328, 2013.

[176] S. Arnon, "The effect of clock jitter in visible light communication applications," *Journal of Lightwave Technology*, vol. 30, no. 21, pp. 3434–3439, 2012.

[177] M. H. A. Davis, "Capacity and error exponent for the direct detection photon channel - part II," *IEEE Transactions on Information Theory*, vol. 40, pp. 1462–1471, November 1988.

[178] R. M. Gagliardi and S. Karp, *Optical Communication, 2nd Ed.* Wiley, 1995.

[179] W. Gappmair and M. Flohberger, "Error performance of coded FSO links in turbulence atmospheric modeled by gamma-gamma distributions," *IEEE Transactions on Wireless Communications*, vol. 8, pp. 2209–2213, May 2009.

[180] T. T. Nguyen and L. Lampe, "Coded multipulse pulse-position modulation for free-space optical communications," *IEEE Transactions on Communications*, vol. 58, no. 4, pp. 1036–1041, 2010.

[181] H. Sugiyama and K. Nosu, "MPPM: A method for improving the band-utilization efficiency in optical PPM," *IEEE Journal of Lightwave Technology*, vol. 7, pp. 1969–1979, March 1989.

[182] J. Hamkins and B. Moision, "Multipulse pulse-position modulation on discrete memoryless channels," *IPN Progress Report*, vol. 42, pp. 1–13, May 2005.

[183] M. K. Simon and V. A. Vilnrotter, "Performance analysis and trade-offs for dual-pulse ppm on optical communications," *IEEE Transactions on Communications*, vol. 52, pp. 1969–1979, November 2004.

[184] H. M. H. Shalaby, "Performance of uncoded overlapping PPM under communication constraints," in *IEEE International Communication Conference Proceeding*, (Geneva, Switzerland), pp. 512–516, May 1993.

[185] D. Shiu and J. M. Kahn, "Differential pulse-position modulation for power-efficient optical communication," *IEEE Transactions of Communications*, vol. 47, pp. 1201–1210, August 1999.

[186] D. Shiu and J. M. Kahn, "Shaping and nonequiprobable signalling for intensity-modulated signals," *IEEE Transactions of Information Theory*, vol. 45, pp. 2661–2668, November 1999.

[187] Z. Ghassemlooy, A. R. Hayes, N. L. Seed, and E. D. Kaluarachchi, "Digital pulse interval modulation for optical communications," *IEEE Communications Magazine*, vol. 36, pp. 95–99, December 1998.

[188] T. Ohtsuki, "Multiple-subcarrier modulation in optical wireless communications," *IEEE Communications Magazine*, vol. 41, pp. 74–79, March 2003.

[189] Y. Fan and R. H. Green, "Comparison of pulse position modulation and pulse width modulation for application in optical communications," *Optics Engineering*, vol. 46, p. 065001, June 2007.

[190] M. Biagi, A. M. Vegni, S. Pergoloni, P. M. Butala, and T. D. C. Little, "Trace-orthogonal PPM-space time block coding under rate constraints for visible light communication," *IEEE Lightwave Technology*, vol. 33, pp. 481–494, January 2015.

[191] R. J. Drost and B. M. Sadler, "Constellation design for color-shift keying using Billiards algorithms," in *IEEE Globecom Workshop on Optical Wireless Communications*, (Miami, FL), pp. 980–984, December 2010.

[192] E. Monterio and S. Hranilovic, "Constellation design for color-shift keying using Interior point methods," in *IEEE Globecom Workshop on Optical Wireless Communications*, (Anaheim, CA), pp. 1224–1228, December 2012.

[193] B. Bai, Q. He, Z. Xu, and Y. Fan, "The color shift key modulation with non-uniform signaling for visible light communication," in *2012 1st IEEE International Conference on Communications in China Workshops (ICCC)*, pp. 37–42, Aug 2012.

[194] E. Monteiro and S. Hranilovic, "Design and implementation of color-shift keying for visible light communications," *Journal of Lightwave Technology*, vol. 32, pp. 2053–2060, May 2014.

[195] R. Drost and B. Sadler, "Constellation design for channel precompensation in multi-wavelength visible light communications," *IEEE Transactions on Communications*, vol. 62, pp. 1995–2005, June 2014.

[196] R. Singh, T. O'Farrell, and J. David, "An enhanced color shift keying modulation scheme for high-speed wireless visible light communications," *Journal of Lightwave Technology*, vol. 32, pp. 2582–2592, July 2014.

[197] J. M. Luna-Rivera, R. Perez-Jimenez, J. Rabadan-Borjes, J. Rufo-Torres, V. Guerra, and C. Suarez-Rodriguez, "Multiuser CSK scheme for indoor visible light communications," *Optical Express*, vol. 22, pp. 24256–24267, October 2014.

[198] S. Pergoloni, M. Biagi, S. Rinauro, S. Colonnese, R. Cusani, and G. Scarano, "Merging color shift keying and complementary pulse position modulation for visible light

illumination and communication," *IEEE Journal of Lightwave Technology*, vol. 33, pp. 192–200, December 2014.

[199] J. Jiang, R. Zhang, and L. Hanzo, "Analysis and design of three-stage concatenated colour shift keying," *IEEE Transactions on Vehicular Technology*, vol. (Accept), 2015.

[200] S. X. Ng, J. Y. Chung, and L. Hanzo, "Turbo-detected unequal protection MPEG-4 wireless video telephony using multi-level coding, trellis coded modulation and space-time trellis coding," *IEE Proceedings - Communications*, vol. 152, pp. 1116–1124, December 2005.

[201] C. Zhu, Y. Huo, R. G. Maunder, S. Kawade, and L. Hanzo, "Iterative joint source-channel decoding aided transmission of losslessly compressed video," in *IEEE Wireless Communications and Networking Conference (WCNC)*, (Shanghai, China), pp. 4629–4634, April 2013.

[202] R. Drost and B. Sadler, "Constellation design for color-shift keying using billiards algorithms," in *2010 IEEE GLOBECOM Workshops (GC Wkshps)*, pp. 980–984, Dec 2010.

[203] J. Jiang, Y. Huo, F. Jin, P. Zhang, Z. Wang, Z. Xu, H. Haas, and L. Hanzo, "Video streaming in the multiuser indoor visible light downlink," *IEEE Access*, vol. 3, pp. 2959–2986, 2015.

[204] V. Balakirsky, "Joint source-channel coding with variable length codes," in *IEEE International Symposium on Information Theory*, p. 419, June 1997.

[205] Y. Takishima, M. Wada, and H. Murakami, "Reversible variable length codes," *IEEE Transactions on Communications*, vol. 43, pp. 158 –162, Feb/Mar/Apr 1995.

[206] C.-W. Lin, Y.-J. Chuang, and J.-L. Wu, "Generic construction algorithms for symmetric and asymmetric RVLCs," in *The 8th International Conference on Communication Systems, 2002. ICCS 2002.*, vol. 2, pp. 968 – 972 vol.2, November 2002.

[207] R. Maunder, J. Wang, S. Ng, L.-L. Yang, and L. Hanzo, "On the performance and complexity of irregular variable length codes for near-capacity joint source and channel coding," *IEEE Transactions on Wireless Communications*, vol. 7, pp. 1338 –1347, April 2008.

[208] S. Ahmed, R. Maunder, L.-L. Yang, and L. Hanzo, "Iterative detection of unity-rate precoded FFH-MFSK and irregular variable-length coding," *IEEE Transactions on Vehicular Technology*, vol. 58, pp. 3765 –3770, September 2009.

[209] Y. Wang and S. Yu, "Joint source-channel decoding for H.264 coded video stream," *IEEE Transactions on Consumer Electronics*, vol. 51, pp. 1273–1276, November 2005.

[210] Y.-J. Wu and H. Ogiwara, "Symbol-interleaver design for turbo trellis-coded modulation," *IEEE Communications Letters*, vol. 8, pp. 632–634, October 2004.

[211] D. Huffman, "A method for the construction of minimum-redundancy codes," *Proceedings of the IRE*, vol. 40, pp. 1098 –1101, September 1952.

[212] R. Bauer and J. Hagenauer, "Symbol-by-symbol MAP decoding of variable length codes," *in Proc. 3rd ITG Conf. Source Channel Coding*, pp. 111–116, January 2000.

[213] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "A soft-input soft-output APP module for iterative decoding of concatenated codes," *IEEE Communications Letters*, vol. 1, pp. 22 –24, January 1997.

[214] D. Chase, "A combined coding and modulation approach for communication over dispersive channels," *IEEE Transactions on Communications*, vol. 21, pp. 159–174, March 1973.

[215] I. Stanojev, O. Simeone, Y. Bar-Ness, and D. H. Kim, "Energy efficiency of non-collaborative and collaborative hybrid-ARQ protocols," *IEEE Transactions on Wireless Communications*, vol. 8, pp. 326–335, January 2009.

[216] J. Ramis and G. Femenias, "Cross-layer design of adaptive multirate wireless networks using truncated HARQ," *IEEE Transactions on Vehicular Technology*, vol. 60, pp. 944–954, March 2011.

[217] J. Ramis, G. Femenias, F. Riera-Palou, and L. Carrasco, "Cross-layer optimization of adaptive multi-rate wireless networks using truncated chase combining HARQ," in *IEEE Global Telecommunications Conference (GLOBECOM)*, pp. 1–6, December 2010.

[218] T. Kwon and D.-H. Cho, "Adaptive-modulation-and-coding-based transmission of control messages for resource allocation in mobile communication systems," *IEEE Transactions on Vehicular Technology*, vol. 58, pp. 2769–2782, July 2009.

[219] Q. Zhang and S. Kassam, "Hybrid ARQ with selective combining for fading channels," *IEEE Journal on Selected Areas in Communications*, vol. 17, pp. 867–880, May 1999.

[220] A. Majumda, D. Sachs, I. Kozintsev, K. Ramchandran, and M. Yeung, "Multicast and unicast real-time video streaming over wireless LANs," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, pp. 524–534, June 2002.

[221] Q. Zhang, Q. Guo, Q. Ni, W. Zhu, and Y.-Q. Zhang, "Sender-adaptive and receiver-driven layered multicast for scalable video over the internet," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, pp. 482–495, April 2005.

[222] Z. Liu, Z. Wu, P. Liu, H. Liu, and Y. Wang, "Layer bargaining: multicast layered video over wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 28, pp. 445–455, April 2010.

[223] M. Wu, S. Makharia, H. Liu, D. Li, and S. Mathur, "IPTV multicast over wireless LAN using merged hybrid ARQ with staggered adaptive FEC," *IEEE Transactions on Broadcasting*, vol. 55, pp. 363–374, June 2009.

[224] I. Bajic, "Efficient cross-layer error control for wireless video multicast," *IEEE Transactions on Broadcasting*, vol. 53, pp. 276–285, March 2007.

[225] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Transactions on Information Theory*, vol. 20, pp. 284–287, March 1974.

[226] "Ieee standard for local and metropolitan area networks part 16: Air interface for broadband wireless access systems," *IEEE Std 802.16-2009 (Revision of IEEE Std 802.16-2004)*, pp. 1–2080, May 2009.

[227] L. Cheng, B. Henty, D. Stancil, F. Bai, and P. Mudalige, "Mobile vehicle-to-vehicle narrow-band channel measurement and characterization of the 5.9 ghz dedicated short range communication (dsrc) frequency band," *IEEE Journal on Selected Areas in Communications*, vol. 25, pp. 1501–1516, October 2007.

[228] L. Hanzo, J. Blogh, and S. Ni, *3G, HSPA and FDD versus TDD networking : smart antennas and adaptive modulation.* Chichester: John Wiley & Sons [s.l.], 2008.

[229] L. L. Hanzo, Y. Akhtman, L. Wang, and M. Jiang, *MIMO-OFDM for LTE, WiFi and WiMAX: Coherent versus Non-coherent and Cooperative Turbo Transceivers.* John Wiley & Sons, IEEE press, October 2010.

[230] H. Gacanin, M. Salmela, and F. Adachi, "Performance analysis of analog network coding with imperfect channel estimation in a frequency-selective fading channel," *IEEE Transactions on Wireless Communications*, vol. 11, pp. 742–750, February 2012.

[231] H. Xiao, Q. Dai, X. Ji, and W. Zhu, "A novel JSCC framework with diversity-multiplexing-coding gain tradeoff for scalable video transmission over cooperative MIMO," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, pp. 994–1006, July 2010.

[232] A. Detti, G. Bianchi, C. Pisa, F. S. Proto, P. Loreti, W. Kellerer, S. Thakolsri, and J. Widmer, "SVEF: an open-source experimental evaluation framework for H.264 scalable video streaming," in *IEEE Symposium on Computers and Communications, ISCC*, (Sousse, Tunisia), pp. 36–41, July 2009.

[233] H. Burchardt, N. Serafimovski, D. Tsonev, S. Videv, and H. Haas, "VLC: Beyond point-to-point communication," *Communications Magazine, IEEE*, vol. 52, pp. 98–105, July 2014.

[234] X. Li, J. Cucic, V. Jungnickel, and J. Armstrong, "On the capacity of intensity-modulated direct-detection systems and the information rate of ACO-OFDM for indoor optical wireless applications," *IEEE Transactions on Communications*, vol. 60, pp. 799–809, March 2012.

[235] T. Komine and M. Nakagawa, "Performance evaluation of visible-light wireless communication system using white LED lighting," in *International Symposium on Computers and Communications*, pp. 258–263, 2004.

[236] K. Lee and H. Park, "Channel model and modulation schemes for visible light communications," in *IEEE International Midwest Symposium on Circuits and Systems*, pp. 1–3, 2011.

[237] J. Jiang, R. Zhang, and L. Hanzo, "Analysis and design of three-stage concatenated colour-shift keying," *IEEE Transactions on Vehicular Technology*, vol. PP, no. 99, pp. 1–1, 2015.

[238] L. Zhang, G. Tech, K. Wegner, and S. Yea, "Test model 6 of 3D-HEVC and MV-HEVC," vol. N13940, ISO/IEC JTC-1/SC29/WG11, November 2013.

[239] J. Boyce, W. Jang, D. Hong, S. Wenger, Y.-K. Wang, and Y. Chen, "High level syntax hooks for future extensions," in *JCT-VC document*, vol. JCTVC-H0388, (San JosÃ©, CA, USA), February 2012.

[240] A. Segall, "BoG report on SHVC," in *JCT-VC document*, vol. JCTVC-K0354, (Shanghai, China), October 2012.

[241] P. Cataldi, M. Grangetto, T. Tillo, E. Magli, and G. Olmo, "Sliding-window raptor codes for efficient scalable wireless video broadcasting with unequal loss protection," *IEEE Transactions on Image Processing*, vol. 19, pp. 1491–1503, June 2010.

[242] N. Thomos, J. Chakareski, and P. Frossard, "Prioritized distributed video delivery with randomized network coding," *IEEE Transactions on Multimedia*, vol. 13, pp. 776–787, Aug 2011.

[243] Nasruminallah, M. El-Hajjar, N. Othman, A. Quang, and L. Hanzo, "Over-complete mapping aided, soft-bit assisted iterative unequal error protection H.264 Joint source and channel decoding," in *Vehicular Technology Conference, 2008. VTC 2008-Fall. IEEE 68th*, pp. 1–5, 2008.

[244] S. Arslan, P. Cosman, and L. Milstein, "Coded hierarchical modulation for wireless progressive image transmission," *IEEE Transactions on Vehicular Technology*, vol. 60, pp. 4299–4313, Nov 2011.

[245] Y. Noh, H. Lee, W. Lee, and I. Lee, "Design of unequal error protection for MIMO-OFDM systems with hierarchical signal constellations," *Journal of Communications and Networks*, vol. 9, pp. 167–176, June 2007.

[246] M. Morimoto, M. Okada, and S. Komaki, "Joint on-board resource sharing and hierarchical modulation scheme for satellite communication," in *Global Telecommunications Conference, 1995. GLOBECOM '95., IEEE*, vol. 3, pp. 1662–1666, 1995.

[247] M. Morimoto, M. Okada, and S. Komaki, "A hierarchical image transmission system for multimedia mobile communication," in *First International Workshop on Wireless Image/Video Communications*, pp. 80–84, 1996.

[248] M. K. Chang and S. Y. Lee, "Performance analysis of cooperative communication system with hierarchical modulation over Rayleigh fading channel," *IEEE Transactions on Wireless Communications*, vol. 8, pp. 2848–2852, Jun. 2009.

[249] H. Sun, C. Dong, S. Ng, and L. Hanzo, "Five decades of hierarchical modulation and its benefits in relay-aided networking," *IEEE Access*, vol. 3, pp. 2891–2921, 2015.

[250] L. Hanzo and P. Cherriman and J. Streit, "Wireless Video Communications: From Second to Third Generation Systems, WLANs and Beyond." IEEE Press, 2001. (For detailed contents please refer to http://www-mobile.ecs.soton.ac.uk.).

[251] D. Karunatilaka, F. Zafar, V. Kalavally, and R. Parthiban, "LED based indoor visible light communications: State of the art," *IEEE Communications Surveys and Tutorials*, vol. 17, pp. 1649–1678, Aug. 2015.

[252] Sony, "F55 CineAlta 4K the future, ahead of schedule." http://pro.sony.com/bbsccms/assets/files/show/highend/pdf/F55_Camera.pdf, 2012. Retrieved on 15/02/2016.

[253] S.-H. Chen and C.-W. Chow, "Color-shift keying and code-division multiple-access transmission for RGB-LED visible light communications using mobile phone camera," *IEEE Photonics Journal*, vol. 6, pp. 1–6, Dec 2014.

[254] Y. Hong, J. Chen, Z. Wang, and C. Yu, "Performance of a precoding MIMO system for decentralized multiuser indoor visible light communications," *IEEE Photonics Journal*, vol. 5, August 2013.

[255] M. F. Guerra-Medina, O. González, B. Rojas-Guillama, J. A. Martín-González, F. Delgado, and J. Rabadán, "Ethernet-OCDMA system for multi-user visible light communications," *Electronics Letters*, vol. 48, pp. 227–228, Feb. 2012.

[256] M. Jafar, D. C. O'Brien, C. J. Stevens, and D. J. Edwards, "Evaluation of coverage area for a wide line-of-sight indoor optical free-space communication system employing coherent detection," *IET Communications*, vol. 2, pp. 18–26, Jan. 2008.

[257] C. CIE, "Commission internationale de l'eclairage proceedings, 1931," 1932.

[258] P. H. Pathak, X. Feng, P. Hu, and P. Mohapatra, "Visible light communication, networking, and sensing: A survey, potential and challenges," *IEEE Communications Surveys Tutorials*, vol. 17, pp. 2047–2077, Fourthquarter 2015.

[259] H. Xiao, Q. Dai, X. Ji, and W. Zhu, "A novel JSCC framework with diversity-multiplexing-coding gain tradeoff for scalable video transmission over cooperative MIMO," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, pp. 994–1006, July 2010.

[260] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Transactions on Information Theory*, vol. 22, pp. 1–10, January 1976.

[261] X. Guo, Y. Lu, F. Wu, D. Zhao, and W. Gao, "Wyner-Ziv-based multiview video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, pp. 713–724, June 2008.

Author Index

Subject Index