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

Adaptive Unequal Error Protection for Scalable Video Streaming

by

Yanqing Zhang
B.Eng., M.Sc.

*A doctoral thesis submitted in fulfilment of
the requirement for the award of Doctor of Philosophy
at the Univesity of Southampton*

August 2020

SUPERVISORS:
Prof. Lajos Hanzo
Dip Ing, MSc, PhD, DSc, FIEEE, FREng, FIET, EIC of IEEE Press
Dr. Mohammed El-Hajjar

School 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

Adaptive Unequal Error Protection for Scalable Video Streaming

by Yanqing Zhang

To accommodate the growing demand for video streaming, the concept of Scalable Video Coding (SVC), also referred to as layered video coding, became the center of research efforts, which generates multiple inter-dependent video layers, including a Based Layer (BL) and multiple Enhancement Layers (ELs). Due to the fact that the ELs are encoded and decoded with reference to the BL, the decodability of the ELs is directly conditioned on that of the BL. This requirement benefits from Unequal Error Protection (UEP) techniques in terms of improving the attainable video quality by offering the strongest protection for the most important video bits.

In this treatise, we mainly focus our attention on the physical layer based UEP techniques, including physical layer Forward Error Correction (FEC) and modulation techniques. Specifically, we consider a pair of FEC techniques, namely Inter-Layer FEC (IL-FEC) and variable-rate FEC schemes. Additionally, physical layer based UEP can also be achieved with the aid of modulation techniques. For example, the Hierarchical Quadrature Amplitude Modulation (HQAM) scheme is capable of providing simple UEP by mapping the more important bits to the specific constellation points having a larger Euclidean distance. Similarly, typical UEP relying on Multiple-Input Multiple-Output (MIMO) schemes feeds the more important bits to the specific MIMO subchannel experiencing better channel quality.

Against the aforementioned background, in this thesis, we design physical layer based adaptive UEP for layered video streaming. Firstly, we exploit modulation based UEP by invoking optimal power-sharing assisted Superposition Coding (SC) schemes for video broadcasting services, where the bits of the various video layers are conveyed by different modulation modes relying on optimized power in order to maximize the overall video performance. Additionally, the UEP potential of the Index Modulation (IM) is investigated relying on a sophisticated Multi-Set Space-Time Shift Keying (MS-STSK), which provides a bit mapping assisted UEP scheme by feeding the source bits into the MS-STSK subchannels according to their importance. Moreover, the family of channel FEC based UEP techniques is also discussed. Specifically, we modify the conventional IL-FEC technique to provide UEP for the ELs, where an adaptive IL-FEC scheme is proposed for providing a gracefully decaying video performance erosion in the face of channel quality degradation. Finally, a variable FEC code-rate scheme is conceived for scalable panoramic video streaming.

Declaration of Authorship

I, **Yanqing Zhang**, declare that the thesis entitled **Adaptive Unequal Error Protection for Scalable Video Streaming** and the work presented in it are my own and has been generated by me as the result of my own original research. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University;
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- Where I have consulted the published work of others, this is always clearly attributed;
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- Parts of this work have been published.

Signed: .... Date: ....

Acknowledgements

I would like to express my heartfelt gratitude to Professor Lajos Hanzo and Dr. Mohammed El-Hajjar for their outstanding supervision and support throughout my research. Without their patient instruction, insightful criticism and expert guidance, my research would not have been done successfully, meanwhile he gives me that “Tomorrow will be able to be better”. Thank Prof. Hanzo for not only offering me valuable suggestions in the academic studies but also teaching me the attitude of “step by step”. In a word, their guidance, inspiration and encouragement have greatly benefited me not only in work but also in life. Most importantly, I would like to thank them for their invaluable friendship.

Many thanks also to my colleagues and the staff of the Next Generation Wireless Research Group, both past and present, for their support, help and discussions throughout my research. Special thanks to my colleagues, Professor Sheng Chen, Professor Lie-Liang Yang, Professor. Rob Maunder, Dr. Soon Xin Ng, Dr. Chao Xu, Dr. Jiankang Zhang, Dr. Ibrahim A. Hemadeh, Dr. Yongkai Huo, Dr. Luping Xiang, Dr. Xiaoyu Zhang, Dr. Siyao Lu, Yusha Liu and Shuai Shao for their kindly provided technical support and collaborative work. Thanks to my friends I have during my PhD years, too numerous to mention here explicitly.

I would also like to express my appreciation to my parents Ruijie Zhang and Qian Liu, for their unconditional love and support. Finally, I would like to greatly thank the financial support of China Scholarship Council.

List of Publications

Journals:

1. **Y. Zhang**, I. A. Hamadeh, M. El-Hajjar and L. Hanzo, “Multi-Set Space-Time Shift Keying Assisted Adaptive Inter-Layer FEC for Wireless Video Streaming”, IEEE Access, vol.7, pp.3592-3609, 2018.
2. **Y. Zhang**, C. Xu, I. A. Hemadeh, M. El-Hajjar and L. Hanzo, “Near-Instantaneously Adaptive Multi-Set Space-Time Shift Keying for UAV-Aided Video Surveillance”, IEEE Transactions on Vehicular Technology, doi: 10.1109/TVT.2020.3012208.
3. **Y. Zhang**, J. Zhang, Y. Huo, C. Xu, M. El-Hajjar and L. Hanzo, “Scalable Panoramic Wireless Video streaming Relying on Optimal-Rate FEC-Coded Adaptive QAM”, IEEE Transactions on Vehicular Technology, doi: 10.1109/TVT.2020.3008384.
4. **Y. Zhang**, J. Zhang, C. Xu, M. El-Hajjar and L. Hanzo, “Optimal-Power Superposition Modulation for Scalable Video Broadcasting”, submitted to IEEE Transaction on Vehicular Technology.

Contents

List of Symbols

Chapter 2

$\hat{b}_i$:	The received version of source bit of i-th modulation mode in superposition coding.
$\hat{c}_i$:	The received version of the FEC encoded bit of i-th modulation mode in superposition coding.
$\tilde{c}_i$:	The re-encoded version of the FEC encoded bit of i-th modulation mode in superposition coding.
d_1:	Minimum Euclidean distance between the quarters.
d_2:	Minimum Euclidean distance between the constellation points.
$f_{l,g}$:	The source bits of the g-th frame of the l-th video layer.
$\lvert f_{l,g} \rvert$:	The frame length of $f_{l,g}$.
$\hat{f}_{l,g}$:	The decoded version of $f_{l,g}$.
G:	Group of pictures.
h:	The fading channel coefficient.
$h_l(\gamma,\kappa)$:	LookUp table for the l-th layer.
$i_{\max}$:	The maximum number of generations.
L:	The maximum number of the scalable layer.
n:	Addictive white Gaussian noise coefficient.
P:	Total transmit power.
P_s:	The population size for evolutionary algorithm.

$p(\cdot)$: The PLR taking into account the error propagation.

$\tilde{p}(\cdot)$: The PLR taking into account interference.

$Q(\cdot)$: The reconstructed video quality according to the specific parameters.

$q_{l,g}$: The image quality of the g-th frame of the l-th layer.

s: Superimposed symbol of superposition coding.

S: The coverage of the BS.

s_g: Superimposed signal of the g-the frame of SVC.

$\hat{s}_g$: The received version of s_g.

T: The TDM period.

V: Source video.

V: Decoded version of V.

x_i: Modulated symbol of the i-th modulation mode in superposition coding.

y: Received superimposed signal.

α: Power coefficient assigned to the first modulation mode.

α_l: Power coefficient assigned to the l-th layer.

β: Power coefficient assigned to the second modulation mode.

$\beta(s)$: The path loss at distance s.

η: Power allocation ratio in HQAM.

π_l: The bit interleaver for the l-th video layer.

π_l^{-1}: The deinterleaver of π_l.

$\Delta q_{l,g}$: Image quality increment of the g-th frame of the l-th layer compared to its $l-1$-th layer.

$\Psi_{l,g}(\cdot)$: The equivalent decodability of $f_{l,g}$.

$\psi_{l,g}(\cdot)$: The decodability of $f_{l,g}$.

$\gamma_l(s)$: The received SNR of the l-th video at distance s.

$\kappa_l(s)$: The received SNR of the interference signal for the l-th video at distance s.

$\mathcal{X}$: slow fading coefficient.

μ_{Cr}:	Mean of crossover.
μ_{λ}:	Mean of mutation.
λ:	The packet bit length.
$\mathcal{O}(\cdot)$:	The computational complexity order.

Chapter 3

$\mathbf{A}$:	Dispersion matrices set.
$\mathbf{A}_q$:	The source bits.
b:	The q-th dispersion matrix.
b_{ASU}:	The bits feeding into the ASU of the MS-STSK.
b_{ASU}^{DAC}:	The bits feeding into the ASU of the DAC assisted MS-STSK.
b_{BL}:	The bit used to convey the BL.
b_{EL}:	The bit used to convey the EL.
$b_{MS\text{-}STSK}$:	The MS-STSK codeword.
b_M:	The bits feeding into the $\mathcal{L}$-PSK/QAM modulator of the MS-STSK.
b_Q:	The bits feeding into the dispersion matrices generator of the MS-STSK.
b_{STSK}:	The STSK codeword.
c:	The FEC encoded bits.
$\mathcal{C}$:	The complex value set.
$\tilde{\mathbf{C}}$:	The channel characteristic matrix.
$\hat{c}$:	The estimated AC index.
$\tilde{\mathbf{e}}$:	The channel cross-correlation vector.
f_d:	The Doppler frequency.
h_{ij}:	The channel coefficient of the signal sent from the j-th TA to the i-th RA.
$\mathbf{H}_n$:	The AG channel matrix.
$\mathbf{H}_n^D$:	The LoS component of the AG channel matrix.
$\mathbf{H}_n^S$:	The non-LoS component of the AG channel matrix.

$\mathbf{I}$: The identity matrix.

$J_0(\cdot)$: The zero-order Bessel function.

$I_{i,a}(u)$: The MI of the intrinsic LLR of the demapper.

$I_{i,e}(u)$: The MI of the extrinsic LLR of the demapper.

$I_{i,a}(c)$: The MI of the intrinsic LLR of the decoder.

$I_{i,e}(c)$: The MI of the extrinsic LLR of the decoder.

K: The Ricean factor.

$\mathbf{K}$: The LDC modulated symbol matrix.

N_c: The number of antenna combinations.

N_c^{DAC}: The number of antenna combinations in the DAC assisted MS-STSK.

n_i: Gaussian noise imposed on i-th RA.

N_{OW}: The observation window size.

N_{OW}^a: The observation lower boundary.

N_{OW}^b: The observation upper boundary.

N_{PS}: The pilot insertion period.

N_t: The number of transmit antennas.

N_r: The number of receive antennas.

N_{RF}: The number of RF chains.

N_0: The variance of Gaussian noise.

l_i: The i-th scalable video layer.

$\hat{l}_i$: The decoded version of the i-th scalable video layer.

$L_{i,a}(u)$: The intrinsic LLR for the inner demapper.

$L_{i,e}(u)$: The extrinsic LLR for the inner demapper.

$L_{o,a}(c)$: The intrinsic LLR for the outer decoder.

$L_{o,e}(c)$: The extrinsic LLR for the inner decoder.

$\hat{l}$: The estimated constellation index.

$\mathcal{L}$:	The order of the classic modulation.
$P(l_i)$:	The PLR of the i-th video layer.
$P(l_i)_e$:	The equivalent PLR of the i-th video layer.
Q:	The number of dispersion matrices.
r:	The received bits.
$\hat{q}$:	The estimated dispersion matrices index.
s:	The MS-STSK encoded bits.
s_{ij}:	The signal sent from the j-th TA to the i-th RA.
$S_{n_t,l}$:	The modulated signal matrix of the SM symbol.
t_{f_d}:	The threshold value for f_d.
T:	The number of time block.
u:	The interleaved bits.
$\mathbf{w}$:	The file tap.
x_i:	The modulated signals of the i-th layer.
$\mathcal{X}$:	The number of time block.
$\tilde{X}$:	The transmitted block signal.
y_i:	The received symbol of the i-th layer.
$\mathbf{Y}$:	The received block-based signal.
π:	The interleaver operation.
π^{-1}:	The deinterleaver operation.
γ:	The instantaneous channel SNR.
Δf_{LOS}:	The frequency offset on the LoS path.
Δf_{offset}:	The LoS frequency offset.
ϕ_t:	The value of AOD.
ϕ_r:	The value of AOA.
ϕ_0:	The angle between the LoS and the direction of movement.
σ_D^2:	The variance of the LoS component.
σ_S^2:	The variance of the non-LoS component.

Chapter 4

b_{ASU}: The bits feeding into the ASU of the MS-STSK.

b_M: The bits feeding into the $\mathcal{L}$-PSK/QAM modulator of the MS-STSK.

$b_{MS\text{-}STSK}$: The MS-STSK codeword.

b_Q: The bits feeding into the dispersion matrices generator of the MS-STSK.

f_i: The mode-switching threshold.

$\mathcal{L}$: The order of the modulation mode.

l_i: The bits of the i-th layer.

$\hat{l}_i$: The decoded version of l_i.

$L_a(s_i)$: The *a priori* information of s_i.

$L_e(s_i)$: The *a posteriori* information of s_i.

N_t: The number of transmit antenna.

N_r: The number of receive antenna.

N_{RF}: The number of the RF chains.

p_i: The parity bit stream of l_i generated.

$P(l_i)$: The PLR of the i-th video layer.

$P(l_i)_e$: The equivalent PLR of the i-th video layer.

P_t: The PLR threshold.

Q: The number of dispersion matrices.

s_i: The systematic bit stream of l_i.

s_{ij}: The mixed bit stream by implanting i-th bit stream into j-th bit stream s_j.

T: The time block.

x_i: The IL-FEC encoded bit stream of l_i.

y_{p_i}: The received parity bit stream of l_i.

y_{s_i}: The received systematic bit stream of s_i.

$y_{s_{ij}}$: The received mixed bit stream of s_{ij}.

y_i: The received bit stream of l_i.

V: The captured source video.

$\hat{V}$: The reconstructed source video.

γ: The instantaneous channel SNR.

σ: The standard deviation of shadow fading.

Chapter 5

$C_{r_{ps}}$: The crossover probability.

$f_{i,j}^{l}$: The i-th tile of the l-th video layer of the j-th video frame.

$|f_{i,j}^{l}|$: The frame length of $f_{i,j}^{l}$.

$\hat{f}_{i,j}^{l}$: The decoded version of $f_{i,j}^{l}$.

G: The GoP size.

$G_{\max}$: The maximum number of generations.

H: The video height.

$h(r,\Gamma)$: The PLR stored in a LUT using Monte-Carlo experiment.

$\hbar(r,\Gamma)$: The fitted mathematical model for $h(r,\Gamma)$.

h_p: The position of the point P on the height of the projected plane.

$J(\cdot)$: The Jacobian determinant.

M: The modulation mode.

N: The number of tiles.

Ps: The population size.

$p(\cdot)$: The PLR according to the specific parameters.

$\mathcal{R}$: The FEC code-rate set.

R: The overall FEC code-rate.

$R_{\Gamma,M}$: The highest FEC code-rate for modulation M at Γ.

$r_{i,j}^{l}$: The FEC code-rate assigned to $f_{i,j}^{l}$.

$r_{\min}$: The lowest FEC code-rate in $\mathcal{R}$.

$r_{\max}$: The highest FEC code-rate in $\mathcal{R}$.

$\hat{\mathbf{r}}_{g,p_s}$: The initial FEC code-rate set of p_s-th population in the g-th generation.

$\tilde{\mathbf{r}}_{g,p_s}$: The mutated $\hat{\mathbf{r}}_{g,p_s}$.

$\check{\mathbf{r}}_{g,p_s}$: The FEC code-rate set $\tilde{\mathbf{r}}_{g,p_s}$ after crossover.

SR: The stretching ratio.

T: The throughput upper bound.

$T_{i,j}^{l}$: The uncoded source rate contributed by $f_{i,j}^{l}$.

T_M: The uncoded source rate for modulation mode M.

T_ω: The overall weighted uncoded source rate.

T_ω^c: The weighted uncoded source rate for class c.

V: The panoramic video.

$\hat{V}$: The reconstructed panoramic video.

v_i: The i-th tile of the panoramic video.

$\hat{v}_i$: The reconstructed v_i.

v_i^l: The l-th layer of the i-th tile of the panoramic video.

W: The video width.

w_p: The position of the point P on the width of the projected plane.

x_p: The position of the point P on x-axis.

y_p: The position of the point P on y-axis.

$y(w,h)$: The original pixel at (w,h).

$\hat{y}(w,h)$: The received pixel at (w,h).

z_p: The position of the point P on z-axis.

$\bar{\alpha}_i$: The weighting coefficient of the i-th tile.

α_i: The normalized weighting coefficient of the i-th tile.

β_c': The weighting coefficient of class c.

β_c: The normalized weighting coefficient of class c.

ρ:	The radius of the 3D spherical video.
ϕ:	The latitude.
θ:	The longitude.
$\vartheta(w,h)$:	The weighting coefficient at position (w,h).
Γ:	The instantaneous channel SNR.
η:	The greedy factor.
λ:	The packet bit length.
λ_{p_s}:	The scaling factor in mutation.

Chapter 1

Introduction

The growing demand for high quality video streaming remains a serious challenge for the existing network due to its coding complexity and the data rate required. To reduce the bandwidth requirements, a whole family of video compression standards has been developed [1–7]. However, the compressed video packets become sensitive to transmission errors, since even a single bit error may result in an undecodable picture, if it corrupts the so-called picture start code, for example. Hence, the concept of Unequal Error Protection (UEP) was proposed to provide different level of protection for the bits having different importance. Furthermore, in order to enhance the scalability of video streaming in the transmission network, adaptive modulation may be invoked, which judiciously strikes the most appropriate throughput versus error-resilience trade-off, depending on the instantaneous channel Signal-to-Noise Ratio (SNR). Against this background, in this treatise, we conceive and investigate physical layer based adaptive UEP designed for scalable video streaming.

1.1 Video Compression

Uncompressed video clips contain a sequence of frames captured from a real-world scene, which requires large memory for data storage and a high bitrate for video transmission. Considering a 2-hour 1920 × 1080-pixel (8-bit) black-white movie having a 24 Frame Per Second (FPS) temporal scanning rate, the bitrate required to flawlessly transmit the movie is given by:

$$1920 \times 1080 \times 8 \times 24 \approx 400 Mbps, \tag{1.1}$$

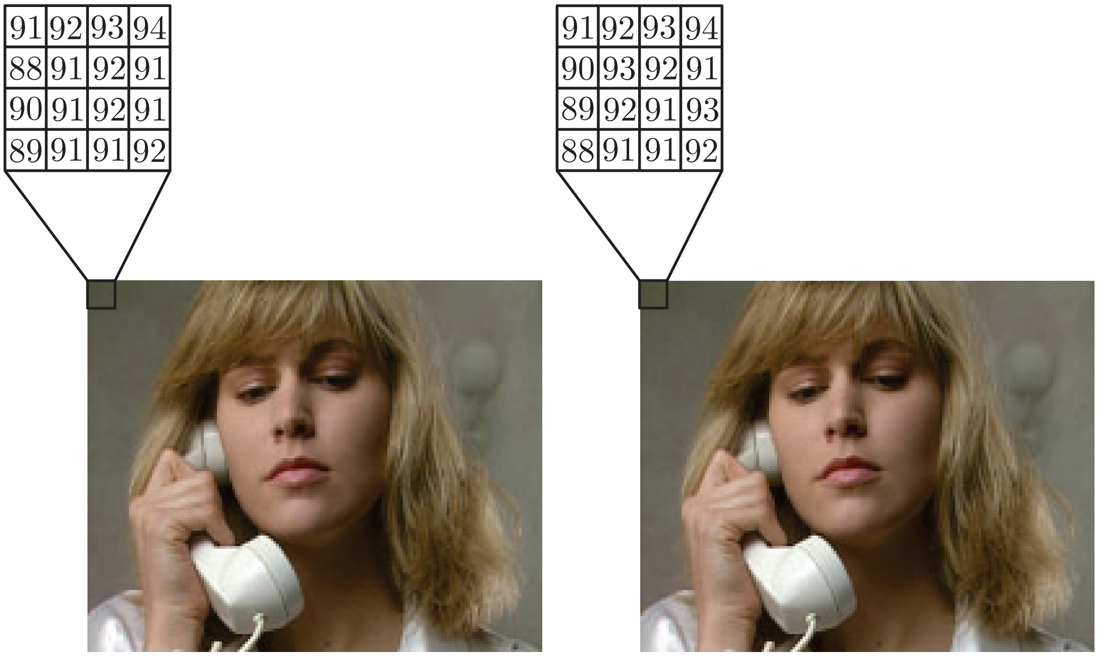
while the storage space of a hard disk used to restore this uncompressed video is:

$$1920 \times 1080 \times 8 \times 24 \times 2 \times 60 \times 60 \approx 460 GBytes. \tag{1.2}$$

Although uncompressed video is capable of providing flawless viewing experience for the users, the demanding requirements in storage space and transmission rate challenges the existing network. It is imperative to find a solution to reduce the number of bits, while maintaining near-flawless video quality. In this section, we commence by introducing the development of video coding techniques and the associated video coding standards. First, we highlight the simple frame-differencing based video codec, which exploits the temporal correlation between the consecutive frames. Then, the history of video coding standards is illustrated, culminating the so-called Scalable Video Coding (SVC) extension of the H.265 standard.

1.1.1 Simple Frame-Differencing Video Codec

A pair of consecutive uncompressed frames of the Suzie sequence is exemplified in Figure 1.1. Let us first focus on the spatial correlation of a single video frame by observing Figure 1.1(a) or Figure 1.1(b), where the gray-scale pixel luminance values in the specific 4×4 block of Figure 1.1(a) or Figure 1.1(b) exhibit a strong spatial similarity among the neighbouring pixels. This is a manifestation of intra-frame correlation. Furthermore, by comparing pairs of pixel values between Figures 1.1(a) and 1.1(b), the two consecutive video frames are also correlated temporally, which is referred to as inter-frame correlation. Figure 1.2 illustrates the Probability Density Function (PDF) of the difference between the co-located pixels, which was shown to have a Laplace-like distribution [8]. It can be seen in Figure 1.2 that the difference of two pixels belonging to the same location of different frames is likely to be close to zero, and the PDF drops dramatically, as the pixel-displacement is increased. This implies that the current frame can be approximated or pre-



91	92	93	94
88	91	92	91
90	91	92	91
89	91	91	92

(a) The first frame of Suzie clip

91	92	93	94
90	93	92	91
89	92	91	93
88	91	91	92

(b) The second frame of Suzie clip

Figure 1.1: The two consecutive frames of the Suzie clip and their associated pixel values of the top-left 4×4 block.

dicted from the previous frames. Below, we firstly review a simple so-called frame-differencing based video codec, which employs basic predictive coding. Then, we review the history of video compression standards, followed by the state-of-the-art in scalable video compression.

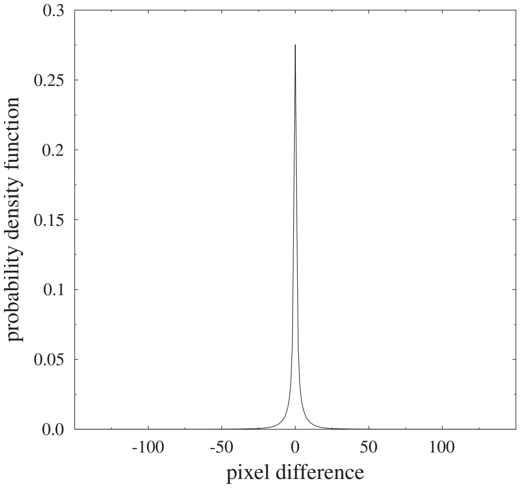
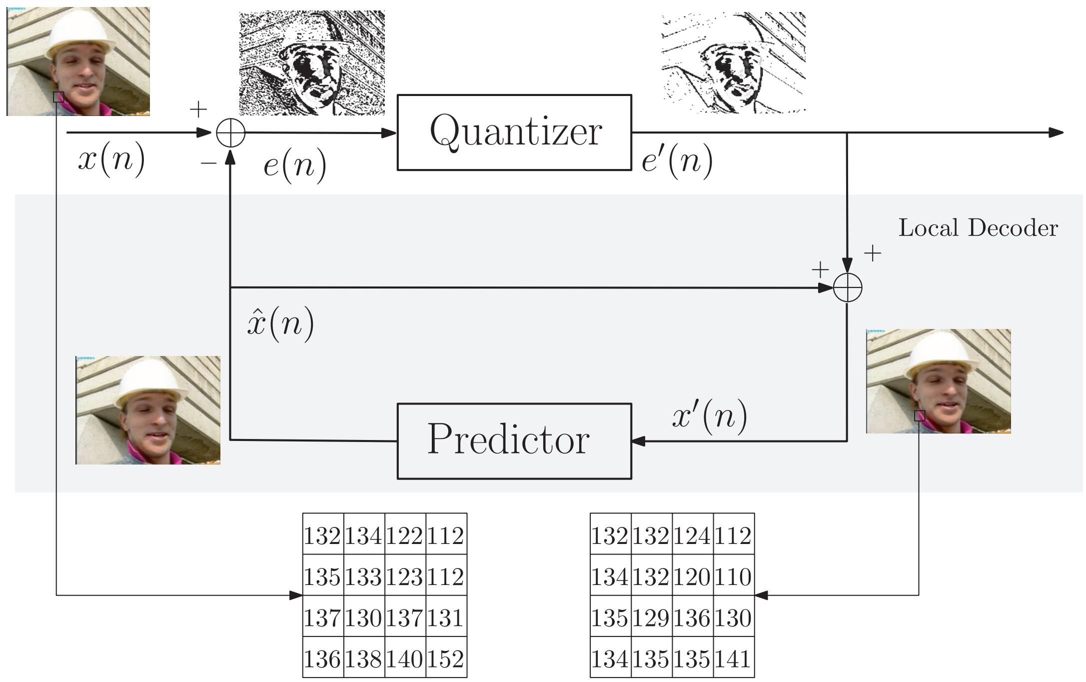


Figure 1.2: The distribution of the numerical error of the co-located pixel between Figures 1.1(a) and 1.1(b).

Figure 1.3(a) depicts a simple encoding process exploiting the inter-frame correlation discussed in the context of Figure 1.1. Observe in Figure 1.3(a) that the current video frame $x(n)$ is predicted by the reference video frame $\hat{x}(n)$ constructed from the most-recent reconstructed video frame $x'(n-1)$, resulting in a "line-drawing-like" difference frame given by $e(n) = x(n) - \hat{x}(n)$. Statistically speaking, $e(n)$ requires a lower number of bits than the original frame $x(n)$, since the associated frame-differencing operation attempts to compensate for the motion of the objects between the consecutive video frames. This difference frame $e(n)$ is then fed into a quantizer, where the insignificantly small noise-like textures in the error frame $e(n)$ are set to zero by the quantization, which generates a coarse version of the frame difference $e'(n)$ and hence results in a further reduced coding rate. Observe in Figure 1.3(a) that most of the areas in the quantized frame difference $e'(n)$ are "flat", having values close to zero, since the texture details are removed during quantization, which results in reducing the bitrate required for coding compared to the original frame $x(n)$ and the difference frame $e(n)$. This quantized version is then forwarded to entropy coding for further compression.

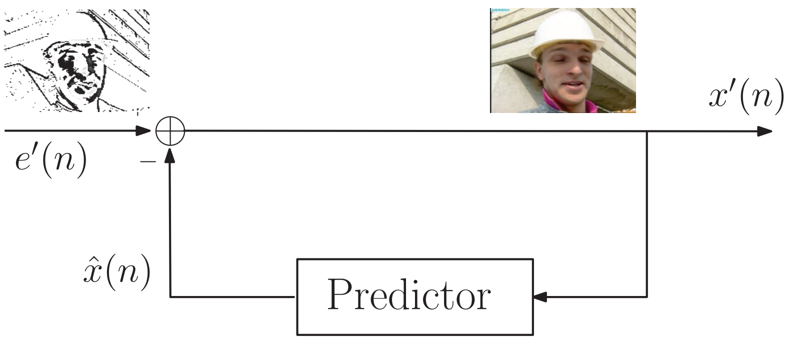
Additionally, a copy of $e'(n)$ is used for recovering the locally reconstructed current frame with the aid of the predicted version of the current frame $\hat{x}(n)$, which can be modelled according to $x'(n) = \hat{x}(n) + e'(n)$. The comparison of the numerical pixel values between the original frame $x(n)$ and reconstructed frame $x'(n)$ is also exemplified in Figure 1.3(a), which shows that the impact of video quality reduction introduced by quantization is fairly minor. Furthermore, the reconstructed frame $x'(n)$ is then used for generating the predicted version of the next video frame,



132	134	122	112
135	133	123	112
137	130	137	131
136	138	140	152

132	132	124	112
134	132	120	110
135	129	136	130
134	135	135	141

(a) Encoder



(b) Decoder

Figure 1.3: Basic video codec schematic using frame-differencing, where the corresponding video encoder and video decoder are displayed in (a) and (b), respectively.

namely $\hat{x}(n+1)$ in the predictor block of Figure 1.3(a), which requires the previously encoded frames.

It is worth noting that there is a local decoder within the encoder, which is identical to the remote decoder of Figure 1.3(b). This ensures that the decoder of Figure 1.3(b) uses the same reconstructed frame $x'(n-1)$ for synthesizing the image, as the one used by the encoder of Figure 1.3(a). To be more explicit, instead of using the previous original frame as the reference frame, the so-called locally reconstructed frame is used for motion compensation, where the “locally decoded” phrase implies “local decoding” at the encoder. This local decoding yields an exact replica of the video frame reconstructed at the distant decoder, which is required for the frame-differencing operation, because the previous original frame is not available at the distant decoder. To elaborate a little further, without the local decoding operation carried out at the encoder, the distant decoder would have to use the reconstructed version of the previous frame in its attempt to reconstruct the current frame, while the encoder would rely on the original unquantized frame, which would result in perceptual objectionable video artefact, especially in the presence of transmission errors.

1.1.2 History of Video Compression Standards

Fueled by the popularity of smart phones and tablets, the data rate demands have been escalating, especially with the growing availability of video services [9]. However, flawless video transmission in error-prone wireless environments is quite challenging [10], not only because delivering video steams requires a high bandwidth, but also a high channel quality. To reduce the required bitrate, sophisticated video compression standards have been designed, such as the H.261 [2], H.262 [3], H.263 [4], the Moving Picture Experts Group (MPEG)-4 [4], H.264/Advanced Video Coding (AVC) [5] and H.265/High Efficiency Video Coding (HEVC) [6], as well as the impending H.266/Versatile Video Coding (VVC) [7, 11] video compression recommendations. H.266 mainly aims for improving the coding efficiency of 4K or even higher-resolution video, including Virtual Reality (VR) compression. The timeline of these standards is summarized in Table 1.1. To elaborate a little further, these standards are briefly discussed as follows.

Table 1.1: Development of the video compression standards.

Year	Developer	Standard
1984	ITU	H.120 [1]
1988	ITU	H.261 [2]
1992	JPEG	JPEG [12]
1993	ISO/IEC	MPEG-1 [13]
1994	ISO/IEC& ITU-T	MPEG-2/H.262 [3]
1994	IEC	DV [14]
1995	ITU-T	H.263 [4]
1997	RealNetworks	RealVideo [15] [16]
1998	ISO	MPEG-4 [9]
2003	ISO/IEC & ITU-T	AVC [5]
2003	AVS Workshop	AVS [17] [18]
2012	ISO/IEC & ITU-T	HEVC [6]
2020	ISO/IEC &ITU-T	VVC [7] [19]

H.120: the International Telecommunication Union (ITU) H.120 [1] scheme represents the first ever digital video coding standard, which relies on Differential Pulse Code Modulation (DPCM) based quantization of pixels and Discrete Cosine Transform (DCT) transformed domain based entropy coding.

H.261: H.261 [2] was designed for transmitting Quarter Common Intermediate Format (QCIF) and Common Intermediate Format (CIF) video clips, which pioneered the era of practical digital video compression techniques. This standard employed a hybrid video coding scheme, which formed the basis of all the following video coding standards.

Joint Photographic Experts Group: the Joint Photographic Experts Group (JPEG) [12] codec was standardized for digital still image compression. The video clip encoded by the JPEG scheme was treated as a sequence independent images. Since the temporal inter-frame redundancy is not exploited, the JPEG compression exhibits a low computational complexity, but a high bitrate.

Moving Picture Experts Group-1: This is the first version of the MPEG compression standard family, which was mainly used for movies stored on Digital Versatile Disc (DVD). However, it was not suitable for the National Television System Committee (NTSC) standard video format.

H.262: The ITU and International Standardization Organization (ISO) jointly developed the MPEG-2/H.262 [3] standard for ensuring compatibility of the NTSC and MPEG-1 standards. H.262 is widely employed as the standard of TeleVision (TV), in terrestrial broadcast, cable and satellites systems.

Digital Video: the Digital Video (DV) coding standard [14] developed by the International Electrotechnical Commission (IEC) was mainly devised for camera recording, whose compression efficiency is higher than that of JPEG, but worse than that of the H.262 standard.

H.263: the ITU Telecommunication Standardization Sector (ITU-T) [4] standardized the H.263 specification in 1995, which aimed for improving the compression efficiency of video conferencing. Later, a pair of more advanced versions were developed for enhancing the compression performance by adding numerous annexes.

Real Video: RealVideo [15, 16] is a successful proprietary video compression format developed by the RealNetworks company, which was first released in 1997. RealVideo is supported by numerous computing platforms, including Windows, Mac, Linux, Solaris and several mobile phones.

Moving Picture Experts Group-4: the MPEG-4 standardization [20] was initiated in 1995 and has been continually updated by a number of new profiles, where the Simple Profile (SP) provides the very similar functions to H.263, while the compression efficiency attained is further improved by the Advanced Simple Profile (ASP).

H.264: the H.264/AVC standard [21], also referred to as MPEG-4 AVC part 10, was completed in 2003, but the associated research continued evolving by adding more extensions. The

H.264/MPEG-4 AVC part 10 was jointly developed by the ITU-T Video Coding Experts Group (VCEG) and the ISO/IEC JTC1 MPEG, which has become one of the most commonly employed standards of video recording, compression and distribution. The goal of this standard was to double the compression efficiency realized by the previous video standards, while maintaining the same video quality. A major extension of H.264/AVC was the SVC [21] scheme, which was completed in 2007, compressing the source frames into multiple sub-bitstreams of different video layers.

Audio Video Coding Standard: the Audio Video Coding Standard (AVS) [17, 18] was initiated by the government of China for replacing MPEG-2, which was standardized in 2005.

H.265: the H.265/HEVC [6] was developed as a successor of the H.264/AVC standard, which was jointly designed by the ISO/IEC MPEG and ITU-T VCEG as ISO/IEC 23008-2 MPEG-H Part 2 and ITU-T H.265. HEVC mainly aims for supporting Ultra High Definition (UHD) videos, which roughly halves the bitrate of the H.264/AVC standard at the same video quality. The emerging H.265 scheme also continued to support the scalable extension by the Scalability extension of HEVC (SHVC) profile, which allows the transmitter to meet multiple users' preferences.

H.266: H.266/VVC [7, 19] constitutes the most recent video compression technique to be standardized, which was developed by the Joint Video Expert Team (JVET) of the ITU-T VCEG and the ISO/IEC MPEG, aiming for doubling the compression efficiency achieved by its HEVC counterpart. JVET was founded as the JVET of ITU-T VCEG and ISO/IEC MPEG in October 2015. The resultant video coding standard was given the acronym VVC and is expected to be finalized in July 2020.

1.1.3 Scalable Video Coding

SVC has drawn researchers' attention for more than 20 years, as a promising technique of enhancing the system's spectral efficiency, ever since the H.262 standard was ratified [21]. Nevertheless, scalable profiles were rarely applied due to their limited coding efficiency, when maintaining a moderate decoding complexity. This problem was mitigated by the H.264 standard, which significantly improved the video compression capability [22] by combining predictive coding, transform coding and entropy coding. Additionally, the scalable video concept was also enhanced and embedded in the HEVC standard, resulting in the SHVC standard [23]. The scalable structure relies on a so-called Base Layer (BL) and several Enhancement Layers (EL) that progressively refine the video quality by incorporating several ELs as and when an increased bitrate is acceptable. The scalability in SVC can be spatial, temporal and quality based (also referred to as SNR scalability).

A layered video scheme is shown in Figure 1.4, where the video sequence captured is encoded with the aid of one or more scalability functionalities into the four layers by a scalable video en-

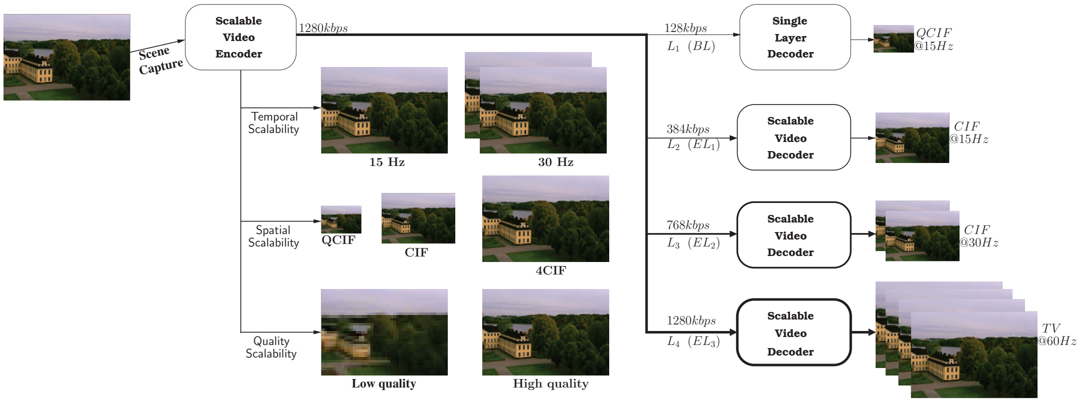


Figure 1.4: Architecture of a SVC scheme.

coder. The four layers in the example shown in Figure 1.4 contain the BL and three ELs. In Figure 1.4, the ELs are encoded progressively using the lower-order ELs as their reference. For example, L_4 shown at the input of the scalable video decoder of Figure 1.4 is encoded using all of teehe three lower layers. Figure 1.4 reveals the basic structure of the SVC, where the layered video coding constitutes a flexible architecture. It is also worth noting that the BL of the SVC can be extracted and decoded independently by a single-layer video decoder [21].

1.1.3.1 Temporal Scalability

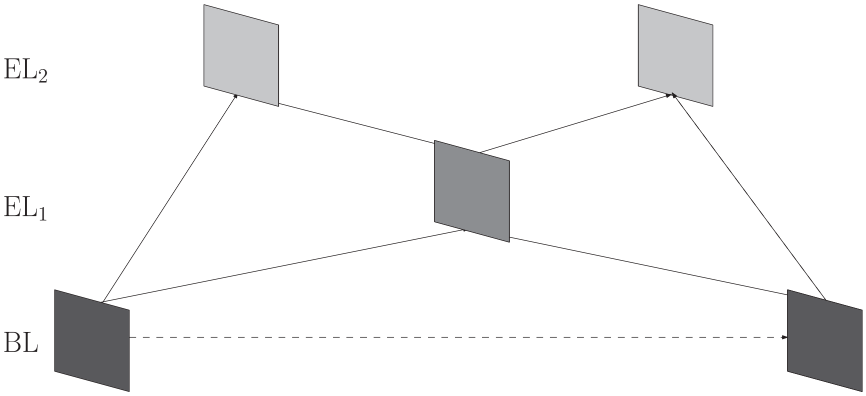


Figure 1.5: Multilayer structure with additional IL prediction for enabling temporal scalability.

The temporal scalability is designed for video services that require different temporal resolutions, namely different frame rates. The target applications include video streaming over wireless channel where the video frame rate may have to be dropped, when experiencing the poor channel condition. In temporally scalable coding, the BL is coded at the lowest frame rate, where the frame rate gradually increases with the layer index.

When relying on the temporal scalability, the set of corresponding video frames can be partitioned into a temporal BL and one or more temporal ELs, as depicted in Figure 1.5. Temporal scalability can generally be facilitated by restricting motion-compensated prediction to reference pictures having a temporal layer identifier that is lower than or equal to the temporal layer identifier of the picture to be predicted. The prior video coding standards such as MPEG-1 [13], H.262 [3] and H.263 [4] etc. are all capable of providing temporal scalability to some extent, but the feasibility of temporal scalability was significantly improved by H.264 [5], which is an explicit benefit of the so-called reference picture memory control technique.

Figure 1.5 illustrates the temporal scalability relying on so-called dyadic temporal EL encoded as Bi-directionally predicted (B)-frames [24] [25] across a Group of Pictures (GoP), where the EL frame that exhibits additional content is encoded by referencing two frames of its lower layers. Moreover, the EL frame can also be used as the reference for encoding its higher layer.

1.1.3.2 Spatial Scalability

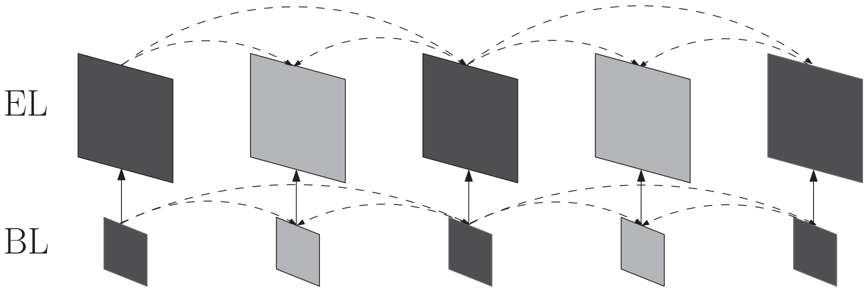


Figure 1.6: Multilayer structure with additional IL prediction for enabling spatial scalable coding.

With the expectation that future applications will support a diverse range of display resolutions and transmission channel capacities, the Joint Video Team (JVT) has developed a scalable extension [21] of the state-of-the-art H.264/AVC video coding standard [5]. This provides support for multiple display resolutions within a single compressed bit stream, also referred to here as spatial scalability [26]. Specifically, larger high-definition displays are becoming common in consumer applications, with displays containing over two million pixels becoming readily available. By contrast, lower-resolution displays having ten to hundred thousand pixels are also popular in applications constrained by size, power and weight. Figure 1.6 shows an example of a twin-layer scalable bitstream relying on spatial scalability containing a low-resolution BL and an EL.

1.1.3.3 Quality Scalability

Quality scalability can be considered as a special case of spatial scalability having identical picture sizes for both the base and for the enhancement layer, which is supported by the general concept for spatially scalable coding, which is usually referred to as Coarse-Grain Scalable (CGS) coding [27].

The same inter-layer prediction mechanisms are employed as for spatially scalable coding, but without using the corresponding upsampling operations. When utilizing inter-layer prediction for coarse-grain quality scalability in SVC, the refinement of texture information is typically achieved by requantizing the residual texture signal in the enhancement layer using a smaller quantization step size than that used for the preceding CGS layer.

1.1.4 Subjective Video Quality Assessment

Subjective video quality evaluation quantifies how it is perceived by an observer, therefore it is also closely related to the Quality of Experience (QoE). Quantifying the subjective video quality is necessary, since objective quality metrics have been shown to correlate badly with QoE ratings.

1.1.4.1 Peak Signal-to-Noise Ratio

For a video sequence containing numerous frames, the Peak SNR (PSNR) is calculated for each frame then averaged across all the frames [28], containing $W \times H$-pixel, formulated as [29]:

$$PSNR = 10\log_{10}\frac{(2^b-1)^2}{MSE}, \tag{1.3}$$

where b represents the number of bits used for presenting a pixel. For example, the majority of existing videos employ $b = $ 8-bit and 10-bit to present a pixel, hence exhibiting a peak signal value of $(2^b - 1) = 255$ and 1023, respectively. Moreover, the Mean Square Error (MSE) indicates the corruption level of a video frame, which is given by:

$$MSE = \frac{1}{WH}\sum_{w=0}^{W-1}\sum_{h=0}^{H-1}[Y(w,h) - Y'(w,h)]^2, \tag{1.4}$$

where $Y(w,h)$ and $Y'(w,h)$ represent the original and compressed pixel value at position (w,h) of the frame. For an N-frame video clip, the average PSNR $\overline{PSNR}$ can be calculated by:

$$\overline{PSNR} = \frac{1}{N}\sum_{i=1}^{N} PSNR_i, \tag{1.5}$$

where $PSNR_i$ represents the PSNR of the i-th frame.

1.1.4.2 Mean Opinion Score

The Mean Opinion Score (MOS) [30] is a widely accepted video and audio metric. The ITU-T has defined several ways of referring to a MOS in Recommendation P.800.1 [31] according to the tested subjects, including audiovisual, conversational, listening, talking, and video quality tests.

The MOS scale is very commonly used, which maps numbers between 1 and 5 to the corresponding ratings between Bad and Excellent, as seen in Table 1.2.

Table 1.2: The MOS ratings.

Rating	Label
5	Excellent
4	Good
3	Fair
2	Poor
1	Bad

1.2 Physical Layer Based Techniques for Providing Unequal Error Protection

From the perspective of the source codec described in Section 1.1, there are mainly two ways of mitigating the impact introduced by transmission errors [32]. A popular class of solutions relies on the so-called error-concealment techniques, which have been investigated for example in [33–36]. These error-concealment techniques reduce the level of annoyance imposed by corrupted video frames at the video frame level for example by replacing a corrupted frame by the previous uncorrected one. However, these methods are usually employed as part of source decoding and increase the decoding complexity and power consumption. A more sophisticated technique of improving the error-resilience is to employ Forward Error Correction (FEC) [37]. In UEP schemes, the more important video content, such as the Intra-frame coded (I) and BL frames, are better protected than the dependent frames, such as the Predicted (P) frames and ELs, for improving the overall reconstructed video quality.

In general, the process of communications is supported by the cooperation of all Open Systems Interconnection (OSI) layers [38], where the video and audio services are part of application layer, but we mainly focus on physical layer based UEP schemes relying on multi-rate FEC and modulation schemes.

1.2.1 FEC Based Unequal Error Protection

In this section, variable code-rate allocation and interlayer (IL) operation-aided FEC (IL-FEC) schemes are considered. Explicitly, the former family employs the most appropriate FEC code-rate assigned to the video frames according to their specific importance, while the latter implants the bits of the more important layer, such the BL, into those of the less important layers, where iterative decoding exchanging soft extrinsic information between layers may be invoked for enhancing the protection of the more important bits.

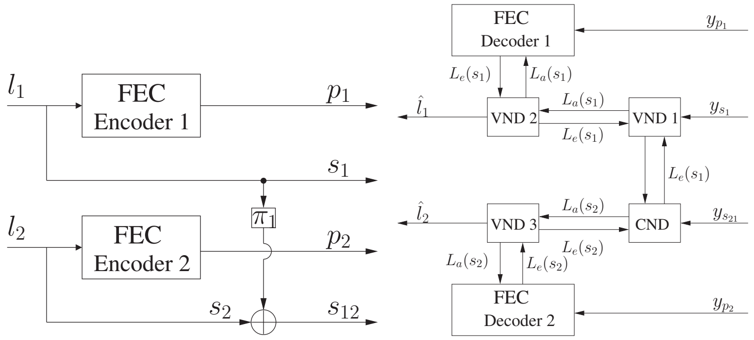
Table 1.3: Major contributions on variable-rate FEC based UEP for video streaming.

Year	Authors	Contributions
2003	Su *et al.* [39]	found an optimal channel coding assignment scheme for scalable video transmission over burst error channel with loss rate feedback.
2003	Yang *et al.* [40]	assigned FEC codes across packets to different frames in a GoP by exploiting the temporal dependency between frames.
2004	Marx & Farah [41]	minimized the mean distortion by non-uniformly distributing the redundancy imposed by a Turbo Codes (TC) between the successive video frames.
2006	Fang & Chau [42]	used a Genetic Algorithm (GA) to find the optimal Reed-Solomon (RS) code-rate for SVC streaming.
2006	Barmada *et al.* [43]	employed TC to provide high protection for the BL under the constraint of the maximum affordable channel coding redundancy.
2008	Chang &*et al.* [44]	employed variable rate coded Convolutional Code (CC) to protect the frames based on the importance of the frame and the macroblocks in a video frame.
2011	Nasruminallah & Hanzo [45]	evaluated the UEP performance of data-partitioned H.264/AVC video streaming systems using various combinations of variable rate coded Short Block Code (SBC) and Recursive Systematic Convolutional (RSC).
2013	Zhang *et al.* [46]	designed a rate-distortion model and assigned the optimal RS to different video packets according to their priority levels as well as the channel conditions.
2013	Xiong *et al.* [47]	proposed a priority encoding transmission for delivering the scalable video over erasure channel.
2016	Huo *et al.* [48]	conceived historical information aware UEP for SHVC streaming over free space optical system.
2017	Chen *et al.* [49]	proposed a joint power and FEC allocation scheme for scalable video streaming, which employs various FEC code-rates to video layers.
2019	Huo *et al.* [50]	designed a specific objective function by carefully considering the tile importance of the panoramic video, where the associated FEC code-rates are optimized by employing Lagrange Multiplier and Newton's down-hill method.
2020	Paudel & Vafi [51]	conceived a UEP algorithm to assign different Low Density Parity Check (LDPC) code-rates to HEVC tiles according to their motion vectors.

1.2.1.1 Variable Code-Rate Allocation

When using FEC, redundant bits, also referred to as parity bits, are incorporated into the raw data steam before modulation. At the receiver side, channel decoding allows the receiver to detect and correct errors with the aid of the parity bits. A high FEC code-rate indicates that only a low fraction of parity bits is introduced during the encoding, which hence requires less extra bandwidth during transmission at the expense of a weak protection capability and vice-versa. Table 1.3 lists the historical evolution of variable code-rate based UEP relying on CC [44], RSC codes [52, 53], SBC [54], RS codes [42], TC [41] and LDPC [55].

1.2.1.2 Inter-layer FEC



(a) Encoder (b) Decoder

Figure 1.7: Basic schematic of the IL-FEC technique, where the corresponding IL-FEC encoder and decoder are displayed in (a) and (b), respectively.

Based on the fact that the ELs have to be encoded and decoded relying on the BL due to the inter-layer coding dependency shown in Figure 1.4, it may be beneficial to utilize the information of the ELs to protect the BL, as originally suggested for an iterative detection aided receiver in [60].

In this treatise, we mainly consider the physical layer based IL-FEC philosophy [60], as exemplified in Figure 1.7 showing a twin-layer scenario as an example. However, the concept of IL-FEC can be readily extended to any number of layers. Observe in Figure 1.7(a) which depicts the encoder of the IL-FEC scheme that the video source bits of the BL are implanted into those of the EL with the aid of modulo-2 connection, but the parity bits generated by the FEC encoder remain unchanged. The video source bits of the BL, the modulo-2 function of the original BL and EL source bits and their parity bits are then multiplexed and modulated before transmission. In order to arrange for the efficient operation of this scheme, so-called systematic FEC codes have to be used, where the video source bits and parity bits are kept separate. Explicitly, the parity bits of FEC Encoder 1 and Encoder 2 are denoted by p_1 and p_2, the systematic video source bits of Encoder 1 are s_1, while the modulo-2 function of s_1 and s_2 is represented by s_{12} in Figure 1.7.

Table 1.4: Major contributions on IL-FEC assisted UEP for video streaming.

Year	Authors	Contributions
2009	Vukobratovic *et al.* [56]	utilized the EL to protect the most important BL using the Expanding Window Fountain (EWF) codes for scalable video multicast.
2010	Nguyen *et al.* [57]	conceived the hierarchical network coding, which used the less important bits to protect their more important counterparts.
2011	Hellge *et al.* [58]	proposed the Raptor code based layer-aware FEC philosophy for layered video streaming over the binary erasure channel.
2011	Halloush & Radha [59]	design multi-generation mixing technique, where the data can be recovered using the encoded into other packets.
2013	Huo *et al.* [60]	designed an IL-FEC techniques for streaming data-partitioned AVC, where the more important bits are implanted into the less important bits.
2016	Zhu *et al.* [61]	proposed a sophisticated Hybrid Automatic Repeat Request (HARQ) assisted IL-FEC scheme for SVC streaming.
2017	Bui *et al.* [62]	improved the layer-aware structure proposed in [58], where the advanced version of Raptor code referred as to RaptorQ was employed.
2020	Chen *et al.* [63]	conceived context-aware RaptorQ assisted layer-aware FEC scheme for HEVC streaming, where the important video packets are protected by redundant packets of the packets of both importance.

At the IL-FEC receiver, the demodulated signals are FEC decoded using the schematic of Figure 1.7(b). Explicitly, the bits of the BL are first decoded, which are then fed into the Check Node Decoder (CND) together with the received mixed bits of the EL representing $s_{12} = s_1 \oplus s_2$ to extract the systematic bits of the EL for FEC decoding. Then, the decoded systematic bits are fed back to the CND, yielding the extrinsic information of the BL. This extrinsic information enhances the Log-Likelihood Ratio (LLR) of the BL using Variable Node Decoder (VND) 1 for improving its Bit Error Rate (BER) performance and hence the overall reconstructed video quality. This iterative decoding continues until the maximum number of iterations is reached. Based on this technique, a novel system for IL coded Self-Concatenated Convolutional Code (SECCC) [64] aided scalable video streaming was conceived by employing a RSC as its constituent codes. Furthermore, this IL-FEC technique may be further extended to the SVC streaming solution of [65], where the concept of variable FEC code-rate described in Section 1.2.1 is also employed using exhaustive search.

1.2.2 Modulation Based UEP

In addition to employing different-rate FEC codes to provide UEP, transceiver based UEP schemes are also conceived for streaming the video signals. Two basic classes are introduced in this section, where the first represents modulation based UEP, including Hierarchical Quadrature Amplitude

Modulation (QAM) (HQAM) and Superposition Coding (SC) [66, 67], while the other relies on transceiver techniques taking into account characteristics of sophisticated Multiple-Input Multiple-Output (MIMO) schemes.

1.2.2.1 Bit Mapping Based UEP

Table 1.5: Major contributions on bit mapping based UEP for video streaming.

Year	Authors	Contributions
2005	Barmada *et al.* [68]	conceived multilevel HQAM arrangements with adaptive constellation distances for data partitioned AVC streaming.
2006	Chang *et al.* [69]	proposed adaptive HQAM for AVC video streaming in macroblock level, where different macroblocks are mapped into different constellation bits.
2006	Ghandi *et al.* [70]	compared the performance of both the SVC and the non-scalable data partitioned AVC streams, both of which achieve UEP using HQAM.
2009	Chang *et al.* [71]	designed a UEP scheme using HQAM, which takes into consideration the non-uniformly distributed importance of I-frame and P-frame in a GoP.
2010	Li *et al.* [72]	conceived an Orthogonal Frequency Division Multiplexing (OFDM), assisted HQAM scheme, where the more important data is protected by both the HQAM and the subcarrier having good channel condition.
2011	Cai *et al.* [73]	proposed scalable modulation for the handheld devices, where the main idea is to remap the constellation points to the bits in different layers.
2012	Alajel *et al.* [74]	conceived an UEP scheme based on HQAM for 3-dimensional video transmission, which exploits the unique characteristics of the color plus depth map stereoscopic video.
2012	Chang *et al.* [75]	proposed a multilevel UEP system using multiplexed HQAM, where an asymmetric HQAM is conceived to reduce the average and peak powers.
2014	Nguyen *et al.* [76]	considered multiple relay, single source and single destination network for video streaming, where HQAM is individually activated at relay nodes.
2017	She *et al.* [77]	designed a logical SC framework at the transmitter, where the modulated signals is generated by mapping successively refined information bits into a single signal constellation.
2018	Lee *et al.* [78]	provided performance analysis of the combination of SC and SHVC, which is compared to other approaches, such as HEVC simulcast broadcasting.

The basic philosophy of modulation based UEP is to map the source information to the different bits of the constellation points of HQAM [69, 76] according to its importance, including the video frame type, such as I, P and B frames, and layer index, representing the BL and EL. Some of the related seminal contributions are listed in Table 1.5.

1.2.2.2 MIMO Based UEP

Table 1.6: Major contributions on MIMO based UEP for video streaming.

Year	Authors	Contributions
2007	Song & Chen [79]	proposed adaptive antenna selection based SVC streaming relying on MIMO technique, where the more important information is fed into the antenna having the higher-SNR.
2010	Xiao *et al.* [80]	conceived a cooperative MIMO framework for SVC streaming, where a sophisticated power allocation strategy is invoked for controlling the relay.
2011	Chang *et al.* [81]	constructed a superimposed MIMO scheme, where the important data is STBC coded and less important data is spatially multiplexed.
2013	Kim *et al.* [82]	investigated the trade-off between error propagation and coding efficiency, where FEC and MIMO are switched according to the Doppler frequency.
2013	Li *et al.* [83]	proposed a scalable resource allocation framework for streaming SVC over multiuser MIMO OFDM networks with the aid of MAXMIN fairness.
2013	Wang & Liao [84]	proposed coded cooperative and opportunistic cooperative schemes for SVC multicasting.
2013	Zhou *et al.* [85]	investigated a channel scheduling problem for SVC broadcasting, where video layers are protected unequally by mapping to appropriate antennas.
2015	Zhou *et al.* [86]	formulated the SVC transmission scheme as a nonlinear optimization problem, exploiting the FEC redundancy and diversity of MIMO antennas.
2016	Cui *et al.* [87]	addressed the antenna heterogeneity in wireless video multicast by the design of multiple similar description video coding and multiplexed STBC.
2017	Choi & Kang [88]	conceived an optimal power control strategy for spatial multiplexing assisted open-loop SVC broadcasting.
2018	Guo *et al.* [89]	proposed a quality-based multi-quality multicast beamforming scheme, which exploits the layered structure and quality information of all users.

The major contributions on MIMO-based UEP for video streaming are summarized in Table 1.6, which also often exploit the different Channel State Information (CSI) of independent by fading antennas. To elaborate a little further, the authors of [79] proposed UEP schemes with the aid of the antenna-based spatial multiplexing, where the channel having the highest received SNR is utilized to transmit the more important information bits. Moreover, Choi *et al* [88] proposed optimal power-allocation assisted spatial multiplexing for layered video broadcasting, where the most important BL is expected to provide wide-ranging coverage, while the least important EL may only be flawlessly detected by the users near the broadcasting transmitter. Furthermore, a superimposed MIMO aided UEP scheme combining both Space-Time Block Codes (STBC) and the Bell Laboratories Layered Space-Time (BLAST) scheme was conceived in [81] for layered video streaming, where the low-rate but more robust STBC is used for conveying the more important BL.

1.3 Adaptive System Design

It has been widely acknowledged that the radio channel imposes a severe challenge on reliable high-speed communication owing to its susceptibility to noise, interference, as well as dispersive fading environments [90]. Although the transmitters are designed for providing a sufficiently high signal level for the distant receivers, due to fading the instantaneous received signal power fluctuates and routinely falls below the sensitivity of the receivers. Hence, the information cannot be decoded correctly [37]. In real-time video transmission, this becomes particularly objectionable. Fortunately, numerous near-instantaneously adaptive techniques have been conceived for improving the robustness of wireless systems [91] by providing users with the best possible compromise among a number of contradicting design factors, such as the power consumption of the mobile station, robustness against transmission errors and so forth [92] [93] [94]. In order to allow the transceiver to cope with the time-variant channel quality of narrowband fading channels, the concept of Adap-

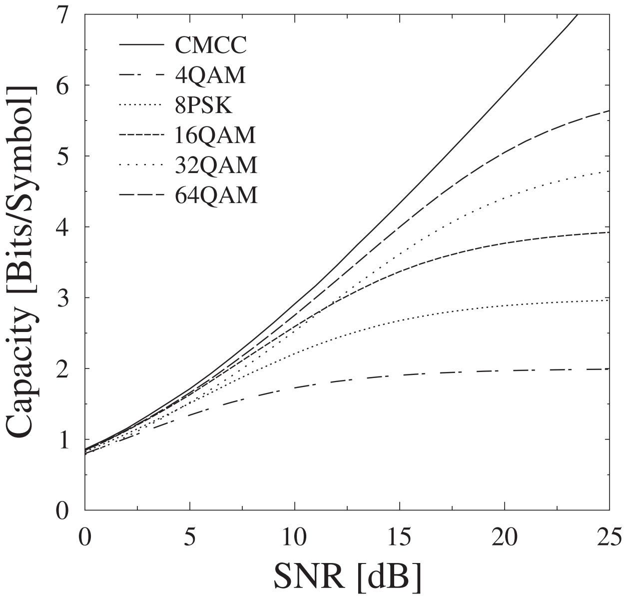


Figure 1.8: The CCMC and DCMC capacity versus SNR for transmission over uncorrelated Rayleigh fading channels.

tive QAM (AQAM) was proposed by Steele and Webb [95], which provides the flexibility to vary both the BER and the bitrate to suit a particular application. AQAM-aided wireless video transmission was conceived for example in [92] and [95], where the Burst-by-Burst (BbB) AQAM assisted system provides a smoother PSNR degradation when the channel SNR is degraded. As a benefit, the subjective image quality erosion imposed by the video artefacts is eliminated and hence the dramatic PSNR degradation of the conventional fixed modulation modes is avoided by the adaptive system.

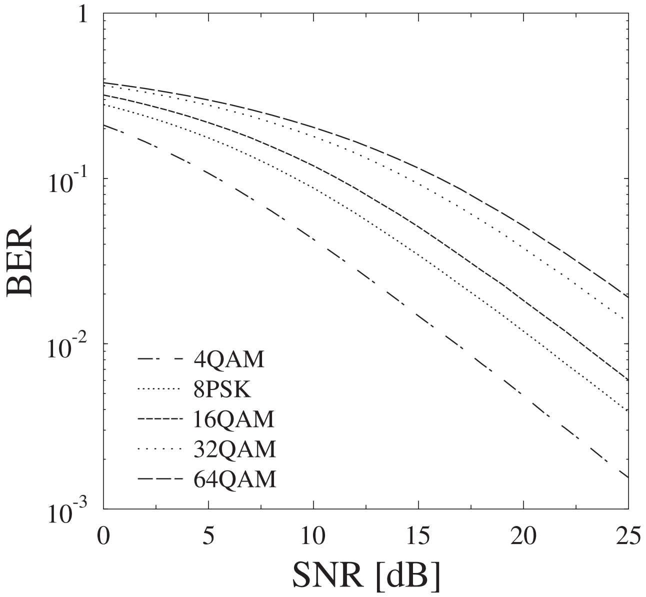


Figure 1.9: BER versus SNR for transmission over uncorrelated STBC 1×1 Rayleigh fading channels.

Figure 1.8 shows the comparison of the Discrete-Input Continuous-Output Memoryless Channel (DCMC) capacity for 4QAM, 8Phase Shift Keying (PSK), 16QAM, 32QAM and 64QAM as well as the Continuous-Input Continuous-Output Memoryless Channel (CCMC) capacity over a Single-Input Single-Output (SISO) fading channel. More specifically, the asymptotic DCMC capacity is 2 Bits Per Symbol (BPS) for 4QAM, 3 BPS for 8PSK, 4 BPS for 16QAM, 5 BPS for 32QAM and 6 BPS for 64QAM. Based on the channel capacity, we can design near-instantaneously adaptive modulation schemes by activating the most appropriate SNR-dependent coding rate and modulation modes. For example, 8PSK modulation may be employed at a SNR of 5 dB, while if the SNR is improved to 15 dB, 64QAM may be adopted. Furthermore, the BER performance of the above-mentioned modulation modes is shown in Figure 1.9, where 4QAM having the lowest throughput upper bound of Figure 1.8 exhibits the best BER performance, while 64QAM having the highest throughput bound requires additional SNR to maintain the BER. To maximize the achiev-

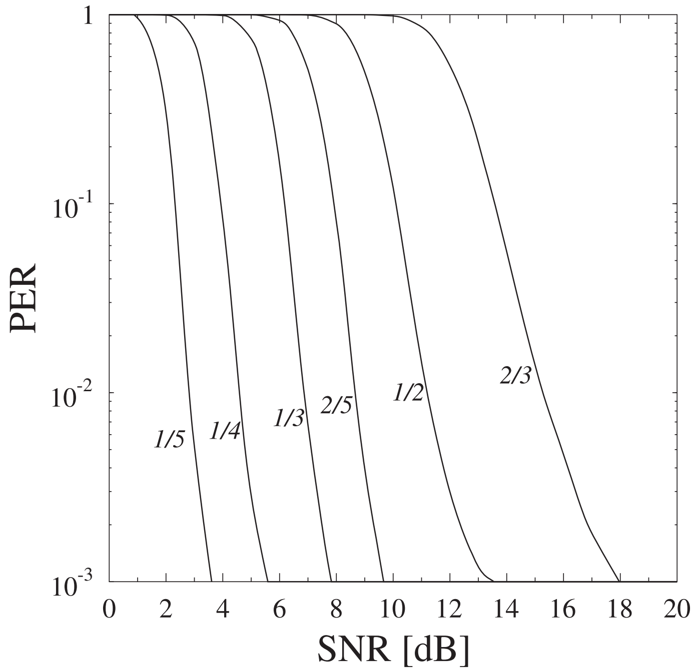


Figure 1.10: The PER versus channel SNR for different TC [7 5 5] code-rates conveyed by 16QAM over STBC 2 × 1 Rayleigh channel.

able system throughput at the specific SNR, while maintaining an acceptable BER performance, the AQAM concept comes to rescue, where the high-capacity modulation modes are only activated only when their BER is below the pre-defined threshold.

Moreover, this near-instantaneous adaptivity can also be readily applied to other physical layer based techniques, such as FEC. Figure 1.10 shows the Packet Error Ratio (PER) performance of various TC code-rates over a STBC 2 × 1 Rayleigh channel, where the code-rate determines the specific proportion of the system capacity that remains available for conveying the desired source information. For example, in the case of an FEC code-rate of 1/4, only 20% of the capacity is used for conveying video bits, while the remaining 80% conveys the FEC parity bits. Observe from Figure 1.10 that as expected the FEC code-rate of 1/5 is capable of providing the best PER performance, outperforming its 1/4 and 1/3 rate counterparts by about 2 dB and 4 dB, respectively, when achieving a PER of 10^{-3}.

1.4 Outline of the Thesis

In this treatise, the focus is on the family of physical layer based UEP techniques designed for SVC streaming, with a special emphasis on FEC and modulation. The outline of this treatise is portrayed in Figure 1.11.

Chapter 1: The basic concept of video coding was introduced in this chapter, followed by the his-

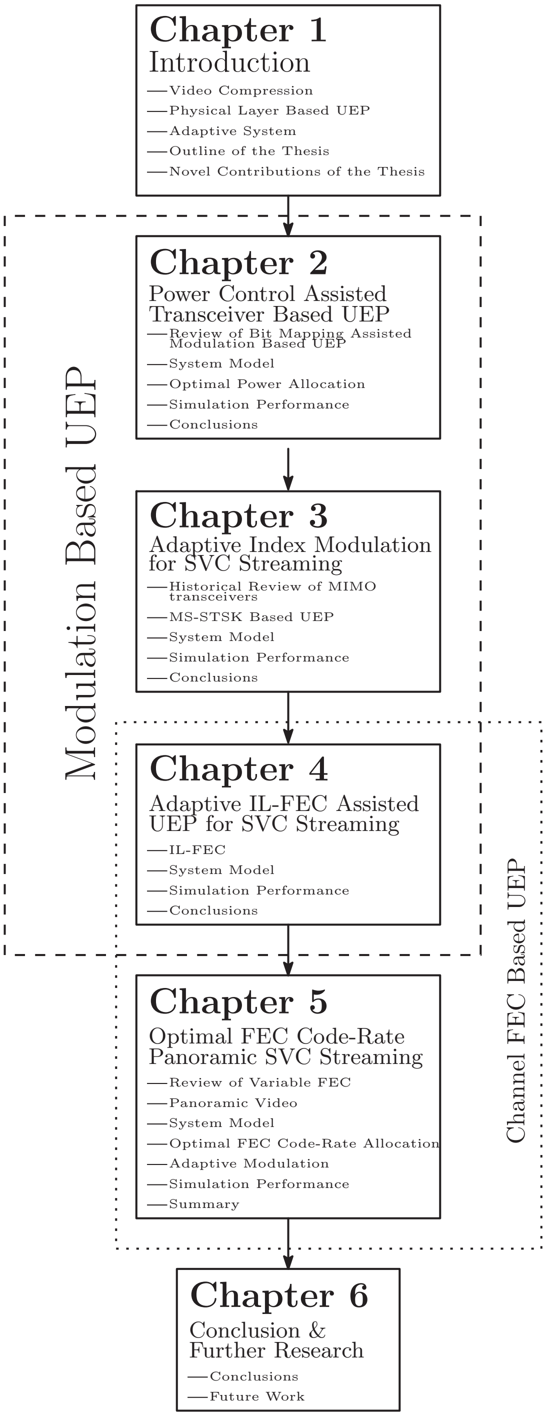


Figure 1.11: Organization of the thesis.

torical evolution of standards. Then, we categorized the physical layer based UEP schemes into channel FEC-coded and transceiver-based techniques. Finally, we briefly discussed the AQAM concept, which supports scalability.

Chapter 2: This chapter considers the family of bit mapping aided modulation based UEP schemes. In Section 2.2, we review the pair of typical bit mapping schemes employed for video transmission, namely HQAM and SC. The former usually maps the source bits to the different-sensitivity subchannels of a single modulation mode, while the latter forwards the source bits to the constellation bits of different modulation modes that are superimposed before transmission. Both schemes require optimizing the power allocation in order to improve the reconstructed video quality. Moreover, to illustrate the SC assisted system a little bit further, Section 2.3 presents the system models of a SC assisted scalable video broadcasting scheme, where the associated power assignment solution is discussed in Section 2.4. The superiority of our proposed SC assisted broadcasting scheme is characterized in Section 2.5, which provides the best quality coverage at the cost of the lowest demodulation complexity.

Chapter 3: This chapter extends the bit mapping aided modulation based UEP scheme of Chapter 2, where a new modulation based UEP scheme relying on Index Modulation (IM) is introduced, which is exemplified using Multi-Set Space-Time Shift Keying (STSK) (MS-STSK) transceiver. In Section 3.2, we review the history of MIMO schemes, including STBC and BLAST, Spatial Modulation (SM) [96], and discuss the recently conceived STSK [97] and the MS-STSK [98] transceivers. Section 3.3 characterizes inherent UEP capability of MS-STSK transceiver, where the iterative decoding is employed for the sake of achieving a near-capacity performance. Inspired by the results of Section 3.3, we propose an adaptive MS-STSK assisted UEP scheme for layered video streaming, where the video bits of the EL may have to be dropped in the low-capacity MS-STSK mode, when experiencing a dramatic SNR degradation due to fluctuating channel conditions. Specifically, we consider a high-mobility Unmanned Aerial Vehicle (UAV) scenario, where the associated Pilot Symbol Assisted Modulation (PSAM) aided channel estimation is specifically designed for the MS-STSK transceiver in order to overcome the fluctuating channel effects. This scheme is then compared to its Equal Error Protection (EEP) counterpart in Section 3.5.

Chapter 4: In this chapter, we focus our attention on the design of a joint transceiver and FEC based UEP scheme for SVC streaming, where, again, the bit mapping aided modulation based UEP of the MS-STSK is exploited as in Chapter 3, but we employ a new IL-FEC technique to provide FEC-based UEP. In Section 4.2, we first present the concept of conventional IL-FEC, which is only capable of improving the robustness of the BL, but not the ELs. We solve this challenge by conceiving the enhanced IL-FEC scheme of Section 4.2.2 to take into account the reference layer of the protected layers. Section 4.3 presents the proposed adaptive IL-FEC assisted MS-STSK aided system model of SVC streaming over fading channels, where the

IL-FEC techniques are adaptively switched according to the near-instantaneous channel SNR based on the pre-recorded SNR thresholds. The system's video performance is quantified, demonstrating the benefits of mode switching in terms of the QoE in Section 4.4. Explicitly, we compare the image quality and the Packet Loss Ratio (PLR) of the enhanced IL-FEC assisted system to those of its conventional IL-FEC assisted counterpart.

Chapter 5: Chapter 5 reports on an attractive variable FEC code-rate assisted UEP scheme specifically designed for scalable panoramic video streaming. Section 5.2 reviews the history of variable FEC code-rate assisted UEP schemes conceived for video streaming. Then, the concept of the panoramic video is introduced in Section 5.3, including the panoramic video testing system, the projection methods and the calculation of the Weighted-to-Spherically-uniform PSNR (WS-PSNR). Armed with these basic concepts, we then present our AQAM assisted optimal FEC code-rate aided system model in Section 5.4, where the associated optimal FEC code-rate assignment is presented. This code-rate assignment relies on the Evolutionary Algorithm (EA) of Section 5.5, the specific importance of the individual video frames, as well as of the mode switching operation of the system, as discussed in Section 5.6. The performance of the proposed system is quantified in Section 5.7, where we compare the optimal FEC code-rate assisted adaptive system to its fixed-mode based and fixed code-rate aided counterparts.

Chapter 6: Chapter 6 summarizes the main findings of the thesis and outlines a range of suggestions for further research.

1.5 Novel Contributions of the Thesis

The thesis is based on the journal papers [99–101] and their novel contributions are listed as follows.

1. We proposed the novel IM assisted UEP scheme for layered video streaming using the design example of an MS-STSK transceiver, where the different subchannels of the IM transceiver exhibit different BER performances. Specifically, the source bits exhibiting different importance are fed into the different-integrity subchannels.

2. Then we further exploit the UEP capability of the IM on a bit-by-bit basis, which is capable of providing a more flexible UEP mapping scheme for layered video streaming. The powerful tool of EXtrinsic Information Transfer (EXIT) charts is employed for characterizing the performance of IM having different configurations. This assists us in designing our bit-level based UEP scheme.

3. We conceive an adaptive IM assisted UEP system for SVC streaming in high-mobility UAV scenarios, where the most appropriate IM configuration is activated according to the near-instantaneous SNR experienced at different mobility levels for the sake of guaranteeing a

specific overall target performance. When the SNR is low, the EL may have to be dropped for the sake of avoiding a high QoE degradation.

4. We conceive an enhanced IL-FEC technique for the sake of providing protection to the EL of the scalable video, which additionally protects the bits of the reference layers relied upon by the IL-FEC protected layer. Then, we amalgamate our IM-based transceiver with this enhanced IL-FEC scheme for SVC streaming, where the index of IL-FEC protected layer varies according to the near-instantaneous channel SNR.

5. A SC assisted layered video broadcasting system is conceived, where in contrast to the conventional orthogonal multiplexing techniques, our power domain based SC superimposes multiple modulation modes conveying different video layers into a single constellation symbol. The optimal power allocation scheme is designed with the aid of a powerful global optimization algorithm, which assigns the most appropriate power for achieving the best possible video coverage.

6. Finally, our research culminates in exploiting our variable FEC code-rate technique for protecting the scalable panoramic video streams. Specifically, the video contents of the panoramic video are categorized into the three importance classes associated with the specific weighting coefficients assigned to them and then our global optimization algorithm is invoked for finding the globally optimal FEC code-rate assigned to the different panoramic video components.

Chapter 2

Superposition Coding Assisted SVC Streaming

2.1 Introduction

Recall from Section 1.2 that the physical layer based UEP schemes can be categorized according to their specific FEC and the modulation scheme. In this chapter, we introduce modulation based UEP, where the source bits having different importance are mapped to different-sensitivity bits of a symbol for the sake of enhancing the overall Successful Decoding Probability (SDP). Explicitly, in this chapter, we focus our attention on bit mapping assisted UEP using the conventional PSK/QAM constellations, which may be classified into HQAM and SC techniques. Specifically, the family of HQAM assisted schemes exploits the associated UEP capability by appropriately mapping the source bits to the constellation points of a single modulation mode, where the bits exhibit different Euclidean distances from the decision boundaries. By contrast, the SC assisted scheme superimposes the signals of multiple modulation modes, which is usually invoked in multi-user scenarios. It is worth noting that both schemes rely on using the optimal power sharing among the bits of different video layers for the sake of improving the robustness of video streaming. By contrast, inappropriate power assignment degrades the SDP of the video bits and hence reduces the quality of the reconstructed video.

The rest of this chapter is organized as follows. Section 2.2 briefly reviews the history of bit mapping assisted UEP schemes employed for video streaming, including both the HQAM and SC families. In Section 2.3, we propose an attractive SC assisted UEP scheme for layered video broadcasting, where the associated power sharing scheme is discussed in Section 2.4. The system performance is quantified in Section 2.5, which illustrates the QoE in terms of the reconstructed video quality as well as the Frame Error Rate (FER) and analyzes the system's complexity at the receiver. Finally, we conclude this chapter in Section 2.6.

2.2 Review of Bit Mapping Assisted Modulation Based UEP

In this section, we review the history of bit mapping aided modulation based UEP schemes designed for video streaming. Low complexity bit allocation algorithms have been widely employed for UEP aided video transmission over wireless channels [102].

2.2.1 Hierarchical QAM Aided UEP

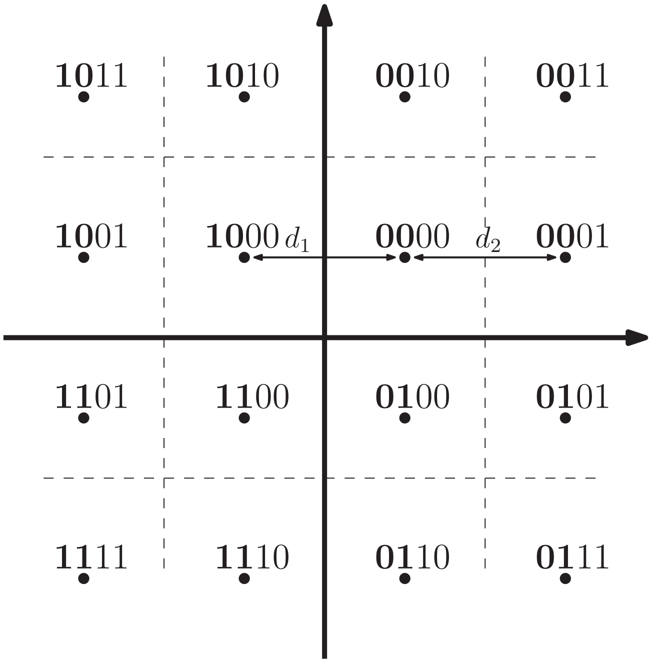


Figure 2.1: Hierarchical 16QAM constellation diagram. The bits in bold face are used to convey the more important bits, while the rest are used for the less important information.

Figure 2.1 depicts the constellation diagram of hierarchical 16QAM, where $d_1/2$ and $d_2/2$ represent the minimum distance from the decision boundaries represented by the dashed lines [103]. To achieve UEP, d_1 and d_2 are adjusted for ensuring that $d_1 > d_2$ [102]. By referring to Figure 2.1, let the Euclidean distance ratio, η, be given by:

$$\eta = \frac{d_1}{d_2}. \tag{2.1}$$

Suffice to note however that if $\eta \neq 1$, the average BER of the classic maximum-minimum-distance constellation is degraded.

Observe from Figure 2.1 that when $\eta = 1$, the signal constellation represents a conventional 16QAM, which also has two different BER subchannels, as detailed in [104]. By controlling η, the associated BER of the different-integrity subchannels can be adjusted.

2.2.2 Superposition Coding Aided UEP

In order to satisfy the ever-increasing data rate requirements of the users, a plethora of research efforts have been focused on improving the bandwidth efficiency, without degrading the user's experience. In a wireless system, broadcasting constitutes a bandwidth-efficient technique, which it allows the users requiring the same content to be served within the same bandwidth.

Additionally, Cover proved that in the broadcast channel the joint transmission rate can be achieved by superimposing a high-rate and a low-rate information stream instead of using time-sharing [66]. This investigation was further generalized by Bergmans [67] to Binary Symmetric Channels (BSC), which is now generally known as the SC concept. The essential aim of SC is to facilitate the transmission of several independent messages over a single wireless channel [77]. The superposition of several independent symbols is analogous to the vectorial addition of their signal constellation diagrams. Observe from Figure 2.2 that the superimposed signal is a vectorial sum of the two modulated signals, which achieves an identical capacity of the sum of both modulation schemes, as represented by:

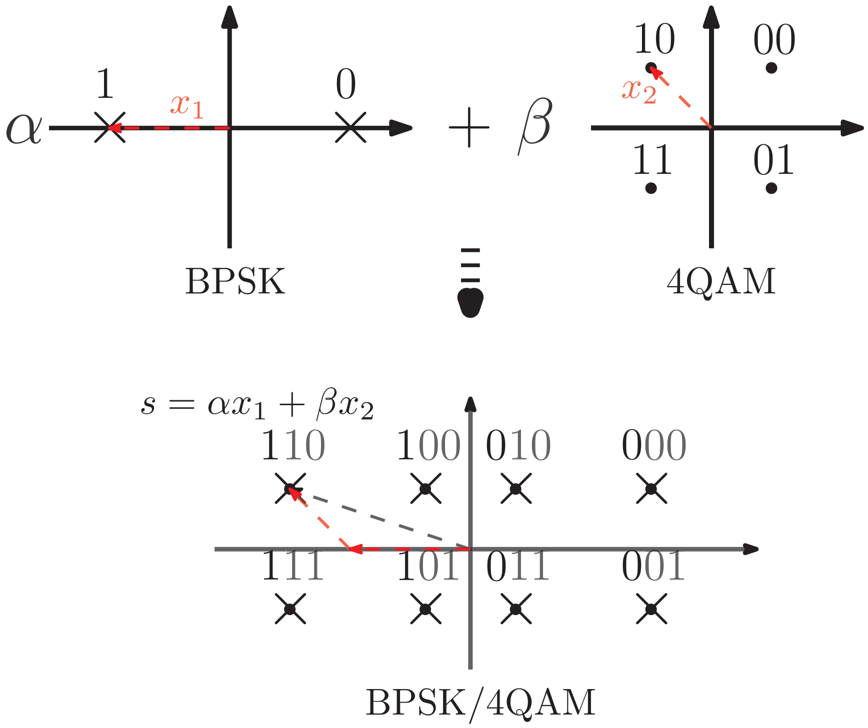


Figure 2.2: SC of a 3-bit symbol block to one of the eight constellation symbols using superimposed BPSK and 4QAM constellations.

$$s = \alpha x_1 + \beta x_2, \tag{2.2}$$

where α and β are a pair of scaling factors, subject to the power constraint of:

$$\alpha^2 + \beta^2 = 1. \tag{2.3}$$

In the example of Figure 2.2, the superimposed constellation point s consists of symbol "1" from BPSK and symbol "10" from 4QAM, which is transmitted as a single wireless transmission block.

Specifically, observe from (2.2) that the position of superimposed constellation point is also determined by the scaling factors. Hence, the associated optimal Maximum Likelihood (ML) and Maximum *a posteriori* (MAP) algorithms used for demodulating the SC signals have to be particularly designed, relying on the knowledge of these factors. Additionally, since the superimposed signals are a summation of the different modulated symbols, an alternative demodulation solution is to progressively peel off the superimposed signals of different modulation modes, which is generally known as Successive Interference Cancellation (SIC) [105]. It is worth noting that the SIC procedure requires the receiver to employ additional re-encoding and re-modulation processes in order to remove the corresponding remodulated signals from the received superimposed signals. Advantage of this SIC process is however that the errors will be propagated to the signals of remaining modulation modes, hence degrading their associated SDP. Fortunately, the development of the near-capacity channel coding, such as TC [106] and LDPC [107], significantly improves the BER performance, even if there is interference imposed by the superimposed signals, which essentially prevents the error propagation and thereby enhances the performance of SC.

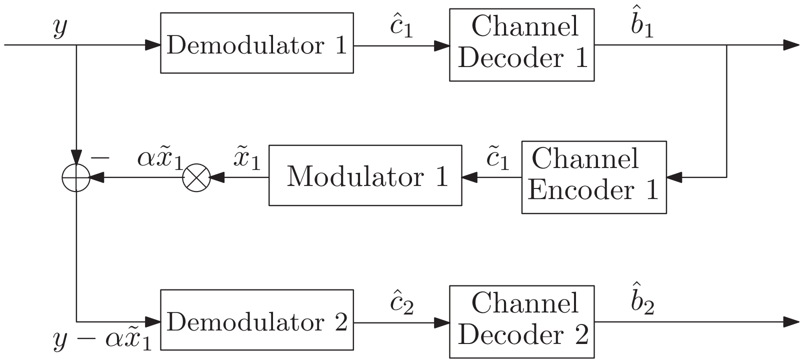


Figure 2.3: The schematic of the SIC decoder used for decoding the superimposed signals of Figure 2.2.

Figure 2.3 depicts the schematic of the SIC employed to decode the superimposed signals of Figure 2.2. In Figure 2.3, the received signal y is given by:

$$y = s \cdot h + n, \tag{2.4}$$

where h and n represent the fading channel coefficient and the Additive White Gaussian Nosie (AWGN), respectively. It can be seen from Figure 2.3 that the received signal y is first demodulated by Demodulator 1, yielding the demodulated version of channel coded bit, namely $\hat{c}_1$, which is then fed into Channel Decoder 1 for generating the decoded source bit $\hat{b}_1$. Then, a copy of $\hat{b}_1$ is forwarded to user 1, while the other is used to recover x_1, as shown in Figure 2.3. Specifically, a copy of $\hat{b}_1$ is re-encoded with the aid of Channel Encoder 1 that is identical to that in the transmitter. The re-encoded bit $\tilde{c}_1$ is then re-modulated by Modulator 1, which is scaled by the assigned coefficient, namely α, and subtracted from the received superimposed signal y for the sake of recovering the source bit $\hat{b}_2$. Figure 2.3 exemplifies a simple two-user SC scenario. For more complex scenarios, the iterative decoding will continue until the signal assigned to the lowest power coefficient is decoded.

The examples shown in Figures 2.2 and 2.3 are mainly applicable to multicast and broadcast scenarios. Specifically, for the multicast scenario, the information to be delivered tp different users are superimposed before transmission and transmitted as a single composite signal. Compared to the conventional Orthogonal Multiple Access (OMA) schemes exploiting the orthogonality in terms of the space-, time- and/or bandwidth-domain, the concept of Non-Orthogonal Multiple Access (NOMA) has been selected as a study item in the 5G Long-Term Evolution (LTE) release 13 standard body. It was termed as Multi-User Superposition Transmission (MUST) [108]. As a result, to extract the user's own information, the SIC scheme of Figure 2.3 has to be invoked for iteratively removing the undesired information pertaining to other users, until the desired message is extracted. Another application scenario is found in the broadcast systems defined by the Advanced Television System Committee (ATSC) 3.0 [109]. In this context, the authors of [78] designed a Layered Division Multiplexing (LDM) assisted SHVC broadcast system, where the bits of the BL and ELs are superimposed and broadcast simultaneously. In the broadcast tower the SNR tends to be high, hence both the BL and EL can be detected without errors.

In the remaining sections of this chapter, we mainly focus on the design of an SHVC aided SC assisted broadcast system, where the bits of various video layers mapped to different modulation modes are SC encoded before their transmission. Furthermore, to benchmark our system, we also consider both its conventional single-layered QAM assisted and its SHVC aided Time Division Multiplexing (TDM) assisted counterparts.

2.3 System Model

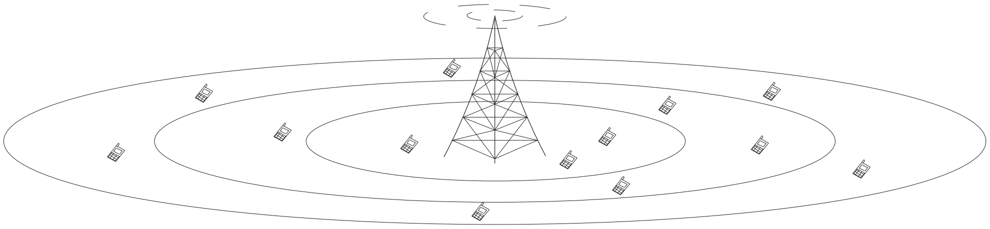


Figure 2.4: Video broadcast system.

Without loss of generality, we mainly consider three video systems designed for video broadcasting scenario, as depicted in Figure 2.4, namely the QAM-HEVC, TDM-SHVC and SC-SHVC arrangements. Specifically, the QAM-HEVC scheme employs a conventional modulation mode to broadcast single-layer video without any additional measures, where the HEVC encoded bit stream is simply FEC encoded and transmitted using conventional QAM. By contrast, in the TDM-SHVC scheme, numerous modulation modes are employed for conveying the video bits of different SHVC layers, which are activated successively. In any time slot only a single modulation mode is activated. By contrast, in the TDM-SHVC scheme, all modulation modes are transmitted using the same transmit power and each modulation mode only transmits the bits of a specific video layer. This scheme has the feasibility of transmit different constellation in different time slots. Finally, similar to its TDM-SHVC counterpart, the SC-SHVC scheme can also employ different modulation modes for conveying the video bits of different SHVC layers. However, instead of allocating different time slots to users, these modulation modes are separated in the power domain, where all video bits of the different layers are transmitted simultaneously and distinguished by the power coefficients assigned. In this section, we briefly discuss the system model of the TDM-SHVC and of the SC-SHVC system.

2.3.1 TDM-SHVC

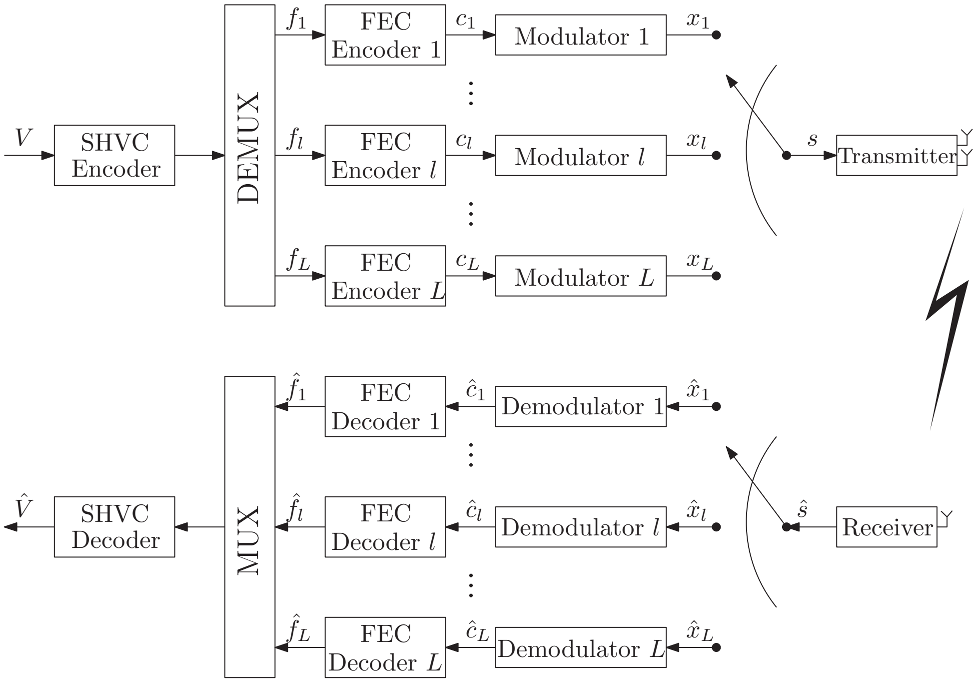


Figure 2.5: The system model of TDM-SHVC scheme.

The associated TDM-SHVC assisted broadcast system is depicted in Figure 2.5, where the captured video source V is compressed by the SHVC encoder, generating a bit stream that contains multiple video layers. Its output bit stream is then demultiplexed into L video bit streams, each for a specific transmission layer, from f_1 to f_L corresponding to the different video layers, which are then respectively encoded by L identical FEC encoders, as shown in Figure 2.5. The FEC encoded bit streams, denoted by c_l, are then fed to the corresponding modulators to generate the modulated signals x_l, where only a single signal stream can be transmitted by the transmitter at a time. Specifically, the selected signals s conveyed by the transmitter can be formulated by:

$$s = \begin{cases} x_1, & t \leq 1/T, \\ x_l, & (l-1)/T < t \leq l/T, \\ x_L, & (L-1)/T < t \leq T, \end{cases} \tag{2.5}$$

where T represents the transmission period required to transmit all the L bit streams. Moreover, before transmission, the selected bit stream s at t is weighted by the transmit power of $\sqrt{P}$.

The receiver of the TDM-SHVC system is also depicted at the bottom of Figure 2.5, where the received signals $\hat{s}$ are forwarded to the corresponding decoder, and $\hat{x}_l$ can be obtained according to (2.5). The demodulated signals $\hat{c}_l$ are then fed to the corresponding FEC decoder, as shown in Figure 2.5, yielding the decoded video source bits of video layer f_l. After T time slots, the multiplexer will rearrange the bits of the layers spanning from f_1 to f_L and convey them to the SHVC decoder for the users' observation.

2.3.2 SC-SHVC

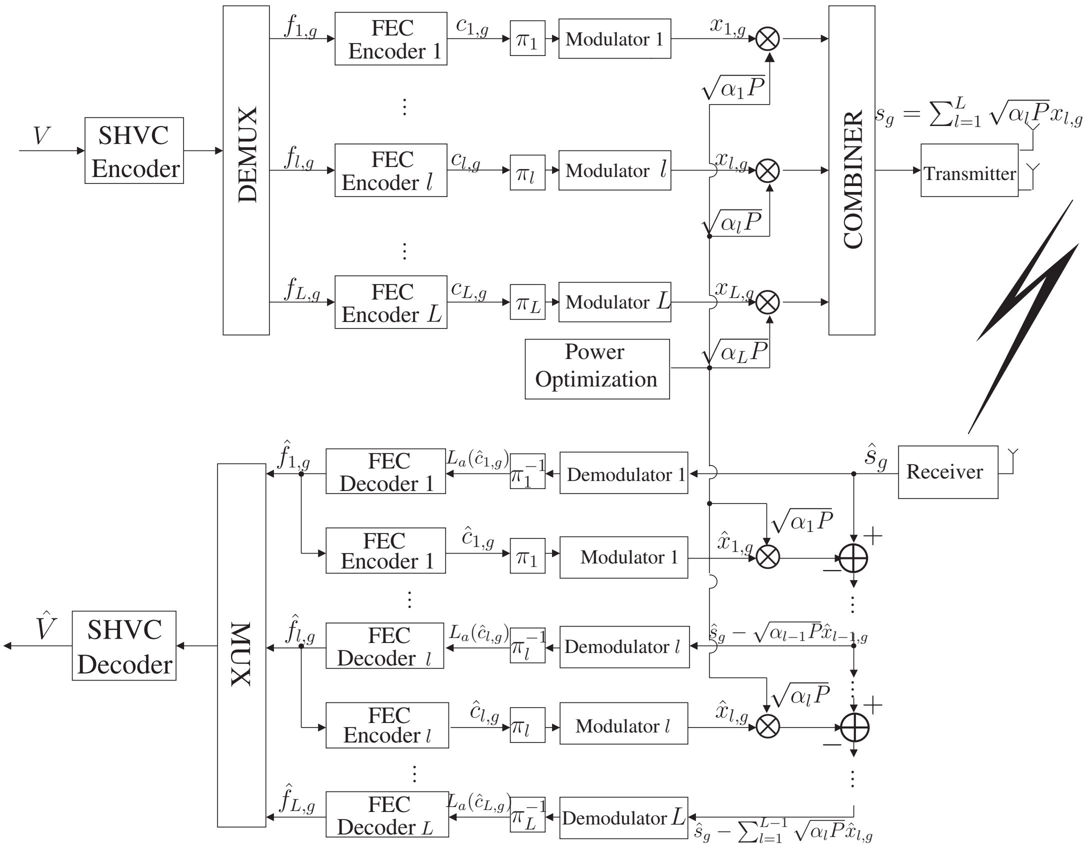


Figure 2.6: The transmitter and receiver model.

In this section we highlight our SC assisted system designed for layered video broadcasting, as shown in Figure 2.6, which depicts both the transmitter and the receiver models. Specifically, the bits of the various video layers are FEC encoded and separately modulated using arbitrary modulation modes, where the modulated symbols are then scaled by associated power scaling coefficients for maximizing the average reconstructed video quality across the Base Station (BS)'s coverage area.

2.3.2.1 Transmitter

As shown in Figure 2.6, the captured video sequence V is encoded using the SHVC encoder that compresses the video clips into L layers $\{f_{l,g}\}_{l=1}^{L}$, where $f_{l,g}$ represents the g-th video frame of the l-th layer. Then, these bit streams are first encoded by their corresponding FEC encoders, which generate the encoded bit streams of $\{c_{l,g}\}_{l=1}^{L}$. The encoded bit streams are then fed into L interleavers $\{\pi_l\}_{l=1}^{L}$, as shown in Figure 2.6, before mapping them to their associated modulators, where L modulated signals $\{x_{l,g}\}_{l=1}^{L}$ are generated. Afterwards, the modulated symbols are scaled by the optimal power coefficients of $\{\alpha_l\}_{l=1}^{L}$, which are also known at the receiver. Finally, these signals are superimposed according to $s_g = \sum_{l=1}^{L} \sqrt{\alpha_l P} x_{l,g}$, before transmission from multiple antennas.

2.3.2.2 Receiver

The receiver model is shown in Figure 2.6, where the SIC technique is invoked for progressively peeling off the superimposed signals of different layers. The received signal $\hat{s}_g$ is first demodulated by Demodulator 1 of Figure 2.6, while the signals of the other layers are treated as interference. Following deinterleaving, the deinterleaved LLR $L_a(\hat{c}_{1,g})$ are then fed into the FEC Decoder 1, generating the decoded version of the BL stream, $\hat{f}_{1,g}$, which can be fed into the video decoder for reconstructing the BL. In order to recover the ELs, we have to remove the interference imposed by the BL and hence a re-encoding process is invoked for reconstructing $x_{1,g}$. Hence we have regenerated the modulated signal $\hat{x}_{1,g}$, as shown in Figure 2.6, where $\hat{x}_{1,g}$ is expected to be identical to $x_{1,g}$ if $\hat{f}_{1,g}$ is successfully decoded. Following power-scaling by $\sqrt{\alpha_1 P}$, the BL signal can be removed from the superimposed signal, and then the receive proceeds with decoding the first EL, which is similar to that of decoding the BL by treating the bits of $[l+1, L]$ layers as interference. The SIC procedure continues until the highest EL $\hat{f}_{L,g}$ is decoded. The decoded video layers $\{\hat{f}_{l,g}\}_{l=1}^{L}$ are then assembled into the SHVC bit stream by the MUX block of Figure 2.6 to reconstruct the decoded video $\hat{V}$.

2.4 Power Optimization

The modulated and scaled signals are superimposed before their transmission, as shown in Figure 2.6, which are uniquely and unambiguously distinguished by the power assigned. This section is focused on the "Power Optimization" block of Figure 2.6, which aims for finding the optimal power allocation scheme for the sake of maximizing the average video quality across the service coverage area of the BS.

2.4.1 Problem Formulation

The problem is formulated for a layered video based open-loop broadcast system operating without requiring any knowledge of the CSI. In this section, we design an Objective Function (OF) that empirically characterizes the relationship between the power assigned to each layer and the average video quality across the BS's coverage area. Having a total transmit power of P, the average received video quality across a GoP period of G can be modelled as:

$$\max \quad \mathbb{E}[Q(P,\boldsymbol{\alpha})], \tag{2.6}$$

$$s.t.: \quad \sum_{l=1}^{L} \alpha_i \leq 1, \tag{2.7}$$

where (2.7) represents the constraint on the power consumption, where $\boldsymbol{\alpha} = \{\alpha_1, \cdots, \alpha_l, \cdots, \alpha_L\}$ indicates the power scaling coefficient set of all the layers. Assuming that the users are uniformly distributed within the maximum cell radius of S covered by the BS, the probability density function of their distances s from the BS can be expressed as $f_s(s) = 2s/S^2$ [110]. The average video quality $\mathbb{E}[Q(P,\boldsymbol{\alpha})]$ can hence be formulated as:

$$\mathbb{E}[Q(P,\boldsymbol{\alpha})] = \int_0^S f_s(s) Q(P,\boldsymbol{\alpha},s)\, \mathrm{d}s, \tag{2.8}$$

where $Q(P,\boldsymbol{\alpha},s)$ is denoted as the video quality at the arbitrary distance $s \in [0,S]$ away from the BS. Additionally, we model the video quality experienced by taking into account the video frame length and its associated SDP, which is expressed by:

$$Q(P,\boldsymbol{\alpha},s) = \sum_{l=1}^{L} \sum_{g=1}^{G} \Delta q_{l,g} \Psi_{l,g}(P,\boldsymbol{\alpha},s), \tag{2.9}$$

where $\Delta q_{l,g}$ represents the video quality improvement experienced by the g-th frame upon receiving the l-th layer, which can be expressed as:

$$\Delta q_{l,g} = \begin{cases} q_{l,g}, & l = 1, \\ q_{l,g} - q_{l-1,g}, & 1 < l \leq L. \end{cases} \tag{2.10}$$

Furthermore, $\Psi_{l,g}(P,\boldsymbol{\alpha},s)$ represents the SDP of the video frame $f_{l,g}$ at distance s, which, in addition to considering its own SDP $\psi_{l,g}$, takes into account the SDP of all its dependent frames, namely $\{\psi_{1,1}, \ldots, \psi_{m,n}, \ldots, \psi_{l,g}\}$. Explicitly, the SDP of $\Psi_{l,g}(P,\boldsymbol{\alpha},s)$ can be expressed as:

$$\Psi_{l,g}(P,\boldsymbol{\alpha},s) = \prod_{m=1}^{l} \prod_{n=1}^{g} \psi_{m,n}(P,\boldsymbol{\alpha},s). \tag{2.11}$$

where $\psi_{m,n}(P,\boldsymbol{\alpha},s)$ denotes the SDP of the frame $f_{m,n}$ at distance s, which will be discussed in the next section.

2.4.2 Successful Video Frame Decoding Probability

In this subsection, we elaborate on modelling the SDP of the variable-length video frame for different modulation modes. It has been shown in [32] that the video FER imposed by the fading channel is independent of the packet length when the distribution of errors may be considered to be uniform. Conditioned on the power constraint P and on the power allocation coefficient set $\boldsymbol{\alpha}$, for the l-th video layer, the PLR of a packet containing λ bits at a specific distance of s is $p_l(P,\boldsymbol{\alpha},s)$. Then the successful decoding probability of video frame $f_{l,g}$ having an arbitrary bit length of $|f_{l,g}|$ can be expressed as:

$$\psi_{l,g}(P,\boldsymbol{\alpha},s) = \Big[1 - p_l(P,\boldsymbol{\alpha},s)\Big]^{\frac{|f_{l,g}|}{\lambda}}. \tag{2.12}$$

As shown in Figure 2.6 the signals of all the layers are superimposed before transmission and $p_l(P,\boldsymbol{\alpha},s)$ can usually be decomposed into: 1) the errors propagated from its dependent layers and, 2) its own errors. The authors of [111] demonstrated that in TC assisted SC broadcasting systems the PLR of the EL can be approximated by the sum of the PLR of the individual ELs assuming perfect SIC plus the PLR of its dependent layers. Let us denote the PLR of the l-th layer by $\tilde{p}_l(P,\boldsymbol{\alpha},s)$, when its dependent layers are assumed to be perfectly decoded and hence no error propagation from the lower layers is considered. The PLR of $p_l(P,\boldsymbol{\alpha},s)$ can thus be modelled as:

$$p_l(P,\boldsymbol{\alpha},s) \approx \begin{cases} \tilde{p}_1(P,\boldsymbol{\alpha},s), & l = 1, \\ \min\left(\sum\limits_{m=1}^{l} \tilde{p}_m(P,\boldsymbol{\alpha},s), 1\right), & 1 < l \leq L, \end{cases} \tag{2.13}$$

where $\tilde{p}_l(P,\boldsymbol{\alpha},s)$ is jointly determined by the received SNR of the l-th signal $\gamma_l(s)$ and the received SNR of the other superimposed signals $\kappa_l(s)$, which are respectively given by:

$$\gamma_l(s) = 10\log_{10} P\alpha_l - \beta(s) - N, \tag{2.14}$$

$$\kappa_l(s) = 10\log_{10} P \sum_{m=l+1}^{L} \alpha_m - \beta(s) - N, \tag{2.15}$$

where N represents the Gaussian noise power on a deciBel (dB) scale. Moreover, $\beta(s)$ represents the expected Path Loss (PL) at s, which is given by:

$$\beta(s) = \overline{\beta}(s) + \chi, \tag{2.16}$$

where $\overline{\beta}(s)$ is the PL given by the COST Hata model, while χ represents the slow fading margin. More explicitly, having a standard deviation of shadow fading σ, in order to compensate 90% shadowing attenuation, χ is set to about 1.3σ [112].

Furthermore, for the sake of modelling the PLR of various SNRs of γ and κ, we first pre-record the PLR with the associated parameters for each layer in a LookUp Table (LUT) using Monte Carlo simulations. More particularly, a pair of LUTs are established, namely $h_L[\gamma]$ and $\{h_l[\gamma,\kappa]\}_{l=1}^{L-1}$, where the former is used for characterizing the PLR of the highest-index EL l_L having no additional interference, while the latter is employed for that of the remaining layers.

Table 2.1: Parameters values of (5.24) for each modulation mode.

M	a_3	a_2	a_1	b_3	b_2	b_1	c_3	c_2	c_1	d
BPSK	0.004684	-0.06528	1.489	-0.003228	-0.07807	-0.8248	-0.006964	0.00579	0.1375	4.646
4QAM	0.009906	-0.3206	4.428	-0.005383	-0.2138	-3.056	-0.0197	0.01607	0.4618	19.61
8PSK	-	-	-	-	-	-	-	-0.01375	1.009	7.321

Alternatively, the LUT can be represented by the empirical model of

$$\log_{10}(h_l[\gamma,\kappa]) \approx \frac{-5}{1+e^{-(a_3\gamma^3+a_2\gamma^2+a_1\gamma+b_3\kappa^3+b_2\kappa^2+b_1\kappa+c_3\gamma^2\kappa+c_2\gamma\kappa^2+c_1\gamma\kappa)+d}}, \tag{2.17}$$

$$\log_{10}(h_L[\gamma]) \approx \frac{-5}{1+e^{-(c_2\gamma^2+c_1\gamma)+d}}, \tag{2.18}$$

where the parameters values are given in Table 2.1. Figure 2.7 illustrates our comparison between the LUT-based and the mathematically fitted empirical model of $\log_{10}(h_l[\gamma,\kappa])$ exemplified using 4QAM. Then, we may solve $\tilde{p}_l(P,\boldsymbol{\alpha},s)$ with the aid of (2.17) as follows:

$$\tilde{p}_l(P,\boldsymbol{\alpha},s) = \begin{cases} h_l[\gamma,\kappa], & 1 \le l < L \\ h_L[\gamma], & l = L. \end{cases} \tag{2.19}$$

Then, assisted by (2.12)-(2.19), the SDP of the video frame $f_{l,g}$ can be expressed as

$$\psi_{l,g}(P,\boldsymbol{\alpha},s) = \begin{cases} \left[1-\min\Big(\sum_{m=1}^{l} h_m[\gamma_m(s),\kappa_m(s)],1\Big)\right]^{\frac{|f_{l,g}|}{\lambda}}, & 1 \le l < L, \\ \left[1-\min\Big(\sum_{m=1}^{L-1} h_m[\gamma_m(s),\kappa_m(s)]+h_L[\gamma_L(s)],1\Big)\right]^{\frac{|f_{l,g}|}{\lambda}}, & l = L. \end{cases} \tag{2.20}$$

To elaborate a little further, (2.20) establishes the connection between the transmit power, scaling coefficients and the FER for the variable-length video frames at an arbitrary distance s away from the BS, which will be exploited for finding the optimal power scaling factors, as detailed in the next section.

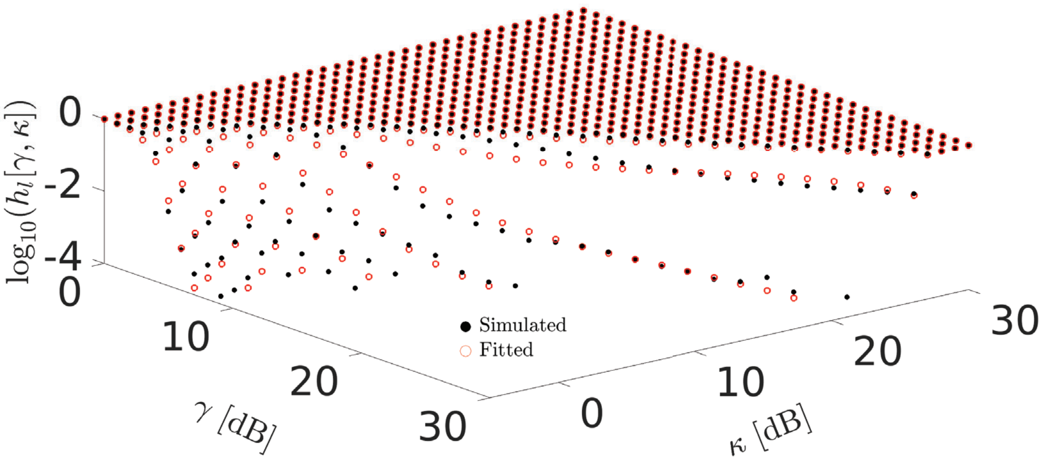


Figure 2.7: LUT values versus their fitted mathematically modeled counterparts for $h_l[\gamma,\kappa]$.

2.4.3 Power Optimization

Armed with the SDP model derived in Section 2.4.2, we can rewrite the received video quality function of (2.7) in discrete form as the summation of the video quality at various distances, as follows:

$$\begin{aligned}\mathbb{E}[Q(P,\boldsymbol{\alpha})] &= \frac{2}{S^2G}\sum_{s=0}^{S}\sum_{l=1}^{L}\sum_{g=1}^{G}\Delta q_{l,g}\Psi_{l,g}(P,\boldsymbol{\alpha},s) \\ &= \frac{2}{S^2G}\sum_{s=0}^{S}\sum_{l=1}^{L}\sum_{g=1}^{G}\Delta q_{l,g}\prod_{m=1}^{l}\prod_{n=1}^{g}\psi_{m,n}(P,\boldsymbol{\alpha},s),\end{aligned} \tag{2.21}$$

where $\psi_{m,n}(P,\boldsymbol{\alpha},s)$ is derived in Equation (2.20). The optimal power scaling coefficient set can then be found by finding the solution of the problem:

$$\{\boldsymbol{\alpha}\}_{opt} = \underset{\boldsymbol{\alpha}=\{\alpha_l\}_{l=1}^{L}\in[0,1]}{\arg\max}\ \mathbb{E}[Q(P,\boldsymbol{\alpha})]. \tag{2.22}$$

Since the values of the power scaling coefficients are continuous, using an exhaustive search becomes infeasible. Instead, we employ EA a ssisted optimization for finding the optimal power scaling coefficients, as described in Algorithm 1. In the mutation subsection of Algorithm 1, $\hat{\boldsymbol{\alpha}}_{i,p_s,best}$ is randomly selected from the population archive, while c_1 and c_2 constitute a pair of random integers fetched from the set $\{1,2,\cdots,p_s-1,p_s+1,\cdots,P_s\}$. Furthermore, μ_{C_r} and μ_λ in the adaptivity subsection of Algorithm 1 are adaptively updated using the arithmetic-mean $\overline{S^A_{C_r}}$ and Lehmer-mean $\overline{S^L_\lambda}$ operators [113], respectively.

2.5 Simulation Results

In this section, we characterize the performance of our proposed optimal power assisted SC system designed for SC-SHVC video streaming. To benchmark our system, we consider both the TDM and the conventional QAM assisted systems. Specifically, both the proposed and the TDM assisted systems, also termed as TDM-SHVC, are used for conveying SVC streams, while the single-layer coded video bits are transmitted by the conventional QAM counterpart, namely by the QAM-HEVC arrangement.

A 32-frame *Basketball* video clip of 4:2:0 YUV format containing 1280 × 720 pixels is used in our simulations, as listed in Table 2.2. In the SHVC scenario, the video clip is encoded using the SHVC scheme into three video layers. The corresponding bitrates are 500 kbps, 1 Mbps and 1.5 Mbps, respectively, resulting in the PSNR values of 32.3 dB, 36.4 dB and 38.3 dB, respectively. In the single-layer HEVC scenario, the video clip is compressed into a 3 Mbps bit stream having a PSNR of 40.5 dB, which exhibits a better PSNR than its SHVC counterpart, as justified in [21]. Furthermore, the GoP is set to 8 in both cases, where the bidirectional predictive frames are deactivated, since they are prone to propagating inter-frame video distortions as well as to imposing extra

Algorithm 1 Power Allocation Algorithm

```
$i \leftarrow 1$                                              ▷ Initialization
$\mu_{C_r} \leftarrow 0.5$
$\mu_{\lambda} \leftarrow 0.5$
$C_{r_{p_s}} \leftarrow randn(\mu_{C_r}, 0.1)$                ▷ Gaussian distribution
$\lambda_{p_s} \leftarrow rand\_cauchy(\mu_{\lambda}, 0.1)$   ▷ Cauchy distribution
randomly generate $P_s$ initial power sets $\{\hat{\boldsymbol{\alpha}}_{i,p_s}\}_{p_s=1}^{Ps}$, which obey uniform distribution of interval [0,1].
for $p_s = 1 : P_s$ do                                        ▷ Satisfy (2.7)
    $\boldsymbol{\alpha}_{i,p_s} \leftarrow \boldsymbol{\alpha}_{g,p_s} / \textbf{sum}(\boldsymbol{\alpha}_{i,p_s})$
    $\hat{\boldsymbol{\alpha}}_{i,p_s} \leftarrow \textbf{sort}(\boldsymbol{\alpha}_{i,p_s})$
while $i \leq i_{\max}$ do
    for $p_s = 1 : P_s$ do                                    ▷ Mutation
        $\tilde{\boldsymbol{\alpha}}_{i,p_s} \leftarrow \hat{\boldsymbol{\alpha}}_{i,p_s} + \lambda_{p_s}(\hat{\boldsymbol{\alpha}}_{i,p_s,best} - \hat{\boldsymbol{\alpha}}_{i,p_s})$
                $+ \lambda_{p_s}(\hat{\boldsymbol{\alpha}}_{i,p_s,c_1} - \hat{\boldsymbol{\alpha}}_{i,p_s,c_2})$
    for $p_s = 1 : P_s$ do                                    ▷ Crossover
        for $l = 1 : L$ do
            if $randu(0,1) \leq C_{r_{p_s}}$ then
                $\check{\alpha}_{i,p_s}^{l} \leftarrow \tilde{\alpha}_{i,p_s}^{l}$
            else
                $\check{\alpha}_{i,p_s}^{l} \leftarrow \hat{\alpha}_{i,p_s}^{l}$
    for $p_s = 1 : P_s$ do                                    ▷ Selection
        if $\mathbb{E}[Q(P, \check{\boldsymbol{\alpha}}_{i,p_s})] \leq \mathbb{E}[Q(P, \hat{\boldsymbol{\alpha}}_{i,p_s})]$ then
            $\hat{\boldsymbol{\alpha}}_{i+1,p_s} \leftarrow \hat{\boldsymbol{\alpha}}_{i,p_s}$
        else
            $\hat{\boldsymbol{\alpha}}_{i+1,p_s} \leftarrow \check{\boldsymbol{\alpha}}_{i,p_s}$
    $\mu_{C_r} = 0.5\mu_{C_r} + 0.5\overline{S_{C_r}^{A}}$    ▷ Adaptation
    $\mu_{\lambda} = 0.5\mu_{\lambda} + 0.5\overline{S_{\lambda}^{L}}$
```

Table 2.2: The parameters employed for the three schemes.

	SC-SHVC	TDM-SHVC	QAM-HEVC
Representation	YUV 4:2:0	YUV 4:2:0	YUV 4:2:0
Bits Per Pixel	8	8	8
GoP	8	8	8
FPS	30	30	30
No. of Frames	32	32	32
Resolution	1280×720	1280×720	1280×720
Video Bitrates	0.5/1/1.5 Mbps	0.5/1/1.5 Mbps	3 Mbps
Y-PSNR	32.3/36.4/38.3 dB	32.3/36.4/38.3 dB	40.5 dB
Channel Coding	TC[7 5]	TC[7 5]	TC[7 5]
Modulation	BPSK/4QAM/8PSK	8PSK/64QAM/512QAM	64QAM

Table 2.3: The channel parameters used in simulations.

Transceiver	STBC 2×1
PL Model	COST Hata
Shadow Fading	8 dB
Transmit Power	40 dBm
BS Height	50 m
UE Height	1.5 m
Bandwidth	1 MHz
BS Coverage	1000 m
Carrier Frequency	2 GHz
Simulation Repeated	100

decoding latency. Furthermore, we employ near-capacity TC having a pair of identical RSC components, each having the octally represented generator polynomials of [7 5]. The resultant $1/3$-rate code is then punctured using the puncturing matrix of [1 1;0 1;1 0] to half-rate.

Moreover, the channel configuration employed for all the three schemes is illustrated in Table 2.3, where we assume that a total transmit power of 40 dBm (10W) is transmitted by a 2×1 STBC[1] scheme over a 1 MHz-bandwidth wireless channel using a 2 GHz carrier frequency for broadcasting the video signals over a maximum coverage radius of 1000 m. The PL is represented by the COST Hata model, where the BS and User Equipment (UE) heights are set to 50 m and 1.5 m. The shadow fading is modelled by the classic log-normal distribution having a standard deviation of 8 dB. Additionally, the population size of P_s and the maximum number of iterations $i_{\max}$

---

[1] We exemplify our system using the simple STBC scheme, which can be readily extended to arbitrary transceiver schemes.

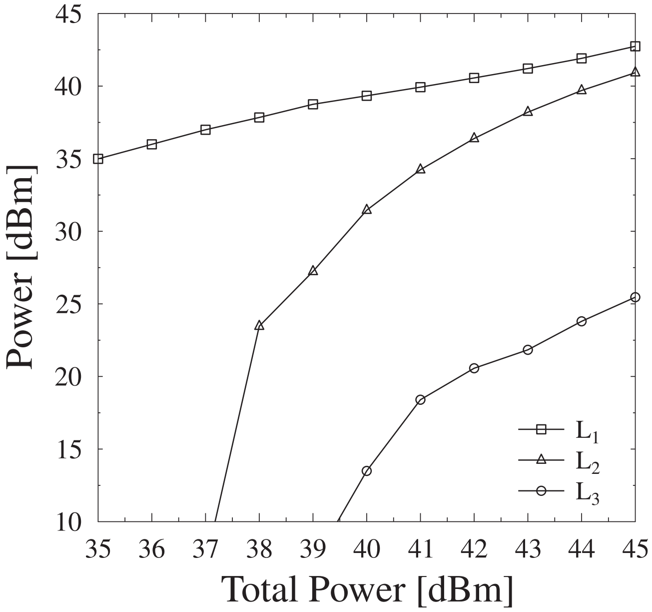


Figure 2.8: The power assigned to the three video layers at different overall transmit power.

of Section 2.4.3 are set to 10 and 30, respectively. Moreover, in the proposed SC-SHVC scheme of Figure 2.6, the modulation modes used for conveying the bits of the three layers are set to BPSK, 4QAM and 8PSK. Furthermore, to provide the same capacity, the TDM-SHVC scheme employs 8PSK, 64QAM and 512QAM for conveying the bits of the three layers, while the QAM-HEVC scheme activates 64QAM. Finally, the simulations are repeated 100 times.

Figure 2.8 shows the optimal power assignment of the L_1, L_2 and L_3 versus the overall transmit power, where the BL accounts for the majority of the overall transmit power to guarantee the received video quality across the BS coverage.

2.5.1 Quality of Experience

The image quality versus distance is depicted in Figure 2.9, which shows that the QAM-HEVC provides the best video quality, when the distance is less than 600 m from the BS, while the two SHVC assisted schemes show a better received quality for the cell-edge users. Additionally, both SHVC schemes experience a similar image quality degradation, when the UE is within 600 m, but the optimal power assisted SC-SHVC outperforms its TDM-SHVC counterpart, especially at the coverage edge of the BS, as shown in Figure 2.9.

Figure 2.10 compares the FER of the three associated schemes, where the FER of the BL of the two SHVC assisted schemes is presented. It can be readily seen in Figure 2.10 that the optimal

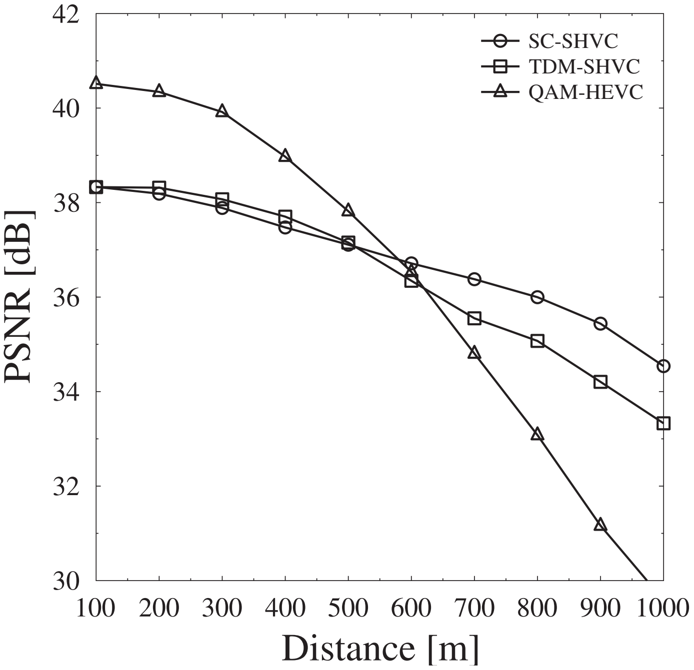


Figure 2.9: PSNR [dB] versus the distance away from the BS between three schemes.

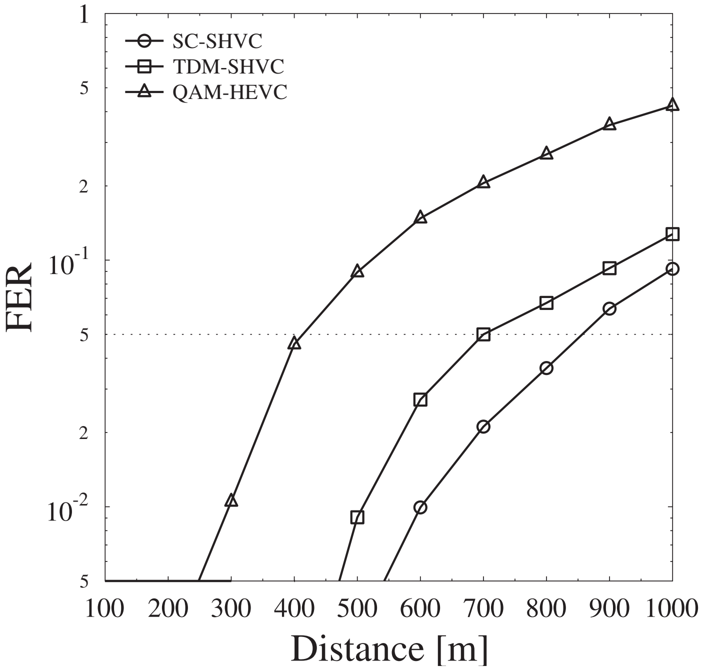


Figure 2.10: FER of the BL versus the distance away from the BS between the three schemes.

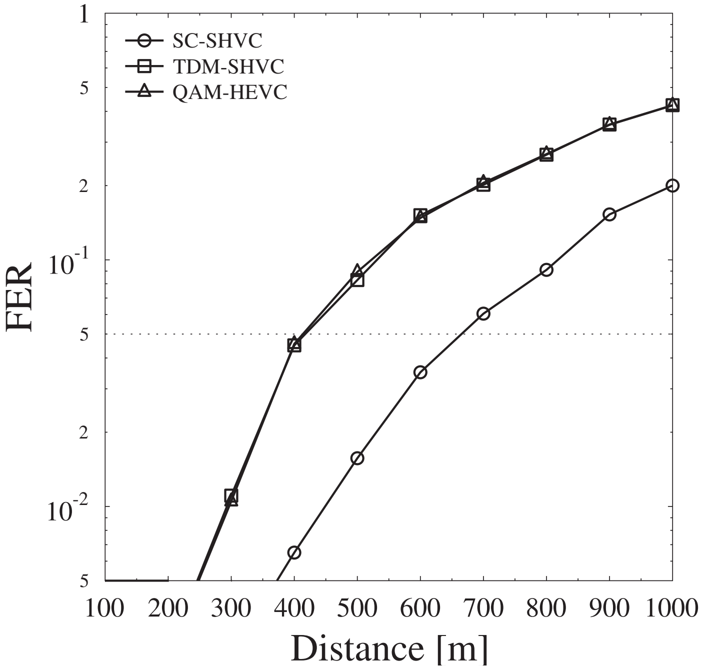


Figure 2.11: FER of the first EL versus the distance away from the BS between the three schemes.

power scaling assisted SC-SHVC scheme guarantees a FER below 5% upto the coverage radius of about 850 m. By contrast, its TDM-SHVC and QAM-HEVC counterparts only cover about 700 m and 400 m coverage radii, when guaranteeing a FER of 5%. Explicitly, our scheme exhibits an improved video experience for the cell-edge users. Moreover, it is worth noting that the proposed optimal SC-SHVC scheme exhibits a FER of approximately 10% at the cell-edge, which is in line with the design objective of (2.16). Figure 2.11 additionally compares the FER of the first EL for the three schemes, where an approximately 300 m coverage extension is provided by the proposed SC-SHVC, when guaranteeing a FER of less than 5% in comparison of the other two schemes that exhibit a similar performance due to the same modulation mode employed.

To elaborate a little further, Figure 2.12 shows the video PSNR versus distance from the BS for the three schemes, when the BS transmit power and coverage are set to 40 dBm and 1000 m, respectively. Specifically, in Figure 2.12, the single-layer assisted QAM-HEVC scheme is capable of ensuring the PSNR over 38 dB across the BS coverage of about 500 m, but the user's QoE drops significantly when the UE is farther away from the BS station. By contrast, the proposed SC-SHVC ensures the basic quality across the coverage of the BS by providing about 35 dB PSNR for the cell-edge users. However, the coverage of 38 dB contour line shrinks to about 250 m due to both the degraded source coding efficiency of the SVC scheme and the less power-sharing coefficients assigned to the higher layers.

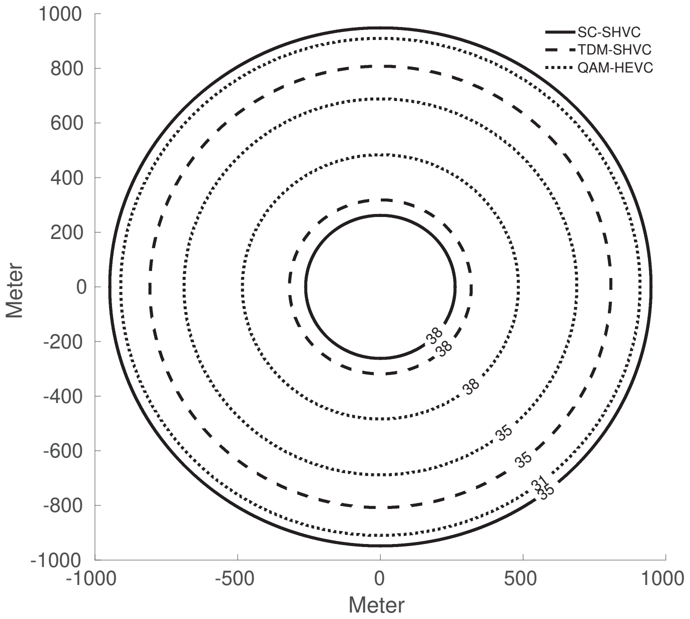


Figure 2.12: Video quality contour plots versus the distance for the three schemes, when $P = 40$ dBm and $S = 1000$ m.

2.5.2 Complexity Analysis

For the specific FEC configuration, the complexity of FEC decoding is dominated by the packet length. Due to the fact that the complexity imposed by FEC encoding and modulation is low compared to its decoding and demodulation counterparts, the re-encoding and the re-modulation complexity of SIC in our SC assisted scheme is neglected. Furthermore, since all the schemes in the simulations employ the same FEC configuration and hence exhibit the same overall FEC encoded rate, we assume that all the schemes share the same FEC decoding complexity. Therefore, we compare the complexity order imposed by demodulation. Table 2.4 compares the demodulation complexity of the three schemes, when the classic ML detection is invoked. Specifically, in our proposed SC assisted system, the superimposed symbols have to be demodulated progressively by using BPSK, 4QAM and 8PSK schemes, yielding the complexity order of $\mathcal{O}(2+4+8)$, while its QAM-HEVC counterpart employing 64QAM yields the complexity order of $\mathcal{O}(64)$. As for TDM-SHVC, due to the fact that the three modulation modes, namely 8PSK, 64QAM and 512QAM, are activated successively, the corresponding computational complexity order becomes their average, namely $\mathcal{O}(\frac{8+64+512}{3})$. It can be seen from Table 2.4 that the SC-SHVC scheme provides a sig-

Table 2.4: Demodulation complexity comparison of the three schemes.

Scheme	SC-SHVC	TDM-SHVC	QAM-HEVC
Complexity	$\mathcal{O}(2+4+8)$	$\mathcal{O}(\frac{8+64+512}{3})$	$\mathcal{O}(64)$

nificant complexity reduction compared to the other two schemes, while the TDM-SHVC scheme exhibits the highest complexity, since it requires a higher-complexity modulation mode for achieving an identical system throughput.

2.6 Chapter Conclusion

In this chapter, we presented an optimal power assisted SC scheme designed for layered video broadcasting. We commenced by first introducing the concept of modulation based UEP schemes in Section 2.2 categorized into HQAM and SC, both of which require appropriate power sharing technique. The SC and TDM assisted system models were illustrated in Section 2.3. The associated power allocation scheme of the SC scheme was further discussed in Section 2.4. The system performance was quantified in Section 2.5, which illustrates the QoE in terms of the reconstructed video PSNR as well as the FER and analyzes the system's complexity at the receiver. Specifically, the proposed SC-SHVC system is capable of providing the best video service over the BS's coverage in terms of both its video quality as well as FER, as shown in Table 2.5, while also imposing the lowest computational complexity at the receiver.

Table 2.5: The coverage comparison of the three schemes for ensuring FER of the BL below 5%.

Scheme	SC-SHVC	TDM-SHVC	QAM-HEVC
Coverage	850 m	700 m	400 m

Chapter 3

Adaptive Index Modulation for SVC Streaming

3.1 Introduction

In Chapter 2, we have designed SC aided modulation based UEP for SVC streaming, where the bits of the various video layers are conveyed by different modulation modes and are superimposed linearly before their transmission with the aid of optimal power sharing. Moreover, at the receiver, the classic SIC technique was invoked for progressively decoding the superimposed signals of different modulation modes. In this chapter, we conceive sophisticated MS-STSK transceivers for attaining UEP. Similar to the HQAM scheme of Section 2.2.1, our IM aided UEP maps the video source bits to the different-integrity MS-STSK subchannels according to their specific importance.

To elaborate a little further, the MS-STSK scheme consists of three main components, including the Antenna Selection Unit (ASU), the $\mathcal{L}$-PSK/QAM modulator and the dispersion matrix generator, each of which is capable of carrying a flexible number of bits and has a different BER. This is exploited as an alternative solution for achieving UEP. Based on this property, we propose an adaptive MS-STSK assisted system for layered video streaming in high-Doppler scenarios, where the video bits of different video layers are mapped to the different-BER MS-STSK subchannels according to pre-designed mapping schemes. We will show that the UEP assisted fixed modes attain an approximately 1 dB SNR gain over its EEP counterpart, at the expense of a modest complexity increase imposed by the associated bit mapping. As a further contribution, our adaptive design significantly improves both the image quality and the video PLR by judiciously activating the most appropriate channel-SNR-controlled transmission mode.

The rest of this chapter is organized as follows. Section 3.2 reviews the historical development of the MIMO-aided transceiver techniques, spanning from SM to STSK, followed by the introduction of the novel MS-STSK transceiver. In Section 3.3, we demonstrate that the MS-STSK transceiver has the potential of providing UEP by mapping the video source bits having different importance and different vulnerability into the corresponding MS-STSK components. Based on the exploitation of the UEP capability provided by the MS-STSK in Section 3.3, Section 3.4 describes a near-instantaneously adaptive MS-STSK assisted SVC streaming system, which may drop the EL, when a low-throughput MS-STSK configuration is selected for operation at low SNRs. Section 3.5 quantifies the performance of the proposed system in terms of its reconstructed video quality, where the proposed UEP scheme on average outperforms its EEP counterpart in both the fixed and adaptive modes. The chapter is concluded in Section 3.6.

3.2 Overview of MIMO transceivers

In this section different MIMO arrangements, such as the SM and the STSK schemes, are reviewed. We then elaborate on the MS-STSK transceiver, which employs an additional ASU component compared to STSK transceiver, which is employed for antenna set selection.

3.2.1 Spatial Modulation

The concept of SM [96, 114, 115] relies on activating one out of a total of N_t antenna elements in a given time slot. This provides an additional means of conveying source information, while eliminating Inter-Antenna Interference (IAI). Figure 3.1 depicts the basic model of a SM scheme, where the input bits are partitioned into two sub-codewords. Explicitly, the first $\log_2(\mathcal{L})$ bits are fed into the classic $\mathcal{L}$-ary modulator, generating the conventional modulated symbols, while $\log_2 N_t$ bits are used to select the specific antenna used for conveying the modulated symbols, which results in a SM codeword of $b = \log_2(\mathcal{L}N_t) = \log_2 \mathcal{L} + \log_2 N_t$ Bit Per Channel Use (BPCU), as shown in Figure 3.1. A particular benefit of this is that only a single Radio Frequency (RF) chain is required, which may be switched between antennas. Furthermore, the modulated signals $\mathbf{S}_{n_t,l} \in \mathcal{C}^{N_t \times 1}$ are given by:

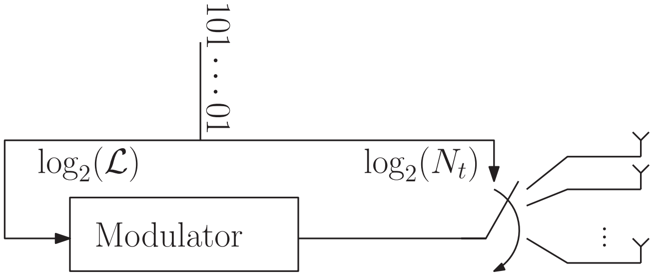


Figure 3.1: Schematic of SM.

$$\mathbf{S}_{n_t,l} = [\underbrace{0, \cdots, 0}_{n_t-1}, s_l, \underbrace{0, \cdots, 0}_{N_t-n_t}]^T, \tag{3.1}$$

where s_l $(1 \leq l \leq \mathcal{L})$ represents the PSK/QAM symbol modulated according to the $\log_2 \mathcal{L}$ input bits, while the integer n_t $(1 \leq n_t \leq N_t)$ corresponds to the remaining input bits. Based on the assumptions in [114] [96] [115], the number of Transmit Antenna (TA)s in SM is usually set to an integer power of 2. Again, it can be seen from (3.1) that only one out of N_t TA elements are activated during each symbol interval. Hence, no symbol-level synchronization between the TAs is needed, and the calibration of the antenna array is also avoided.

Furthermore, in contrast to the BLAST arrangement requiring the excessive-complexity joint detection of multiple antennas' signals, this SM scheme employs single-antenna-based ML detection at the receiver, hence imposing a low detection complexity. More explicitly, when the conditional probability of $P(\mathbf{Y}|\mathbf{S}_{n_t,l},\mathbf{H})$ is expressed as:

$$P(\mathbf{Y}|\mathbf{S}_{n_t,l},\mathbf{H}) = \frac{1}{(\pi N_0)^{N_r}} \exp(\frac{||\mathbf{Y}-\mathbf{H}\mathbf{S}_{n_t,l}||^2}{N_0}), \tag{3.2}$$

where $\mathbf{Y}$ and $\mathbf{H}$ represent the received signal and the channel impulse response matrix, respectively. Additionally, the optimal ML detection is formulated as [115]

$$(\hat{n}_t,\hat{l}) = \arg\max_{n_t,l} P(\mathbf{Y}|\mathbf{S}_{n_t,l},\mathbf{H}) \tag{3.3}$$

$$= \arg\min_{n_t,l} ||\mathbf{Y}-\mathbf{H}\mathbf{S}_{n_t,l}||^2, \tag{3.4}$$

$$= \arg\min_{n_t,l} ||\mathbf{Y}-s_l\mathbf{h}_{n_t}||^2, \tag{3.5}$$

where $\mathbf{h}_{n_t} = [h_{1n_t},\cdots,h_{N_r n_t}]^T$ represents the n_t-th column of the channel matrix $\mathbf{H}$. Observe from (3.5) that for the single-antenna-based ML algorithm the computational complexity increases linearly upon increasing the number of transmit antennas N_t. More particularly, the computational complexity of ML detection expressed in terms of the number of real-valued multiplications is given by:

$$\frac{6N_tN_r\mathcal{L}}{\log_2(N_T\mathcal{L})}. \tag{3.6}$$

As a consequence, SM and its derivatives have special benefits.

3.2.2 Space-Time Shift Keying

Naturally, SM also have some disadvantages, such as its lack of transmit diversity, since it uses single TA. As a remedy, it was argued in [116] that Linear Dispersion Code (LDC) is capable of striking an attractive design trade-off between the diversity and multiplexing gain. Hence it was suggested by Sugiura *et al.* [97] that it may be beneficial to embed the LDC arrangement into the SM scheme for the sake of attaining both diversity and multiplexing gains, which results in the compelling concept of STSK scheme.

More particularly, the STSK scheme is based on the activation of Q number of appropriately indexed space-time dispersion matrices within each STSK block period, rather than that of the TA indices at each symbol duration as in the SM scheme. It can be readily seen in Figure 3.2 that the STSK scheme contains two types of components, namely Q number of dispersion matrices $\{\mathbf{A}_1, \cdots, \mathbf{A}_q, \cdots, \mathbf{A}_Q\}$ ($\mathbf{A}_q \in \mathbb{C}^{N_t \times T}$), where N_t and T represent the number of TAs and the transmission periods plus the $\mathcal{L}$-PSK/QAM modulator. Observe in Figure 3.2 that similar to SM, the STSK bits b_{STSK} are partitioned into a pair of sub-streams, namely b_Q and b_M, as shown in Figure 3.2. Specifically, the bits fed b_Q select the specific dispersion matrix $\mathbf{A}_q$ from the Q pre-defined dispersion matrices, which spreads the conventional PSK/QAM symbol s_l across the TAs and time slots, generating the space-time block $\mathbf{S} \in \mathbb{C}^{N_T \times T}$ given by [116]:

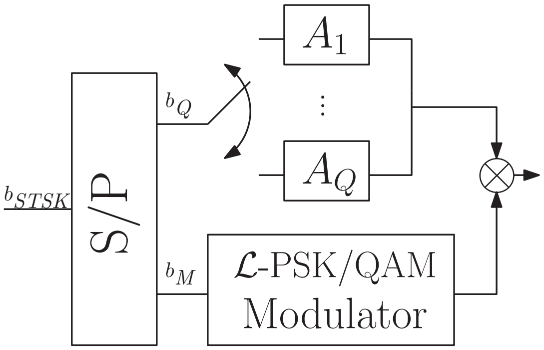


Figure 3.2: Schematic of STSK.

$$
\begin{aligned}
\mathbf{S} &= s_l \mathbf{A}_q \qquad (3.7)\\
&= \overbrace{0 \underbrace{\begin{bmatrix} a_{11}^{(1)} & \cdots & a_{1T}^{(1)} \\ \vdots & \ddots & \vdots \\ a_{N_t 1}^{(1)} & \cdots & a_{N_t T}^{(1)} \end{bmatrix}}_{\mathbf{A}_1} + \cdots + 0 \underbrace{\begin{bmatrix} a_{11}^{(q-1)} & \cdots & a_{1T}^{(q-1)} \\ \vdots & \ddots & \vdots \\ a_{N_t 1}^{(q-1)} & \cdots & a_{N_t T}^{(q-1)} \end{bmatrix}}_{\mathbf{A}_{q-1}}}^{\text{deactivated matrices}} \\
&+ s_l \overbrace{\underbrace{\begin{bmatrix} a_{11}^{(q)} & \cdots & a_{1T}^{(q)} \\ \vdots & \ddots & \vdots \\ a_{N_t 1}^{(q)} & \cdots & a_{N_t T}^{(q)} \end{bmatrix}}_{\mathbf{A}_q}}^{\text{activated matrix}} \\
&+ \overbrace{0 \underbrace{\begin{bmatrix} a_{11}^{(q+1)} & \cdots & a_{1T}^{(q+1)} \\ \vdots & \ddots & \vdots \\ a_{N_t 1}^{(q+1)} & \cdots & a_{N_t T}^{(q+1)} \end{bmatrix}}_{\mathbf{A}_{q+1}} + \cdots \quad + 0 \underbrace{\begin{bmatrix} a_{11}^{(Q)} & \cdots & a_{1T}^{(Q)} \\ \vdots & \ddots & \vdots \\ a_{N_t 1}^{(Q)} & \cdots & a_{N_t T}^{(Q)} \end{bmatrix}}_{\mathbf{A}_Q}}^{\text{deactivated matrices}},
\end{aligned}
$$

where s_l is the complex-valued symbol of the classic $\mathcal{L}$-PSK/QAM modulation scheme employed, corresponding to $b_M = \log_2 \mathcal{L}$ number of input bits. By contrast, the activated dispersion matrix $\mathbf{A}_q$ is selected from the set of Q dispersion matrices, which is associated with $b_Q = \log_2 Q$ number of input bits. Additionally, in order to maintain a unity average transmission power, the dispersion matrix has to comply with the power constraint of

$$\mathrm{tr}[\mathbf{A}_q^{\mathbf{H}}\mathbf{A}_q] = T, \quad 1 \leq q \leq Q, \tag{3.8}$$

where tr[$\cdot$] indicates the trace of the matrix. The power constraint of (3.8) is Q times higher than that of the LDC, since the LDC scheme linearly combines the Q dispersion matrices. Let us elaborate a little further on the STSK scheme by introducing an example for $b_{STSK} = 3$ bits. Based on the fact that $b_{STSK} = b_Q + b_M$, the parameter combinations conveying $b_{STSK} = 3$ can be formulated as $(Q, \mathcal{L} = (1,8;2,4;4,2;8,1))$, as exemplified in Table 3.1. Table 3.1 lists four possible combinations having a total throughput of 3-bits per STSK block. For a specific set of input bits, for example "110", the output is $\mathbf{S} = \mathbf{A}_1 s_7 = \mathbf{A}_1 e^{j\frac{6\pi}{4}}$, when $(Q = 1, \mathcal{L} = 8)$, which becomes $\mathbf{S} = \mathbf{A}_2 s_3 = \mathbf{A}_2 e^{j\frac{2\pi}{2}}$ when $(Q = 2, \mathcal{L} = 4)$. Furthermore, the normalized throughput per time slot of the STSK scheme may be expressed as

Table 3.1: Example of STSK scheme, when employing $b_{STSK} = 3$.

Input bits	Q=1 $\mathcal{L}$=8		Q=2 $\mathcal{L}$=4		Q=4 $\mathcal{L}$=2		Q=8 $\mathcal{L}$=1	
	$\mathbf{A}_q$	s_l	$\mathbf{A}_q$	s_l	$\mathbf{A}_q$	s_l	$\mathbf{A}_q$	s_l
000	$\mathbf{A}_1$	$s_1 = 1$	$\mathbf{A}_1$	$s_1 = 1$	$\mathbf{A}_1$	$s_1 = 1$	$\mathbf{A}_1$	$s_1 = 1$
001	$\mathbf{A}_1$	$s_2 = e^{j\frac{\pi}{4}}$	$\mathbf{A}_1$	$s_2 = e^{j\frac{\pi}{2}}$	$\mathbf{A}_1$	$s_2 = e^{j\pi}$	$\mathbf{A}_2$	$s_1 = 1$
010	$\mathbf{A}_1$	$s_3 = e^{j\frac{2\pi}{4}}$	$\mathbf{A}_1$	$s_3 = e^{j\frac{2\pi}{2}}$	$\mathbf{A}_2$	$s_1 = 1$	$\mathbf{A}_3$	$s_1 = 1$
011	$\mathbf{A}_1$	$s_4 = e^{j\frac{3\pi}{4}}$	$\mathbf{A}_1$	$s_4 = e^{j\frac{3\pi}{2}}$	$\mathbf{A}_2$	$s_2 = e^{j\pi}$	$\mathbf{A}_4$	$s_1 = 1$
100	$\mathbf{A}_1$	$s_5 = e^{j\frac{4\pi}{4}}$	$\mathbf{A}_2$	$s_1 = 1$	$\mathbf{A}_3$	$s_1 = 1$	$\mathbf{A}_5$	$s_1 = 1$
101	$\mathbf{A}_1$	$s_6 = e^{j\frac{5\pi}{4}}$	$\mathbf{A}_2$	$s_2 = e^{j\frac{\pi}{2}}$	$\mathbf{A}_3$	$s_2 = e^{j\pi}$	$\mathbf{A}_6$	$s_1 = 1$
110	$\mathbf{A}_1$	$s_7 = e^{j\frac{6\pi}{4}}$	$\mathbf{A}_2$	$s_3 = e^{j\frac{2\pi}{2}}$	$\mathbf{A}_4$	$s_1 = 1$	$\mathbf{A}_7$	$s_1 = 1$
111	$\mathbf{A}_1$	$s_8 = e^{j\frac{7\pi}{4}}$	$\mathbf{A}_2$	$s_4 = e^{j\frac{3\pi}{2}}$	$\mathbf{A}_4$	$s_2 = e^{j\pi}$	$\mathbf{A}_8$	$s_1 = 1$

$$\frac{\log_2(Q\mathcal{L})}{T} \quad \text{[bits/symbol]}. \tag{3.9}$$

3.2.3 Multi-Set Space-Time Shift Keying

Inspired by the above-mentioned STSK scheme, the concept of MS-STSK that combines the benefits of SM and STSK was proposed in [98], which achieves high multiplexing and diversity gains by conveying additional information over the selected Antenna Combination (AC) index, while also attaining a beneficial diversity gain. In the MS-STSK scheme, the transmitter produces an STSK codeword to be transmitted over a specific combination of N_t TAs using N_{RF} RF chains, as shown in Figure 3.3, where the number of RF chains is identical to the number of the spatial dimensions in the STSK scheme and less or equal to the number of TAs. When the number of RF chains is equal to the number of TAs, namely when $N_{RF} = N_t$, the system becomes the conventional STSK scheme. Specifically, when $N_{RF} < N_t$, the ASU of Figure 3.3 has to be activated, where the additional information bits will be used for selecting the AC index.

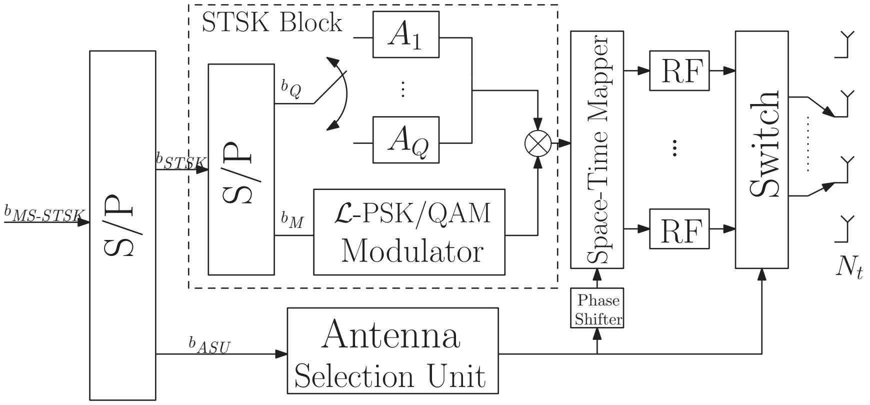


Figure 3.3: Schematic of MS-STSK.

It can be readily seen in Figure 3.3 that in the MS-STSK scheme, the MS-STSK information bits $b_{MS\text{-}STSK}$ are partitioned into two main components, including the STSK block $b_{STSK} = b_M + b_Q$, as illustrated in Section 3.2.2, plus the additional ASU component b_{ASU}, as depicted in Figure 3.3. Explicitly, the STSK block generates the classic space-time block $\mathbf{S} = \mathbf{A}_q s_l = [\mathbf{s}_1, \cdots, \mathbf{s}_{n_{RF}}, \cdots, \mathbf{s}_{N_{RF}}]^T$ of STSK according to its input bits b_{STSK}, where $\mathbf{s}_{n_{RF}} \in \mathbb{C}^{1\times T}$ is the n_{RF}-th row of $\mathbf{S}$, $\mathbf{A}_q \in \mathcal{C}^{N_{RF}\times T}$ is the q-th selected matrix based on b_Q and s_l is the l-th constellation point of the $\mathcal{L}$-PSK/QAM modulator based on b_M. This STSK block $\mathbf{S} \in \mathbb{C}^{N_{RF}\times T}$ is then transmitted over N_{RF} RF chains of the N_t TAs according to the bits fed into the ASU block b_{ASU}. Moreover, the total number of ACs can be defined as $N_c = 2^{b_{ASU}}$, where b_{ASU} can be expressed as:

$$b_{ASU} = \log_2 \frac{N_t}{N_{RF}}. \tag{3.10}$$

It is worth noting that the value of $\frac{N_t}{N_{RF}}$ must be an integer power of 2 as well. To elaborate little further on the ASU, let us consider an *MS-STSK*(4,2,2,2,4,4)$|_{QAM}$ system having the parameters of $N_t = 4$, $N_{RF} = 2$, $N_r = 2$, $T = 2$, $Q = 4$ and $\mathcal{L} = 4$. The STSK block first generates $2^{Q\mathcal{L}} = 16$ different STSK codewords to convey first $\log_2(Q\mathcal{L})$ bits of the input bits. For the so-called Distinct AC (DAC) scheme defined in [98], the next $b_{ASU} = \log_2 \frac{N_t}{N_{RF}} = \log_2 \frac{4}{2} = 1$ bit is used to map the STSK codewords to $N_c = 2^{b_{ASU}} = 2$ ACs. Gavin the STSK codeword $\mathbf{S} = \begin{bmatrix} \mathbf{s}_1 \\ \mathbf{s}_2 \end{bmatrix}$, the two possible ACs are given by:

$$\tilde{\mathbf{S}} = \begin{bmatrix} \mathbf{s}_1 \\ \mathbf{s}_2 \\ 0 \\ 0 \end{bmatrix} \; and \; \begin{bmatrix} 0 \\ 0 \\ \mathbf{s}_1 \\ \mathbf{s}_2 \end{bmatrix}, \tag{3.11}$$

where $\tilde{\mathbf{S}} \in \mathbb{C}^{N_t\times T}$ represents block-based MS-STSK symbol block.

For a $(N_T \times N_R)$-element MS-STSK system, the block-based received vector can be expressed as

$$\mathbf{Y} = \mathbf{H}\tilde{\mathbf{S}} + \mathbf{N}, \tag{3.12}$$

where $\mathbf{Y} \in \mathbb{C}^{N_r \times T}$ represents the received block-based signal. The vectorial stacking operation is also applicable for the MS-STSK system, as given by:

$$\bar{\mathbf{Y}} = \bar{\mathbf{H}}\mathcal{X}\mathcal{I}\mathbf{K} + \bar{\mathbf{N}}, \tag{3.13}$$

where we have

$$\bar{\mathbf{Y}} = vec(\mathbf{Y}) \in \mathbb{C}^{N_r T \times 1} \tag{3.14}$$

$$\bar{\mathbf{H}} = \mathbf{I} \otimes \mathbf{H} \in \mathbb{C}^{N_r T \times N_t T} \tag{3.15}$$

$$\bar{\mathbf{N}} = vec(\mathbf{N}) \in \mathcal{C}^{N_r T \times 1} \tag{3.16}$$

$$\mathcal{X} = [vec(\tilde{\mathbf{A}}_{1,1}) \cdots vec(\tilde{\mathbf{A}}_{q,c}) \cdots vec(\tilde{\mathbf{A}}_{q,N_c})] \in \mathcal{C}^{N_t T \times N_c Q}, \tag{3.17}$$

where $\mathbf{I}$ is a T-order identity matrix and $\otimes$ is the Kronecker operator. Furthermore, the AC is selected by the matrix $\mathcal{I} \in \mathbb{C}^{N_t T \times N_c Q}$, as given by:

$$\mathcal{I} = [\mathbf{0} \cdots \mathbf{I}_c \cdots \mathbf{0}]^T, \tag{3.18}$$

where $\mathbf{I}_c$ is a $(Q \times Q)$-element identity matrix, which is used for selecting the c-th combination of MS-STSK dispersion matrix set. The transmitted symbol used for activating the q-th dispersion matrix $\mathbf{K} \in \mathbb{C}^{Q \times 1}$ is expressed as

$$\mathbf{K} = [\underbrace{0, \cdots, 0}_{q-1}, s_l, \underbrace{0, \cdots, 0}_{Q-q}]^T, \tag{3.19}$$

where s_l represents the $\mathcal{L}$-PSK/QAM constellation points.

Assuming that all antenna elements of TA and Receive Antennas (RA) are sufficiently far apart to exhibit an uncorrelated channel, no Inter-Channel Interference (ICI) experienced in the vectorized system model of (3.13). Hence, the estimated version of the dispersion matrices, the $\mathcal{L}$-PSK/QAM and the AC indices at the receiver can be expressed as:

$$< \hat{q}, \hat{l}, \hat{c} > = \arg\min_{\hat{q},\hat{l},\hat{c}} ||\bar{\mathbf{Y}} - \bar{\mathbf{H}}\mathcal{X}\mathcal{I}_c\mathbf{K}_{q,l}||^2, \tag{3.20}$$

$$= \arg\min_{\hat{q},\hat{l},\hat{c}} ||\bar{\mathbf{Y}} - s_l(\bar{\mathbf{H}}_c\mathcal{X})_q||, \tag{3.21}$$

where $(\bar{\mathbf{H}}_c\mathcal{X})$ is the column vector of $(\bar{\mathbf{H}}\mathcal{X})$ after selecting the c-th AC by activating the c-th identity matrix in $\mathcal{I}$ and nulling the other $(N_c - 1)$ identity matrices.

3.3 Multi-Set Space-Time Shift Keying Based UEP

The above section briefly introduced the concept of the MS-STSK transceiver, where it can be seen from Figure 3.3 that the MS-STSK transceiver is constituted by three basic components. In this section, we focus our attention on using the MS-STSK transceiver to achieve UEP by mapping the source information to different MS-STSK codeword bits. It can be seen from Figure 3.3 that the MS-STSK codeword $b_{MS\text{-}STSK}$ can be considered as a combination of the b_{ASU} ASU bits used to select the antenna set and activated STSK codeword b_{STSK}, where the latter can be further partitioned into b_Q bits mapped to the dispersion matrices and b_M bits conveyed by the $\mathcal{L}$-PSK/QAM modulator, as shown in Figure 3.4 [99]. In the example of Figure 3.4, the input bit stream is partitioned into $b_{MS\text{-}STSK} = 6$-BPCU that contains $b_Q = 2$ bits, $b_M = 2$ bits and $b_{ASU} = 2$ bits. In fact, by varying the parameters of each component of the MS-STSK transceiver, the $b_{MS\text{-}STSK} = 6$-BPCU can be formed by different combinations. For example, apart from the example constitution exemplified in Figure 3.4, $b_{MS\text{-}STSK} = 6$-BPCU can also be achieved by $b_Q = 1$ bits, $b_M = 2$ bits and $b_{ASU} = 3$ bits, or $b_Q = 3$ bits, $b_M = 2$ bits and $b_{ASU} = 1$ bits, and so forth.

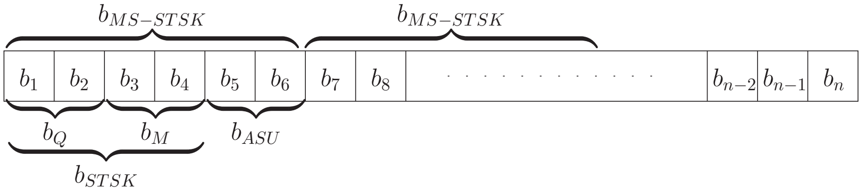


Figure 3.4: Bit Structure of MS-STSK.

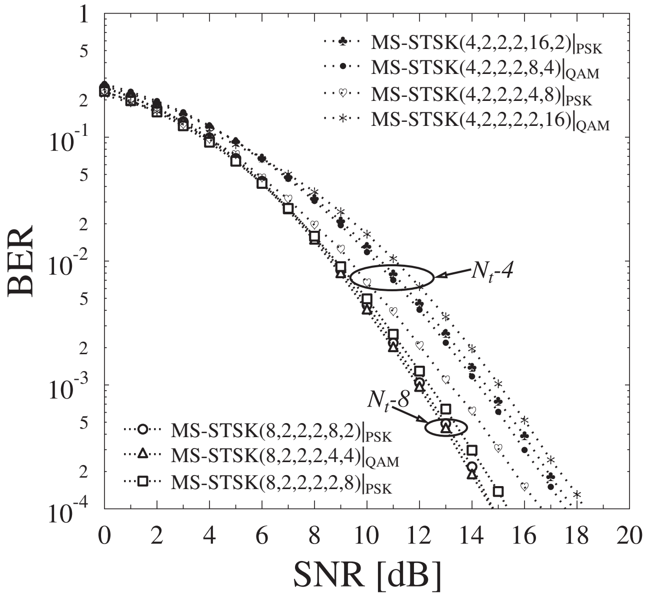


Figure 3.5: BER performance of MS-STSK under various configurations at a fixed throughput of b_{MSSTSK} = 6-BPCU.

Figure 3.5 shows the MS-STSK BER performance for various configurations, when transmitting over narrowband Rayleigh fading channels. Observe from Figure 3.5 that although all configurations exhibit the same normalized throughput of $b_{\mathit{MS\text{-}STSK}} = 6$-BPCU the BER performance of the MS-STSK system depends on the specific configuration of the parameters. It can also be observed that for a larger number of TAs, the sysyem is capable of providing an improved BER. For example, the group of curves labeled with N_t-8 outperforms the group labeled with N_t-4, since more antennas are capable of increasing the diversity gain. Moreover, for the MS-STSK configurations having the same N_t, the BER performance is also different. These findings evidence that the three components of the MS-STSK transceiver exhibit different BER performance, which further inspires us to exploit the UEP potential of the MS-STSK transceiver.

Below, we will first verify that MS-STSK transceiver has the potential of providing UEP by feeding the video source bits having different importance into the corresponding MS-STSK modules of Figure 3.3. Then, a more sophisticated iterative decoding assisted UEP is further exploited with the aid of the EXIT chart, where the video source bits having different importance are mapped to the associated MS-STSK bits of Figure 3.4 according to their capacity performance.

3.3.1 Non-iterative MS-STSK Assisted UEP

In this section, we aim for finding the BER performance of the three components of the MS-STSK transceiver, where we focus our attention on the BER performance of each component bits, namely b_{ASU}, b_Q and b_M. We exhibit our simulation results over Rayleigh fading channel, which is presented using the example configurations of $\mathit{MS\text{-}STSK}(4,2,2,2,8,4)|_{QAM}$, $\mathit{MS\text{-}STSK}(8,2,2,2,4,4)|_{QAM}$ and $\mathit{MS\text{-}STSK}(16,2,2,2,8,8)|_{QAM}$. It can be seen in Figure 3.6 that the ASU is capable of

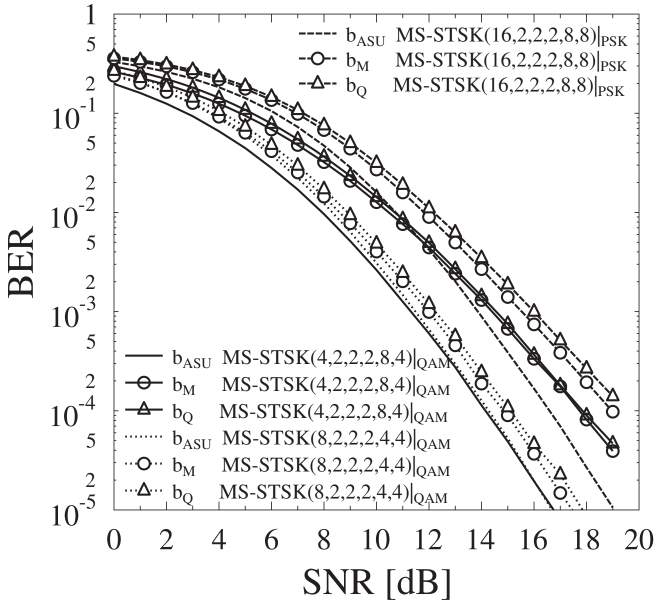


Figure 3.6: BER performances of different blocks of MS-STSK over Rayleigh channel.

attaining a lower BER than the $\mathcal{L}$-QAM/PSK modulator, while both outperform the dispersion matrix component. It is worth noting that in Figure 3.6 the curves of the $\textit{MS-STSK}(8,2,2,2,4,4)|_{QAM}$ and $\textit{MS-STSK}(16,2,2,2,8,8)|_{PSK}$ systems constitute $\textit{MS-STSK}(8,2,2,2,4,4)|_{QAM}$ system, we have $b_{ASU} = b_Q = b_M = 2$ bits, while the $\textit{MS-STSK}(16,2,2,2,8,8)|_{PSK}$ system relies on $b_{ASU} = b_Q$ $= b_M = 3$ bits. However they exhibit different BER performances, as shown Figure 3.6. Therefore, we conclude that the MS-STSK transceiver has the potential of providing UEP by feeding the video source bits having different importance into the different MS-STSK modules of Figure 3.3.

3.3.2 Iterative Decoding Assisted MS-STSK Based UEP

It can bee seen Figure 3.6, all the ASU index bits exhibit the best error-sensitivity, while the b_Q-bits and b_M-bits of the dispersion matrix indices and of the QAM-bits defined in Figure 3.9 tend to have different error-sensitivity. Hence, we have to deal with the bit-sensitivities on a bit-level basis for the sake of finding the sophisticated MS-STSK assisted optimal UEP mapping scheme. Additionally, to fully exploit the potential UEP using the novel MS-STSK transceiver, we introduce the EXIT chart to analyze the iterative decoding performance, which is capable of visualising the exchange of extrinsic information between the concatenated decoders.

3.3.2.1 Historical Review of EXIT Chart

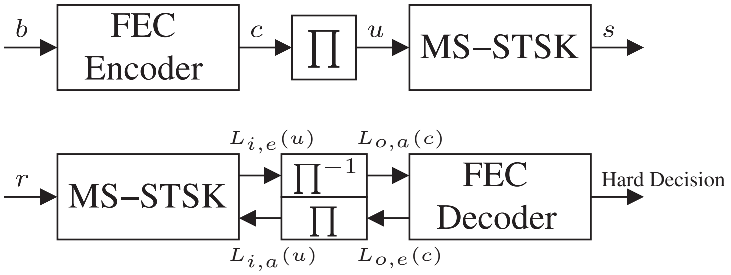


Figure 3.7: Block diagram of a two-stage iteratively decoded system using MS-STSK and FEC.

Figure 3.7 portrays a two-stage concatenated system, where the source bit stream b is encoded by a FEC codec, outputting the bit stream c, followed by a random bit interleaver Π. The interleaved sequence u of Figure 3.7 is then mapped to the MS-STSK symbol stream s. Let us denote the LLR of the bits concerned by $L(\cdot)$, while the notation i and o correspond to the inner and outer serially concatenated component constituted by the MS-STSK transceiver and the FEC codec, which are hence also referred to as the inner and the outer serially concatenated components codecs, respectively. Furthermore, the subscripts a and e used in Figure 3.7 indicate *a priori* and *a posteriori* information, respectively. The received signal stream r of Figure 3.7 is demapped first by the MS-STSK decoder with the aid of the Logarithmic Maximum *a posteriori* (Log-MAP) algorithm to its LLR representation, namely to $L_{i,e}(u)$, which is then deinterleaved by the soft-bit deinterleaver block Π^{-1} of Figure 3.7 to generate the soft bit stream $L_{o,a}(c)$. This stream is then passed to the FEC decoder as *a priori* information in order to generate extrinsic LLR values $L_{o,e}(c)$. As seen in Figure 3.7, the *a posteriori* information $L_{o,e}(c)$ generated by the FEC decoder is then fed back to the MS-STSK transceiver as *a priori* information after reorganizing it using the interleaver. Additionally, to quantify the information content of the LLR values, the classic Mutual Information (MI) I is used for quantifying the relationship between two simultaneously-sampled random variables [117].

3.3.2.2 EXIT Chart Assisted Performance on Bit-Level Basis

Briefly, EXIT charts constitute powerful tools of visualising the exchange of extrinsic information between the concatenated decoders of Figure 3.7, which are capable of accurately predicting the convergence behaviour of the iterative receiver based on the MI exchanged between its components [118] [119]. Figure 3.8 exhibits the exchange of the MI performance for the system shown in Figure 3.7, where a higher value of the *a posteriori* MI $I_{o,a}(c)/I_{i,e}(u)$ usually represents a lower BER. It can be seen in Figure 3.8 that the MI of the MS-STSK bits vary with the *a priori* MI $I_{i,a}(u)/I_{o,e}(e)$. Intriguingly, in Figure 3.8 b_{ASU} exhibits the highest $I_{o,a}(c)/I_{i,e}(u)$ MI when no *a priori* information is provided, namely $I_{i,a}(u)/I_{o,e}(e) = 0$, but it is overtaken by b_{Q_1} that has the highest MI when full knowledge of *a priori* information is obtained, namely $I_{i,a}(u)/I_{o,e}(e) = 1$. Figure 3.8 illustrates the UEP potential of the MS-STSK transceiver in terms of the bit-level basis

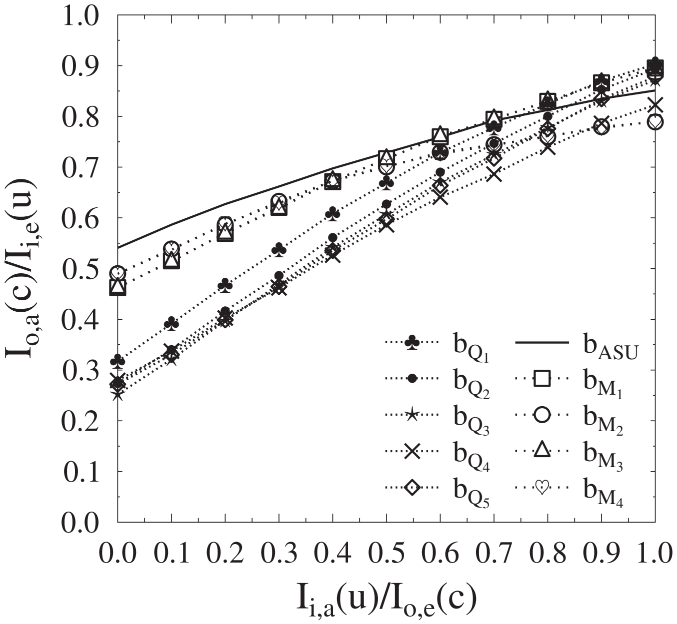


Figure 3.8: The *MS-STSK(4,2,2,2,32,16)*$|_{QAM}$ scheme's MI curve for transmission over Ricean fading channels ($K = 0$ dB) at the channel SNR of 6 dB.

visualised with the aid of the EXIT chart, where varied MI of various MS-STSK bits enhances the flexibility when designing the UEP scheme. Based on the above exploitation, in the following sections we aim for building video transmission system over wireless channel, where a sophisticated bit mapping schemes are particularly designed for employed MS-STSK configurations with the aid of the exploitation in Figure 3.8.

3.4 Proposed System Model for SVC Streaming

Inspired by the results found in Section 3.3.2.2, let us now embark on designing the MS-STSK assisted UEP scheme for scalable video streaming over wireless channel. Specifically, the proposed video streaming system is particularly designed for high-mobility UAV-aided surveillance, where the video captured is transmitted to the destination taking into account the video resolution required, the estimated CSI and the UAVs mobility. More explicitly, the transmission of the ELs may have to be abandoned even if they are requested by the destination, if the near-instantaneous channel conditions are momentarily unsuitable for supporting the required bitrates. In the following, we will commence by introducing the MS-STSK transceiver, including the transmitter and receiver models. Then, the PSAM-aided channel estimation is specifically introduced for the MS-STSK transceiver in order to precisely acquire the varying CSI in the high-Doppler scenario. Finally, we exhibit our design guideline for the bit mapping scheme to provide UEP for the scalable video bits exhibiting different importance on the bit-level basis.

3.4.1 Transmitter Model

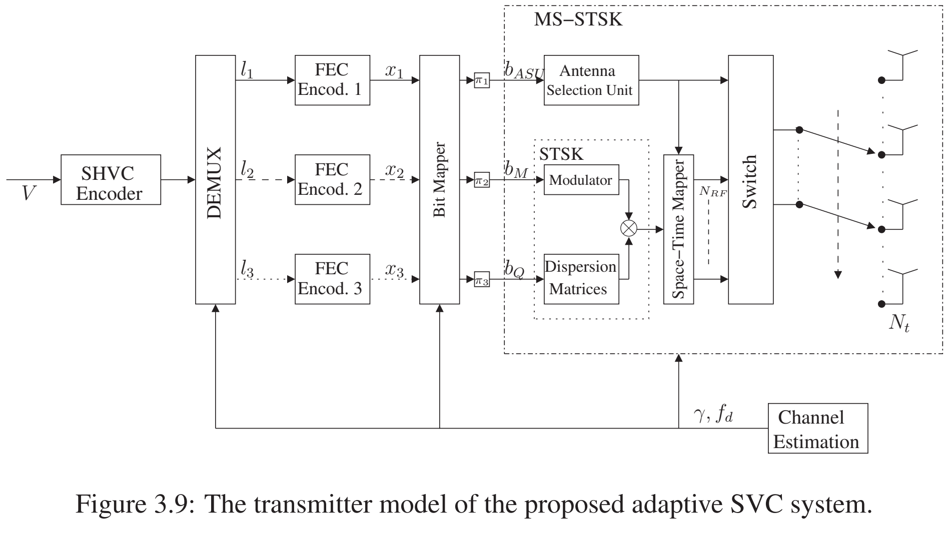


Figure 3.9: The transmitter model of the proposed adaptive SVC system.

Figure 3.9 depicts the transmitter architecture of our proposed MS-STSK assisted adaptive system conceived for layered video streaming. It can be seen from Figure 3.9 that the captured video V is compressed by a SHVC Encoder, generating a compressed video stream that consists of three layers, namely l_1, l_2 and l_3 corresponding to the BL, first EL and second EL, respectively, which are then fed into a demultiplexer.

We assume that in the proposed system the maximum number of SVC layers is identical to that of the adaptation modes supported. Explicitly, we have three candidate adaptation modes for layered video streaming, where *Mode* 1 is utilized for the BL's streaming, *Mode* 2 for that of the BL and first EL, and *Mode* 3 for the bits belonging to all layers. The mode-switching operation is jointly controlled by both the estimated instantaneous SNR γ and the normalized maximum Doppler frequency f_d. The thresholds that determine the mode-switching operation are denoted by t_{f_d}, hence the mode selection operation is given by:

$$Mode = \begin{cases} Mode\ 1 & \gamma_{f_d} \leq t_{1,f_d} \\ Mode\ 2 & t_{1,f_d} < \gamma_{f_d} \leq t_{2,f_d} \\ Mode\ 3 & t_{2,f_d} < \gamma_{f_d}, \end{cases} \tag{3.22}$$

where the notation γ_{f_d} denotes the estimated channel SNR γ under the normalized maximum Doppler frequency f_d, which means that for different Doppler spreads the threshold values may be different.

The thresholds t_{f_d} are determined by the maximum tolerable PLR of the video packets. More explicitly, due to the dependency between the EL and its associated reference layers, where the EL cannot be successfully decoded without the flawless recovery of its reference layers, we now introduce the concept of Equivalent PLR (e-PLR) $P(l_i)_e$ as the metric of quantifying the associated protection capability, which is defined as follows:

$$\begin{cases} P(l_1) & i = 1 \\ \sum\limits_{m=2}^{i} P(l_m) \prod\limits_{n=1}^{m-1} [1 - P(l_n)] + P(l_1) & i > 1. \end{cases} \tag{3.23}$$

Since no reference layer is used by the BL, the e-PLR value of the BL $P(l_1)_e$ is simply equivalent to its PLR value $P(l_1)$. The PLR thresholds to be satisfied at a given normalized maximum Doppler frequency f_d are recorded when the e-PLR $P(l_i)_e$ value dips below the maximum tolerable e-PLR of 5% [93], where the video artifacts introduced by the lost packets become perceptually tolerable.

Furthermore, the operating procedure of each mode is detailed as follows:

1. *Mode* 1: *Mode* 1 is designed to provide the most robust streaming, where the configuration of *MS-STSK(2,2,2,2,2,2)*$|_{PSK}$ that achieves the most conservative throughput, namely $b_{\textit{MS-STSK}} = 2$ bits, is activated for the BL streaming only. In this mode, both the ELs of the layered bit streams output by the SHVC Encoder are truncated and discarded. Hence, only FEC Encoder 1 of Figure 3.9 is activated for the protection of the BL bit stream, hence resulting in the FEC encoded bit stream of x_1. Since no ELs are transmitted in this mode, the Bit Mapper of Figure 3.9 is deactivated for equally protecting the bits of the BL.

2. *Mode* 2: *Mode* 2 provides a higher throughput assisted by the transceiver configuration of *MS-STSK(4,2,2,2,4,4)*$|_{QAM}$ having $b_{MS\text{-}STSK} = 5$ bits for transmitting the additional bit stream of the first EL l_2. In this mode, the FEC Encoders 1 and 2 of Figure 3.9 are activated for generating the encoded bit streams of x_1 and x_2 for l_1 and l_2, respectively. Owing to the dependency between l_1 and l_2, the Bit Mapper of Figure 3.9 utilizes the bits that are capable of exhibiting the best BER performance in the MS-STSK codeword of Figure 3.4 to convey the bits of the BL x_1, while their weaker counterparts are allocated for the less important bit stream x_2.

3. *Mode* 3: With the aid of the *MS-STSK(4,2,2,2,32,16)*$|_{QAM}$ configuration that transmits $b_{MS\text{-}STSK}$ $= 10$ bits, *Mode* 3 becomes capable of conveying the bit streams of all the three video layers. In this mode, all the three FEC Encoders of Figure 3.9 are activated for encoding l_1, l_2 and l_3 into x_1, x_2 and x_3, respectively. The Bit Mapper of Figure 3.9 then allocates the bits in the MS-STSK codeword to convey the source bits having different importance according to the video layer index, which is quite similar to the operating procedure of *Mode* 2.

The detailed procedure of mapping the unequally protected source bits to the MS-STSK subchannels is discussed in Section 3.4.4. The modulated bits are then transmitted by the MS-STSK transceiver over the Air-to-Ground (AG) channel to be detailed in Section 3.4.3.1. For the sake of limiting the power consumption in our UAV scenario, the maximum number of TAs N_t is set to 4, but for the configuration of *MS-STSK(2,2,2,2,2,2)*$|_{PSK}$ in *Mode* 1, only two TAs are activated.

3.4.2 Receiver Model

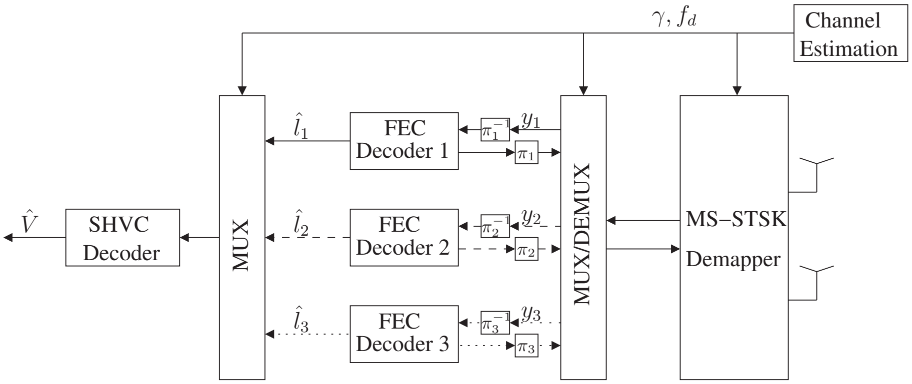


Figure 3.10: The receiver model of the proposed adaptive SVC system.

Figure 3.10 depicts the receiver model of our proposed system, where iterative decoding is used for exchanging extrinsic soft information between the parallel FEC Decoders and the MS-STSK Demapper. The received signals are first detected by the MS-STSK Demapper of Figure 3.10 for generating the soft LLR output of the MS-STSK scheme. Then, the MUX/DEMUX of Figure 3.10 reorganizes the soft information received from the MS-STSK Demapper according to the activated mode and forwards the outputs to the FEC Decoders, where y_1, y_2 and y_3 are the received version of the encoded bit streams of x_1, x_2 and x_3, respectively. It is worth noting that the number of demultiplexed layers is determined by the activated mode, which means that, for example, no y_2 and y_3 are received in *Mode* 1, while no y_3 is received in *Mode* 2 of Figure 3.10.

The extrinsic information generated by the FEC Decoders is then forwarded to the MS-STSK Demapper via the MUX/DEMUX that reorganizes the codewords for the demapper to be used as its *a priori* information in preparation for the next iteration. This iterative decoding process of Figure 3.10 continues until the maximum number of iterations is reached. The hard-decoded video streams, namely $\hat{l}$, are then reorganized by the MUX, as shown in Figure 3.10, which are then forwarded to the SHVC Decoder of Figure 3.10 in order to reconstruct the video.

3.4.3 PSAM Assisted Channel Estimation

In this section we incorporate PSAM-aided channel estimation into our MS-STSK transceiver. First, we present the AG channel model used in our system, since the PSAM configuration is conditioned on the channel characteristics. Due to the high mobility of UAVs, we have a high Doppler frequency. Hence, the conventional training based CSI estimation assuming that the complex-valued channel envelope is slowly varying becomes inaccurate and hence the system would suffer from an inevitable error floor. Therefore, we insert the pilot symbols periodically into the transmitted data frame instead of transmitting the entire training sequence as a preamble before data transmission ensues.

3.4.3.1 Air-to-Ground Channel Model

The 3rd Generation Partnership Project (3GPP) is currently developing both the LTE Advanced Pro and the 5th Generation (5G) networks, making them suitable for supporting Vehicle-to-Everything (V2X) [120] and for UAV scenarios [121], including the UAV and BS heights, Angle of Arrival (AoA), Angle of Departure (AoD), Angle Spread of Arrival (ASA) and Angle Spread of Departure (ASD) and so forth. However, since the impact of path loss is omitted in this treatise, the practical values, such as the height of the UAV and BS, exemplified in [121] are hence not considered. Additionally, we assume that the parameters determining fast fading, such as AoA and AoD, are assumed to be periodically updated and known for the BS. Thereby, the high-mobility aeronautical Ricean fading channels considered in this treatise are explicitly characterized by the following distinctive features [122]:

1. High frequency offset Δf_{LOS} on the strong Line of Sight (LoS) path.
2. High normalized maximum Doppler frequency f_d for the diffuse scattering component.
3. The AoD ϕ_t and the AoA ϕ_r in the context of employing multiple TAs and RAs.

These challenges call for the 'clean-state' consideration of a variety of high-mobility communication techniques. In AG communications, the channel can be expressed as [123]:

$$\mathbf{H}_n = \mathbf{H}_n^D + \mathbf{H}_n^S, \tag{3.24}$$

which requires $E\{tr[\mathbf{H}_n(\mathbf{H}_n)^H]\} = N_t N_r$ for power normalization. Furthermore, the $(N_t \times N_r)$-element matrices of $\mathbf{H}_n^D$ and $\mathbf{H}_n^S$ represent the LoS and the scattered components, with power of σ_D^2 and σ_S^2, respectively. The LoS matrix is given by $\mathbf{H}_n^D = \sigma_D e^{j2\pi\Delta f_{LOS} n} \mathbf{a}_t \mathbf{a}_r^T$, where $\Delta f_{LOS} = f_d \cos(\Delta f_{offset}) \le f_d$ [122]. We denote the normalized maximum Doppler frequency by f_d and the LoS frequency offset by Δf_{offset}. The angle between the LoS and the direction of movement ϕ_0 is assumed to be uniformly distributed in the interval $[-\pi, \pi]$. The two directional signal vectors, namely $\mathbf{a}_t$ and $\mathbf{a}_r^T$, are defined as:

$$\mathbf{a}_t = [1, e^{j2\pi d cos(\phi_t)}, ..., e^{j2\pi d(N_t-1)cos(\phi_t)}]^T, \tag{3.25}$$

and

$$\mathbf{a}_r = [1, e^{j2\pi d cos(\phi_r)}, ..., e^{j2\pi d(N_r-1)cos(\phi_r)}]^T, \tag{3.26}$$

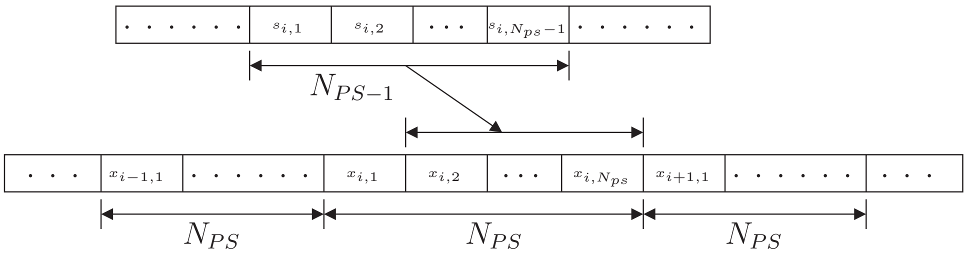
respectively, where ϕ_t, ϕ_r and d refer to the associated AoD, AoA and the antenna spacing, respectively. The Ricean K-factor is defined as $K = \frac{\sigma_D^2}{\sigma_S^2}$, which results in $\sigma_D = \sqrt{\frac{K}{K+1}}$, $\sigma_S = \sqrt{\frac{1}{K+1}}$ and $\sigma_D^2 + \sigma_S^2 = 1$. Furthermore, the LoS autocorrelation is given by $E\{\mathbf{H}_{n+k}^D(\mathbf{H}_n^D)^H\} = \frac{K}{K+1} e^{j2\pi\Delta f_{LOS} k} \mathbf{R}_{AA}$, where the $(N_t \times N_t)$-element matrix $\mathbf{R}_{AA}$ is formulated as $\mathbf{R}_{AA} = \mathbf{a}_t[n+k]\mathbf{a}_r[n+k]^T \mathbf{a}_r[n]^* \mathbf{a}_t[n]^H$, while the autocorrelation matrix of the scattered component can be expressed as $E\{\mathbf{H}_{n+k}^S(\mathbf{H}_n^S)^H\} = \frac{N_r}{K+1} J_0(2\pi f_d k)\mathbf{I}_M$ [122]. This autocorrelation matrix is eminently suitable for PSAM assisted channel estimation, as described in the next section.

3.4.3.2 Preliminaries of PSAM

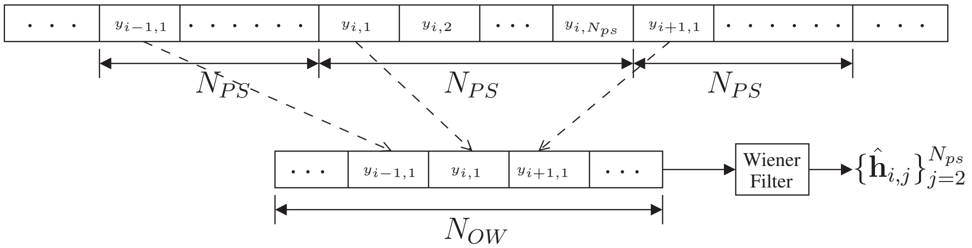
The PSAM frame-structure proposed by Cavers in [124] is portrayed in Figure 3.11, which is suitable for a SISO fading channel [122]. We denote the j-th data symbol of the i-th PSAM frame by $\{s_{i,j}\}_{j=1}^{N_{PS}\text{-}1}$, which is identical to the symbol of $\{x_{i,j}\}_{j=2}^{N_{PS}}$, and denote the pre-defined pilot symbol also known to the receiver for the i-th PSAM frame by $x_{i,1}$. Figure 3.11(a) illustrates that the pilot symbols, namely $x_{i,1}$, are inserted into the symbol streams with the period of N_{PS}. Then, the received signals of this SISO example can be modelled as:

$$y_{i,j} = x_{i,j} h_{i,j} + v_{i,j}, \tag{3.27}$$

with v representing the AWGN. Then, in order to estimate the channel state of $\{h_{i,j}\}_{j=2}^{N_{PS}}$ for the data symbols of $\{x_{i,j}\}_{j=2}^{N_{PS}}$, the nearby N_{OW} number of pilot samples are utilized, where N_{OW} indicates the observation window size. Figure 3.11(b) depicts the PSAM detection process, in which a set of N_{OW} pilot symbols $\{y_{i,1}\}_{i=-N_{OW}^a}^{N_{OW}^b}$ are extracted from the received PSAM symbol streams to estimate



(a) PSAM at the transmitter.



(b) PSAM at the receiver.

Figure 3.11: The frame structure of PSAM.

the channel. The multiplicative channel gain of each detected symbol is derived by interpolation, yielding:

$$\begin{aligned}\{\hat{h}_{i,j}\}_{j=2}^{N_{PS}} &= \sum_{\bar{i}=-N_{OW}^a}^{N_{OW}^b} w_{\bar{i},j} y_{i+\bar{i},1}/x_{i+\bar{i},1} \\ &= \sum_{\tilde{i}=1}^{N_{OW}} w_{\tilde{i},j} y_{i-N_{OW}^a-1+\tilde{i},1}/x_{i-N_{OW}^a-1+\tilde{i},1} \\ &= \mathbf{w}_j^T(\mathbf{X}_i^P)^H \mathbf{y}_i^P, \end{aligned} \tag{3.28}$$

where the notations $N_{OW}^a = \lfloor \frac{N_{OW}}{2} \rfloor - 1$ and $N_{OW}^b = \lfloor \frac{N_{OW}+1}{2} \rfloor$ represent the observation boundaries. The filter taps are expressed by $\mathbf{w}_j = [w_{1,j}, w_{2,j}, ..., w_{N_{OW},j}]^T$, while the transmitted pilot symbols and received pilot samples are given by $\mathbf{X}_i^P = \text{diag}[x_{i-N_{OW}^a,1}, ..., x_{i,1}, ..., x_{i+N_{OW}^b,1}]$ and $\mathbf{y}_i^P = [y_{i-N_{OW}^a,1}, ..., y_{i,1}, ..., y_{i+N_{OW}^b,1}]^T$, respectively. The filter taps $\mathbf{w}_j$ may be optimized for the sake of minimizing the MSE between the Wiener filter output $\hat{h}_{i,j}$ and $h_{i,j}$ [122]. The corresponding MSE cost function may be formulated as:

$$\sigma_{MSE}^2 = E\{\|h_{i,j} - \hat{h}_{i,j}\|^2\} = 1 - 2\mathbf{w}_j^T \tilde{\mathbf{e}} + \mathbf{w}_j^T \tilde{\mathbf{C}} \mathbf{w}_j^*. \tag{3.29}$$

The channel characteristic matrix $\tilde{\mathbf{C}}$ can be expressed as:

$$\tilde{\mathbf{C}} = \begin{bmatrix} \psi[0]+N_0 & \psi[-1] & \dots & \psi[-(N_{OW}-1)] \\ \psi[1] & \psi[0]+N_0 & \dots & \psi[-(N_{OW}-2)] \\ \psi[2] & \psi[1] & \dots & \psi[-(N_{OW}-3)] \\ \vdots & \vdots & \ddots & \vdots \\ \psi[-(N_{OW}-1)] & \psi[-(N_{OW}-2)] & \dots & \psi[0]+N_0 \end{bmatrix}, \tag{3.30}$$

where we have $\{\psi[kN_{PS}] = \frac{K}{K+1}e^{j2\pi\Delta f_{LOS}kN_{PS}} + \frac{1}{K+1}J_0(|2\pi f_d kN_{PS}|)\}_{k=-(N_{OW}-1)}^{N_{OW}-1}$. The channel's cross-correlation vector is expressed as:

$$\begin{aligned}\tilde{\mathbf{e}} = \big[&\psi[-N_{OW}^{a}N_{PS}-j+1], \psi[-(N_{OW}^{a}-1)N_{PS}-j+1], ..., \\ &\psi[-j+1], ..., \psi[N_{OW}^{b}N_{PS}-j+1]\big]^T. \end{aligned} \tag{3.31}$$

Therefore, by invoking the Minimum MSE (MMSE) solution derived by setting the derivative of the channel estimation error with respect to the estimator coefficients to zero according to $\frac{\partial\sigma_{MSE}^2}{\partial\mathbf{w}_j} = 0$, the associated Wiener-Hopf equation can be formulated as:

$$\mathbf{w}_j^* = \tilde{\mathbf{C}}^{-1}\tilde{\mathbf{e}}^T. \tag{3.32}$$

Hence, the channel estimates $\{\hat{h}_{i,j}\}_{j=2}^{N_{PS}}$ can be acquired by invoking (3.28).

3.4.3.3 PSAM for MS-STSK

Let us now consider an MS-STSK system employing N_t TAs and N_r RAs, where we denote the channel by $\mathbf{H}_{i,j} \in \mathbb{C}^{N_r\times N_t}$. In multi-antenna transmissions, since the signals sent from multiple TAs are superimposed at the receiver, the pilot symbols inserted are spread over N_t symbol periods [122] [125] [126]. This requires the transmitted block to be a square matrix. More explicitly, to satisfy the above constraint in the MS-STSK transceiver, we assume that the pilot block of each frame becomes a square diagonal matrix represented by $\bar{\mathbf{X}}_{i,1} = x_{i,1}\mathbf{I}_{N_t}$ and that the channel's complex-valued envelope remains unchanged for $n = N_t/T$ MS-STSK symbol durations. We denote the PSAM assisted MS-STSK symbol block by $\{\bar{\mathbf{X}}_{i,j} \in \mathbb{C}^{N_t\times(nT)}\}_{j=2}^{N_{PS}}$. Hence, the block-based received signal vector can be expressed as:

$$\mathbf{Y}_{i,j} = \mathbf{H}_{i,j}\bar{\mathbf{X}}_{i,j} + \mathbf{V}_{i,j}, \tag{3.33}$$

where $\mathbf{Y}_{i,j} \in \mathbb{C}^{N_r\times(nT)}$ represents the received block-based signal, while $\mathbf{V}_{i,j} \in \mathbb{C}^{N_r\times(nT)}$ denotes the zero-mean AWGN of power N_0. Therefore, the PSAM channel estimation can be formulated for the MS-STSK system as:

$$\{\hat{\mathbf{H}}_{i,j}\}_{j=2}^{N_{PS}} = \sum_{\bar{i}=-N_{OW}^{a}}^{N_{OW}^{b}} \mathbf{W}_{\bar{i},j}^T\bar{\mathbf{X}}_{i+\bar{i}}^H\mathbf{Y}_{i+\bar{i},1} = \mathbf{W}_j^T(\bar{\mathbf{X}}_i^P)^H\mathbf{Y}_i^P. \tag{3.34}$$

The filter taps $\{\mathbf{W}_{\tilde{i},j}\}_{\tilde{i}=1}^{N_{OW}} = [\mathbf{W}_{1,j}^T, \mathbf{W}_{2,j}^T, ..., \mathbf{W}_{N_{OW},j}^T]$ then become $(N_t \times N_t)$-element matrices. The pilots of the MS-STSK transceiver become $\bar{\mathbf{X}}_i^P = \mathrm{diag}\{\bar{\mathbf{X}}_{i-N_{OW}^a,1}, ..., \bar{\mathbf{X}}_{i,1}, ..., \bar{\mathbf{X}}_{i+N_{OW}^b,1}\}$, while the received pilot samples are $\mathbf{Y}_i^P = [\mathbf{Y}_{i-N_{OW}^a,1}^T, ..., \mathbf{Y}_{i,1}^T, ..., \mathbf{Y}_{i+N_{OW}^b,1}^T]^T$. The MSE cost function can then be modelled as:

$$\begin{aligned}\sigma_{MSE}^2 &= E\{\|\mathbf{H}_{i,j} - \hat{\mathbf{H}}_{i,j}\|^2\}/(N_tN_r) \\ &= 1 - 2\mathrm{tr}(\mathbf{W}_j^T\tilde{\mathbf{e}}) + \mathrm{tr}(\mathbf{W}_j^T\tilde{\mathbf{C}}\mathbf{W}_j^*). \end{aligned} \tag{3.35}$$

Substituting $\psi[kN_{PS}]$ of (3.30) by the $(N_t \times N_t)$-element matrix $\{\Psi[k] = \frac{K}{N_t N_r (K+1)} e^{(j2\pi \Delta f_{LOS} k N_{PS})} \mathbf{R}_{AA} + \frac{1}{N_t(K+1)} J_0(|2\pi f_d k N_{PS}|) \mathbf{I}_{N_t}\}_{k=-(N_{Ow}-1)}^{N_{Ow}-1}$, where $J_0(\cdot)$ is the zero-order Bessel function of the first kind, the resultant MMSE solution is given by [122]:

$$\mathbf{W}_t^* = \tilde{\mathbf{C}}^{-1}\tilde{\mathbf{e}}. \tag{3.36}$$

Therefore, the channel matrices $\{\hat{\mathbf{H}}_{i,j}\}_{j=2}^{N_{PS}}$ of the MS-STSK symbol blocks can be obtained by substituting (3.36) into (3.34).

3.4.4 Bit Mapping Design

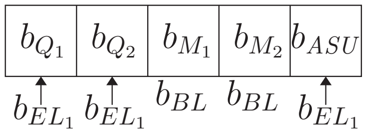
Briefly, EXIT charts constitute powerful tools of visualising the exchange of extrinsic information between the concatenated decoders of Figure 3.7, which are capable of accurately predicting the convergence behaviour of the iterative receiver based on the MI exchanged between its components [118] [119]. Let us now embark on designing EXIT-chart-aided the source-bit to MS-STSK-bit allocation of the Bit Mapper unit of Figure 3.9 for *Mode* 2 and *Mode* 3 of Figure 3.9 with the objective of assigning the most appropriate protection according to the specific importance of the source bits, which is detailed as follow:

1. Choose the system parameters, such as channel model and MS-STSK configurations.
2. Set up the system model and testify the system performance via Monte-Carlo experiments, as illustrated in Figure 3.8.
3. Estimating the capacity capability of the MS-STSK bits by calculating the area $\mathcal{A}$ under its MI curve.
4. Design the source-bit to MS-STSK bit mapping allocation of the Bit Mapper unit of Figure 3.9 according to the estimated $\mathcal{A}$ of MS-STSK bits, as shown in Figure 3.12.

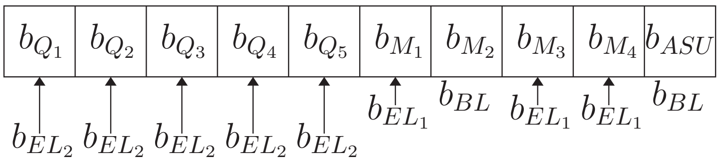
Explicitly, the more important bits of the BL are mapped to that specific MS-STSK subchannel, which has the largest area $\mathcal{A}$ under its MI curve representing the highest MI, as discussed in [127] [106]. The other video bits are mapped to the remaining MS-STSK subchannels that may exhibit lower MI. As a result, the bit mapping methods of the Bit Mapper shown in Figures 3.9 for *Mode* 2 and *Mode* 3 are given in Figure 3.12(a) and 3.12(b), respectively.

3.5 Simulation Results

In this section, we present our results for characterizing the near-instantaneously adaptive MS-STSK system proposed for UAV surveillance, where three MS-STSK configurations having distinct throughputs support triple-layer video streaming.



(a) *MS-STSK(4,2,2,2,4,4)*$|_{QAM}$.



(b) *MS-STSK(4,2,2,2,32,16)*$|_{QAM}$.

Figure 3.12: Bit mapping allocation.

The parameters of the video clip are listed in Table 3.2. A video clip represented in (352×288)-pixel CIF and 4:2:0 YUV format was encoded using the SHVC reference software's SHM codec model.

The scanning rate for the clip was set to 30 FPS and the GoP interval was set to 8 for all video simulations, which means that the Instantaneous Decoding Refresh (IDR)/Clean Random Access (CRA) [6] frames are inserted every 8 frames. No B frames were activated in our simulations, since they potentially lead to the inter-frame video distortions owing to inter-frame error propagation. Additionally, due to the fact that the B frames may introduce additional decoding latency, hence eroding flawless lip-synchronization, our encoded test video sequence only contains I frames and P frames. Our research-objective is to investigate the video-quality scalability by our simulations. For detecting the presence of residual errors, Cyclic Redundancy Check (CRC) codes are concatenated to the tail of each packet to ascertain the flawless SDP of the received packets. If the CRC detection fails, the corrupted packets are discarded and the corresponding abandoned video frames are replaced by "frame-copy" based error concealment relying on the most recent flawless video-frame.

Furthermore, Table 4.1 also tabulates the systematic configurations. The classic TCs consisting of a pair of identical RSCs having the generator polynomials of [111 101] is employed in the simulations, as given in Table 3.2, which is punctured to half rate. The number of inner iterations of the TC between the inner RSCs is fixed to 8. Recall that the channel model was given in Section 3.4.3.1, where the Ricean K-factor is set to $K = 0$ dB. Additionally, the MS-STSK configurations supporting the three different video modes of our adaptive system are listed in Table 3.3. Explicitly, *MS-STSK(2,2,2,2,2,2)*$|_{PSK}$, *MS-STSK(4,2,2,2,4,4)*$|_{QAM}$ and *MS-STSK(4,2,2,2,32,16)*$|_{QAM}$ are used for *Mode* 1, *Mode* 2 and *Mode* 3, respectively.

Let us now briefly focus our attention on the additional complexity imposed on the UAV by our proposed UEP design, which is mainly imposed by the Bit Mapper block of Figure 3.9. We partition the bit-mapping procedure into offline preparation and online operation. As part of offline

Table 3.2: Parameters employed of the simulation.

Parameters	
Representation	YUV 4:2:0
Format	CIF(352×288)
Bits Per Pixel	8
FPS	30
No. of Frames	64
No. of Layers	3
Video Codec	SHVC
GoP	8
Error-Free Y-PSNR (dB)	33.45, 36.27, 39.6
Bitrate (kbps)	239, 332, 654
Error Concealment	Frame-Copy
Channel Coding	TC [111 101]
TC Inner Iteration	8
Outer Iteration	2
Channel	Ricean (K = 0 dB)
Simulation Repeated	200

design, we illustrate the preparation required for designing the mapping scheme, while the online operation carries out the bit-mapping for achieving UEP.

To evaluate the BER of the individual MS-STSK bits of Figure 3.4, an empirical Monte Carlo simulation is employed using the parameters of Table 3.3, used in our simulations before carrying on the simulations. Then, the MI of all the MS-STSK bits of Figure 3.4 is plotted in Figure 3.8, which is then exploited for designing the bit-mapping scheme of Figure 3.12. Specifically, Section 3.4.4 provided our offline design guidelines for our MS-STSK assisted UEP scheme for a particular channel model and MS-STSK configuration. Hence, the offline generated mapping is specific for a particular channel, and MS-STSK configuration, but our method is readily applicable for different near-capacity FEC techniques, such as TC and LDPC codes.

The online complexity imposed by our proposed mapping design is dominated by the Bit Mapper block of Figure 3.9. The FEC encoded bits are fed into the Bit Mapper block, which rearranges the bits according to our mapping guideline of Figure 3.12, imposing negligible online complexity. Explicitly, the complexity of this block is similar to that of the classic bit mapping scheme of HQAM shown in [69].

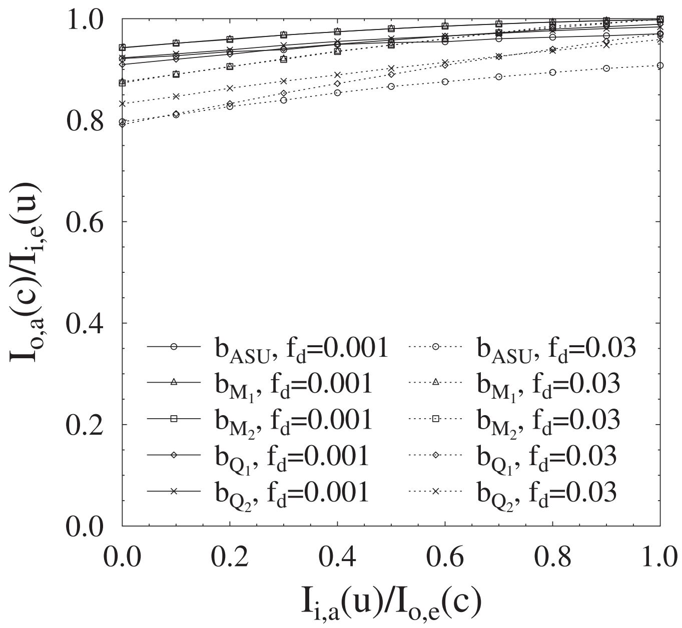


Figure 3.13: MI performance of our *MS-STSK(4,2,2,4,4)*$|_{QAM}$ under two different normalized Doppler frequencies, when $K = 0$ dB at a channel SNR of 7 dB.

Table 3.3: Bit allocations of the MS-STSK configurations.

Configurations	frame length (bits)	$b_{MS\text{-}STSK}$	b_{ASU}	b_M	b_Q
MS-STSK(2,2,2,2,2,2)$|_{PSK}$	1800	2	0	1	1
MS-STSK(4,2,2,2,4,4)$|_{QAM}$	4500	5	1	2	2
MS-STSK(4,2,2,2,32,16)$|_{QAM}$	9000	10	1	4	5

3.5.1 Doppler Effect on MS-STSK

Based on the AG model of Section 3.4.3.1, where the parameter of $\Delta f_{LOS} = f_d \cos(\phi_0)$ is determined by both f_d and AoA/AoD, for the sake of simplicity, we consider the scenario of $\phi = \phi_0 = \phi_r = \phi_t \in [-\pi, \pi]$ [122]. These values remain unchanged for $N_{PS}N_{OW}$ number of MS-STSK symbol durations. Figure 3.13 illustrates our MI comparison between the normalized maximum Doppler frequencies f_d of 0.001 and 0.03 based on our PSAM assisted channel estimator for transmission over Ricean channels, where the AoA/AoD angle ϕ is updated across GoP period in order to have accurate CSI. By comparing the scenarios of $f_d = 0.001$ and 0.03, we observe that the MI degradation becomes more severe, as f_d is increased, which is in line with our expectation. In the following simulations we consider the worst-case scenario of $\Delta f_{LOS} = f_d$ for the sake of simplicity.

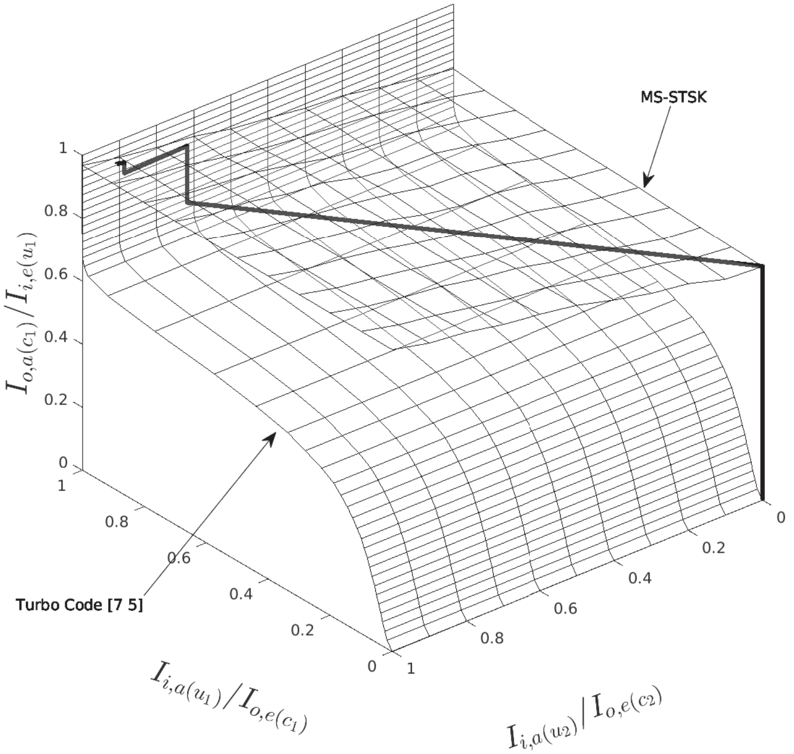


Figure 3.14: The trajectory of the BL between the MS-STSK configured with *MS-STSK(4,2,2,2,4,4)*$|_{QAM}$ and the half rate TC [7 5] for *Mode* 2 at the channel SNR of 3 dB.

3.5.2 MI Exchange Performance

Figure 3.14 portrays the 3D EXIT chart [128] based iterative decoding trajectory of the l_1 between the MS-STSK transceiver and the half rate TC when *Mode* 2 is selected, where only two video sub-layers, namely l_1 and l_2, are transmitted, as shown in Figure 3.12(a). Explicitly, $I_{i,a(u_1)}$ and $I_{i,a(u_2)}$ represents the *a priori* information of l_1 containing b_{M_1} and b_{M_2} and of l_2 containing b_{Q_1}, b_{Q_2} and b_{ASU}, as shown in Figure 3.12(a), respectively. The MS-STSK plane of $I_{i,e(u_1)}$ in Figure 3.14 only represents the MI of the bit stream of l_1, which is determined by the *a priori* information of both l_1 and l_2. Observe in Figure 3.14 that the increase of $I_{i,e(u_1)}$ is determined by both $I_{i,a(u_1)}$ and $I_{i,a(u_2)}$. The trajectory shown in Figure 3.14 exhibits the MI transfer between the TC decoder and the MS-STSK transceiver for l_1. In Figure 3.14, the received signals are demapped by the MS-STSK first, where no *a priori* information $I_{i,a(u_1)}$ and $I_{i,a(u_2)}$ is provided, yielding the *a posteriori* MI $I_{i,e(u_1)}$ of about 0.74, which is then fed into the TC decoder as its *a priori* information of $I_{o,a(c_1)}$. With the aid of the *a priori* information of $I_{o,a(c_1)}$, the TC carries out the first outer iteration and returns the *a posteriori* MI of $I_{o,e(c_1)}$, entering it into the MS-STSK demapper as the *a priori* information of $I_{i,a(u_1)}$. Then the second outer iteration starts. Observe in Figure 3.14 that the improvement after the second outer iteration becomes fairly minor. Therefore, the number of outer iterations between the TC and the MS-STSK transceiver is set to 2, based on our observation of

Figure 3.14, namely that the improvements gleaned by more outer iterative decoding operations become negligible.

3.5.3 Mode Switching Performance

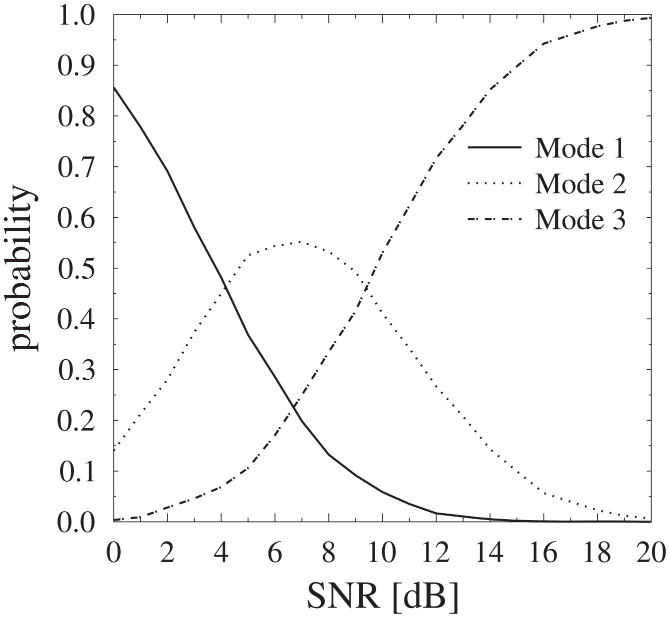
Table 3.4: Parameter sets for different mobilities.

Mobility Scenario	1	2	3	4
f_d	0.001	0.01	0.02	0.03
σ_{sf} (dB)	4	5	6	7
$t_{1_{UEP}}$ (dB)	3.3	3.5	3.9	4.5
$t_{1_{EEP}}$ (dB)	4	4.3	4.8	5.2
$t_{2_{UEP}}$ (dB)	9.2	9.5	9.8	10.3
$t_{2_{EEP}}$ (dB)	10.2	10.5	11	11.4

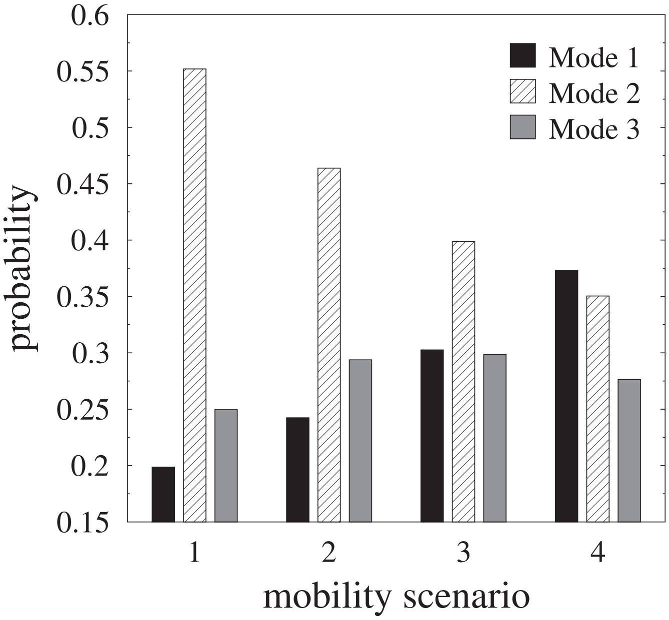
Table 3.4 lists the parameters for different mobility scenarios, where f_d and σ_{sf} represent the normalized maximum Doppler frequencies and the standard deviation of shadow fading that increases with the mobility of the UAV, as observed in [129] [130]. Furthermore, t_1 and t_2 are the pre-recorded thresholds used for controlling the mode-switching operation by comparing the estimated instantaneous channel SNRs to the minimum threshold in the different mobility scenarios. Since the increased Doppler frequency degrades the accuracy of channel estimation, unless the pilot density is increased proportionally, a higher Doppler frequency requires higher SNR for switching to a higher-throughput video mode in order to maintain the robustness of the system. The configuration of *MS-STSK(2,2,2,2,2,2)*$|_{PSK}$ seen in Table 3.3 is set as the default mode, while the configurations of *MS-STSK(4,2,2,2,4,4)*$|_{QAM}$ and *MS-STSK(4,2,2,2,32,16)*$|_{QAM}$ are activated, when the channel SNR exceeds t_{1,f_d} and t_{2,f_d}, respectively. To benchmark our proposed UEP-aided near-instantaneously adaptive system, the EEP based adaptive MS-STSK counterpart is considered as the benchmarker, where the bit-to-MS-STSK subchannel mapping of Figure 3.12 employed for the UEP scheme is disabled. The near-instantaneously adaptive mode-switching thresholds designed for this adaptive EEP system are also recorded in Table 3.4, requiring an additional 1.1 dB of channel SNR for switching to the high-throughput video modes.

3.5.4 Quality of Experience Performance

Figure 3.15(a) depicts the PDF of the three fixed modes versus the channel SNR for the thresholds listed in Table 3.4. Observe in Figure 3.15(a) that for the mobility Scenario 1 described in Table 3.4, *Mode* 1 is the most frequently used one at an average channel SNR below 4 dB, followed by *Mode* 2, while *Mode* 3 is the least used mode. By contrast, at a channel SNR of 10 dB, the probability of activating *Mode* 3 is higher than that of *Mode* 1, but *Mode* 2 becomes the most



(a) PDF of the three operating modes versus channel SNR for the mobility Scenario 1 of Table 3.4.



(b) PDF of the three operating modes versus the mobility Scenarios at SNR of 7 dB.

Figure 3.15: PDF of the three fixed modes of operation.

frequent mode. To elaborate a little further, Figure 3.15(b) shows that upon increasing the mobility, the robust *Mode* 1 becomes more dominant, when aiming for guaranteeing the best possible

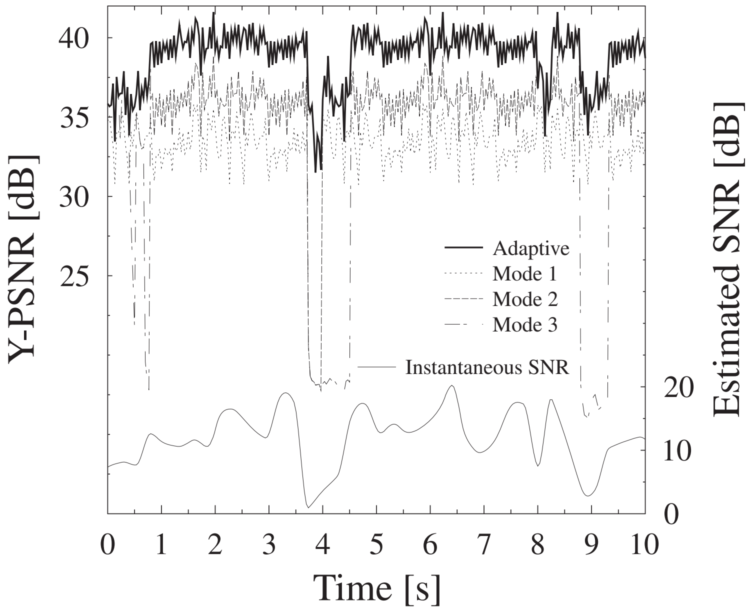


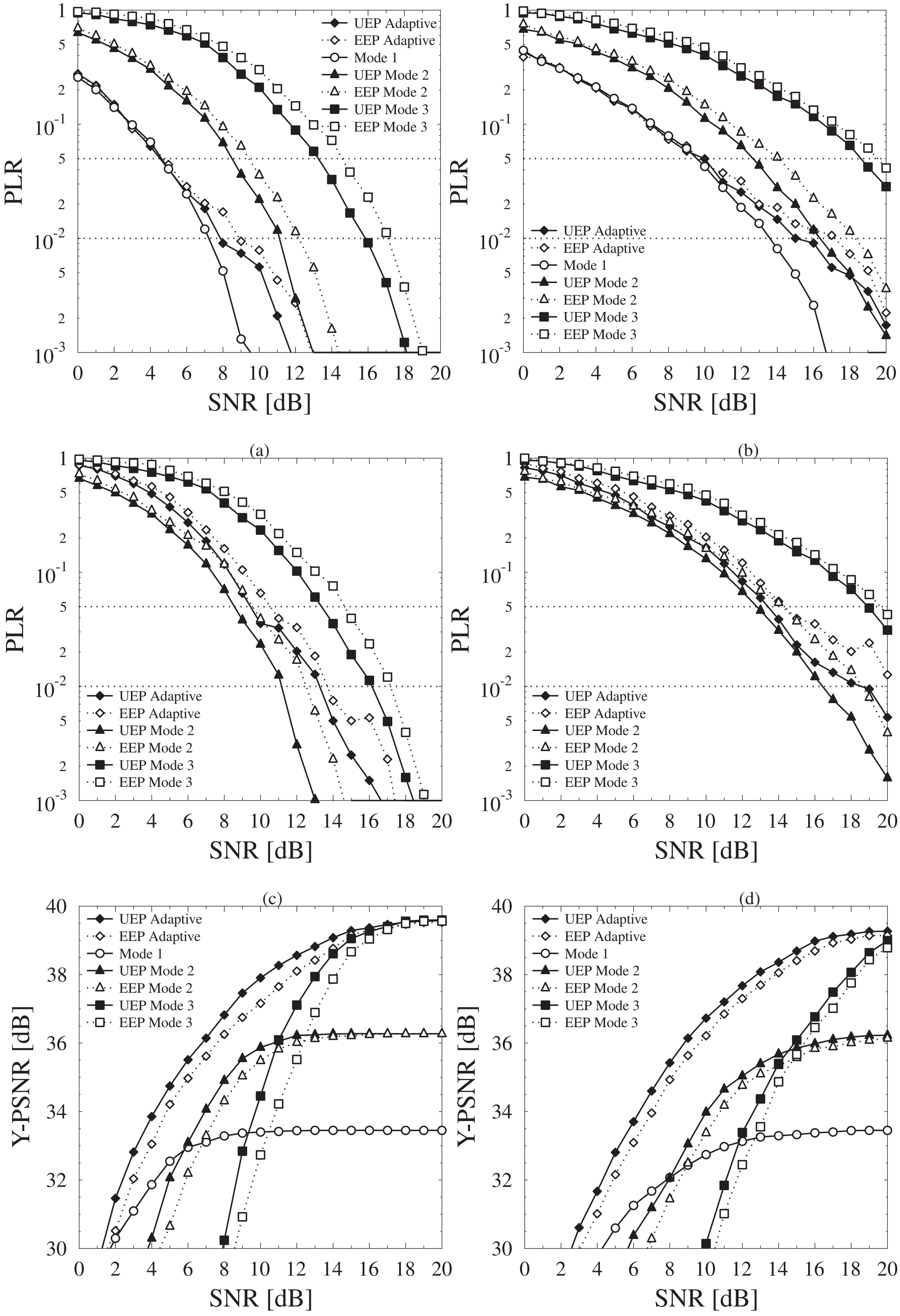
Figure 3.16: Video quality in Y-PSNR versus time for UEP assisted adaptive design, Mode 1, Mode 2 and Mode 3, respectively, for mobility Scenario 1 at the average channel SNR of 12 dB.

performance of the system.

In the time interval shown in Figure 3.16, the sudden PSNR reductions of the video quality suffered by the higher-rate fixed Mode 2 and Mode 3 indicate originally higher but gravely error-infested video quality. This grave perceptual video degradation is circumvented by the activation of the inherently lower-PSNR, but more robust Mode 1 of our adaptive design. As a benefit, the adaptive scheme exhibits the highest PSNR, as seen in Figure 3.17 in more detail.

To elaborate, Figures 3.17(a), (b), (c) and (d) show the simulation results in terms of the PLR, while Figures 3.17(e) and (f) in terms of the PSNR versus channel SNR for the mobility Scenarios 1 and 3. More explicitly, the three columns from left to right represent the PLR value of l_1 and l_2 as well as the image quality (PSNR), respectively, while mobility Scenarios of 1 and 3 are set for the first row ((a)(c)(e)) and the second row ((b)(d)(f)), respectively. Since no UEP is employed for $Mode$ 1 only transmitting l_1, this fixed mode is shared by both the UEP and EEP based adaptive system.

It can be seen in Figure 3.17 that the system performance of mobility Scenario 3, shown in the second row of Figure 3.17 is on average worse than that of mobility Scenario 1, shown in the first row. This is because the higher mobility of the UAV not only results in more dramatically fluctuating channels but also degrades the accuracy of the PSAM-aided channel estimation, therefore

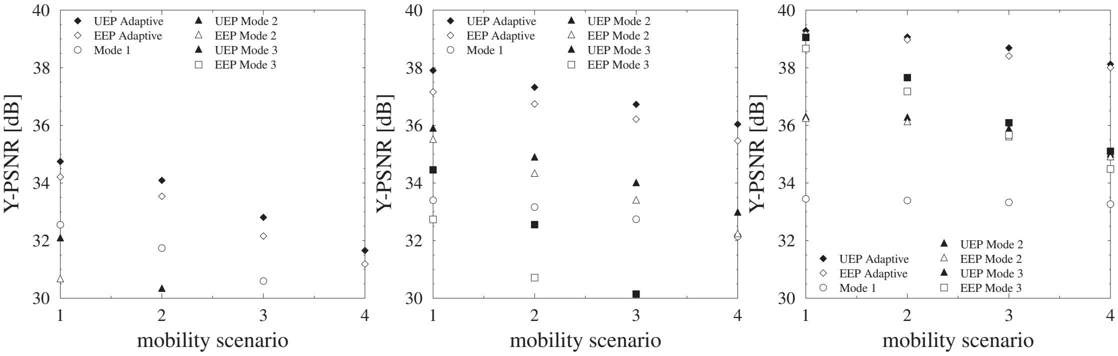


(a) (b)

(c) (d)

(e) (f)

Figure 3.17: The comparison between the three fixed modes and the adaptive system, where the first, second and third rows represent the PLR value of the l_1, the e-PLR value of l_2 and the image quality, respectively, for mobility Scenario 1 (first column) and 3 (second column).



(a) $SNR = 5\ dB$.

(b) $SNR = 10\ dB$.

(c) $SNR = 15\ dB$.

Figure 3.18: Image quality versus mobility scenarios for the EEP and UEP schemes.

significantly degrading the overall system performance.

Figure 3.17 shows that our proposed UEP allocation methods of Figure 3.12 invoked for the MS-STSK configuration improves the system performance both in terms of the PLR and the reconstructed video quality (Y-PSNR) compared to their EEP counterparts for both the fixed modes and the adaptive designs. The PLR value $P(l_1)$ of l_1 is portrayed in Figures 3.17(a) and (b) for mobility Scenarios of 1 and 3, respectively, which indicates that our proposed UEP design attains about 1 dB channel SNR gain at $P(l_1) = 5\%$ recorded for the fixed modes. Furthermore, the $P(l_1)$ difference of the adaptive systems between the UEP and EEP becomes distinct in the high-channel-SNR region, where *Mode* 2 and *Mode* 3 become more frequently selected, as shown in Figure 3.15. Having said that, both the UEP based or the EEP based adaptive system exhibit a similar $P(l_1)$ value to that of the fixed *Mode* 1. Ultimately, the UEP based design shows some improvement at the high SNR region.

The $P(l_2)_e$ value is portrayed in Figures 3.17(c) and (d), which represents the equivalent PLR of l_2 by taking into account its reference layer, namely l_1, as well. Observe in Figures 3.17(c) and (d) that the $P(l_2)_e$ improvement of the UEP *Mode* 2 is visible in both mobility scenarios, while the UEP assisted adaptive system yields a lower PLR than its EEP assisted counterpart. A slight PLR degradation can be found in Figures 3.17(c) and (d) for the adaptive designs compared to that of their associated fixed Mode 2, since the bit streams of l_2 may be occasionally dropped when Mode 1 is selected.

Figures 3.17(e) and (f) illustrate the Y-PSNR performance comparison between UEP and EEP for both the fixed modes and the adaptive schemes, where the reconstructed video quality of the UEP scheme is better than that of its EEP counterparts. It can be seen in Figures 3.17(e) and (f) that the UEP assisted scheme is capable of improving the average Y-PSNR by requiring about 1 dB lower channel SNR in comparison to its corresponding EEP assisted counterparts in both mobility scenarios.

(a) Mode 1 (b) Error in Mode 1

(c) Adaptive (d) Error in Adaptive Scheme

Figure 3.19: The subjective image comparison at the instantaneous channel SNR of 4 dB

Additionally, Figure 3.17 also depicts the improvements attained by our adaptive designs that judiciously activate the most appropriate mode compared to their corresponding fixed counterparts, which is outlined in Figures 3.17(e) and (f). The gain of the image quality (Y-PSNR) improved by the adaptive system becomes more distinct in the mobility Scenario 3, where the instantaneous channel quality fluctuates more dramatically due to the higher shadow fading.

Figure 3.18 illustrates the reconstructed video quality of the associated mobility scenarios at the average channel SNRs of 5 dB, 10 dB and 15 dB, respectively. It can be seen in Figure 3.18 that the adaptive systems relying either on UEP or EEP are capable of efficiently mitigating the variations of the channel quality imposed by the motion of the UAVs by judiciously selecting the appropriate operating modes, especially in high-mobility scenarios. This results in a more graceful video quality erosion than that of the other fixed modes. In a nutshell, the proposed UEP assisted system further improves the video quality by using the mapping designs of Figure 3.12 at the expense of a modest complexity increase.

Finally, the subjective image quality recorded at the instantaneous channel SNR of 4 dB is shown in Figure 3.19. Explicitly, observe in Figure 3.19 that the adaptive system is capable of switching to the higher-throughput, higher-PSNR modes, hence exhibiting a more detailed video texture in Figure 3.19(c) than that of its Mode 1 counterpart in Figure 3.19(a). To make the visual comparison more explicit, in Figure 3.19(b) and (d) we portrayed the error between the original and received video frames, where the latter exhibits a lower error.

3.6 Chapter Conclusion

Table 3.5: The channel SNR required for $p(l_1) \leq 5\%$ of the test sequence.

Mobility Scenario	1	3
Mode 1	4.6 dB	9.5 dB
UEP Mode 2	8.5 dB	12.7 dB
EEP Mode 2	9.5 dB	14 dB
UEP Mode 3	13.2 dB	18.6 dB
EEP Mode 3	14.7 dB	19.5 dB
UEP Adaptive	4.6 dB	9.5 dB
EEP Adaptive	4.6 dB	9.5 dB

Table 3.6: The channel SNR required for ensuring Y-PSNR=32 dB of the test sequence.

Mobility Scenario	1	3
Mode 1	4.2 dB	7.9 dB
UEP Mode 2	4.9 dB	7.9 dB
EEP Mode 2	5.9 dB	8.5 dB
UEP Mode 3	8.6 dB	11.1 dB
EEP Mode 3	9.5 dB	11.7 dB
UEP Adaptive	2.3 dB	4.3 dB
EEP Adaptive	3 dB	4.9 dB

In this chapter, we exploited the UEP capability of the MS-STSK on the component- and bit-level basis, where a corresponding adaptive MS-STSK system was conceived for layered video streaming over Ricean channel. In Section 3.2, we reviewed the development of the MIMO transceiver techniques, including SM and STSK, followed by the introduction of the novel MS-STSK transceiver. Section 3.3 verified that the MS-STSK transceiver is capable of providing UEP by feeding the video source bits having different importance into the corresponding MS-STSK components, which is then extended to the near-capacity research with the aid of the EXIT chart.

Assisted by the exploitation in Section 3.3, we conceived an adaptive MS-STSK assisted SVC streaming system in Section 3.4 capable of combating the high Doppler spread, where the ELs have to be dropped when the low-capacity MS-STSK configuration is selected at the low SNR. The proposed system was then quantified in Section 3.5, which shows that the proposed UEP scheme improve the PLR and video quality performance compared to its EEP counterpart in both the fixed and adaptive modes. Observe in Table 3.5 that the proposed UEP Mode 2 and Mode 3 attain about 1-1.5 dB SNR gain when achieving the PLR of 5%. Additionally, both the adaptive designs show about 2-3.5 dB channel SNR gain compared to their fixed counterparts, when reaching the PSNR of 32 dB, as shown in Table 3.6.

Chapter 4

Adaptive IL-FEC Assisted SVC Streaming

4.1 Introduction

In Chapters 2 and 3, we designed that the bit mapping assisted UEP schemes for layered video streaming over dispersive wireless channels. More explicitly, in Chapter 3, we conceived IM based bit mapping assisted UEP schemes for providing stronger protection for the more important BL. Based on this UEP philosophy, in this chapter a sophisticated IL-FEC technique relying on channel coding based UEP is embedded into IM for layered video streaming, where UEP is joint provided both by IM and IL-FEC. Specifically, the principle of the IL-FEC technique is to embed the bits of the BL stream into those of another independent bit stream, such as the EL for example by using their modulo-2 connection. Then this embedded information can be used at the decoder side as extrinsic information for enhancing the robustness of the original BL bit stream. Hellge *et al.* [58] conceived a Layer-Aware Forward Error Correction (LA-FEC) scheme for two-layer video streaming, where the source bits of the more important BL are embedded into the parity bits of the EL with the aid of a predefined pattern. As further development, a more sophisticated physical layer based IL-FEC was proposed by Huo *et al.* [65], where the source bits of the BL are implanted into those of its ELs, so that the soft iterative MAP decoding algorithms can be invoked for enhancing the SDP of the BL.

However, all these methods focus on improving the robustness of the BL at the expense of the EL, which cannot be directly employed in our IM assisted UEP scheme due to its inability to protect the EL. Hence, we propose an enhanced IL-FEC technique for enhancing the protection of the EL, where the bits of an EL are implanted into those of the other layers, provided that the bits of these reference layers are well protected. The simulation results show that about 2 dB channel SNR gain can be achieved by the proposed enhanced IL-FEC compared to its conventional counterpart, provided that the PLR of their protected layer does not exceed 5%.

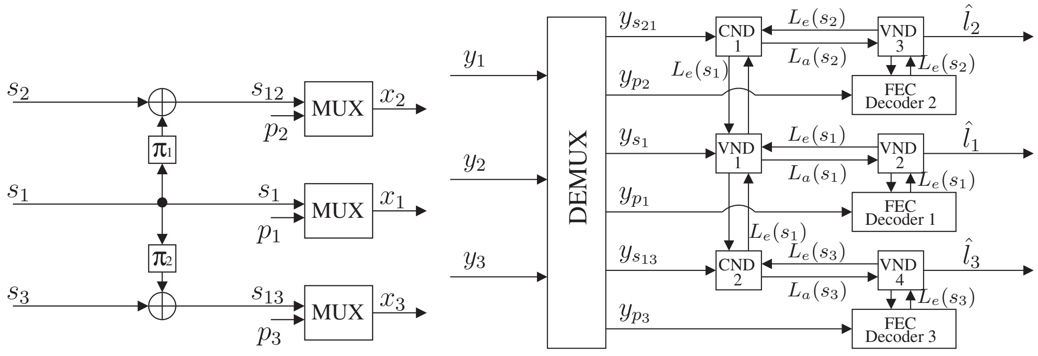
The chapter is organized as follows. Firstly, in Section 4.2, we commence by introducing the conventional IL-FEC technique proposed in [65]. Then, the enhanced IL-FEC scheme capable of providing protection not only for the BL but also for the EL is conceived, followed by the system model proposed in Section 4.3. The system performance is quantified in Section 4.4, where the IL-FEC assisted adaptive system is used as a benchmarker for illustrating the improvement attained by the proposed design. Finally, we conclude in Section 4.5.

4.2 Inter-layer FEC

In this section, we commence by introducing the concept of the IL-FEC technique conceived in [65], followed by our enhanced IL-FEC scheme that is capable of providing protection for the ELs.

4.2.1 Conventional IL-FEC

Figure 4.1(a) illustrates the schematic of the IL-FEC encoder conceived in [60], where the three bit stream l_1, l_2 and l_3 are input into their associated FEC Encoders, generating the corresponding systematic bit streams s_1, s_2 and s_3 as well as their associated parity streams p_1, p_2 and p_3. Then, in order to enhance the robustness of l_1, the systematic bit sequence s_1 is first scrambled by the interleavers π_1 and π_2, respectively, whose outputs are then implanted into the other two streams, namely s_2 and s_3, by a modulo-2 operation, resulting in the associated pair of mixed bit streams, namely s_{12} and s_{13}, respectively. It is worth noting that all the three parity streams remain unchanged and they are transmitted together with the three IL-FEC encoded systematic streams.



(a) Encoder

(b) Decoder

Figure 4.1: The (a) Encoder and (b) Decoder of the conventional IL-FEC technique.

At the decoder depicted in Figure 4.1(b), there are six inputs, y_{s_1}, y_{p_1}, $y_{s_{12}}$, y_{p_2}, $y_{s_{13}}$ and y_{p_3}, which represent the received version of s_1, p_1, s_{12}, p_2, s_{13} and p_3. To iteratively exploit the IL-FEC dependencies amongst all the layers, the classic VND and CND concepts are invoked for exchanging extrinsic information [60], as illustrated in Figure 4.2. Explicitly, assuming that u_1, u_2 are the random binary variables, where $u_3 = u_1 \oplus u_2$, the VND sums two LLR inputs for generating a more reliable LLR output, as formulated as $L_{o3}(u_1) = L_{i1}(u_1) + L_{i2}(u_1)$. The boxplus operation [131] of $L(u_3 = u_1 \oplus u_2) = L(u_1) \boxplus L(u_2)$ contributes to improving the reliability of the bit u_3, given the fact that the reliability of the bits u_1 and u_2 is known, which was defined by Hagenauer as [132]

$$\begin{aligned} L(u_3) &= \log \frac{1 + e^{L(u_1)} e^{L(u_2)}}{e^{L(u_1)} + e^{L(u_2)}} \\ &\approx sign(L(u_1)) \cdot sign(L(u_2)) \cdot \min(|L(u_1)|, |L(u_2|) \end{aligned} \tag{4.1}$$

under the additional rules of

$$L(u) \boxplus \pm\infty = \pm L(u) \tag{4.2}$$

$$L(u) \boxplus 0 = 0. \tag{4.3}$$

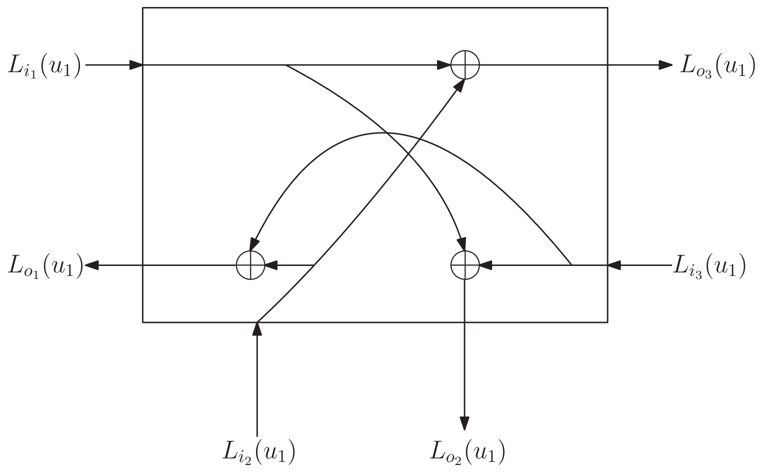
Therefore, to generate the soft information representing u_3, the operation of the CND can be formulated as $L_o(u_3) = L_i(u_1) \boxplus L_i(u_2)$, assuming that the soft LLRs of u_1 and u_2 are known.

Due to the fact that the systematic bit streams s_2 and s_3 are implanted into s_1, they cannot be independently decoded by their corresponding FEC Decoders, as shown in Figure 4.1(b). Specifically, to decode both bit streams successfully, the decoding process has to obey a specific order. For example, it has to decode s_1 before decoding s_2 and s_3, since decoding s_2 and s_3 requires the extrinsic information gleaned from s_1. As shown in Figure 4.1(b), the decoding process is described as follows

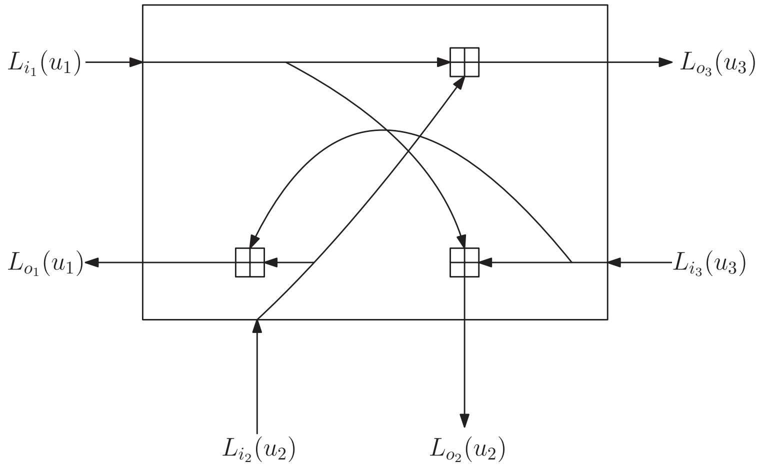
1. The LLR of y_{s_1} is fed into VND 1 of Figure 4.1(b). As at this stage no extrinsic information is provided by the CNDs, y_{s_1} is simply forwarded to VND 2 as $L_a(s_1)$. Then FEC Decoder 1 generates the extrinsic information $L_e(s_1)$ by also taking into account the associated received soft parity y_{p_1}, which is passed back to VND 1 via VND 2.

2. At this stage, VND 1 of Figure 4.1(b) updates the LLR of y_{s_1} with the aid of extrinsic information and forwards $L_e(s_1)$ to CND 1 and CND 2, hence enabling them to generate the *a priori* information $L_a(s_2)$ and $L_a(s_3)$ by additionally taking into account $y_{s_{12}}$ and $y_{s_{13}}$, respectively. FEC Decoder 2 and Decoder 3 of Figure 4.1(b) receive soft information of $L_a(s_2)$ and $L_a(s_3)$ as well as their associated parity streams y_{p_2} and y_{p_3} in order to generate the extrinsic information $L_e(s_2)$ and $L_e(s_3)$, respectively, which is then passed back to the CND 1 and CND 2 for enhancing the *a priori* information of $L_a(s_1)$.

3. With the aid of the extrinsic information passed to it by CNDs, VND 1 of Figure 4.1(b) updates the *a priori* information furnished for s_1 and hence the corresponding confidence is enhanced, resulting in an improved BER performance, when the first iteration is completed. The iterative decoding process continues until the maximum number of iterations is reached.



(a) Conventional VND



(b) Conventional CND

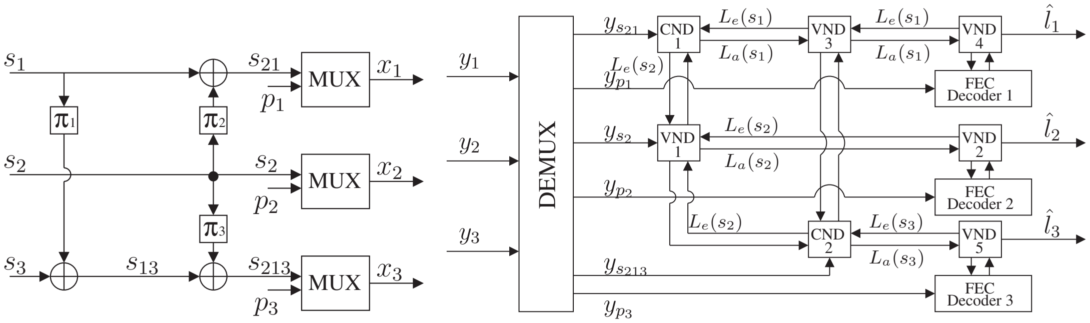
Figure 4.2: The (a) VND and (b) CND of the conventional IL-FEC technique.

4. When the affordable number of iterations is exhausted, the decoded bit streams $\hat{l}_1$, $\hat{l}_2$ and $\hat{l}_3$ are generated by VND 2, VND 3 and VND 4 of Figure 4.1(b), respectively. To obtain $\hat{l}_1$, VND 2 of Figure 4.1(b) adds up $L_a(s_1)$ gleaned from VND 1 and $L_e(s_1)$ arriving from FEC Decoder 1, while $\hat{l}_1$ and $\hat{l}_2$ are generated by VND 3 and VND 4 of Figure 4.1(b), respectively.

4.2.2 Proposed Enhanced IL-FEC

In Section 4.2.1, we presented the conventional IL-FEC scheme proposed by Huo *et al.* [32], where the systematic bits of the BL are implanted into the ELs. The implanted bits extracted from the ELs at the decoder contribute to enhancing the LLRs of the BL, hence improving the overall reconstructed video quality. However, this method is only capable of providing protection for the BL, but it cannot be directly employed for protecting the ELs. In this section, we propose an enhanced IL-FEC technique for providing improved protection for the ELs, where in addition to protecting the EL, the robustness of the reference layer of the EL is also carefully taken into account. As a result, the decoder of Figure 4.1(b) has to be particularly designed.

Figure 4.3 illustrates our enhanced IL-FEC scheme, which is exemplified using the first EL, namely l_2. Similar to the notations defined in Section 4.2.1, s_i, p_i and x_i represent the systematic bits, the parity bits and the final concatenated bits of the i-th layer, respectively, while π is the interleaver between the different layers employed for guaranteeing the interleaved bits remain independent of each other. In Figure 4.3, we can see that before implanting the bits of s_2 into the other layers, the systematic bits of l_1 are first implanted into s_3 after the interleaving operation of π_1, generating the mixed bits of s_{13}. Then, the systematic bits of l_2, namely s_2, are implanted both into s_1 and into the mixed bit stream s_{13}, respectively, which outputs a new pair of systematic streams of s_{21} and s_{213}, as shown in Figure 4.3(a). Finally, the three systematic streams and the corresponding parity bits are concatenated, yielding the bits of x_1, x_2 and x_3 for modulation, as shown in Figure 4.3.

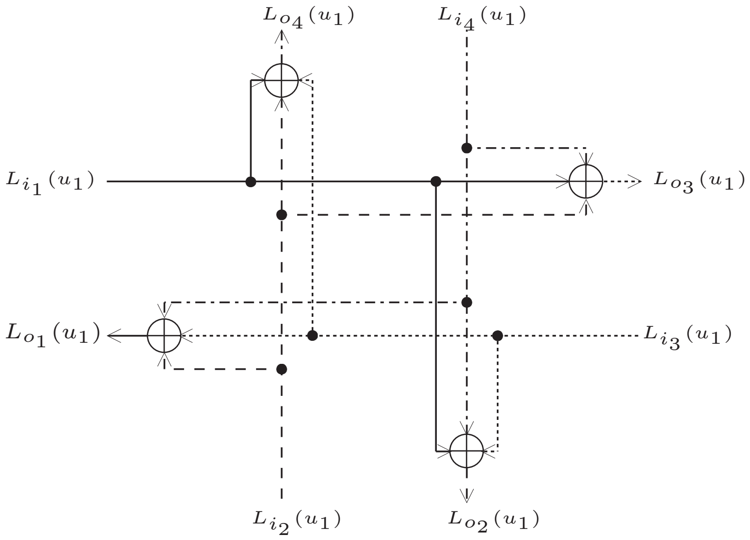


(a) Encoder (b) Decoder

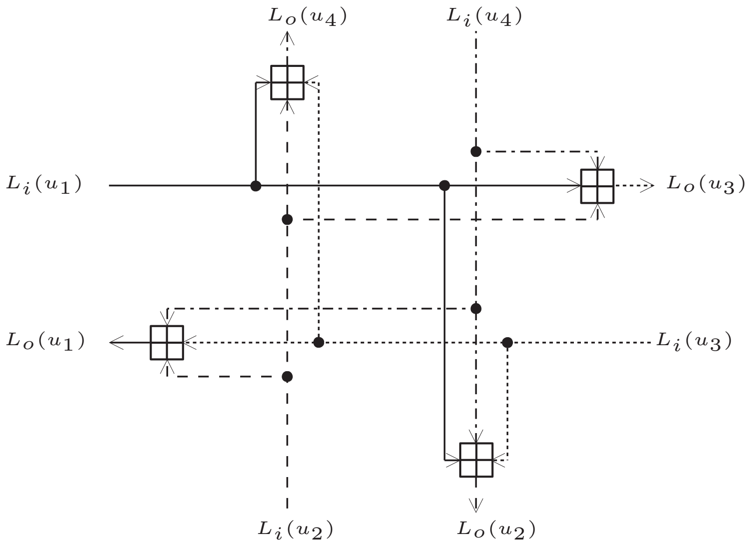
Figure 4.3: The (a) Encoder and (b) Decoder of the enhanced IL-FEC technique.

The proposed decoding procedure of the enhanced IL-FEC is illustrated in Figure 4.3, where y_1, y_2 and y_3 represent the received version of x_1, x_2 and x_3, respectively. The three bit streams are demultiplexed by the DEMUX block of Figure 4.3(b) to generate the systematic information, namely $y_{s_{21}}$, y_{s_2} and $y_{s_{213}}$, as well as their corresponding parity bits, namely, y_{p_1}, y_{p_2} and y_{p_3}, for the layers l_1, l_2 and l_3, respectively. Since the systematic bits s_1 and s_2 are implanted into s_3 while s_2 is implanted into s_1, the received version of s_2, namely y_{s_2}, can be decoded independently, while decoding s_1 requires information about s_2 and decoding s_3 requires the assistance of both s_1 and s_2. Thus, the decoding process of Figure 4.3 must follow the sequential order: y_{s_2}, $y_{s_{12}}$ and $y_{s_{213}}$. Therefore, as depicted in Figure 4.3(b), the decoding process obeys the following steps

1. The systematic bit of the IL-FEC protected layer, namely y_{s_2}, as shown in Figure 4.3(b), is fed into VND 1 to generate the corresponding *a priori* information of $L_a(s_2)$ assisted by the extrinsic information gleaned from both CND 1 and CND 2. Furthermore, $L_a(s_2)$ is generated by simply duplicating the soft LLR of y_{s_2}, since at this stage no extrinsic information



(a) VND



(b) CND

Figure 4.4: Structures of (a) VND and (b) CND for enhanced IL-FEC.

is provided by the CNDs. Then, $L_a(s_2)$ is fed to the FEC Decoder 2 of Figure 4.7(b) along with its corresponding parity y_{p_2}, hence generating the extrinsic information $L_e(s_2)$. This extrinsic LLR is then fed back to VND 1 via VND 2 to produce the extrinsic information for CND 1 and CND 2 with the aid of y_{s_2}, as seen in Figure 4.3(b).

2. CND 1 of Figure 4.3(b) receives the extrinsic LLR of $L_e(s_2)$ from VND 1 and then extracts y_{s_1} from $y_{s_{21}}$. The extrinsic information obtained is then fed into VND 3 to generate $L_a(s_1)$, which is equal to y_{s_1}, since the bits in l_3 have not as yet been processed, hence no extra information is provided by CND 2 for VND 3. The extrinsic information $L_e(s_1)$ generated by FEC Decoder 1 is passed back to VND 3 and CND 1 for providing extra information both for CND 2 and VND 1, respectively.
3. With the aid of the output of VND 1 and VND 3, CND 2 becomes able to glean $L_a(s_3)$ from y_{213} and then forwards it to the FEC Decoder 3 of Figure 4.3(b) via VND 5 in order to generate the extrinsic information $L_e(s_3)$. As seen in Figure 4.3(b), CND 2 uses this extrinsic information together with $y_{s_{213}}$ and either $L_e(s_2)$ or $L_e(s_3)$ to generate the feedback information for VND 1 and VND 3, respectively, in order to prepare for the next iteration.
4. Then, the decoding process of Figure 4.3(b) starts again from VND 1. However, in contrast the procedure in Step 1, the extrinsic information gleaned from CND 1 and CND 2 is no longer zero, since the related soft information has been exchanged among the three FEC Decoders of Figure 4.3(b), hence improving the soft information $L_a(s_2)$. Similarly, the *a priori* information $L_a(s_1)$ is enhanced by exploiting the extrinsic information of VND 3, which results in an enhanced BER performance for l_1. When the maximum number of iterations is reached, VND 2, 4 and 5 output the final LLR generated by considering both $L_a(s_i)$ and $L_e(s_i)$, which is then hard-decoded to $\hat{l}_2$, $\hat{l}_1$ and $\hat{l}_3$, respectively.

Furthermore, due to the fact that the enhanced IL-FEC decoder takes into account the reference bits of the protected layer, the corresponding VND and CND operations need to be updated as well, as shown in Figure 4.4. Assuming that u_1, u_2 and u_3 are random binary variables and that we have $u_4 = u_1 \oplus u_2 \oplus u_3$, the VND now is updated to sum the three LLR inputs for generating a more reliable LLR output, which may be formulated as $L_{o_4}(u_1) = L_{i_1}(u_1) + L_{i_2}(u_1) + L_{i_3}(u_1)$. The boxplus operation of $L_o(u_4 = u_1 \oplus u_2 \oplus u_3) L_{i_1}(u_1) + L_{i_2}(u_1) + L_{i_3}(u_1) = L_i(u_1) \boxplus L_i(u_2) \boxplus L_i(u_3)$ contributes to improving the reliability of the bits, given that the reliability of the bits u_1, u_2 and u_3 is known, which also can be calculated from (4.1).

4.3 System Model

In this section, we elaborate on our proposed layered video streaming architecture, where the protected video layer in the enhanced IL-FEC technique of Figure 4.3 is adaptively selected according to the instantaneous channel SNR. The IL-FEC encoded bits are then fed into different-integrity MS-STSK subchannels according to their importance, where the bits of the most important layer are mapped to the specific component exhibiting the lowest BER.

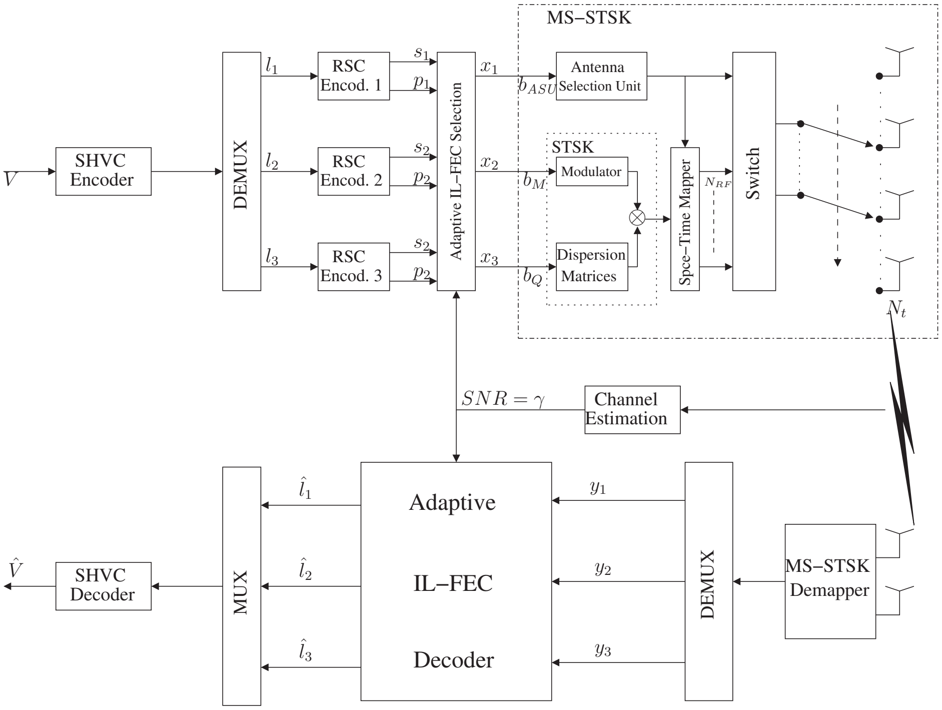
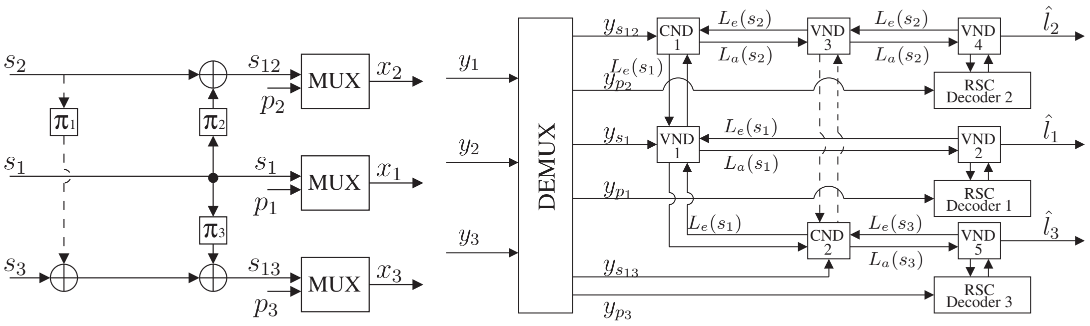


Figure 4.5: Architecture of proposed MS-STSK aided adaptive system for SVC streaming.

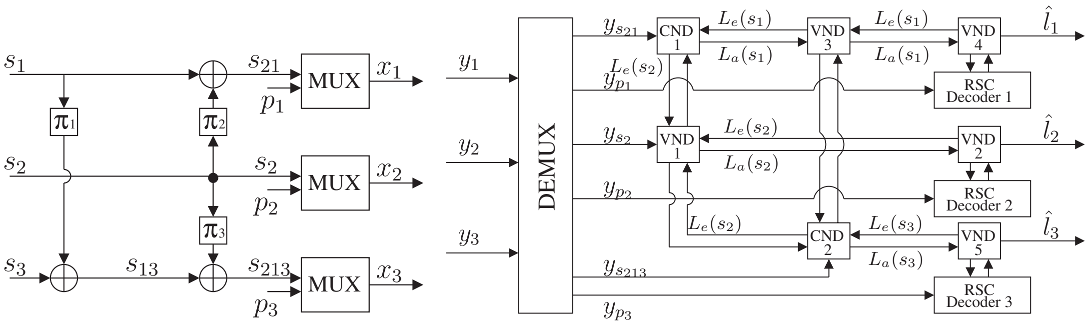
4.3.1 Proposed Transmitter Model

At the transmitter shown in Figure 4.5, the captured video source V is first compressed by the SHVC encoder, generating a bit stream containing multiple layers, which is then demultiplexed into the three bit streams each for a specific layer, namely l_1, l_2 and l_3 corresponding to the BL, first EL and second EL, respectively. These bit streams are then separately encoded by the three identical RSC encoders, as shown in Figure 4.5. The output of the RSC encoders results in six bit streams, including three systematic streams referred to as s_1, s_2 and s_3 as well as three parity streams, p_1, p_2 and p_3. The adaptive IL-FEC selection unit and decoder of Figure 4.5 aim for adaptively configuring the IL-FEC scheme to judiciously assign protection to the layers. There are three fixed-mode candidates provided for the adaptive IL-FEC selection unit, namely *Mode* 1, *Mode* 2 and *Mode* 3, which aim for protecting the BL, first EL and second EL, respectively. However, due to the inability of the conventional IL-FEC to protect the ELs, we propose an enhanced IL-FEC solution to be detailed later in this section. The modified *Mode* 1, *Mode* 2 and *Mode* 3 are depicted in Figures 4.6, 4.7 and 4.8, respectively, where the most appropriate mode is activated according to the instantaneous channel SNR, which simply implies switching the implantation mode. Then, the bit streams generated by the adaptive IL-FEC selection unit shown in Figure 4.5 are then fed into



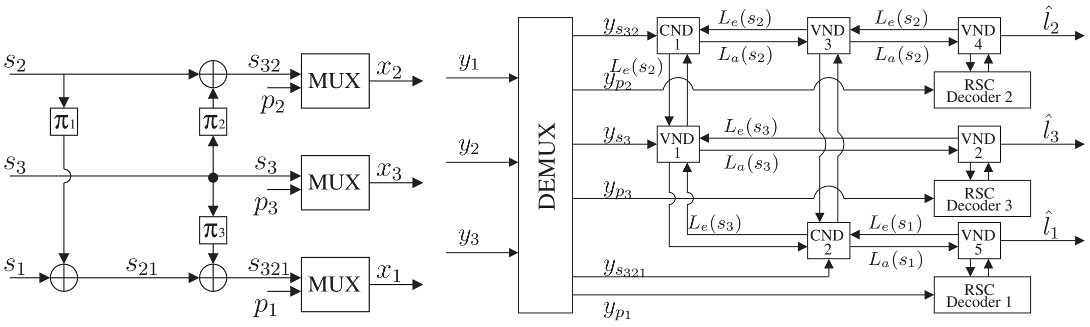
(a) Encoder (b) Decoder

Figure 4.6: The (a) Encoder and (b) Decoder of the IL-FEC technique for *Mode* 1, which achieves the identical function as that of Figure 4.1.



(a) Encoder (b) Decoder

Figure 4.7: The (a) Encoder and (b) Decoder of the IL-FEC technique for enhanced *Mode* 2.



(a) Encdoer (b) Decoder

Figure 4.8: The (a) Encoder and (b) Decoder of the IL-FEC technique for enhanced *Mode* 3.

the MS-STSK transceiver of Figure 3.3, where the BL, first EL and second EL bits are forwarded to the ASU block, to the classic modulator and to the dispersion matrices generator, respectively, bearing in mind their different BER performances shown in Figure 3.6. Again, a frequency-flat Rayleigh fading plus shadow fading channel is considered. The protection modes conceived for the adaptive IL-FEC selection unit are described as follows

1. *Mode* 1: In order to protect the BL, the IL based protection applied to l_1 in our system is identical to that proposed in [29] [65], as shown in Figure 4.6(a), where the dotted line indicates that this implantation function is disabled, since the BL is independent of any other layers. It can be seen in Figure 4.6 that two copies of the systematic bit stream of the BL s_1 are interleaved and implanted into s_2 and s_3 using the conventional XOR operation according to $s_{12}^k = s_1^k \oplus s_2^k$ and $s_{13}^k = s_1^k \oplus s_3^k$, respectively. This results in the mixed bit-streams of s_{12}^k and s_{13}^k seen in Figure 4.6(a). The outputs become s_1, s_{12} and s_{13}, complemented by the three corresponding parity bit streams.
2. *Mode* 2: Figure 4.7 illustrates the enhanced *Mode* 2, where l_2 becomes the IL-FEC protected layer. Considering the dependency between l_2 and l_1, apart from assigning IL-FEC protection to l_2, the robustness of l_1 is also taken into consideration. Thus, first the systematic bits of l_1 are interleaved by π_1 and then implanted into s_3 for the sake of guaranteeing the performance of the BL, yielding the mixed sequence of $s_{13} = s_1 \oplus s_3$. Then, two copies of the systematic sequence of the protected layer s_2 are interleaved by π_2 and π_3, as shown in Figure 4.7, and then implanted into s_1 and s_{13}, respectively. This operation results in generating two new sequences, namely $s_{21} = s_2 \oplus s_1$ and $s_{213} = s_2 \oplus s_{13}$, while the other copy of bit stream s_2 is output directly. As shown in Figure 4.7, the IL-FEC-processed systematic bit streams become s_{21}, s_2 and s_{213}.
3. *Mode* 3: The process of assigning IL-FEC protection to l_3 is quite similar to that of the enhanced *Mode* 2, with the IL-FEC protected layer becoming l_3 instead of l_2, as shown in Figure 4.8. Note that instead of guaranteeing the BER performance of l_1 as in *Mode* 2, the system provides extra protection for l_2 by implanting the bit stream of s_2 into that of s_1. Therefore, the system first interleaves s_2 and then implants it into s_1, hence resulting in a new bit sequence of $s_{21} = s_2 \oplus s_1$. Then, as observed in Figure 4.8, two copies of the systematic bit stream s_3 are interleaved and implanted into s_{21} and s_2, hence resulting in the new mixed streams of $s_{321} = s_3 \oplus s_{21}$ as well as $s_{32} = s_3 \oplus s_2$, while the other copy remains unchanged. Finally, the new outputs representing the systematic bits become s_{321}, s_{32} and s_3.

We emphasize that all the three proposed modes can be realized using the same circuit, as shown in Figures 4.6-4.8, where the IL-FEC protected layer is adaptively selected from these three modes according to the instantaneous channel SNR. Thus, the complexity order of our enhanced adaptive IL-FEC system is identical to that of its counterpart in [60]. Since for *Mode* 2 the bit stream of the BL is independent of the other layers, the additional implantation is no longer required and hence it is deactivated, as shown in Figure 4.6(a). Furthermore, the three RSC Decoders shown in Figures 4.5-4.8 are essentially identical and have the same function.

The adaptive IL-FEC selection unit of Figure 4.5 selects the appropriate IL scheme to assign the most appropriate protection based on one of the above three modes by taking into account the estimated channel SNR γ, as follows:

$$IL - FEC = \begin{cases} Mode\,1 & \gamma \leq f_1, \\ Mode\,2 & f_1 < \gamma \leq f_2, \\ Mode\,3 & f_2 < \gamma, \end{cases} \tag{4.4}$$

where the threshold f_i is defined in terms of the e-PLR $P(l_i)_e$ of the corresponding layer l_i. Owing to the dependency between the BL and ELs, the PLR of the EL also has to take into account that of its reference counterparts. For example, the value of f_1 is decided by $P(l_1)_e$, which is equivalent to $P(l_1)$, while $P(l_2)_e$ requires both $P(l_1)$ as well as $P(l_2)$ and determines the threshold value of f_2. Thus, the e-PLR can be treated as the conventional PLR, where its reference layers are error free. For a given layer l_i, $P(l_i)_e$ can be expressed as:

$$P(l_i)_e = \begin{cases} P(l_1) & i = 1, \\ P(l_1) + \sum_{m=2}^{i} P(l_m) \prod_{n=1}^{m-1} [1 - P(l_n)] & i > 1. \end{cases} \tag{4.5}$$

The values of the threshold f_i are set to γ dB, namely to the specific SNR at which the PLR remains 'just' below a certain threshold value of $P(l_i)_e <= P_t$. Explicitly, the value of f_1 is the SNR where we have $P(l_1)_e <= P_t$, while that of f_2 can be found when $P(l_2)_e <= P_t$.

We emphasize that $s_{i,j}$ is different from $s_{j,i}$, when implanting systematic bits of l_i into those of l_j, unless l_i and l_j have the same number of bits. When the length of l_i is higher than that of l_j, a feasible solution is to merge several bits of l_i by an XOR operation before implanting it into l_j for the sake of bit-length matching. By contrast, if l_j has more bits, some of the bits in l_i may have to be used more than once. More details about unequal-length cross-layer interleaving techniques can be found in [60].

The remaining parity bits of all the three layers are then multiplexed with the newly generated systematic codes, leading to the three new bit streams of x_1, x_2 and x_3, which are then fed into the MS-STSK transceiver of Figure 4.5. Again, Figure 3.6 has demonstrated that in general the BER performance of the ASU of MS-STSK is better than that of the classic QAM/PSK sub-channel as well as that of the dispersion matrix-based sub-channel, hence resulting in the lowest BER among these components at a given SNR. Therefore, we feed x_1 to the ASU sub-channel for guaranteeing the best protection for the BL. We then feed x_2 into the $\mathcal{L}$-QAM/PSK modulator and x_3 into the dispersion matrix sub-channel, hence providing inherent UEP for the three video layers. More details about the MS-STSK scheme can be found in [98, 133–135]. The encoded MS-STSK codewords are then transmitted over narrowband Rayleigh fading channels.

4.3.2 Proposed Receiver Model

In this section, we detail the decoding process of our proposed adaptive wireless video system. As illustrated in Figure 4.5, the MS-STSK transceiver first receives the signal and translates it into the LLR representation of the MS-STSK codewords by the Log-MAP algorithm. Then, the soft MS-STSK codewords are forwarded to a demultiplexer to generate the three bit streams, namely y_1, y_2 and y_3 of Figure 4.5, which are forwarded to the adaptive IL-FEC decoder.

Figures 4.7 and 4.8 depict the enhanced IL-FEC *Mode* 2 and *Mode* 3, while no modification is performed for *Mode* 1. Hence the encoder and decoder devised for *Mode* 1 shown in Figure 4.6 are identical to those in [60] [65], as discussed in Section 4.2.1. Moreover, the decoding process illustrated in Section 4.3 specifically details the philosophy of the enhanced *Mode* 2 of Figure 4.7, when l_2 is the IL-FEC protected layer. The decoding process of *Mode* 1 is slightly different, where the systematic bits of the BL s_1 are implanted into s_2 and s_3, as shown in Figure 4.6. Thus, the received stream y_{s_1} can be decoded independently, while $y_{s_{12}}$ and $y_{s_{13}}$ can be in parallel decoded with the aid of $L_e(s_1)$. The decoding process of the enhanced *Mode* 3 is fairly similar to that of the enhanced *Mode* 2. Since in the enhanced *Mode* 3 the systematic bit stream s_3 is implanted into s_2, while s_2 and s_3 are implanted into s_1, as shown in Figure 4.8(a), they have to obey a certain decoding order similar to that of *Mode* 2 at the receiver: namely s_3, s_2 and s_1, as depicted in Figure 4.8(b). The multiplexer MUX of Figure 4.5 reorganizes the three bit streams, namely $\hat{l}_1$, $\hat{l}_2$ as well as $\hat{l}_3$, and the SHVC decoder reconstructs the video $\hat{V}$. By iteratively exchanging soft extrinsic information with the RSC decoder of the other layers, as seen in Figures 4.6, 4.7 and 4.8, the IL-FEC protected layer benefits from an improved BER and PSNR performance.

4.4 Simulation Performance

In this section, we present our simulation results for characterizing the proposed MS-STSK assisted adaptive IL-FEC aided system. Again, the SHVC reference software SHM is utilized for encoding the *Foreman* video clip. The GoP period is set to 4 for all video simulations, which means that the IDR/CRA frames are inserted every 4 frames. No B frames are used in our simulations due to the fact that they are prone to propagating inter-frame video distortions. Similarly, the bidirectional predictive B frames propagate video distortion and increase the latency, hence preventing flawless lip-synchronization. As a consequence, the video sequence in our simulations simply consists of I frames and P frames. Furthermore, we disable the spatial and temporal scalability functionalities, when encoding the video sequence into three different-quality layers, where the quality of the layers is controlled by setting the bitrate for each layer. The bit stream of each video frame is mapped to an MS-STSK packet, whose length is defined in Table 4.1. The receiver checks if the received

Table 4.1: Parameters for transmitting the employed *Foreman* and *Football* sequences.

Parameters	Foreman	Football
Representation	YUV 4:2:0	YUV 4:2:0
Format	QCIF(176×144)	CIF(352×288)
Bits Per Pixel	8	8
FPS	30	15
No. of Frames	30	30
No. of Layers	3	3
Video Codec	SHVC	SHVC
GoP	4	4
Bitrate (kbps)	126.7, 259.7, 385.3	514.3, 749.1, 1025.7
Error-Free YPSNR (dB)	31.97, 40.39, 42.11	29.88,36.75,40.24
Error Concealment	Frame-Copy	Frame-Copy
FEC	RSC [15 17]	RSC [15 17]
No. of IL-FEC Iterations	2	2
MS-STSK Configuration	MS-STSK(4,2,2,2,8,4)$\vert_{\text{QAM}}$	MS-STSK(8,2,2,2,16,8)$\vert_{\text{PSK}}$
Packet Length (bits)	2000×$b_{MS\text{-}STSK}$(6)	2000×$b_{MS\text{-}STSK}$(9)
Channel	narrowband Rayleigh	narrowband Rayleigh
Normalized Doppler	0.03	0.03
Shadow Fading (σ dB)	2	2
Simulations Repeated	100	100

packet has any bit errors using the associated CRC. If the CRC detection fails, the corrupted frames are dropped and replaced by “frame-copy” based error concealment.

Apart from the above source configuration, the FEC-aided MS-STSK transceivers are configured as follows. The three RSC codecs are configured by the binary generator polynomials of [1101 1111]. Additionally, the *MS-STSK(4,2,2,2,8,4)*$|_{QAM}$ and *MS-STSK(8,2,2,2,16,8)*$|_{PSK}$ configurations are used by the MS-STSK transceiver. The bit allocations of two MS-STSK configurations are listed in Table 4.2. The MS-STSK transceiver configured as *MS-STSK(4,2,2,2,8,4)*$|_{QAM}$ has 4 TAs and 2 RAs as well as 2 RF chains, hence resulting in a 6-bit $b_{MS\text{-}STSK}$ sequence that consists of 1 bit for the ASU, 2 bits for the $\mathcal{L}$-QAM/PSK modulator and 3 bits for the dispersion matrices. As for the configuration of *MS-STSK(8,2,2,2,16,8)*$|_{PSK}$, there are 8 TAs, 2 RAs and 2 RF chains, yielding a 9-bit $b_{MS\text{-}STSK}$ sequence associated with 2 bits for the ASU, 3 bits for the modulator and 4 bits for the dispersion matrix index.

A part of a video sequence, namely the *Foreman* sequence, which has 30 frames and is scanned at 30 FPS, is encoded into three layers, having bitrates of 126.7, 259.7 and 385.3 kbps respectively and using the MS-STSK configuration of *MS-STSK(4,2,2,2,8,4)*$|_{QAM}$, as shown in Table 4.1. In the

Table 4.2: Bit allocations of the MS-STSK configurations.

Configurations	$b_{MS\text{-}STSK}$	b_{ASU}	b_M	b_Q
$MS\text{-}STSK(4,2,2,2,8,4)\|_{QAM}$	6	1	2	3
$MS\text{-}STSK(8,2,2,2,16,8)\|_{\text{PSK}}$	9	2	3	4

EEP, the number of bits in each layer is split into three streams on average, which are then fed into the three modules of MS-STSK seen in Figure 3.3 while for the UEP the bits of different layers are fed into the three corresponding MS-STSK modules. Therefore, compared to EEP, the ASU of MS-STSK in UEP only contains the bits of the BL of the scalable video stream, namely l_1, while the $\mathcal{L}$-QAM/PSK modulator only has the bits of l_2. Finally, the dispersion matrix index only has the bits of l_3. The system considered here extracts the sub-layers and feeds them into different blocks, when using the $MS\text{-}STSK(4,2,2,2,8,4)|_{QAM}$ configuration of MS-STSK.

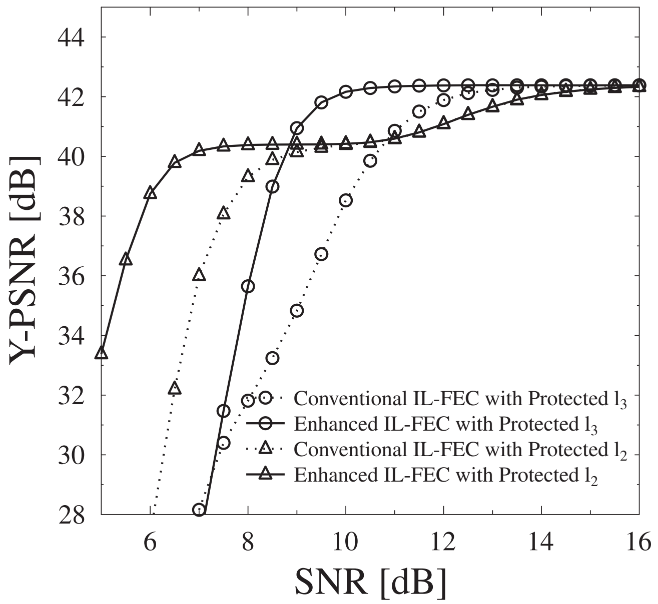


Figure 4.9: The image quality comparison between the enhanced and the conventional IL-FEC for the *Foreman* test sequence associated with $MS\text{-}STSK(4,2,2,2,8,4)|_{QAM}$.

4.4.1 Enhanced IL-FEC Performance

Figure 4.9 compares the reconstructed video quality of the conventional and of the enhanced IL-FEC techniques, where the enhanced IL-FEC on average outperforms its conventional counterpart by about 2-3 dB channel SNR. For example, when the protected layer is l_2, the proposed enhanced IL-FEC attains about 2 dB channel SNR gain compared to its conventional counterpart, as seen by comparing the continuous and dotted lines marked by triangles. When the channel SNR reaches about 14 dB, their PSNR improves further from about 40 dB to about 42 dB. As for the protected l_3 scenario, both the enhanced and the conventional IL-FEC schemes require a higher channel SNR, as shown in Figure 4.9 by the curves marked by circles. Suffice to say that the enhanced IL-FEC based protected l_2 scheme has a flawless error free PSNR of about 40 dB for channel SNR in excess of 6 dB, which its protected l_3 based counterpart requires about 8.5 dB at their cross-over point in Figure 4.9. However, the scheme represented by triangles is not contaminated by transmission errors, whereas the one marked by circles is error-infested at PSNR=40 dB, which results in a perceptually degraded QoE.

4.4.2 Mode Switching Performance

Table 4.3: Thresholds for the systems.

Candidate	f_1	f_2
Foreman-enhanced	4.9 dB	9.2 dB
Foreman-conventional	7.5 dB	11.3 dB
Foreman-UEP	-	-
Foreman-EEP	-	-
Football-enhanced	8.8 dB	12.6 dB
Football-conventional	11.5 dB	15.2 dB
Football-UEP	-	-
Football-EEP	-	-

In order to characterize our system, we compare the performance of our adaptive system that invokes the enhanced IL-FEC technique to that of the conventional adaptive system as well as to that of the UEP scheme and to that of the EEP scheme. Explicitly, the difference between two adaptive schemes is presented in the adaptive IL-FEC selection unit, as shown in Figure 4.5, while the other parameters set for the video sequences, the RSC codecs and the MS-STSK transceiver are identical. For the enhanced adaptive system, *Mode* 1, *Mode* 2 and *Mode* 3 are illustrated in Figures 4.6, 4.7 and 4.8, respectively, while those of the conventional adaptive counterpart only use the schematic of Figure 4.1, associated with implanting s_2 into s_1 and s_3 for conventional *Mode* 2 and implanting s_3 into s_1 and s_2 for conventional *Mode* 3. It is worth noting that *Mode* 1 is identical for both adaptive schemes, while the difference between two adaptive schemes occurs in *Mode* 2 and *Mode* 3, respectively. Additionally, both UEP and EEP are realized by the MS-STSK transceiver, both of which use only RSC codecs instead of the IL-FEC techniques, hence no adaptivity activated in these two schemes. We consider a P_t value of 5% as recommended in [93] in order to adaptively determine the mode-switching thresholds, since the perceptual image degradation imposed on the reconstructed video at the receiver by a PLR value of less than or equal to 5% becomes fairly minor.

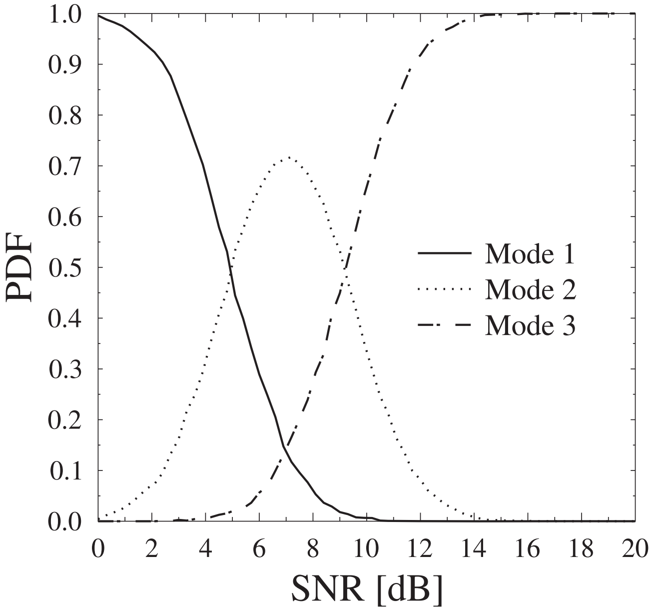


Figure 4.10: PDF of three enhanced modes versus channel SNR for the *Foreman* test sequence.

By recalling (4.5), the e-PLR can be expressed as:

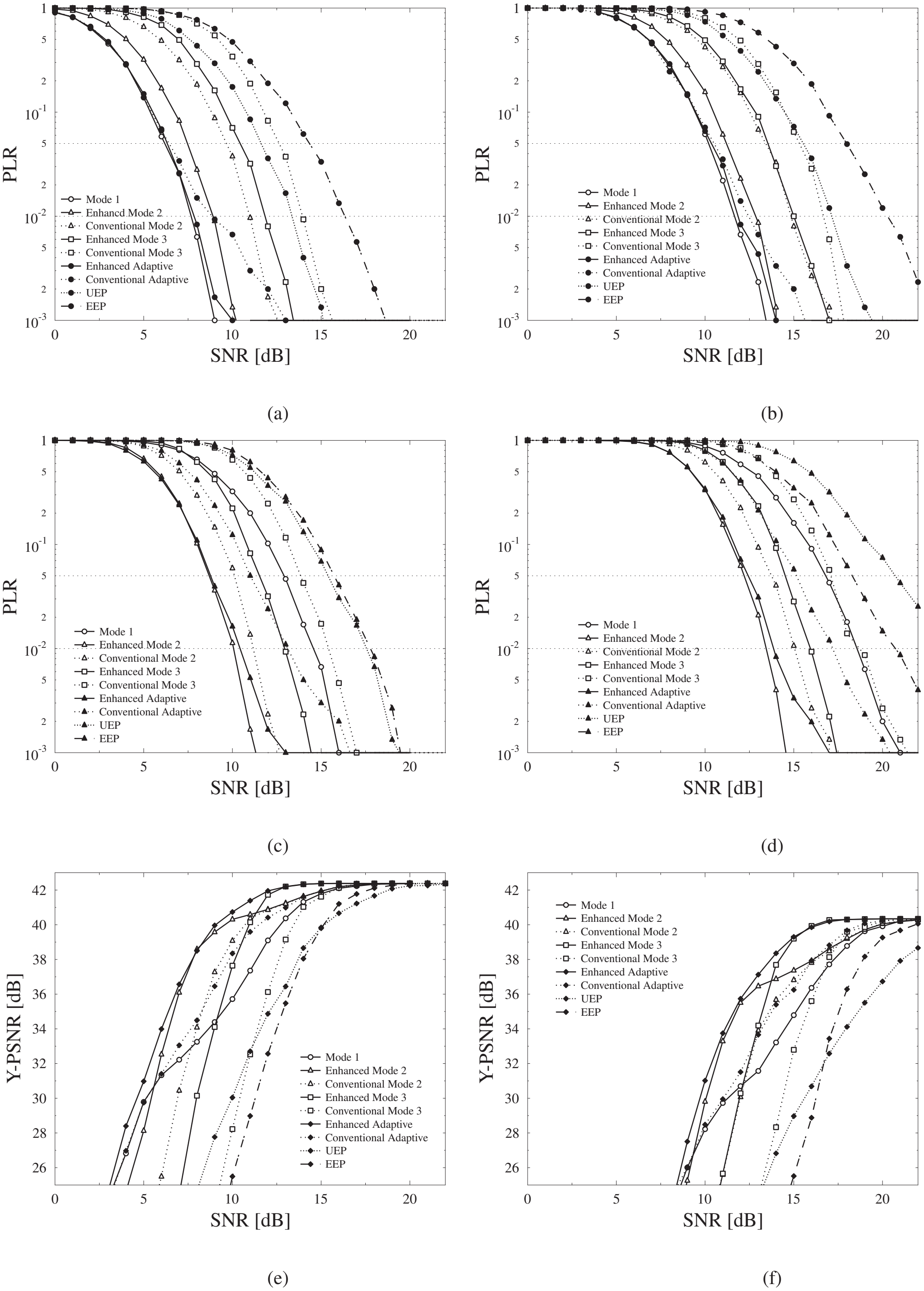
$$P(l_1)_e = P(l_1), \tag{4.6}$$

$$P(l_2)_e = [1 - P(l_1)] \cdot P(l_2) + P(l_1). \tag{4.7}$$

The thresholds set for selecting modes are given in Table 4.3, where the terms *enhanced* and *conventional* represent the specific type of the IL-FEC technique applied for the adaptive system. Again, all other parameters specifying the systems can be found in Table 4.1. Note that no threshold value is set for the EEP and UEP schemes, since the IL-FEC mode is deactivated for both schemes. Figure 4.10 depicts the probability density function of three enhanced modes versus channel SNR in the *Foreman* test scenario, where at a channel SNR of $\gamma < 4.9$ dB, *Mode* 1 is the frequently used mode, while at $\gamma > 9.2$ dB, the probability of using *Mode* 3 is higher than that of *Mode* 2 and *Mode* 3. Figure 4.11 compares the e-PLR and the image quality (PSNR) performances of our enhanced scheme to other counterparts for both the *Foreman* and *Football* clips.

4.4.3 Quality of Experience Performance

Figures 4.11(a) and 4.11(b) depict the $P(l_1)$ value versus the channel SNR, where we can observe that $P(l_1)$ of our enhanced *Mode* 2 and *Mode* 3 outperforms the other counterparts. Additionally, the enhanced adaptive scheme is capable of protecting the BL almost as well as *Mode* 1. This is be-



(a) (b)

(c) (d)

(e) (f)

Figure 4.11: The comparison between the enhanced and the conventional adaptive system for the *Foreman* (left column) and the *Football* (right column) test sequences, where the first, second and third rows present $P(l_1)_e$, $P(l_2)_e$ and the image quality versus channel SNR, respectively.

Figure 4.12: Comparison of frames at the channel SNR of 11 dB for the *Football* sequence. The five columns (from left to right) represent the original video, the enhanced adaptive scheme, the conventional adaptive scheme, the UEP scheme and the EEP scheme, respectively.

Table 4.4: The channel SNR required for $P(l_2)_e$ of the *Foreman* test sequence.

Candidate	$\leq 5\%$	$\leq 1\%$
Mode 1	12.9 dB	14.7 dB
Enhanced *Mode* 2	8.8 dB	10.1 dB
Conventional *Mode* 2	10.2 dB	11.3 dB
Enhanced *Mode* 3	11.6 dB	12.9 dB
Conventional *Mode* 3	13.8 dB	15.6 dB
Enhanced Adaptive	8.8 dB	10.5 dB
Conventional Adaptive	11 dB	13.1 dB
UEP	15.5 dB	17.5 dB
EEP	15.8 dB	17.8 dB

cause the enhanced *Mode* 2 shown in Figure 4.7 also takes into account the BL by implanting the s_1 bit stream into its s_3 counterpart and hence the mode-switching between *Mode* 1 and *Mode* 2 only imposes a modest degradation on the BL compared to the conventional adaptive scheme. However, the difference between two adaptive schemes becomes more distinguishable in Figures 4.11(c) and 4.11(d). The channel SNRs required to achieve $P(l_2)_e \leq 5\%$ and $\leq 1\%$ for the *Foreman* test sequence can be found in Table 4.4, where the enhanced *Mode* 2 substantially outperforms the other modes and results in a 1.4 dB SNR gain compared to its conventional counterpart, hence improving the image quality (PSNR) performance of the corresponding adaptive system. Furthermore, since our enhanced *Mode* 3 also takes into account l_2, as shown in Figure 4.8, it also yields a better $P(l_2)_e$ value than its conventional counterpart, hence imposing a low $P(l_2)_e$ degradation during mode-switching of the adaptive system. Therefore, it can be observed from Figures 4.11(c) and 4.11(d) that the $P(l_2)_e$ value of our enhanced adaptive system is more close to that of the enhanced *Mode* 2, while an obvious video degradation is observed for its conventional adaptive counterpart.

Figures 4.11(e) and 4.11(f) show the image quality (PSNR) performance versus the channel SNR for two video sequences. Observe that the enhanced *Mode* 2 and *Mode* 3 are capable of yielding a better PSNR performance than their conventional fixed counterparts, namely the conventional *Mode* 2 and *Mode* 3. Observe in Figure 4.11(e) that to reach a PSNR of 38 dB the enhanced *Mode* 2 outperforms its conventional counterpart by about 0.8 dB, while a 2.2 dB power reduction is achieved by the enhanced *Mode* 3 compared to the conventional *Mode* 3. It can be seen from Figures 4.11(e) and 4.11(f) that our enhanced adaptive scheme is capable of yielding an improved PSNR over a large proportion of the channel SNR range over that of its corresponding fixed-mode counterparts. A subjective video-quality comparison of the benchmarkers recorded for the *Football* test sequence is presented in Figure 4.12, where both the adaptive schemes ensure an unimpaired subjective video quality, but the proposed adaptive system is capable providing a better video quality, as shown in Figure 4.11(f).

4.5 Chapter Conclusion

In Section 4.2.1, we first discussed the conventional IL-FEC technique employed for layered video streaming, which is only capable of providing protection for the BL bits, but it is incapable of protecting the ELs. We solve this by additionally protecting reference layer of the IL-FEC protected layer. Specifically, in Section 4.2.2, in addition to implanting the bits of the protected layer into the other layers, our enhanced technique further implants the reference bits of the protected layer into the other non-IL-FEC protected layer. As a result, the proposed enhanced IL-FEC reduces the channel SNR required compared to its conventional counterpart by approximately 2 dB, as shown in Figure 4.9. Furthermore, to circumvent the limitation that the fixed IL-FEC scheme is incapable of providing the best reconstructed video quality over a wide SNR range, we proposed an MS-STSK assisted adaptive IL-FEC based design that judiciously activates the most appropriate enhanced IL-FEC mode according to the channel SNR. The proposed MS-STSK assisted system adaptively selects the most appropriate enhanced IL-FEC schemes by comparing the channel SNR

Table 4.5: The channel SNR required when reaching $P(l_1)_e$ and $P(l_2)_e$ below 5% for the *Foreman* test sequence.

Candidate	$P(l_1)_e$	$P(l_2)_e$
Mode 1	6.2 dB	12.9 dB
Enhanced *Mode* 2	7.4 dB	8.8 dB
Conventional *Mode* 2	9.6 dB	10.2 dB
Enhanced *Mode* 3	10.5 dB	11.6 dB
Conventional *Mode* 3	12.7 dB	13.8 dB
Enhanced Adaptive	6.2 dB	8.8 dB
Conventional Adaptive	6.2dB	11 dB
UEP	11.6 dB	15.5 dB
EEP	14.2 dB	15.8 dB

experienced with the pre-recorded mode-switching thresholds, as shown in Figure 4.10. The simulation results shown in Table 4.5 and Figure 4.11 demonstrate that the proposed enhanced IL-FEC assisted modes outperform their conventional counterpart by approximately 2 dB. Additionally, the proposed adaptive system was capable of striking the best video PSNR versus error-resilience trade-off amongst all the schemes.

Chapter 5

Optimal FEC Code-Rate Assisted SVC Streaming

5.1 Introduction

In Chapter 4, we proposed an IL-FEC assisted UEP scheme that implants the bits to be protected into the other layers for scalable video streaming. By contrast, in this chapter, another FEC based UEP scheme is conceived, which has a variable FEC code-rate. Specifically, the proposed variable FEC code-rate assisted scheme aims for assigning different FEC code-rates providing carefully tuned protection for the different-sensitivity video bits according to their influence on the video degradation, without imposing any bandwidth expansion on the existing system. For example, the most robust FEC code-rate is assigned to the more important video frames, such as the I-frame, while weaker protection can be assigned to the less important P-frames. In this chapter, we aim for finding the optimal FEC code-rate allocation scheme for scalable panoramic video relying on the feedback of the near-instantaneous channel SNR, where the AQAM technique is invoked to extend the operating SNR range of the system.

This chapter is organized as follows. We first review the concept of variable FEC code-rate assignment in Section 5.2, which is categorized into frame-level and macroblock-level schemes. Section 5.3 presents the concept of panoramic video schemes, including the projection conversion methods and the existing schemes designed for panoramic video streaming. Then, we illustrate our optimal FEC code-rate assisted transmission system conceived for scalable panoramic video streaming in Section 5.4, where the associated optimization process and the associated adaptive mode switching operation are discussed in Sections 5.5 and 5.6. Then, we quantify the performance of the proposed system in Section 5.7 and conclude in Section 5.8.

5.2 Review of Variable FEC Rate Assisted UEP for Video Streaming

The published video standards, such as AVC and HEVC, specify the coded representation of the video data. In order to achieve efficient compression, the predictive coding structure is wide employed. Owing to the temporal horizontal coding dependency inherent in the GoP structure shown in Figure 5.1, we can see that more video frames will be affected due to error propagation, if an earlier encoded reference frame is corrupted. It is worth noting that in Figure 5.1, we do not rely on B-frames for simplicity, but without loss of generality. However, the predictive coding techniques of Figure 5.1 introduce error propagation between the inter-encoded frames upon communicating over error-prone channels, where the corruption imposed on an I-frame may be propagated across the GoP and can only be eliminated by the next error-free I-frame. Hence, in order to provide the video stream with increased robustness, while communicating over error-prone channels, it is necessary to employ powerful error control.

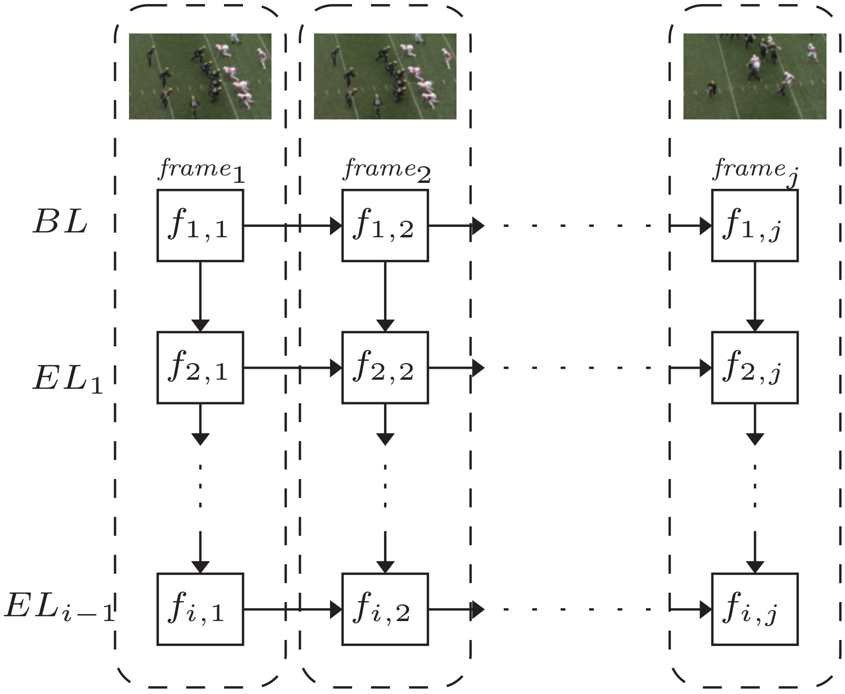


Figure 5.1: SVC coding structure.

To minimize the impact of transmission errors, an appropriate choice of FEC rates for the video frames in a GoP is necessary, while maintaining a near-constant transmission rate for a GoP. Explicitly, more FEC parity bits are expected to be provided for the former frames of a GoP, while fewer parity bits are added to the latter ones within a GoP cycle. Furthermore, in each frame, the frames of the ELs are encoded and decoded with the aid of low-layer counterparts. In this section, we focus our attention on the family of variable-rate FEC assisted UEP schemes, which again assign a lower FEC code-rate to the more important bits of the video parameters that are frequently referenced by the dependent parameters and vice versa. However, a lower FEC code-rate results in more redundant parity bits, hence the less source bits can be delivered within a given bandwidth. In this section, we review a set of three different FEC schemes in the context of video streaming, including the conventional EEP scheme that treats all video bits equally and a pair of UEP schemes. Specifically, we categorize our variable FEC code-rate assisted UEP scheme into frame-level and macroblock-level arrangements.

5.2.1 Equal Error Protection

Figure 5.2 illustrates the conventional EEP scheme, where we can see that all the video frames have an identical FEC code-rate, as indicated by the same color, in other words, all bits are protected equally, regardless of the frame type. Hence, this scheme exhibits the lowest complexity. However, due to the coding dependency shown in Figure 5.1, a P frame may be occasionally dropped due to the corruption of its dependent frames, even if its own bits are received perfectly.

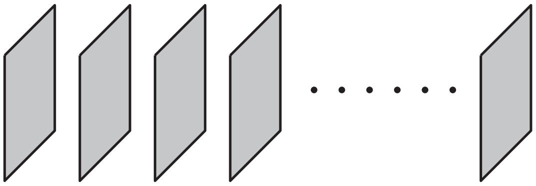


Figure 5.2: The frame level based EEP, where all video frames share the identical FEC code-rate.

5.2.2 Frame-Level Based UEP

The principle of frame-level based FEC code-rate allocation assigns different code-rates to different frames according to their importance, i.e. their position in the GoP cycle. It can be seen from Figure 5.3 that different protection is provided for the video frames according to their importance, where a darker color represents that the frame benefits from better protection, namely from a lower FEC code-rate. The approach presented in this subsection for the protection of video transmitted over wireless networks is based on the following idea: the gravity of error propagation imposed by a certain corrupted P-frame in the video sequence depends on the particular position of this P-frame within the GoP sequence [41]. If the P-frame is positioned immediately after an I-frame,

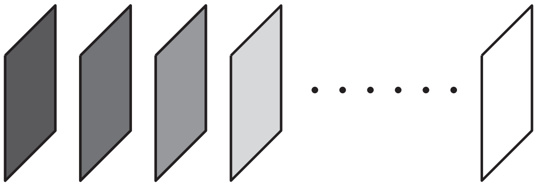


Figure 5.3: The frame level based UEP, where various FEC code-rates are assigned to different video frames according to their importance.

its influence on the rest of the sequence will be much more grave than that of a P-frame positioned right before an I-frame. Therefore, it is plausible that the overall video quality may be improved by enhancing the protection of the first few P-frames of a GoP sequence, with respect to the rest of the sequence. Since the overall amount of redundancy in this approach must be identical to that of the classical EEP, this means that the protection of the last few P-frames of each sequence must be "sacrificed" to a degree. However, the only difference with regard to the EEP scheme is that the redundancy is non-uniformly distributed across the frames for minimizing the mean distortion across the entire transmitted sequence. Furthermore, this operation is optimized using an analytical technique presented in Section 5.5.

5.2.3 Macroblock-Level Based UEP

The macroblock-level based variable FEC code-rate assisted UEP scheme is illustrated in Figure 5.4, where we can see that each video frame is partitioned into multiple macroblocks. For simplicity, Figure 5.4 assumes that the macroblocks have identical size, so that all video frames have the same number of macroblocks. The size of the macroblock may depend on the specific video coding configuration and on the content of each video frame, but this issue is beyond the scope of our discussions. It can be seen from Figure 5.4 that the macroblock-level based UEP scheme is capable of supporting a more flexible designs compared to its frame-level based counterpart. In addition to the non-uniformly distributed importance of I and P frames in a GoP, the macroblock level based UEP scheme also takes into account the non-uniformly distributed importance of macroblocks within a video frame [44]. More explicitly, based on the fact that due to Variable Length Coding (VLC) the coding procedure usually starts from the top-left corner of a frame, when transmission errors occur in the macroblocks at the top-left of a video frame, the decoder may fail synchronization, which results in aborting the decoding process of the current video frame. On the other hand, if transmission errors occur within the macroblocks located near the bottom-right corner of the video frame, only a few remaining macroblocks are affected and the associated image quality degradation becomes minor. Therefore, the macroblocks at the beginning of every video frame are expected to be allocated more powerful protection than the macroblocks at the end of the video frame, since macroblocks at the beginning of a video frame will inflict more sever error propagation, if errors occur in them.

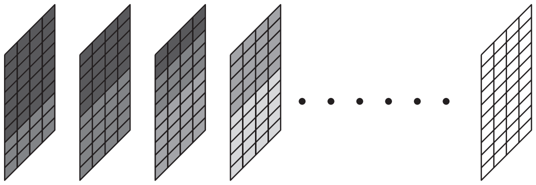


Figure 5.4: The frame level based UEP, where various FEC code-rates are assigned to different video frames according to their importance.

5.3 Panoramic Video

Figure 5.5: Sphere format.

The panoramic video representation provides the users with flawless 360 degree immersive experiences, which has fascinated the researchers [11, 136–138]. The original spherical panoramic video observed by the users cannot be encoded and transmitted directly, which has hence inspired the development of the so-called projection methods. In the rest of this section, we will elaborate on the specific compression procedure of panoramic video clips. Specifically, we first introduce the panoramic video testing model, which discusses the fundamental procedures of evaluating the performance of panoramic video processing. Then, we illustrate the projection methods that convert the 3D spherical format to the conventional 2D plane in order to be able to employ the existing coding standards. The Equirectangular Projection (ERP) [139] and the Cubemap Projection (CMP) [140] are detailed. Then, the corresponding WS-PSNR concept is introduced for both schemes, which takes into account the inevitable 3D to 2D conversion distortion. Finally, at the end of this section, we introduce the existing transmission schemes conceived for panoramic video streaming.

5.3.1 Panoramic Video Testing System

The panoramic video testing system is depicted in Figure 5.6, where the high-fidelity test material must be in the 4:2:0 YUV format for representing 360 degree panoramic video in ERP format [139], which then has to be converted to the spherical format[1], using the detailed procedure shown in Figure 5.5. Due to the fact that the existing coding standards are incapable of encoding this 3D spherical format directly, prior to the encoding, the 3D format is projected to the conventional 2D format according to one of the formats listed in Figure 5.6, including the ERP of [139], the Equal-Area Projection (EAP) of [141], CMP [140], Icosahedral Projection (ISP) [142], Octahedron

---

[1] It is worth noting that the ERP testing material is only employed during the testing procedure, while in the real scenario the captured panoramic video is displayed in the original 3D spherical format.

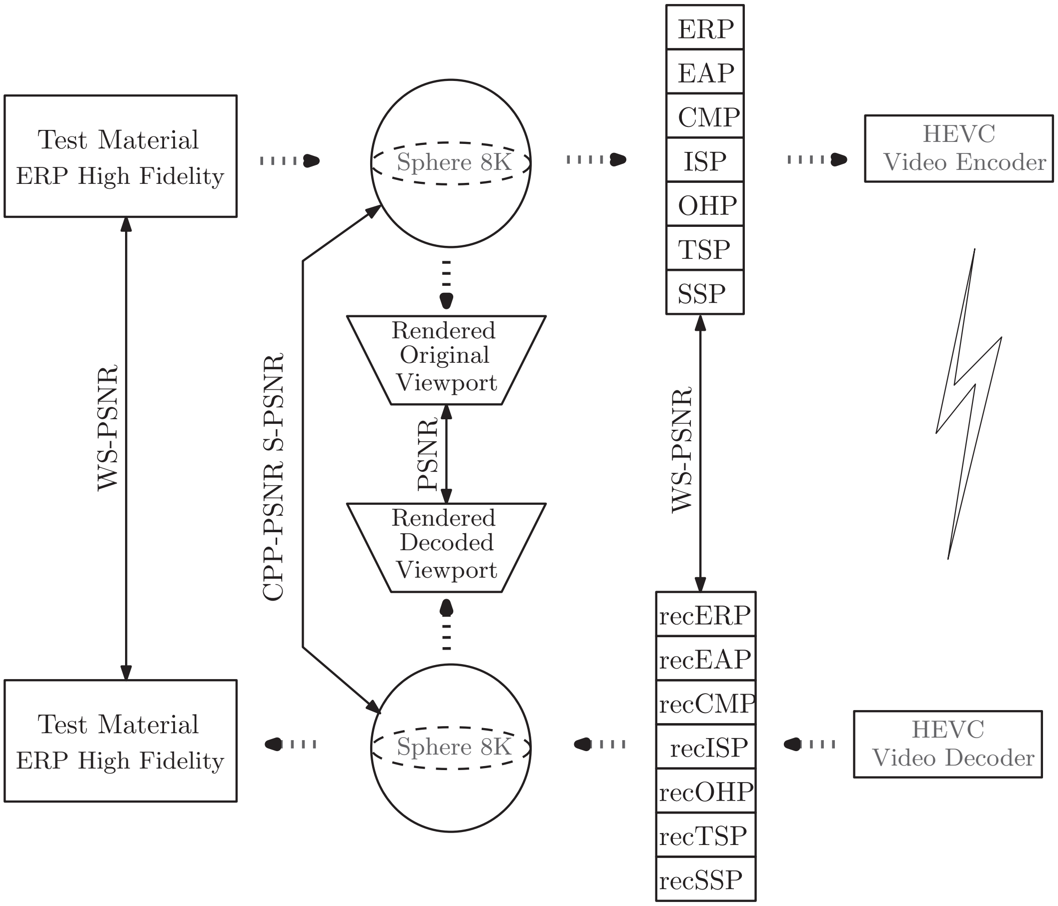


Figure 5.6: The panoramic video testing procedure.

Projection (OHP) [143], Truncated Square Pyramid Projection (TSP) [144] and Segmented Sphere Projection (SSP) [145], which exhibit different coding efficiency and conversion distortion. Then, the projected 2D format is compressed into bit streams using one of the existing video coding standards, such as the HEVC [6] seen in Figure 5.6.

The decoding procedure is illustrated in the bottom half of Figure 5.6, where the received bits are first decoded with the aid of the HEVC decoder, generating the decoded version of the projected format. Then, the inverse conversion is invoked to convert the 2D projection to the 3D spherical format, as shown in Figure 5.6, which displays the immersive 3D image for the users.

Additionally, in Figure 5.6, we also illustrate the evaluation metrics of the subjective video quality for both the 3D spherical and the 2D formats. For example, considering the fact that converting spherical 3D format to the various 2D projections introduces different distortions, the WS-PSNR was introduced by Sun *et al.* [146] for quantifying the subjective quality. Furthermore, the so-called Craster Parabolic Projection PSNR (CPP-PSNR) [142] and the Spherical PSNR using the Nearest Neighbor position without interpolation (S-PSNR-NN) [147] metrics were conceived for evaluating the spherical image quality. However, these spherical format based metrics also have their specific drawbacks and hence are not widely employed. Explicitly, CPP-PSNR requires pixel-interpolation across the sphere, which imposes additional computational complexity. Furthermore, since the S-PSNR-NN metric only contains 653662 sample positions, it is unable to process all pixels in ultra-high definition videos. Hence, it has limited evaluation accuracy, since only 8.9% of the pixels are taken into account when employing a 4K(3840×1920=7372800)-pixel image. Additionally, the rendered viewport, also referred to as Field of View (FoV) represents the observation window of the user, where the conventional PSNR metric is employed to evaluate the subjective video quality.

5.3.2 2D Projections

In order to employ the existing video coding standards, numerous projection schemes have been conceived to convert the 3D spherical video to the 2D presentations, which exhibit different conversion distortion and projection complexity. In this subsection, we mainly focus on exhibiting various typical projection schemes in details.

5.3.2.1 Equirectangular Projection

Figure 5.7: The ERP projection.

Figure 5.7 shows a conversion example from the spherical format of Figure 5.5 to the ERP projection of Fig. 5.7. Figure 5.8 shows the mapping procedure of ERP, where for a sphere radius of ρ, the width W and height H of the rectangular projection is given by:

$$\begin{aligned} W &= 2\pi\rho \\ H &= \pi\rho. \end{aligned} \tag{5.1}$$

Assuming that $P(x_p, y_p, z_p)$ denotes an arbitrary point on the sphere of Figure 5.8, the latitude ϕ can be expressed as:

$$\phi = \arcsin\frac{z_p}{\rho}, \quad -\rho \leq z_p \leq \rho, \tag{5.2}$$

while the longitude θ is expressed as:

$$\theta = \arctan 2(y_p, x_p), \quad -\rho \leq x_p, y_p \leq \rho, \tag{5.3}$$

where $\arctan 2(\cdot)$ returns the result in the interval of $[0, \pi]$ when $y_p \geq 0$ and in $(-\pi, 0)$ when $y_p < 0$. Hence, the coordinates associated with the projected point $P'(w_p, h_p)$ on the rectangular plane become:

$$\begin{aligned} w_p &= \theta \cdot \rho \\ h_p &= \phi \cdot \rho. \end{aligned} \tag{5.4}$$

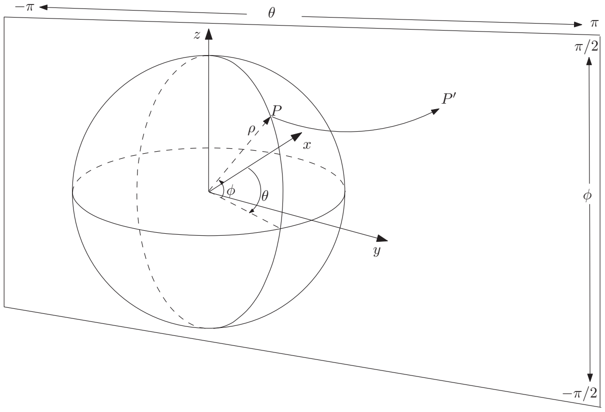


Figure 5.8: The ERP mapping procedure.

5.3.2.2 Cubemap Projection

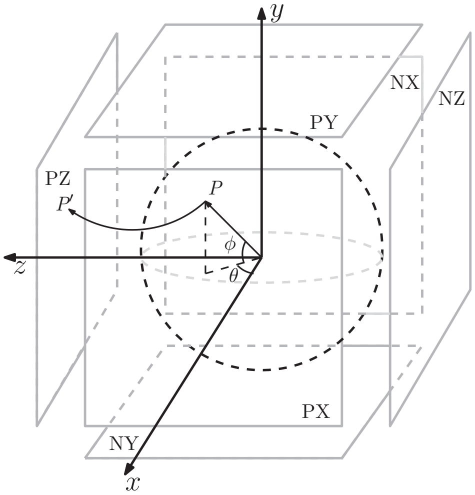


Figure 5.9: The CMP mapping procedure.

The above ERP mapping exhibits the lowest projection complexity using simple mathematical equations. However, this projection is incapable of accurate pixel conversion from the 3D spherical format to the 2D presentation at the North and South poles, where non-negligible distortions are introduced. To mitigate the projection distortion of the ERP mapping, the CMP [140] can be employed, which is depicted in Figure 5.9, where the acronyms PX, PY, PZ, NZ, NX and NZ are

Table 5.1: Face notations of CMP.

Face Label	Notes
PX	Front face with positive x axis value
NX	Back face with negative x axis value
PY	Top face with positive y axis value
NY	Bottom face with negative y axis value
PZ	Right face with positive z axis value
NZ	Left face with negative z axis value

defined in Table 5.1. We assume that P is an arbitrary point on the sphere's surface with coordinates of (θ, ϕ) that are defined by (5.2) and (5.3). The coordinates of the associated projected point $P'(w_p, h_p)$ on the corresponding cube face of Figure 5.9 are:

$$\begin{aligned} w_p &= \rho \tan\theta, \\ h_p &= \rho \frac{\tan\phi}{\cos\theta}, \end{aligned} \tag{5.5}$$

where ρ represents the sphere radius. The converted CMP projection is illustrated in Figure 5.10(b), where the top and bottom 25% of the ERP area of Figure 5.10a constitute the top and bottom face of the projected cube, namely PY and NY, while the 50% of the ERP area in the middle forms the other four faces, as seen in Figure 5.10(a).

5.3.3 WS-PSNR

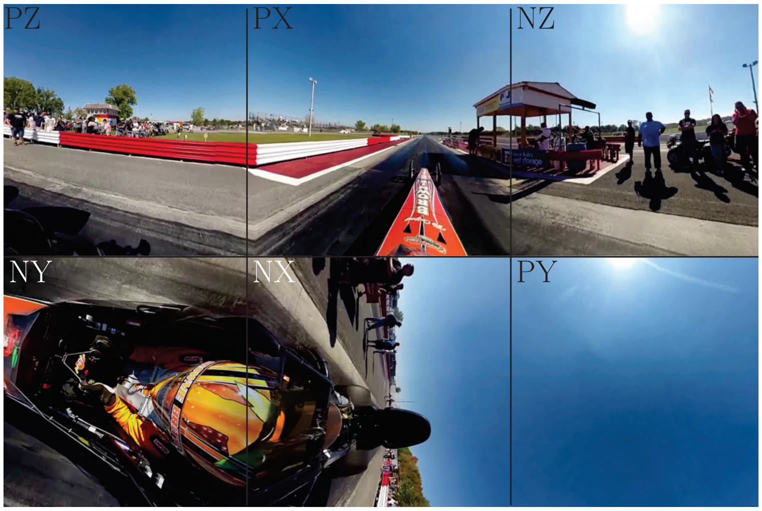
Mapping from the representation space, such as the two-dimensional planes of Figures 5.7 and 5.10b, to the spherical surface of Fig. 5.5 for observation is usually performed using a non-linear transformation, which means that the distortion imposed is also non-linear. For example, a pixel at the south and north pole of the spherical sphere may be repeatedly used in the 2D ERP projected format, which stretches the pole pixels and introduces more distortion at the poles. Hence, the conventional metric that uniformly weights the pixels on representation plane fails to provide an accurate objective quality assessment for panoramic video. The concept of WS-PSNR was proposed by Sun *et al.* [146], where the distortion of the samples in the representation space is further weighted by the corresponding projection area in the observation space.

Given the maximum possible intensity level of the image MAX_I, the WS-PSNR of panoramic video can be expressed as [146]:

$$WS\text{-}PSNR = 10 \log_{10} \frac{MAX_I}{WMSE}, \tag{5.6}$$

where Weighted Mean Square Error (WMSE) takes into account the pixel weights according to its location. Specifically, having a resolution of $W \times H$ and weighting coefficient of $\vartheta(w, h)$ at pixel

(a) ERP paritition



(b) CMP projection

Figure 5.10: The CMP projection and the associated ERP partitions.

position (w,h), the WMSE can be expressed as:

$$WMSE = \frac{\sum_{w=0}^{W-1}\sum_{h=0}^{H-1}\left[(y(w,h)-y'(w,h)\right]^2\vartheta(w,h)}{\sum_{w=0}^{W-1}\sum_{h=0}^{H-1}\vartheta(w,h)}, \tag{5.7}$$

where where $y(w,h)$ and $y'(w,h)$ are the original and received pixel values at the position (w,h). It is worth noting that the weighting coefficient of $\vartheta(w,h)$ varies according to the projection techniques.

To elaborate a little further on the weighting coefficients, let us introduce the concept of Stretching Ratio (SR) to quantify the distortion imposed by the nonlinear projection. Using (x,y) and (θ,ϕ) as the coordinates on the 2D projected plane and as the spherical position, the relationship between (x,y) and (θ,ϕ) can be given by:

$$\begin{aligned} x &= f(\theta,\phi), \\ y &= g(\theta,\phi). \end{aligned} \tag{5.8}$$

Figure 5.11: Weighting coefficient of ERP.

Then, for each (x, y) laying in the center of a micro-unit of $dxdy$ whose area is $\Delta A(x, y)$, we map $dxdy$ to the associated spherical surface, mainly to the spherical micro-unit $d\theta d\phi$ having the area of $\Delta S(\theta, \phi)$, where the relationship between the two micro-units $dxdy$ and $d\theta d\phi$ can be expressed as $dxdy = J(\theta, \phi) d\theta d\phi$ [148], with $J(\theta, \phi)$ representing the Jacobian determinant, which can be derived from (5.8):

$$J(\theta, \phi) = \frac{\partial(x, y)}{\partial(\theta, \phi)} = \begin{vmatrix} \frac{\partial(x)}{\partial(\theta)} & \frac{\partial(x)}{\partial(\phi)} \\ \frac{\partial(y)}{\partial(\theta)} & \frac{\partial(y)}{\partial(\phi)} \end{vmatrix}. \tag{5.9}$$

Hence, the SR can be calculated by:

$$SR = \frac{\Delta S(\theta, \phi)}{\Delta A(x, y)} = \frac{\cos\phi}{|J(\theta, \phi)|}. \tag{5.10}$$

5.3.3.1 Equirectangular Projection

For the ERP shown in Figure 5.7, where (x, y) and (θ, ϕ) represent the ERP and the spherical coordinates, according to (5.4), the normalized projection relationship of (5.8) for the ERP can be formulated as $x = f(\theta, \phi) = \theta$ and $y = g(\theta, \phi) = \phi$. Therefore, the SR of the ERP can be expressed as:

$$SR(x, y)_{ERP} = \cos\phi = \cos y. \tag{5.11}$$

In ERP having a resolution of $W \times H$ and $W = 2H$, the weighting coefficient of the position (w, h) on the ERP plane can be expressed as [146]:

$$\vartheta(w, h) = \cos\frac{(h + 0.5 - H/2)\pi}{H}, \tag{5.12}$$

where the relationship between (w, h) and (x, y) can be represented as $x = \frac{2(w+0.5-W/2)\pi}{W}$ and $y = \frac{(h+0.5-H/2)\pi}{H}$. The weighting coefficient of the ERP is depicted in Figure 5.11, where the white colour represents higher importance, while the dark region represents lower importance.

5.3.3.2 Cubemap Projection

As for the CMP illustrated in Figure 5.10b, since the geometric conversion is identical for all the six faces, we only have to derive the weighting coefficients of a single cube face. The normalized relationship between (x,y) on the cube face and (θ,ϕ) on the spherical surface can be formulated as $x = f(\theta,\phi) = \tan\theta$ and $y = g(\theta,\phi) = \frac{\tan\phi}{\cos\theta}$. Then, the SR of the cube face can be expressed as [146]:

$$SR(x,y)_{CMP} = (\cos\theta\cos\phi)^3 = (1+x^2+y^2)^{\frac{-3}{2}}. \tag{5.13}$$

Considering a cube face of the CMP having a resolution of $W \times W$, with the aid of the relationship between (w,h) and (x,y) given by $x = \frac{w+0.5-W/2}{W/2}$ and $y = \frac{h+0.5-W/2}{W/2}$, the weighting coefficients for an arbitrary cube face can be expressed as [146]:

$$\vartheta(w,h) = (1+\frac{d^2(w,h)}{\rho^2})^{\frac{-3}{2}}, \tag{5.14}$$

where $\rho = \frac{W}{2}$ represents the radius of the sphere, and $d^2(w,h) = (w+0.5-W/2)^2 + (h+0.5-W/2)^2$ is the squared distance from the centre of the cube face to position (w,h). The weighting coefficients of the CMP of Figure 5.10b are illustrated in Figure 5.12, where the six faces exhibit identical weighting coefficients.

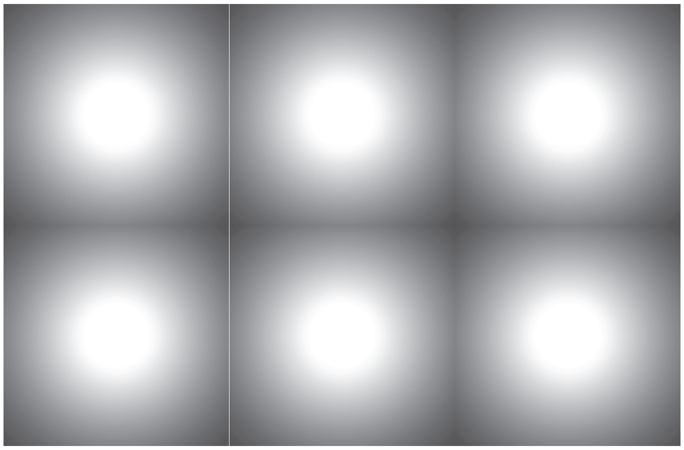


Figure 5.12: Weighting coefficient of CMP.

5.3.4 Panoramic Video Transmission

To reduce the demanding bitrate requirement of panoramic video streams, the JVET has been dedicated for years to the conception of the next-generation video coding standard, namely the H.266/VVC standard [7], whose objective is to provide a significant improvement in compression performance over the existing H.265/HEVC standard [6].

Furthermore, a method that spatially partitions the 360° video into multiple independent rectangular segments, also referred to as tiles, was proposed in [149], which allows both the video quality and the rate to be adapted locally. The development of the tile-based strategy has significantly advanced the quality of interactive panoramic video systems [150], where only the specific portions of the scene observed by the users is transmitted [136], which is termed as the viewport-adaptive approach. Naturally, this philosophy critically relies on the accuracy of Head Movement Prediction (HMP). Fortunately, numerous experiments have verified that the user's head movement can be accurately predicted [137]. However, since the success of the FoV-guided interactive panoramic video system is highly conditioned on the accuracy of the HMP, the video playback becomes prone to image freezes, hence potentially degrading the users' experience. To solve this problem, Bao *et al.* proposed a HMP based system that additionally transmits the video-frame portions that deviated from the predicted areas for guaranteeing the users' experience as well as minimizing the bitrate required [151]. The authors of [152] utilized both the saliency of video, as well as the prediction of the FoV and the status information of users to obtain an optimal association of the users for maximizing the system utility.

Additionally, to reduce the bitrate necessitated for panoramic video streaming, the concept of SVC [23] that compresses video clips into multiple layers according to the requirement of the clients was conceived. He *et al.* [153] devised a Region of Interest (RoI) based scalable video coding scheme, which encodes the specific video frame region, where users are interested in having high resolution with the aid of the EL. They were able to reduce the bitrate by approximately 87% at a modest quality degradation in the RoI. A triple-layer scalable video coding scheme was conceived in [154] for improving the coding efficiency of VR applications, where the particular video frame portions outside the FoV were encoded as the BL only, while the FoV portions were further enhanced by the ELs.

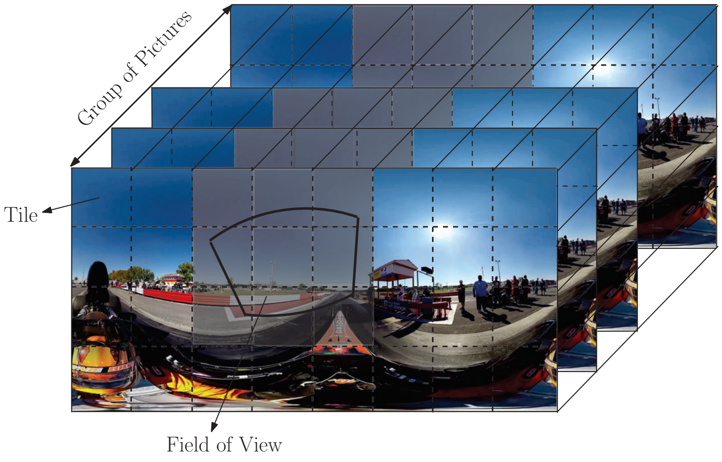


Figure 5.13: Tile based ERP.

Figure 5.13 illustrates the tile based 2D projections exemplified using ERP mapping, where the panoramic video frame is spatially partitioned into 4×8 square tiles. It is worth noting that all the panoramic video frames follow the same tiling scheme, in which the tiled frames are encoded and decoded independently to the other tiles. Additionally, the shading area of Figure 5.13 showing the user's FoV covers 9 out of 32 tiles, which directly affects the user's experience and are expected to be better protected.

5.4 System Model

In this section, we introduce the proposed adaptive scheme conceived for layered panoramic video communications over wireless channels. It is worth noting that our system is exemplified using the ERP method illustrated in Section 5.3.2.1, which is readily extended to other projection schemes. Before elaborating on the system model, let us briefly introduce the importance classes of the different video tiles using Figure 5.13, where the entire panoramic video frame is spatially partitioned into 32 tiles that are independently encoded and transmitted to the clients. Each video tile of Figure 5.13 is compressed into two layers with the aid of the SHVC encoder, namely a BL and an EL, where the BL of all tiles and the EL of the FoV are expected to be sent to the client, which are further categorized into three priority classes according to their subjective importance, as detailed as follows:

1. Class 1 (C1) is constituted by the all-important BL of the FoV, namely $\{v_i^1\}_{i \in FoV}$, which directly affects the subjective experience of the client, but also affects the corresponding ELs. Hence, it is expected to be strongly protected.
2. Class 2 (C2) contains the BL of the video tiles outside the FoV $\{v_i^1\}_{i \notin FoV}$. Bearing in mind the fact that there may be HMP errors that may lead to video freezes or artefacts, the BL of the video tiles outside the FoV represent the second priory class, which mitigates the potential video quality erosion imposed by HMP errors.
3. Finally, Class 3 (C3) is formed of the EL of the FoV video tiles $\{v_i^2\}_{i \in FoV}$ used for further enhancing the FoV image quality.

Furthermore, we also proposed powerful FEC coding rate optimization algorithms, which are part of the Rate Optimization block of Figure 5.14 and will be detailed in Section 5.5.

5.4.1 Transmitter

The transmitter model of the proposed design is depicted in Figure 5.14. The panoramic video V is first projected into the two-dimensional plane, which is then split into N tiles $\{v_i\}_{i=1}^N$ of Figure 5.7 ready for compression. The SHVC encoder then compresses them into two bit streams, namely the BL and the EL, which are then forwarded to the FEC encoders after demultiplexing. The demultiplexer of Figure 5.14 splits the video streams into three sub-streams, namely 1) the BL of the predicted FoV $\{f_{i,j}^1\}_{i \in FoV}$, 2) the BL outside the FoV tiles $\{f_{i,j}^1\}_{i \notin FoV}$, and 3) the EL of the predicted FoV $\{f_{i,j}^2\}_{i \in FoV}$, corresponding to the Class 1, 2 and 3, respectively, where $f_{i,j}^l$ indicates

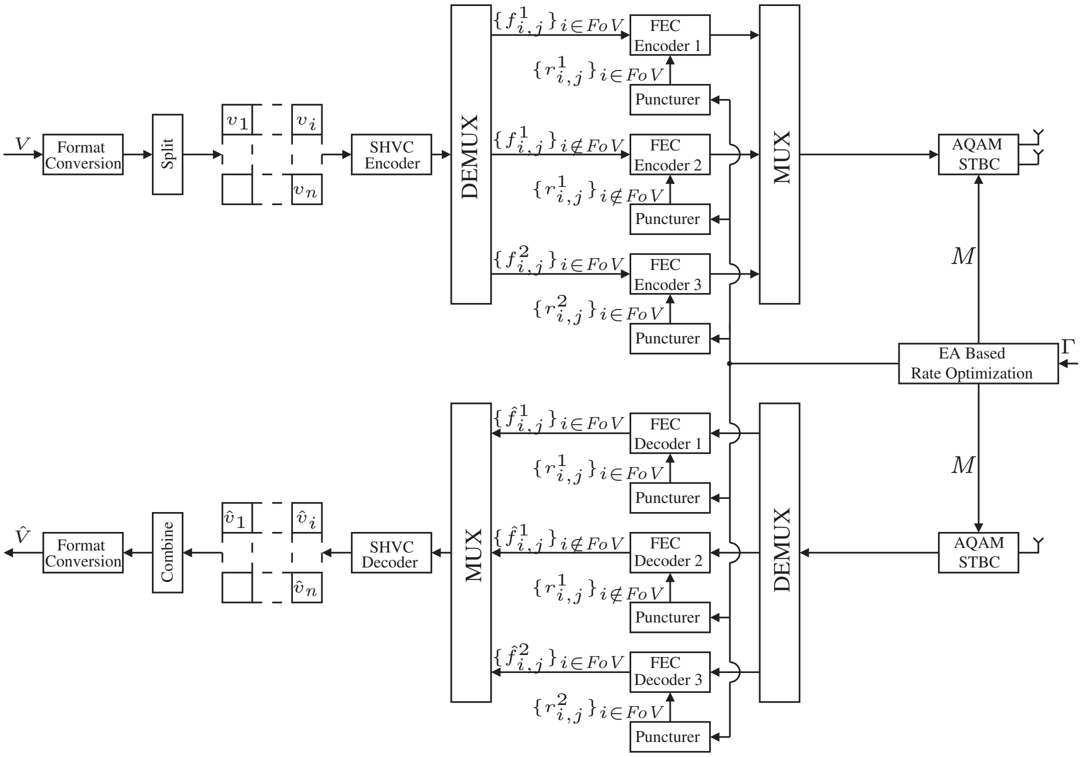


Figure 5.14: Block diagram of the proposed system.

the bits of i-th tile of the l-th video sublayer of the j-th video frame. The three sub-streams are then encoded by the FEC Encoders 1, 2, 3 according to the code-rate sets of $\{r^1_{i,j}\}_{i\in FoV}$, $\{r^1_{i,j}\}_{i\notin FoV}$ and $\{r^2_{i,j}\}_{i\in FoV}$, respectively, as depicted in Figure 5.14, which are optimized by the Coding Rate Allocation block of Figure 5.14 depending on the instantaneous channel SNR Γ.

Furthermore, the AQAM STBC selects the most appropriate modulation mode M from the three fixed modulation candidates, namely BPSK, 16QAM, 128QAM, for robust video streaming and then transmits the modulated data using space-time block coding. Let us now consider the throughput of the three modulation modes. Specifically, BPSK conveys the C1 BL bit streams of the FoV tiles $\{f^1_{i,j}\}_{i\in FoV}$ only, while 16QAM transmits the C1 and C2 BL of all tiles, $\{f^1_{i,j}\}_{i\in FoV} \cup \{f^1_{i,j}\}_{i\notin FoV}$. Finally, all the C1 and C2 BLs and the C3 EL of the tiles within the FoV, namely $\{f^1_{i,j}\}_{i\in FoV} \cup \{f^1_{i,j}\}_{i\notin FoV} \cup \{f^2_{i,j}\}_{i\in FoV}$, are transmitted, when 128QAM is activated. More explicitly, the most robust BPSK modulation mode only conveys the C1 bit stream, 16QAM delivers the bit streams of C1 and C2, while 128QAM contains all the three bit streams. Hence, in BPSK the coding rate allocation block of Figure 5.14 explores the most appropriate coding rate allocation for $\{f^1_{i,j}\}_{i\in FoV}$ only, while in the 16QAM and 128QAM scenarios, a more complex code-rate optimization algorithm may be invoked for the sake of exploiting the globally optimal FEC coding rates for achieving the highest uncoded source-rate, whilst progressively offering a stronger protection to the more important information. Additionally, we emphasize that the AQAM design can be substituted by the family of sophisticated multi-functional MIMO techniques, such as SM [96], STSK [97] and

MS-STSK [98].

5.4.2 Receiver

The processing of the received panoramic signals is portrayed in Figure 5.14, which are first demapped by the AQAM STBC decoder block of Figure 5.14 and then fed into the FEC Decoders. Naturally, the FEC Decoders associated with the untransmitted sub-streams have to be deactivated. More explicitly, FEC Decoders 2 and 3 are deactivated when BPSK is invoked, since $\{f_{i,j}^1\}_{i \notin FoV}$ and $\{f_{i,j}^2\}_{i \in FoV}$ are not transmitted, while only FEC Decoder 3 is deactivated for 16QAM in the absence of the discarded $\{f_{i,j}^2\}_{i \in FoV}$ bits. The multiplexer then reorganizes the decoded bit streams and forwards them to the SHVC Decoder, as illustrated in Figure 5.14, where the received tile-based videos $\{\hat{v}_i\}_{i=1}^N$ are reconstructed, which are then converted back to the panoramic video $\hat{V}$ displayed.

5.5 Optimal FEC Code-Rate Allocation

This section aims for finding the most appropriate FEC code-rate allocation expected to exhibit the best received video quality when both the system bandwidth and the channel conditions are constrained. Recall from Figure 5.7 that the tiles of a specific frame are encoded independently of each other, but the tiles in the same portion of the consecutive frames obey inter-frame coding dependency, representing I-frames and P-frames. Hence, the FEC code-rate allocation algorithm is designed for ensuring that even the least important video bits of the tiles in C1 are better protected than the most important bits of the tiles in C2 and C3. Moreover, due to the projection mapping distortion detailed in Figure 5.11, the different tiles in the same class may also exhibit different importances and hence are protected unequally. This unequal protection may be arranged by introducing a pair of weighting coefficients, which will be detailed in Section 5.5.3.

The rest of this section is organized as follows. Our code-rate optimization problem is formulated in Section 5.5.1 based on a Cost Function (CF) maximizing the uncoded source-rate by selecting the most appropriate code-rate providing adequate resilience. Furthermore, a pair of weighting coefficients are introduced for judiciously adjusting the relative importance of the bit-protection classes, and the inter-frame as well as the inter-layer coding dependencies. The following sections then elaborate on maximizing the uncoded source-rate by finding the optimal FEC coding rates at a given channel SNR and fixed modulation mode. Specifically, in Section 5.5.2, we mathematically model the relationship between the channel SNR and the FEC coding rates for the three modulation modes. Then, the weighting coefficients of the CF will be estimated in Section 5.5.3, while the optimal FEC coding rates are discussed in Section 5.5.4. To clearly present the code-rate optimization algorithm, the most complex 128QAM conveying all the three classes is exemplified in the following sections, but the CF can be readily modified for the other modulation modes.

5.5.1 Problem Formulation

In this section, we mainly formulate the FEC coding-rate allocation problem for the sake of maximizing the overall weighted uncoded source-rate, where the problem formulation takes into account the tile weighting coefficients of Figure 5.11, the inter-frame and inter-layer coding dependencies. Assuming that the panoramic video is partitioned into N tiles and each has the GoP size of G, the overall weighted uncoded source-rate across a GoP period can be formulated as:

$$\begin{aligned} \max \quad & T_\omega\left(\{r_{i,j}^l\}, \Gamma\right), && (5.15)\\ s.t.: \quad & \sum_{i=1}^{N}\sum_{j=1}^{G}\frac{|f_{i,j}^1|}{r_{i,j}^1} + \sum_{i\in FoV}\sum_{j=1}^{G}\frac{|f_{i,j}^2|}{r_{i,j}^2} \leq T, && (5.16)\\ & r_{i,j}^l \in \mathcal{R}. && (5.17)\end{aligned}$$

(5.16) imposes a constraint on the FEC code-rate allocation for having an FEC coded rate below the upper bound of T, where $|f_{i,j}^l|$ represents the frame length of the video frame $f_{i,j}^l$ in terms of the number of bits. Additionally, (5.17) represents the constraint that the FEC code-rate has to be selected from the legitimate code-rate set $\mathcal{R}$. (5.15) can be reformulated as the weighted uncoded source-rate summation of the three classes, as given by:

$$T_\omega\left(\{r_{i,j}^l\}, \Gamma\right) = \sum_{c=1}^{3} T_\omega^c\left(\{r_{i,j}^l\}, \Gamma\right), \tag{5.18}$$

where $T_\omega^c\left(\{r_{i,j}^l\}, \Gamma\right)$ represents the weighted uncoded source-rate contributed by Class $c \in \{1, 2, 3\}$, which can be further formulated as:

$$T_\omega^c\left(\{r_{i,j}^l\}, \Gamma\right) = \begin{cases} \beta_1 \sum\limits_{i\in FoV} \alpha_i \sum\limits_{j=1}^{G} T_{i,j}^1\left(r_{i,j}^1, \Gamma\right), & c = 1,\\ \beta_2 \sum\limits_{i\notin FoV} \alpha_i \sum\limits_{j=1}^{G} T_{i,j}^1\left(r_{i,j}^1, \Gamma\right), & c = 2,\\ \beta_3 \sum\limits_{i\in FoV} \alpha_i \sum\limits_{j=1}^{G} T_{i,j}^2\left(r_{i,j}^2, \Gamma\right), & c = 3.\end{cases} \tag{5.19}$$

Moreover, β_c represents the weighting coefficients of the protection priority, assigning stronger protection to the more important classes, while α_i represents the tile importance within the class by referring Figure 5.11, both of which will be detailed in Section 5.5.3. Furthermore, $T_{i,j}^l\left(\{r_{i,j}^l\}, \Gamma\right)$ indicates the uncoded source-rate attributed to the corresponding video frame $f_{i,j}^l$ at a given channel SNR Γ, which is formulated as:

$$T_{i,j}^l\left(r_{i,j}^l, \Gamma\right) = |f_{i,j}^l| \prod_{j=1}^{j}\prod_{\ell=1}^{l} \psi\left(|f_{i,j}^\ell|, r_{i,j}^\ell, \Gamma\right). \tag{5.20}$$

In (5.20), $\psi\left(|f_{i,j}^\ell|, r_{i,j}^\ell, \Gamma\right)$ represents the SDP of video frame $f_{i,j}^\ell$ and $T_{i,j}^l\left(r_{i,j}^l, \Gamma\right)$ is obtained by taking into account the inter-layer and inter-frame coding dependency. Substituting (5.19) and (5.20) into

Table 5.2: Parameters values of (5.24) for each modulation mode.

M	a_4	a_3	a_2	a_1	b_4	b_3	b_2	b_1	c
BPSK	-0.0001581	0.003889	0.03512	-1.942	380.3	-745.8	501.6	-103.3	-1.299
16QAM	-0.0001072	0.006789	-0.1048	-1.149	375.4	-736	492.6	-99.17	-14.42
128QAM	0	0.0009328	-0.03699	-0.888	495.6	-993.5	692.8	-165.8	-28.77

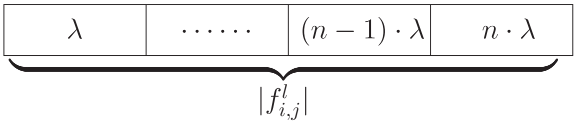


Figure 5.15: A video frame $f^l_{i,j}$ having a frame length of $|f^l_{i,j}| = n \times \lambda$ bits may be partitioned into n shorter packets withe a length of λ bits.

(5.18), we can have the expression of the weighted uncoded source-rate, as given by:

$$\begin{aligned} T_\omega(\{r^l_{i,j}\},\Gamma) = \beta_1 &\sum_{i\in FoV} \alpha_i \sum_{j=1}^{G} |f^1_{i,j}| \prod_{j=1}^{j} \psi(|f^1_{i,j}|, r^1_{i,j}, \Gamma) \\ + \beta_2 &\sum_{i\notin FoV} \alpha_i \sum_{j=1}^{G} |f^1_{i,j}| \prod_{j=1}^{j} \psi(|f^1_{i,j}|, r^1_{i,j}, \Gamma) \\ + \beta_3 &\sum_{i\in FoV} \alpha_i \sum_{j=1}^{G} |f^2_{i,j}| \prod_{j=1}^{j} \prod_{\ell=1}^{2} \psi(|f^\ell_{i,j}|, r^\ell_{i,j}, \Gamma). \end{aligned} \tag{5.21}$$

5.5.2 Successful Video Frame Decoding Probability

In this subsection, we elaborate on mathematically modelling the SDP of the variable-length video frame for the three modulation modes. It is acknowledged that the FEC performance varies with the input packet length regardless of the FEC configuration [155]. Additionally, the video frames are usually compressed to the bit streams exhibiting different frame lengths in terms of number of bits. These constraints impose further difficulty on predicting the FER of the variable-length video frame. Fortunately, the authors of [32] [48] proposed the FER estimation method for the video frame exhibiting variable bit length, which is based on the assumption that the video frame error probability imposed by the fading channel is independent of the packet length when the distribution of error is considered to be uniform. For characterizing the video frame error distributions, we assume that the PLR of a packet containing λ bits is p. Then the SDP of an $|f^l_{i,j}|$-bit video frame, where we have $|f^l_{i,j}| = n \times \lambda$, can be obtained by:

$$\psi = \left[1-p\right]^n, \tag{5.22}$$

which is depicted in Figure 5.15. Similarly, the SDP of an arbitrary video frame containing different number of bits can be found by exploiting the PLR of a packet having a fixed length of λ, which is given by:

$$\psi(|f^{\ell}_{i,j}|, r^{\ell}_{i,j}, \Gamma) = \left[1 - p\left(r^{\ell}_{i,j}, \Gamma\right)\right]^{\frac{|f^{\ell}_{i,j}|}{\lambda}}, \tag{5.23}$$

where $p\left(r^{\ell}_{i,j}, \Gamma\right)$ denotes the PLR of the λ-bit packet at a given code-rate $r^{\ell}_{i,j}$ and at an instantaneous channel SNR Γ. Furthermore, for the sake of modelling the PLR under various SNR values Γ and FEC code-rate r, we first pre-recorded the PLR associated with those parameters for a long video streaming session and stored them in a LUT $h[r, \Gamma]$ based on our Monte-Carlo experiments. This LUT is then used for creating the following model:

$$\begin{aligned} p\left(r^{\ell}_{i,j}, \Gamma\right) = h[r, \Gamma] \approx \hbar[r, \Gamma] = \\ \frac{1}{1 + e^{-(a_4\Gamma^4 + a_3\Gamma^3 + a_2\Gamma^2 + a_1\Gamma + b_4 r^4 + b_3 r^3 + b_2 r^2 + b_1 r) + c}}, \end{aligned} \tag{5.24}$$

as exemplified in Figure 5.16, where $\hbar[r, \Gamma]$ represents the fitted mathematical model. Specifically, Figure 5.16(a) compares the scatter-diagrams of the Monte-Carlo simulated and the mathematically fitted curves, while Figure 5.16(b) depicts the absolute difference between them, namely $|h[r, \Gamma] - \hbar[r, \Gamma]|$. The numerical values of the parameters used for modelling the different modulation modes are given in Table 5.2. It is worth noting that (5.24) and the values included in Table 5.2 are evaluated empirically.

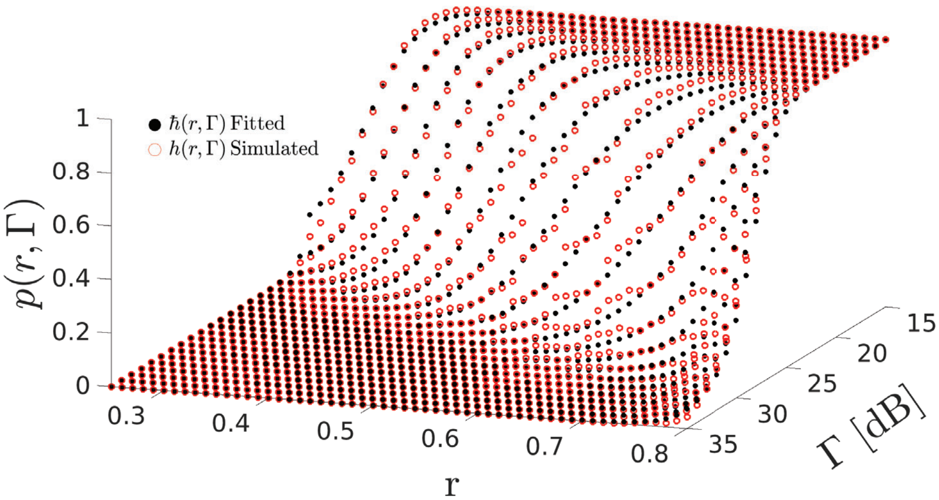
5.5.3 Estimation of the Weighting Factors

The weighting coefficients α and β in (5.21) indicate the importance of each class of the tile-based panoramic video, whose aim is to ensure that the resources are judiciously allocated, commensurately to the importance of the information. This subsection explicitly details the determination of the weighting factors α and β introduced in (5.21). Explicitly, the weighting coefficient α tends to guarantee that within the same class the tile exhibiting the lowest transformation distortion of Figure 5.11 is strongly protected, while β is used for ensuring the robustness of the bit streams of the more important classes.

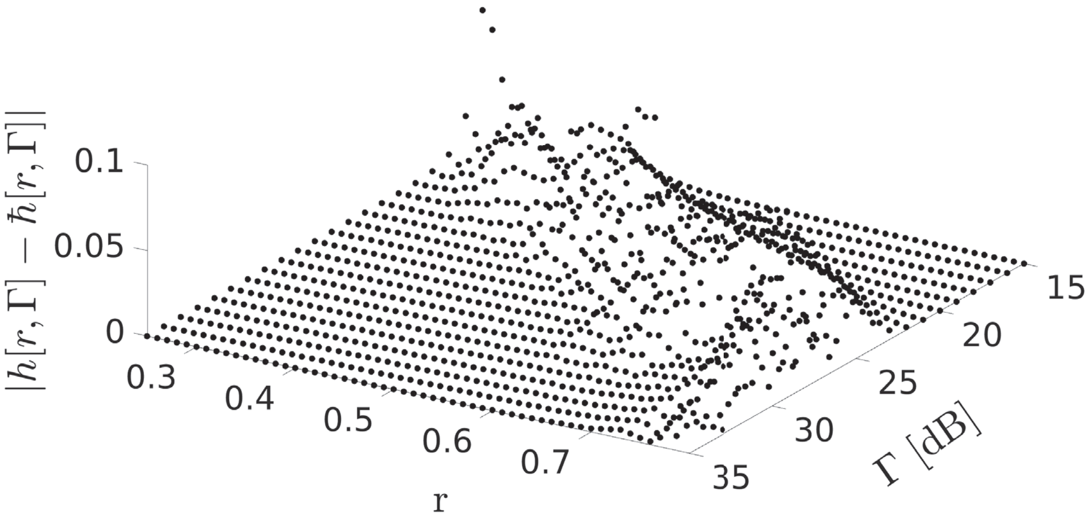
5.5.3.1 Estimation of the tile weighting factor α

Since the panoramic video is converted from the spherical format to the 2D format, according to the ERP projection ready for compression and streaming, the importance of the individual pixels on the original sphere may become different in the projected format, as illustrated in Figure 5.11.

In this section, we elaborate on calculating the weighting factor of α_i for each tile v_i. Our methodology to generate α_i is to first partition the weighting coefficients obtained in (5.12) of Figure 5.14 into the same tiles of Figure 5.7, yielding $\vartheta_i(w, h)$. Then, the weighting factor of α_i of tile v_i can be obtained by averaging all the weighting coefficients of the partitioned $\vartheta_i(w, h)$, which



(a) Simulated scatter-diagram and its fitted mathematical counterpart for 128QAM, where the parameter are given in Table 5.2 and $\lambda = 1000$ bits.



(b) Absolute difference between the Simulated and Fitted scatter points of Figure 5.16a.

Figure 5.16: Model comparison between the simulated scatter-diagram and its fitted mathematical counterpart for 128QAM.

is given by:

$$\bar{\alpha}_i = \frac{\sum\limits_{w=1}^{W_i} \sum\limits_{h=1}^{H_i} \vartheta_i(w,h)}{W_i \times H_i}, \tag{5.25}$$

where W_i and H_i denote the width and height of the tile v_i, while w and h represent the horizontal and vertical pixel positions within the tile v_i. The tile weighting factors α_i are then normalized as:

$$\{\alpha_i\}_{i=1}^{N} = \frac{\bar{\alpha}_i}{\sum\limits_{i=1}^{N} \bar{\alpha}_i}. \tag{5.26}$$

5.5.3.2 Estimation of class weighting factor β

In this section, we aim for obtaining the optimized weighting factors β_1, β_2 and β_3 that maximize the weighted uncoded source-rate of (5.21), while taking into account the priority classes of C1, C2 and C3. More explicitly, the video bits of the tiles in C1 that most substantially affect the user's perceptual experience are expected to be of the highest priority, followed by the tiles in C2 and C3, respectively. Therefore, the weighting factors β_1, β_2 and β_3 are selected based on our working assumption that even the least important video bits of the tiles in C1 must be better protected than the most important bits of the tiles in C2. This can be achieved by defining the tentative weights β'_1 and β'_2, which ensure that the weighted uncoded source-rate contributed by the least important bits in C1, namely $\min\limits_{i\in FoV}\{|f^1_{i,G}|\}$, always remains higher than that of the most important uncoded source bits in C2, namely $\max\limits_{i\notin FoV}\{|f^1_{i,1}|\}$, which is formulated as:

$$\beta'_1 \min_{\forall i\in FoV}\{|f^1_{i,G}|\} \geq \beta'_2 \max_{\forall i\notin FoV}\{|f^1_{i,1}|\}. \tag{5.27}$$

Similarly, for C2 and C3, the tentative factors β'_2 and β'_3 endeavour to ensure that the least important video bits of C2, i.e. $\min\limits_{\forall i\notin FoV}\{|f^1_{i,G}|\}$, are better protected than the most important video bits of C3, namely $\max\limits_{\forall i\in FoV}\{|f^2_{i,1}|\}$, which is formulated as:

$$\beta'_2 \min_{\forall i\notin FoV}\{|f^1_{i,G}|\} \geq \beta'_3 \max_{\forall i\in FoV}\{|f^2_{i,1}|\}. \tag{5.28}$$

Finally, setting β'_3 representing the weight of the least important information to 1.0, the class weighting factors are expressed as:

$$\begin{aligned}
\beta'_1 &= \frac{\max\limits_{\forall i\notin FoV}\{|f^1_{i,1}|\}}{\min\limits_{\forall i\in FoV}\{|f^1_{i,G}|\}} \cdot \frac{\max\limits_{\forall i\in FoV}\{|f^2_{i,1}|\}}{\min\limits_{\forall i\notin FoV}\{|f^1_{i,G}|\}},\\
\beta'_2 &= \frac{\max\limits_{\forall i\in FoV}\{|f^2_{i,1}|\}}{\min\limits_{\forall i\notin FoV}\{|f^1_{i,G}|\}},\\
\beta'_3 &= 1.0,
\end{aligned} \tag{5.29}$$

which are then normalized according to:

$$\begin{aligned}\beta_1 &= \frac{\beta_1'}{\beta_1' + \beta_2' + \beta_3'},\\ \beta_2 &= \frac{\beta_2'}{\beta_1' + \beta_2' + \beta_3'},\\ \beta_3 &= \frac{\beta_3'}{\beta_1' + \beta_2' + \beta_3'}.\end{aligned} \tag{5.30}$$

5.5.4 Coding Rate Optimization

By substituting (5.20), (5.23), (5.24), (5.26) and (5.30), into (5.21), the weighted uncoded source-rate can be further detailed as:

$$\begin{aligned} T_\omega\left(\{r_{i,j}^l\}, \Gamma\right) = \beta_1 &\sum_{i \in FoV} \alpha_i \sum_{j=1}^{G} |f_{i,j}^1| \prod_{j=1}^{j} \left[1 - p(r_{i,j}^1, \Gamma)\right]^{\frac{|f_{i,j}^1|}{\lambda}} \\ + \beta_2 &\sum_{i \notin FoV} \alpha_i \sum_{j=1}^{G} |f_{i,j}^1| \prod_{j=1}^{j} \left[1 - p(r_{i,j}^1, \Gamma)\right]^{\frac{|f_{i,j}^1|}{\lambda}} \\ + \beta_3 &\sum_{i \in FoV} \alpha_i \sum_{j=1}^{G} |f_{i,j}^2| \prod_{j=1}^{j} \prod_{\ell=1}^{2} \left[1 - p(r_{i,j}^\ell, \Gamma)\right]^{\frac{|f_{i,j}^\ell|}{\lambda}}, \end{aligned} \tag{5.31}$$

where in addition to accounting for the inter-frame coding dependency, the third item of (5.31) takes into account the inter-layer dependencies as well.

Assuming that the instantaneous channel SNR Γ, the modulation mode M and the encoded video frame lengths $\{|f_{i,j}^l|\}_{\forall i,j,l}$ expressed in bits are known in the system of Figure 5.14, the maximization of the weighted uncoded source-rate of (5.31) is equivalent to exploiting the optimal coding rate $\{r_{i,j}^l\}_{\forall i,j,l}$ for the associated video frame $\{f_{i,j}^l\}_{\forall i,j,l}$, which can be expressed as:

$$\{r_{i,j}^l\}_{opt} = \underset{\{r_{i,j}^l\}_{\forall i,j,l} \in \mathcal{R}}{\arg\max} \left\{ T_\omega\left(\{r_{i,j}^l\}\right) \right\}. \tag{5.36}$$

The optimal coding rate set $\{r_{i,j}^l\}_{\forall i,j,l}$ is restricted to the interval $\mathcal{R}$ of $[r_{min}, r_{max}]$ constrained in (5.17), where r_{min} denotes the lowest coding rate, which is capable of providing the best protection for the source information, while r_{max} the highest available code-rate.

The EA assisted coding rate optimization is detailed in Algorithm 2, which aims for maximizing the weighted uncoded source-rate based CF of (5.31), whilst maintaining the throughput below the upper bound constraint of (5.16). It is worth noting that for different modulation modes the CF of (5.31) and the upper bound constraint of (5.16) may be slightly different. For example, in BPSK that only transmits C1 bits, the class weighting factor β is deactivated and the CF used in Algorithm 2 is simplified to the first term at the right hand side of (5.31).

Algorithm 2 Coding Rate Allocation Algorithm

1: **Initialization**: Set the number of generations to $g = 1$ and randomly generate the initial coding rate set $\{\hat{\mathbf{r}}_{g,p_s}\}_{p_s=1}^{Ps}$, where $\hat{\mathbf{r}}_{g,p_s} = \{r_{i,j}^{l}\}_{g,p_s}$, whose elements must be gleaned from the set $\mathcal{R}$ of $[r_{min}, r_{max}]$ and have to satisfy the constraint condition of (5.16). The crossover probability C_r of the EA is initialized to $\mu_{C_r} = 0.5$, whereas the scaling factor to $\mu_\lambda = 0.5$. The archive of the EA is initialized to be empty.

2: **Population evaluation**: Evaluate the CF value given in (5.31) for each $\hat{\mathbf{r}}_{g,p_s}$, where $1 \leq p_s \leq P_s$. The archive of EA contains the P_s best candidate solutions for $\{\hat{\mathbf{r}}_{g,p_s}\}_{p_s=1}^{P_s}$, and is updated every generation by adding the $\lfloor P_s \cdot \eta \rfloor$ parent solutions in the top $100 \cdot \eta\%$ of the high fitness to it, where η is the greedy factor. If the archive size exceeds P_s, some candidates have to be randomly removed from it.

3: **Mutation**: The mutation perturbs the candidate solutions by adding randomly selected and appropriately scaled difference-vectors to each base population vector $\hat{\mathbf{r}}_{g,p_s}$ as follows:

$$\begin{aligned}\tilde{\mathbf{r}}_{g,p_s} =& \hat{\mathbf{r}}_{g,p_s} + \lambda_{p_s}(\hat{\mathbf{r}}_{g,p_s,best} - \hat{\mathbf{r}}_{p_s}) \\ &+ \lambda_{p_s}(\hat{\mathbf{r}}_{g,p_s,c_1} - \hat{\mathbf{r}}_{g,p_s,c_2}),\end{aligned} \tag{5.32}$$

where the scaling factor $\lambda_{p_s} \in (0, 1]$ is randomly generated for each individual according to the Cauchy distribution having a mean of μ_λ and standard deviation of 0.1. Then $\hat{\mathbf{r}}_{g,p_s,best}$ is randomly selected from the archive, while c_1 and c_2 are two random integers fetched from the set $\{1, 2, \cdots, p_s - 1, p_s + 1, \cdots, P_s\}$.

4: **Crossover**: Substituting the target $\hat{\mathbf{r}}_{g,p_s}$ by the corresponding elements of the donor vector $\tilde{\mathbf{r}}_{g,p_s}$, a trial vector $\check{\mathbf{r}}_{g,p_s}$ is generated. Specifically, $\check{r}_{i,j,g,p_s}^{l}$ of the p_s-th trial gleaned from $\check{\mathbf{r}}_{g,p_s}$ can be obtained by:

$$\check{r}_{i,j,g,p_s}^{l} = \begin{cases} \tilde{r}_{i,j,g,p_s}^{l}, rand(0,1) \leq C_{r_{p_s}} \\ \hat{r}_{i,j,g,p_s}^{l}, \text{otherwise}, \end{cases} \tag{5.33}$$

where $C_{r_{p_s}} \in [0, 1]$ is the randomly generated crossover probability for each individual based on the normal distribution with location parameter μ_{C_r} and the scaled parameter 0.1.

5: **Selection**: If $T_\omega(\check{\mathbf{r}}_{g,p_s}) \geq T_\omega(\hat{\mathbf{r}}_{g,p_s})$, the trail vector survives to the next generation and we set $\hat{\mathbf{r}}_{g+1,p_s} = \check{\mathbf{r}}_{g,p_s}$. Otherwise, the target vector survives and we set $\hat{\mathbf{r}}_{g+1,p_s} = \hat{\mathbf{r}}_{g,p_s}$.

6: **Adaptation**: The mean of the crossover probability μ_{C_r} and the scaling factor μ_λ are updated according to [156]:

$$\mu_{C_r} = (1 - c) \cdot \mu_{C_r} + c \cdot mean_A(S_{C_r}) \tag{5.34}$$

$$\mu_\lambda = (1 - c) \cdot \mu_\lambda + c \cdot mean_L(S_\lambda), \tag{5.35}$$

where $c \in (0, 1]$ is the adaptive update factor, $mean_A(\cdot)$ and $mean_L(\cdot)$ denote the arithmetic-mean and Lehmer-mean operators [156], while S_{C_r} and S_λ denote the sets of successful crossover probabilities C_r, and scaling factors λ_i in generation g.

7: **Termination**: The procedure is terminated, when the maximum number of generation G_{max} is reached. Otherwise, set $g = g + 1$ and go back to Step 2.

5.6 Adaptive QAM

Again, the proposed AQAM system of Figure 5.14 employs three modulation modes that exhibit different throughputs and robustness in order to provide a flexible solution striking the best possible compromise between them. Potent AQAM assisted designs have been investigated ever since the concept of AQAM modem was proposed [95], which has also been employed for video streaming [92–94]. In this section, we elaborate on exploiting the most appropriate mode-switching operation for our AQAM system for the sake of maximizing the received video quality.

Explicitly, the mode-switching operation has to select the most appropriate modulation mode capable of providing the best performance among all candidate modes at the given near-instantaneous channel conditions. At low channel SNRs the most robust modulation mode is invoked for providing low-rate, but potentially error-free video service, while at high channel SNRs the highest-throughput modulation mode improves the video quality. To elaborate a little further, our AQAM design judiciously selects the most appropriate modulation mode having the highest throughput at a given channel SNR, whilst guaranteeing that the most important BL is flawlessly decoded.

Our objective is to activate the highest coding rate $R_{\Gamma,M}$ for modulation mode M at a specific channel SNR Γ that is capable of achieving error-free protection. Hence, the modulation mode selection used for the proposed UEP scheme can be designed relying on a pair of conditions: (1) the coding rate $R_{\Gamma,M}$ exists and (2) the modulation mode is capable of yielding the highest uncoded source-rate. The first condition guarantees that the most important information, such as the BL of the FoV, can be successfully decoded, while the second one ensures that the modulation mode activated maximizes the uncoded source-rate. Therefore, at the specific channel SNR Γ the modulation mode selection procedure first verifies the existence of a legitimate coding rate $\exists R_{\Gamma,M}$ for each modulation mode and the modulation candidates are then compared based on the criterion:

$$M = \underset{M\in\{BPSK,16QAM,128QAM\}}{\arg\max}\{T_M\} \quad , \tag{5.37}$$

where T_{BPSK}, T_{16QAM} and T_{128QAM} represent the uncoded source-rate provided by BPSK, 16QAM and 128QAM, respectively, which can be predicted from (5.31), provided that the FEC coding rates are optimized. For example, the uncoded source-rate for 128QAM is expressed as:

$$\begin{aligned} T_{128QAM} = & \sum_{i\in FoV}\sum_{j=1}^{G}|f^1_{i,j}|\prod_{\jmath=1}^{j}\left[1-p(r^1_{i,\jmath},\Gamma)\right]^{\frac{|f^1_{i,\jmath}|}{\lambda}} \\ & + \sum_{i\notin FoV}\sum_{j=1}^{G}|f^1_{i,j}|\prod_{\jmath=1}^{j}\left[1-p(r^1_{i,\jmath},\Gamma)\right]^{\frac{|f^1_{i,\jmath}|}{\lambda}} \\ & + \sum_{i\in FoV}\sum_{j=1}^{G}|f^2_{i,j}|\prod_{\jmath=1}^{j}\prod_{\ell=1}^{2}\left[1-p(r^\ell_{i,\jmath},\Gamma)\right]^{\frac{|f^\ell_{i,\jmath}|}{\lambda}}, \end{aligned} \tag{5.38}$$

where $p(r^\ell_{i,\jmath},\Gamma)$ can be found in (5.24) and Table 5.2. It can be seen in (5.38) that obtaining the uncoded source-rate requires the information of the optimal FEC code-rate assignment, which is

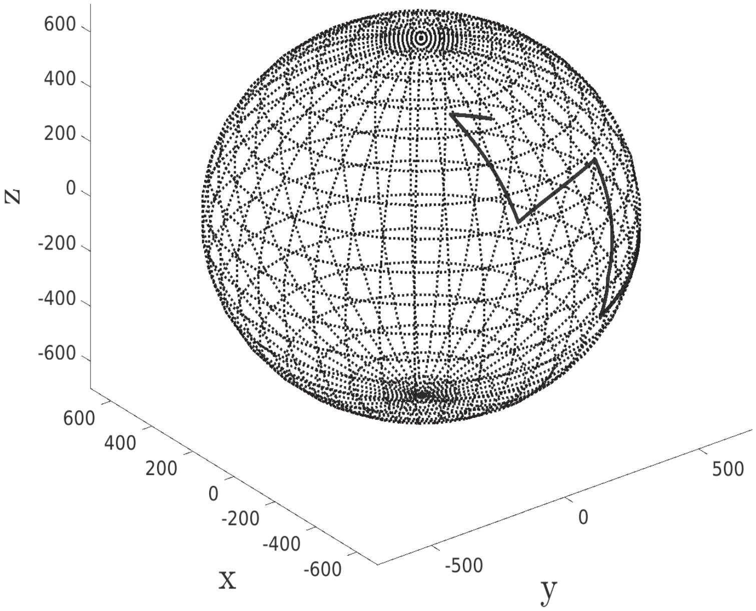


Figure 5.17: FoV track prediction.

exemplified in Section 5.5. Hence, the modulation selection (5.37) has to be invoked only when the optimal FEC code-rates are obtained for each modulation mode.

Additionally, to benchmark our UEP assisted AQAM system, the modulation mode assignment of the EEP assisted AQAM system is also detailed in this section, which employs the same mode-switching operation as the above-mentioned EA assisted UEP system. Fundamentally, the highest-order modulation mode is activated when it is capable of flawlessly conveying the most important FoV BL bits. The EEP assisted AQAM scheme requires that all bit streams are sufficiently well protected, namely that $R_{\Gamma,M} \geq R$, where R is the EEP coding rate. For example, for 128QAM this is given by:

$$R = \frac{\sum_{i=1}^{N} \sum_{j=1}^{G} |f_{i,j}^{1}| + \sum_{i \in FoV} \sum_{j=1}^{G} |f_{i,j}^{2}|}{T}, \tag{5.39}$$

where N and T represent the total number of video tiles and the throughput upper bound across a GoP, respectively.

Table 5.3: Parameters used in simulations.

Parameters	Value
Representation	YUV 4:2:0
Bits Per Pixel	8
GoP	8
FPS	30
No. of Frames	64
FoV	90°
Projection	ERP
Tiles	$4 \times 8 = 32$
Resolution	4096×2048
Resolution Per Tile	512×512
Video Codec	SHVC
FEC Generator Polynomials	RSC[17 15 13 11]
Channel	uncorrelated block Rayleigh
Shadow Fading Standard Deviation	4 dB
Modulation	BPSK/16QAM/128QAM
STBC	2×1
Bandwidth	3 MHz
Simulations Repeated	100

5.7 Simulation Performance

In this section, we benchmark our proposed AQAM assisted variable rate FEC aided system against its EEP counterpart. The parameters of the *RaceVR* sequence employed in the simulations are listed in Table 5.3. The ERP having a resolution of 4096×2048 pixels, as depicted in Figure 5.7, is partitioned into 32 tiles with each consisting of 512×512 pixels.

Specifically, the *RaceVR* video clip of 4:2:0 YUV format was encoded by the SHVC reference software into numerous Network Abstraction Layer Unit (NALU) packets at a scanning rate of 30 FPS, each having CRC bits attached to its tail to check the error-free SDP of the received video frame. If the CRC detection fails, the corrupted packets are discarded and the corresponding abandoned video frames are replaced by “frame-copy” based error concealment using the most recent error-free video frame. The GoP is set to 8, which means that the so-called IDR/CRA frames of SHVC are inserted every 8 frames. The bidirectionally predicted B frames are deactivated in our simulations, since they are prone to propagating inter-frame video distortions as well as to imposing decoding latency. As a consequence, the video sequence in our simulations only consists of I- and P-frames.

In addition to the video parameters, Table 5.3 also lists the transmission parameters of our system. The RSC code having octally represented generator polynomials of [17 15 13 11] is employed, which exhibits a coding rate r_{org} of 1/4. The RSC code used for UEP is punctured to the higher coding rates till 0.8, which exhibits the FEC-rate set having 56 candidates of [0.25:0.01:0.8], while the RSC code-rate used for EEP assisted system is fixed to 0.5. Furthermore, the AQAM assisted 2×1 STBC scheme is employed for generating the signals for wireless transmission. The population size P_s and the maximum number of generation used by our EA algorithm are set to 300 and 1000, respectively. Furthermore, uncorrelated block Rayleigh fading is employed for all our simulations. Additionally, shadow fading having a standard deviation of 4 dB is also imposed. The bandwidth is set to 3MHz, imposing the maximum bit rates of 3 Mbps, 12 Mbps and 21 Mbps for BPSK, 16QAM and 128QAM, respectively, when assuming a Nyquist roll-off factor of 0. Additionally, the spherical FoV track is exemplified in Figure 5.17 with the aid of the ERP mapping illustrated in Figure 5.8, which can be estimated using the methods of [157] [158].

5.7.1 Complexity Analysis

Recall from Section 5.5 that the optimization is carried out across the GoP period, exhibiting an optimization period of $\frac{G}{\mathcal{F}}$, where $\mathcal{F}$ represents the FPS of the video sequence. Particularly, in the example of Table 5.3, where $G = 8$ and $\mathcal{F} = 30$ the modulation mode and its associated FEC code-rate are optimized every $\frac{8}{30} \approx 0.267$ second. The optimization frequency decreases with the growing GoP period at the specific $\mathcal{F}$, for example the optimization period increasing to 0.533 second when doubling the G. However, this is achieved at the expense of additional optimization complexity imposed, since the increased number of parameters to be optimized across a GoP period requires more generations to converge and hence degrades the convergence performance.

Furthermore, in the optimization, a low-complexity termination criterion is constituted by the number of CF evaluations K, which may be readily used for evaluating the loose computational complexity imposed. For a specific population size P_s terminated after G_{max} generations, the number of K employed by the EA for finding the optimal FEC code-rate set is equal to P_sG_{max}. By contrast, the number of K using exhaustive search for modulation mode M is equivalent to $(N_MG)^{|\mathcal{R}|}$, where N_M represents the number of the tile conveyed when modulation mode M is activated, and $|\mathcal{R}|$ denotes the number of FEC code-rate candidate set $\mathcal{R}$. Hence, the complexity imposed by the EA is $\frac{P_sG_{max}}{(N_MG)^{|\mathcal{R}|}}$ of that required by the exhaustive search. In fact, the values of N_M, G and $|\mathcal{R}|$ also affect the value of K employed by the EA, since a larger G_{max} has to be employed in order to maintain the convergence. However, since the treatise aims to propose an EA assisted FEC code-rate allocation for layered panoramic video streaming, the balance between the above parameters, namely N_M, G and $|\mathcal{R}|$, and the convergence performance G_{max} is beyond the scope of this treatise and this is left for our future work.

5.7.2 Mode Switching Performance

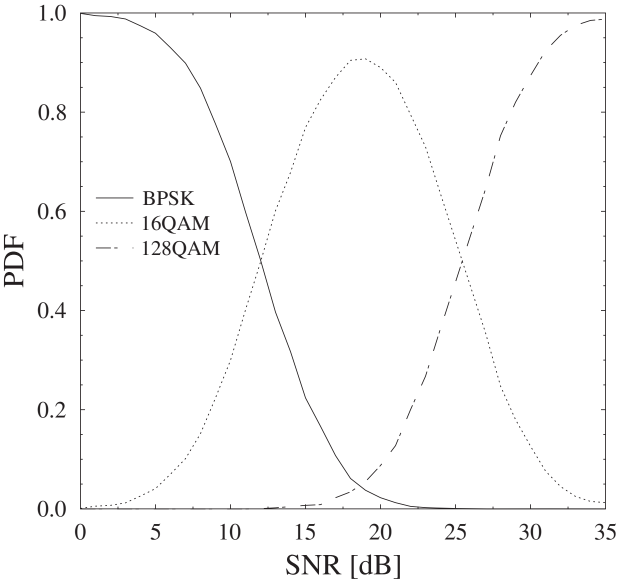


Figure 5.18: PDF of proposed UEP assisted adaptive modem modes versus the channel SNR.

Figure 5.18 depicts the PDF of the three modulation modes versus the average channel SNR according to the mode-switching criteria detailed in Section 5.6. Observe that at a channel SNR of $\Gamma <12$ dB, BPSK is the most frequently used modulation mode, while at $\Gamma >26$ dB, the probability of activating 128QAM is higher than that of BPSK and 16QAM. Furthermore, Figure 5.19 shows how the AQAM system changes its modulation modes based on the near-instantaneous channel conditions at an average channel SNR of 16 dB, which shows that 16QAM is more frequently selected, while the other two modulation modes are only occasionally activated based on the criteria of Section 5.6.

5.7.3 Quality of Experience Performance

Figure 5.20 compares the simulation results of our proposed UEP assisted AQAM and of its EEP assisted counterparts. To elaborate, Figure 5.20(a)-(d) portray the PSNR of the FoV, the FER of the BL of the FoV, the WS-PSNR of the ERP projection and the FER of the BL of the ERP projection, respectively. Observe from Figure 5.20 that the EA assisted UEP designs on average outperform their EEP counterparts. The BPSK mode that only transmits the FoV bits is characterized in Figure 5.20(a) and (b), where about 1.5 dB channel SNR gain is provided by the UEP design at the FER of 5%. The gains exhibited in the UEP assisted BPSK mode may be attributed to the fact that

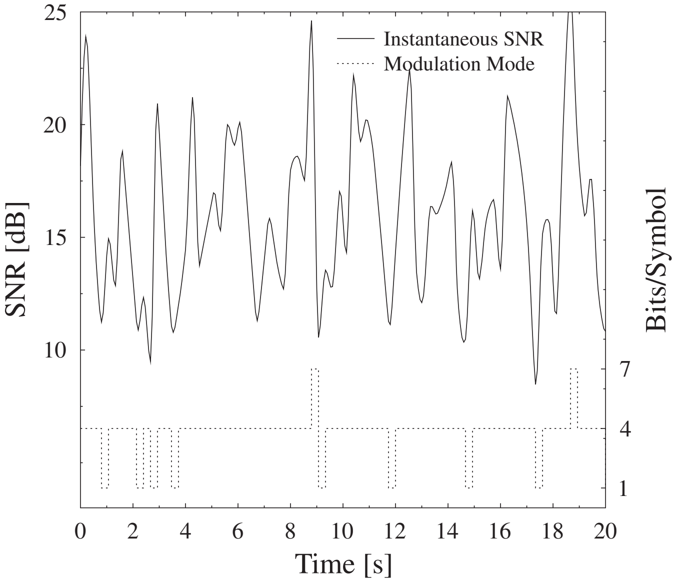


Figure 5.19: Modulation mode switching at an average channel SNR of 16 dB.

the CF of (5.31) optimized by the EA takes into account the video coding dependencies.

Additionally, the UEP assisted 16QAM significantly improves the FoV performance, whilst maintaining a similar WS-PSNR and FER performance to its EEP assisted counterpart. Specifically, the UEP and EEP assisted 16QAM modes achieve a FoV FER of 5% at approximately 16.5 and 21.5 dB channel SNR, respectively, as illustrated in Figure 5.20(b). Furthermore, Figure 5.20(a) shows that the PSNR of the FoV exhibited by the UEP assisted 16QAM design is about 3 dB higher than that of its EEP counterpart at the channel SNR of 18 dB.

Furthermore, as a benefit of optimizing the CF of (5.31), the proposed EA assisted UEP system outperforms its EEP assisted counterpart by approximately 9 dB channel SNR at a 5% FER value of the FoV, as seen in Figure 5.20(b). Quantitatively, the UEP and EEP schemes require channel SNRs of 22 and 31 dB at FER=5%, respectively. A range of further insights may be gleaned upon scrutinizing Figure 5.20, which are left for the reader to explore.

A subjective video quality comparison of the benchmarkers recorded for the FoV of the *RaceVR* test sequence is presented in Figure 5.21 at a channel SNR of 12 dB. The UEP assisted BPSK of Figure 5.21(a) and the EEP assisted BPSK of Figure 5.21(b) as well as UEP assisted 16QAM of Figure 5.21(c) are capable of reconstructing the image, while the UEP assisted BPSK in Figure 5.21(a) exhibits the smallest distortion. To make the visual comparison more explicit, in Figure 5.21(d), (e) and (f) we portrayed the error between the original and received video frames, where a darker region represents a higher error.

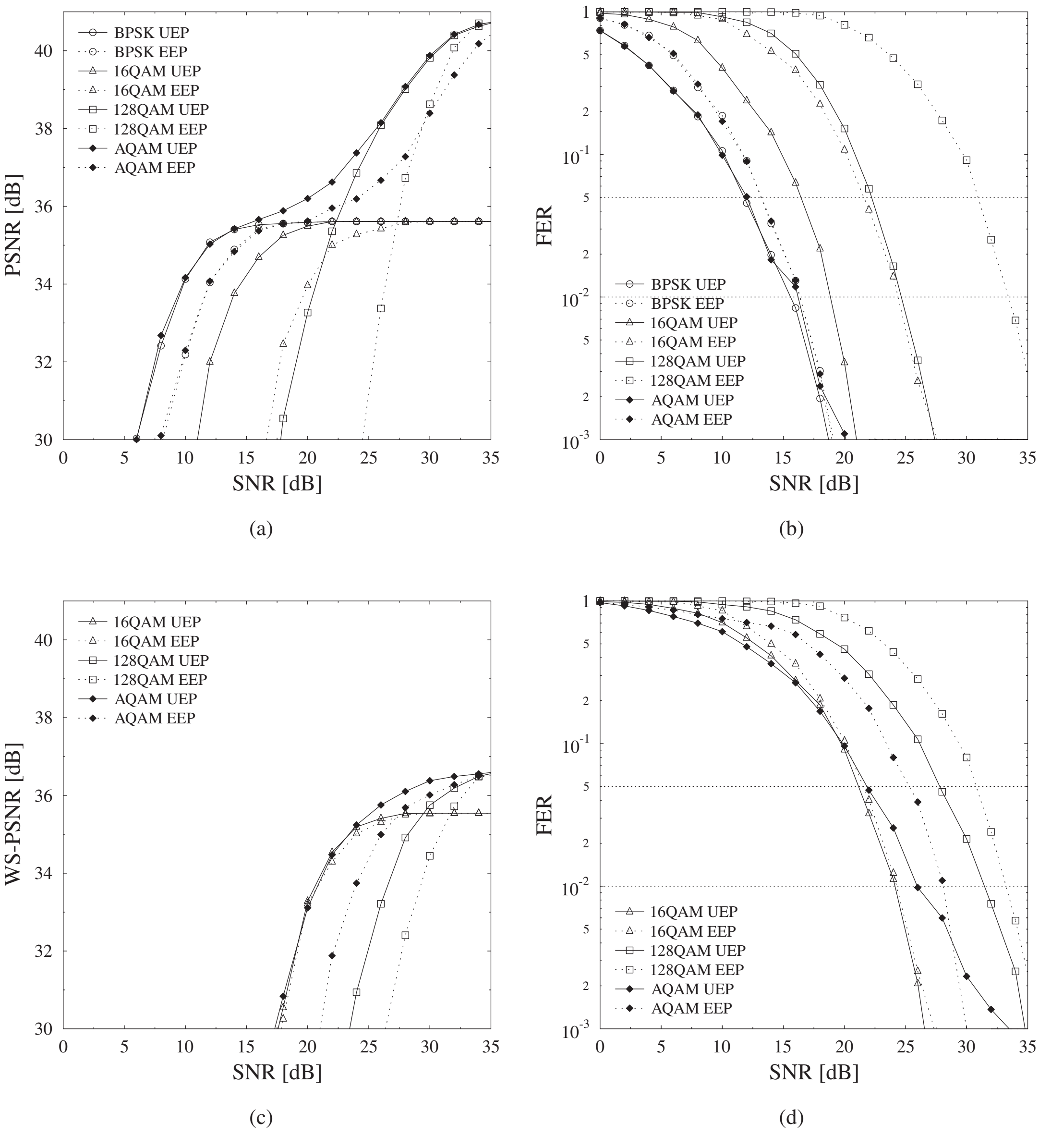
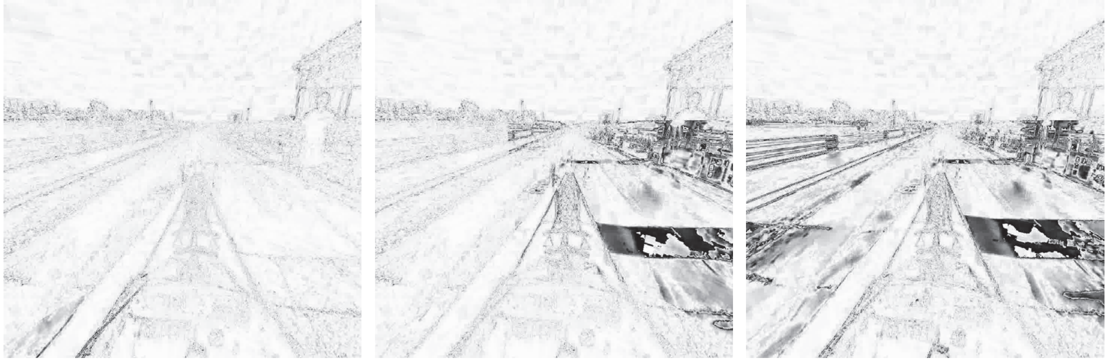


Figure 5.20: The comparisons between the proposed systems and their EEP counterparts for the *RaceVR* test sequence against channel SNR, which are (a) the PSNR of the FoV, (b) the FER of the FoV, (c) the WS-PSNR of the ERP and (d) the FER of the BL of the ERP, respectively.

(a) BPSK UEP (35.1 dB) (b) BPSK EEP (34.0 dB) (c) 16QAM UEP (32.1 dB)



(d) BPSK UEP (e) BPSK EEP (f) 16QAM UEP

Figure 5.21: Reconstructed FoV frame comparison of different modulation modes at the channel SNR of 12 dB, where (b), (d) and (e) portray the error between the original and received video of (a), (c) and (e), respectively.

5.7.4 Uncoded Source-Rate Performance

Figure 5.22 shows the uncoded source-rate of the dynamically reconfigured AQAM for both the UEP and EEP assisted schemes, compared to their three fixed modes, where the effective source-rate of the UEP assisted designs is on average better than that of their EEP assisted counterparts. Explicitly, our proposed UEP assisted AQAM scheme is capable of providing the best uncoded source-rate by judiciously selecting the modulation mode, while its EEP assisted AQAM counterpart is unable to adaptively maximize the source-rate.

5.8 Chapter Conclusion

In this chapter, we proposed an optimal FEC code-rate aided AQAM system for layered panoramic video streaming. In Section 5.2, we first reviewed the concept of variable FEC code-rate based UEP applied both at a frame-level and macroblock-level. Then, we discussed the development of the panoramic video representation in Section 5.3, including its testing model, the 3D to 2D

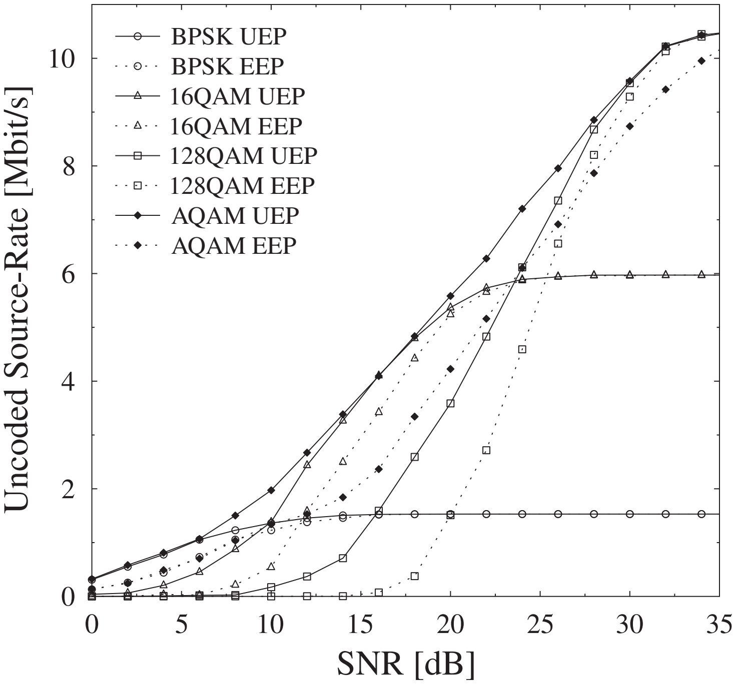


Figure 5.22: Uncoded source-rate versus channel SNR comparison of the three fixed modulation modes and the adaptive AQAM in terms of both the UEP and EEP assisted schemes.

Table 5.4: The channel SNR required at a 5% FER value of the BL of the FoV.

Candidate	channel SNR
BPSK UEP	12 dB
BPSK EEP	13.1 dB
16QAM UEP	16.5 dB
16QAM EEP	21.7 dB
128QAM UEP	22.2 dB
128QAM EEP	31 dB
AQAM UEP	12 dB
AQAM EEP	15.8 dB

projection methods, the corresponding subjective quality metrics and the associated transmission techniques. Then, in Section 5.4, we outlined our optimal FEC code-rate assisted AQAM system designed for scalable panoramic video streaming. The associated optimization process employed for FEC code-rate allocation and adaptive mode switching decisions were discussed in Sections 5.5 and 5.6, respectively.

As for the simulation results, we first analyzed the complexity of the system imposed by the global optimization algorithm, which depends on the length of the GoP sequence as well as the population size and the number of generations. Figures 5.20(a)-(d) show the video quality and FER performance, where the proposed system significantly outperforms its EEP counterpart both for the adaptive and fixed modes. Specifically, it can be seen in Figure 5.20(b) that the proposed UEP assisted 16QAM attains about 5 dB SNR gain over its EEP assisted counterpart at the 5% FER of the BL, where the UEP assisted 128QAM outperforms its EEP counterpart by approximately 9 dB, as listed in Table 5.4. Figure 5.22 further compares the uncoded source-rate of the proposed system to that of its EEP counterpart, which demonstrates that the proposed UEP assisted adaptive system is capable of better exploiting the system bandwidth.

Chapter 6

Conclusions and Future Research

In this concluding chapter, the overall thesis summary and conclusions will be presented in Section 6.1, followed by potential future research ideas in Section 6.2.

6.1 Thesis Conclusion

In this thesis, we presented several physical layer based UEP designs conceived for SVC streaming with the aid of both modulation as well as FEC based techniques. Specifically, in the modulation based designs, we employed the bit mapping based UEP scheme for layered video streaming, where the bits of various video layers are mapped to the corresponding constellation bits according to their specific importance. Additionally, the FEC based UEP scheme employed for SVC streaming was also taken into account, including both IL-FEC and variable FEC code-rate based schemes, where we designed the FEC code-rate allocation using a bio-inspired algorithm to find the optimal assignment scheme capable of providing the best reconstructed video quality without imposing any bandwidth expansion.

Chapter 1: In Chapter 1, we provided a generic overview of physical layer based SVC transmission. More specifically, in Section 1.1, we reviewed the development of video compression techniques, where a simple frame-differencing based video codec was used for presenting the basic concept of video encoding. Then, we briefly discussed the development of the historical video compression standards in Section 1.1.2, followed by the introduction of the SVC principle, including temporal, spatial as well as quality based scalability, and presented a pair of video assessment methods in Section 1.1.4. In Section 1.2, we described the family of physical layer based UEP schemes, including its FEC and modulation based representatives, where the former was then categorized into IL-FEC and variable FEC code-rate arrangements, while the latter was classified as bit mapping and MIMO aided schemes. Afterwards, Section 1.3 presented a comprehensive discussion on the adaptive system, which was exemplified using near-instantaneously adaptive AQAM and variable FEC code-rate based arrangements.

Chapter 2: In Chapter 2, a novel SC assisted SVC broadcasting system was proposed in the context of modulation based UEP schemes. This design philosophy relied on superimposing different modulation modes used for conveying the bits of various video layers, which required an optimal power allocation scheme in order to extract the different layers' bits from the received superimposed signals. Furthermore, we formulated the associated power optimization as a mathematical problem of maximizing the average video quality within the BS coverage, which takes into account the video quality, the associated frame decodability and the PL model.

In Section 2.2, the concept of modulation based UEP schemes designed for video streaming was reviewed, which was categorized into HQAM and SC schemes. Specifically, the constellation points of the HQAM scheme employ Gray mapping, as shown in Figure 2.1, where the subchannel of the HQAM exhibiting a lower BER is capable of providing better protection for the more important information. By contrast, in the SC scheme, the constellation points of various modulation modes conveying different source bits are superimposed with the aid of the power sharing coefficients, which usually results in non-Gray mapping. Moreover, fuelled by the development of near-capacity FEC techniques, such as LDPC and TC, the superimposed symbols can be demodulated with the aid of SIC, which requires the re-encoding of the demodulated signals, which can then be deducted from the mixed signals for the sake of extracting the symbols of the other superimposed scheme.

In Section 2.3, we discussed our SC assisted SVC video broadcast system in detail, where as a benchmarker, the schematic of TDM that represents the conventional orthogonal multiplexing technique was also briefly introduced. Figure 2.6 shows the proposed SC-SHVC model of scalable video broadcasting. Explicitly, in Figure 2.6, the captured video was first encoded by the SHVC encoder, generating a compressed video bit stream, which was then demultiplexed to multiple sub-streams according to the layer index. These bit streams were then fed into the FEC encoders for channel encoding, where the encoded bit streams were forwarded to the corresponding modulators, respectively. The modulated signals of different video layers were scaled by the power sharing coefficients assigned, as shown in Figure 2.6 in order to achieve the best video service across the entire BS coverage, which were specifically designed by taking into account numerous factors, such as the total transmit power, BS coverage, PL and so forth. Moreover, at the receiver of the proposed system, the SIC technique was invoked, where again, the signals are progressively decoded and removed from the received superimposed signals for the sake of extracting the video layers. Explicitly, the received superimposed signals were first demodulated and FEC-decoded to extract the bits of the BL, where a replica of the bit stream of the decoded BL was then re-encoded, re-modulated and removed from the received signals.

The associated power optimization algorithm was presented in Section 2.4, which was formulated for analytically maximizing the average reconstructed video quality right across the entire BS coverage area. Specifically, the problem setting jointly considered the video quality of each video frame and the corresponding decodability, which takes into account the specific inter-layer and inter-frame video coding dependencies. Moreover, the decodability of different video frames depend on the frame length, the power coefficients assigned and the modulation mode employed. These functions were experimentally pre-recorded using Monte-Carlo simulations and then they were further fitted into the function of (2.17). Due to the fact that the power coefficients to be optimized had continuous values,they could not be optimized by discrete search algorithms, such as exhaustive search. Hence, the bio-inspired EA was invoked to find the globally optimal results.

Finally, the performance of the proposed SC-SHVC video broadcasting system was quantified in Section 2.5. Explicitly, the proposed system exhibited a slight video quality degradation at the cell-center, but it showed a significant image quality and PLR improvement at the cell-edge, which provides a moderate video quality degradation when the UE is away from the BS. Furthermore, we also compared the computational complexity of the schemes considered, where the proposed system imposes the lowest demodulation complexity order amongst all the considered schemes.

Chapter 3: Chapter 2 highlighted the basic schematic of modulation based UEP schemes used for layered video streaming, which map the source bits to the different-integrity subchannels of conventional modulation schemes. BY contrast, in Chapter 3, we conceived a new bit mapping aided modulation based UEP scheme relying on the novel concept of IM.

In Section 3.2, we first briefly reviewed the historic development of SM and STSK techniques, followed by the novel MS-STSK transceiver. SM is capable of providing a high normalized throughput,despite its low complexity. Since only a single antenna is activated, which is selected from a set of multiple antennas, SM is incapable of providing any transmit diversity. A concept referred to as STSK was then presented, where instead of activating the indexed antennas, one out of Q space-time dispersion matrices is activated during each STSK symbol for attaining both diversity and multiplexing gains. This dispersion matrix based scheme offers a high design flexibility, since we can optimize both the dispersion matrix employed as well as the number of transmit and receive antennas, hence striking a beneficial design trade-off between the attainable multiplexing and diversity gains. Furthermore, in MS-STSK, the information is conveyed over two components, namely over the afore-mentioned STSK component as well as over the ASU of Figure 3.3, which selects a single antenna combination for conveying extra bits, hence leading to enhanced multiplexing gains, while simultaneously attaining diversity gains.

Section 3.3 exploited the potential UEP capability of the MS-STSK transceiver. Firstly, Figure 3.5 showed the BER performance versus channel SNR for various configurations, when transmitting over narrowband Rayleigh fading channels, which indicated that although all configurations exhibited the same normalized throughput of BPCU=6-bit, the BER performance of the MS-STSK system depended on the specific configuration of the parameters. In Figure 3.6, we portrayed the BER performance difference of the MS-STSK transceiver components. Explicitly, Figure 3.6 depicted the BER performance of the three MS-STSK components using three different sets of configurations. It can be seen in Figure 3.6 that the ASU is capable of attaining a lower BER than the $\mathcal{L}$-QAM/PSK modulator, while both outperform the dispersion matrix component. Furthermore, we achieved the near-capacity performance for the MS-STSK transceiver with the aid of the powerful design tool of EXIT charts, where the different MS-STSK subchannels exhibited different BERs.

Based on Section 3.3, in Section 3.4, we conceived UEP assisted adaptive MS-STSK for layered video streaming over high-mobility scenarios. Specifically, three MS-STSK configurations exhibiting different capacity and robustness were employed, which were routinely activated based on the instantaneous SNR and the Doppler frequency of the wireless channel. Specifically, a more robust, but low-capacity MS-STSK configuration was activated when the wireless channel experienced deep fading, where the ELs had to be abandoned even though they were requested by the destination, because the channel conditions were hostile for supporting high bit rates. Furthermore, to estimate the rapidly fluctuating channel conditions in the face of a high Doppler frequency, we specifically designed the pilot assisted channel estimation for our MS-STSK transceiver.

The performance of the proposed system was quantified in Section 3.5. First of all, Figure 3.13 illustrated effects of Doppler-frequency on the MS-STSK bits when $f_d = 0.001$ and $f_d = 0.3$. Explicitly, a significant MI degradation was found when the Doppler frequency increased. Then, we characterized the MI exchange between the MS-STSK transceiver components, as shown in Figure 3.14, and the performance benefits of mode switching as portrayed in Figure 3.15. Moreover, Figures 3.16 and 3.17 depicted the reconstructed image quality and the PLR versus the average channel SNR. Specifically, Figure 3.16 illustrated the PSNR of the transceiver modes at the average channel SNR of 12 dB, where the proposed adaptive scheme was capable of providing the best reconstructed video quality by selecting the most appropriate MS-STSK configuration. By contrast, Figure 3.17 portrayed the average performance versus the channel SNR of different MS-STSK modes, which showed that the UEP assisted modes on average outperformed their corresponding EEP counterparts, while the UEP assisted adaptive MS-STSK system provided the best performance both in terms of the reconstructed image quality as well as the video PLR. Finally, we concluded the chapter in Section 3.6.

Chapter 4: In Chapter 4, we turned our attention from the transceiver based UEP technique to its FEC based counterpart, specifically to the IL-FEC technique, which was capable of protecting the BL by implanting its bits into the ELs. To improve its performance, this IL-FEC scheme was further developed by improving the protection of the ELs of the SVC. We then proposed an adaptive IL-FEC scheme for SVC streaming relying on the MS-STSK transceiver, where MS-STSK and our IL-FEC were conceived for jointly providing UEP for enhancing the overall reconstructed video quality.

In Section 4.2, we reviewed the conventional IL-FEC technique used for protecting the BL of the SVC stream. Explicitly, the systematic bits of the BL were implanted into the ELs using XOR operation. However, this conventional IL-FEC was only capable of protecting the BL, but not the EL. We solved this problem by also protecting the reference layer of the EL by implanting the associated bits into the other layers, as shown in Figure 4.3. The resultant enhanced IL-FEC decoder was illustrated in Figure 4.3(b).

Section 4.3 presented our proposed adaptive video streaming system relying on our enhanced IL-FEC technique mentioned in Section 4.2, which was amalgamated with the MS-STSK transceiver of Chapter 3. Specifically, the captured video was encoded by the SHVC scheme into a compressed SVC stream, which was then demultiplexed into the three bit streams shown in Figure 4.5. These video bits were then encoded by their associated RSC encoders, generating three systematic bit streams and their corresponding parity bits, respectively. Then, the adaptive IL-FEC selection unit of Figure 4.5 selected the most appropriate scheme from the three schemes portrayed in Figures 4.6-4.8 according to the instantaneous channel SNR. The IL-FEC encoded bit streams were fed into the corresponding MS-STSK components according to their importance. Explicitly, based on the BER performance of Figure 3.6, the bits of the BL x_1 were fed into the ASU, while x_2 and x_3 were fed into the $\mathcal{L}$-PSK/QAM modulator and dispersion matrices generator, receptively. Finally, these bits were encoded using MS-STSK mapping and transmitted.

Figure 4.9 first compared the image quality of the proposed and the conventional IL-FEC schemes, when the protected layers were set to l_2 and l_3, respectively, where an approximately 2-3 dB gain was provided by the enhanced IL-FEC scheme conceived. The QoE performance of the proposed system was illustrated in Figure 4.11, where to benchmark our system, we also considered the conventional IL-FEC assisted and the EEP assisted systems. Figure 4.11 showed that the proposed adaptive IL-FEC assisted MS-STSK system provided the best image quality.

Chapter 5: In Chapter 4, we have detailed the concept and implementation of our IL-FEC based UEP scheme designed for SVC streaming, while in Chapter 5, we introduced another FEC based UEP scheme relying on variable FEC code-rate. We specifically designed the FEC code-rate allocation scheme for scalable panoramic video streaming, which jointly considered the particular characteristics of the panoramic video and of the video coding dependencies amongst different layers.

Section 5.2 reviewed the concept of the variable FEC code-rate based UEP scheme designed for video streaming on a frame-level and macroblock-level basis. Explicitly, in the frame-level based scheme, the bits pertaining to a single video frame share an identical FEC code-rate, which may however differ from the one assigned to the other video frames. By contrast, the macroblock-level based scheme partitioned a video frame into multiple macroblocks, and assigned different FEC code-rates to them according to their importance, which improved the robustness of the video stream at the cost of a moderate extra complexity.

In Section 5.3, we introduced the concept of panoramic video representations, including their testing system model, the conversion projection methods, the corresponding WS-PSNR and the existing panoramic video streaming techniques. Figure 5.6 depicted the panoramic video testing system, where the ERP high fidelity employed as the testing material was first converted to the 3D spherical format that was observed by the users. Due to the fact that the existing video coding techniques cannot directly process 3D video, it had to be projected onto the 2D planes using for example the ERP and CMP techniques. Moreover, since converting from the 3D spherical format to 2D projection introduces stretching distortions, the conventional PSNR metric is unable to accurately quantify its subjective quality. Hence, we opted for the WS-PSNR both for ERP and for CMP. At the end of Section 5.3, we showed that to save bandwidth, the panoramic video was typically spatially partitioned into multiple rectangular tiles and the FoV tiles were transmitted at high quality, while the remaining tiles were either transmitted at low quality or abandoned.

We then presented our variable FEC code-rate assisted UEP scheme in Section 5.4, where we categorized the video bits of different panoramic content into three importance classes: C1 includes the BL bits of the FoV, which directly affects the user's experience; C2 contains the BL of the tiles outside the FoV for preventing dramatic QoE degradation due to FoV misprediction; and C3 consists of the EL of the FoV tiles that represents the least important class. Explicitly, the spherical video was first converted to the 2D plane and then split into numerous tiles, which were then fed into the SHVC encoder for independent compression, as shown in Figure 5.14. The compressed video bit streams were then demultiplexed according to the specific tile importance of the predefined class, where the classified bits were forwarded to the corresponding FEC encoders, respectively. Moreover, the FEC code-rate assigned to each video frame was determined by the optimization block of Figure 5.14 considering the instantaneous channel SNR. Afterwards, the FEC encoded bit streams were multiplexed into a single bit stream and then transmitted. It is worth noting that three different modulation modes were provided for the system, where the least important information was dropped, when a low-capacity modulation mode was activated. At the receiver, the received signals were demultiplexed according to their class index, which were then FEC decoded. The multiplexer of Figure 5.14 rearranged the decoded source bits and forward them to the SHVC decoder for reconstructing the 2D projection, which was then converted back to the 3D spherical format for observation.

Section 5.5 presented our optimal FEC code-rate allocation. Specifically, we formulated the problem of maximizing the weighted uncoded source-rate of the system determined by a video frame's bit length and the corresponding FER, where the weighting coefficients take into account both the class importance as well as the tile importance. In fact, since the majority of the parameters contributing to the OF are constant, such as the frame length and the weighting coefficients, the only variables to be optimized becomes the FEC code-rate set, which are then compared to (5.37) for selecting the most appropriate modulation mode.

Section 5.7 quantified the performance of the proposed system shown in Figure 5.14, where the simulations results were verified using the FoV track prediction of Figure 5.17. We first analyzed the optimization complexity imposed by the EA algorithm, which was determined by GoP size, the number of FEC code-rates to be optimized, the population size and the number of generations required. Then, we discussed the mode switching performance for the near-instantaneously adaptive AQAM, where Figure 5.18 showed the PDF of the adaptive modulation modes versus the average channel SNR, while Figure 5.19 characterized the modulation mode switching operation at an average channel SNR of 16 dB. The image quality versus the FER was shown in Figure 5.20, where the proposed system achieved the highest image quality for both the FoV and non-FoV areas. Additionally, Figure 5.22 showed the uncoded source-rate versus channel SNR comparison for the both UEP and EEP schemes, where the AQAM assisted UEP scheme achieved a more graceful PSNR erosion upon reducing the channel SNR.

6.2 Future Research

In this section, we briefly discuss a number of future research ideas.

6.2.1 Joint Optimization of FEC and Modulation for SVC Streaming

In Chapter 4, we have shown the potential of amalgamating the IL-FEC based and modulation based UEP scheme for scalable video streaming, where the IL-FEC coded bit streams were fed into the different-integrity MS-STSK subchannels according to their importance. Additionally, we may consider a joint optimization scheme that amalgamates variable FEC code-rate and bit mapping assisted modulation based UEP to provide more flexible video streaming protection. Inspired by Chapters 3 and 5, we may conceive the joint optimization of PSNR, error-resilience, complexity delay etc. This would allow us to determine the pareto front of all optimal solutions, where by definition, none of these parameters can be improved without degrading at least one of the others.

6.2.2 Cross-Layer Assisted UEP for SVC Streaming

In this thesis, we have conceived potent physical layer based UEP schemesrelying on FEC and modulation schemes. This philosophy can be extended by exploiting, the UEP potential of the upper OSI layers, such as the application, transport, network and data link layers, where a cross-layer design may be conceived. For example, the network layer may be expected to find the optimal routing solution in terms of the video QoE. Moreover, Automatic Repeat Request (ARQ) technique may be used for enhancing the error-resilience attained.

6.2.3 Optical Wireless Techniques for SVC Streaming

In Chapter 5, we designed scalable panoramic video streaming over the fading channels. Explicitly, the panoramic video provides immersive viewing experience for the users at the expense of requiring a high bit rate, which remains an open challenge for the existing networks. Fortunately, the optical wireless networks relying on both visible light and infrared light have been proven to be capable of providing a much higher bandwidth than the conventional radio waves, hence they may be more capable of supporting high bit rate applications, such as panoramic video. However, there are numerous open challenges also in optical wireless communications.

6.2.4 SVC Streaming in Multiuser Scenario

This thesis mainly considered video transmission in single user scenarios. This research can be extended to multi-user scenario, where the different users may view different video contents delivered over the same backhaul. Specifically, the typical resource allocation methods, such as max-min and maximum sum optimization, have to be specifically designed in order to cater for the specific characteristics of the video stream.

Bibliography

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

[2] "Video Codec for Audiovisual Services at px64 kbit/s ITU-T Recommendation H.261, Version 1," *ITU-T, ITU-T Recommendation H.261 Version 1*, 1990.

[3] ISO/IEC JTC1/SC29/WG11, "Generic coding of moving pictures and associated audio information-Part 2: Video," *ITU-T Rec. H.262 and ISO/IEC 13818-2 (MPEG-2 Video) and ITU-T Recommendation H.262*, ISO/IEC and ITU-T, Geneva, 1994.

[4] F. H. Fitzek and M. Reisslein, "MPEG-4 and H. 263 video traces for network performance evaluation," *IEEE network*, vol. 15, no. 6, pp. 40–54, 2001.

[5] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, 2003.

[6] G. J. Sullivan, J. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649–1668, 2012.

[7] B. Bross, K. Andersson, M. Blser, V. Drugeon, S. Kim, J. Lainema, J. Li, S. Liu, J. Ohm, G. J. Sullivan, and R. Yu, "General Video Coding Technology in Responses to the Joint Call for Proposals on Video Compression With Capability Beyond HEVC," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 5, pp. 1226–1240, 2020.

[8] X. Li, N. Oertel, A. Hutter, and A. Kaup, "Laplace Distribution Based Lagrangian Rate Distortion Optimization for Hybrid Video Coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, no. 2, pp. 193–205, 2009.

[9] L. Hanzo, P. Cherriman, and J. Streit, *Video compression and communications: from basics to H. 261, H. 263, H. 264, MPEG4 for DVB and HSDPA-style adaptive turbo-transceivers*. John Wiley & Sons, 2007.

[10] T. Stockhammer, M. M. Hannuksela, and T. Wiegand, "H. 264/AVC in wireless environments," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 657–673, 2003.

[11] A. Segall, V. Baroncini, J. Boyce, J. Chen, and T. Suzuki, "Joint Call for Proposals on Video Compression with Capability beyond HEVC," *Joint Video Exploration Team (JVET) of ITU-T SG*, vol. 16, 2017.

[12] G. K. Wallace, "The JPEG still picture compression standard," *Communications of the ACM*, p. 30, 1991.

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

[14] International Electrotechnical Commission and others, "RecordingHelical-scan digital video cassette recording system using 6, 35 mm magnetic tape for consumer use (525–60, 625–50, 1125–60 and 1250–50 systems)Part 4: Pack header table and contents [IEC 61834-4: 1998 (E)]," *Geneva: International Electrotechnical Commission*, 1998.

[15] G. J. Conklin, G. S. Greenbaum, K. O. Lillevold, A. F. Lippman, and Y. A. Reznik, "Video coding for streaming media delivery on the Internet," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 3, pp. 269–281, 2001.

[16] L. Bouchard, "Multimedia software for mobile phones," *IEEE software*, vol. 27, no. 3, pp. 8–10, 2010.

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

[18] L. Fan, S. Ma, and F. Wu, "Overview of AVS video standard," in *2004 IEEE International Conference on Multimedia and Expo (ICME)(IEEE Cat. No. 04TH8763)*, vol. 1, pp. 423–426, IEEE, 2004.

[19] J. Pfaff, H. Schwarz, D. Marpe, B. Bross, S. De-Luxn-Hernndez, P. Helle, C. R. Helmrich, T. Hinz, W. . Lim, J. Ma, T. Nguyen, J. Rasch, M. Schfer, M. Siekmann, G. Venugopal, A. Wieckowski, M. Winken, and T. Wiegand, "Video Compression Using Generalized Binary Partitioning, Trellis Coded Quantization, Perceptually Optimized Encoding, and Advanced Prediction and Transform Coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 5, pp. 1281–1295, 2020.

[20] I. E. Richardson, *H. 264 and MPEG-4 video compression: video coding for next-generation multimedia*. John Wiley & Sons, 2004.

[21] 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, no. 9, pp. 1103–1120, 2007.

[22] T. S. Sector, "ITU-T Recommendation H.264. Advanced video coding for generic audiovisual services," 2009.

[23] J. M. Boyce, Y. Ye, J. Chen, and A. K. Ramasubramonian, "Overview of SHVC: scalable extensions of the high efficiency video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 1, pp. 20–34, 2016.

[24] H. Schwarz, D. Marpe, and T. Wiegand, "Analysis of Hierarchical B Pictures and MCTF," in *2006 IEEE International Conference on Multimedia and Expo*, no. 9112788, pp. 1929–1932, July 2006.

[25] H. Schwarz, D. Marpe, and T. Wiegand, "Hierarchical B Pictures," *Joint Video Team, Doc. JVT-P014*, Jul. 2005.

[26] C. A. Segall and G. J. Sullivan, "Spatial scalability within the H. 264/AVC scalable video coding extension," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1121–1135, 2007.

[27] 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, no. 9, pp. 1103–1120, 2007.

[28] A. K. Moorthy, K. Seshadrinathan, R. Soundararajan, and A. C. Bovik, "Wireless video quality assessment: A study of subjective scores and objective algorithms," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, no. 4, pp. 587–599, 2010.

[29] Y. Huo, M. El-Hajjar, and L. Hanzo, "Wireless video: An interlayer error-protection-aided multilayer approach," *IEEE Vehicular Technology Magazine*, vol. 9, no. 3, pp. 104–112, 2014.

[30] Q. Huynh-Thu, M.-N. Garcia, F. Speranza, P. Corriveau, and A. Raake, "Study of rating scales for subjective quality assessment of high-definition video," *IEEE Transactions on Broadcasting*, vol. 57, no. 1, pp. 1–14, 2010.

[31] J. Bergstra and C. Middelburg, "ITU-T recommendation p. 800.1: Mean opinion score terminology," *International Telecommunication Union, Geneva*, 2006.

[32] Y. Huo, C. Hellge, T. Wiegand, and L. Hanzo, "A tutorial and review on inter-layer FEC coded layered video streaming," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 2, pp. 1166–1207, 2015.

[33] Y. Wang, S. Wenger, J. Wen, and A. K. Katsaggelos, "Error resilient video coding techniques," *IEEE Signal Processing Magazine*, vol. 17, no. 4, pp. 61–82, 2000.

[34] A. Sehgal, A. Jagmohan, and N. Ahuja, “Wyner-Ziv coding of video: an error-resilient compression framework,” *IEEE Transactions on Multimedia*, vol. 6, no. 2, pp. 249–258, 2004.

[35] Y. Zhang, C. Zhu, and K.-H. Yap, “A joint source-channel video coding scheme based on distributed source coding,” *IEEE Transactions on Multimedia*, vol. 10, no. 8, pp. 1648–1656, 2008.

[36] Y. Guo, Y. Chen, Y.-K. Wang, H. Li, M. M. Hannuksela, and M. Gabbouj, “Error resilient coding and error concealment in scalable video coding,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, no. 6, pp. 781–795, 2009.

[37] L. Hanzo, P. J. Cherriman, and J. Streit, *Wireless video communications: second to third generation and beyond*. IEEE Press-John Wiley & Sons, 2001.

[38] H. Zimmermann, “OSI reference model–The ISO model of architecture for open systems interconnection,” *IEEE Transactions on Communications*, vol. 28, no. 4, pp. 425–432, 1980.

[39] Y.-C. Su, C.-S. Yang, and C.-W. Lee, “Optimal FEC assignment for scalable video transmission over burst error channel with loss rate feedback,” *Signal Processing: Image Communication*, vol. 18, no. 7, pp. 537–547, 2003.

[40] X. Yang, C. Zhu, Z. Li, X. Lin, G. Feng, S. Wu, and N. Ling, “Unequal loss protection for robust transmission of motion compensated video over the internet,” *Signal Processing: Image Communication*, vol. 18, no. 3, pp. 157–167, 2003.

[41] 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.

[42] T. Fang and L.-P. Chau, “GOP-based channel rate allocation using genetic algorithm for scalable video streaming over error-prone networks,” *IEEE Transactions on Image Processing*, vol. 15, no. 6, pp. 1323–1330, 2006.

[43] B. Barmada, M. M. Ghandi, E. V. Jones, and M. Ghanbari, “Combined turbo coding and hierarchical QAM for unequal error protection of H. 264 coded video,” *Signal Processing: Image Communication*, vol. 21, no. 5, pp. 390–395, 2006.

[44] Y. C. Chang, S. W. Lee, and R. Komyia, “A fast forward error correction allocation algorithm for unequal error protection of video transmission over wireless channels,” *IEEE Transactions on Consumer Electronics*, vol. 54, no. 3, pp. 1066–1073, 2008.

[45] Nasruminallah and L. Hanzo, “Near-capacity h.264 multimedia communications using iterative joint source-channel decoding,” *IEEE Communications Surveys Tutorials*, vol. 14, no. 2, pp. 538–564, 2012.

[46] Y. Zhang, S. Qin, B. Li, and Z. He, "Rate-distortion optimized unequal loss protection for video transmission over packet erasure channels," *Signal Processing: Image Communication*, vol. 28, no. 10, pp. 1390–1404, 2013.

[47] 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, no. 11, pp. 4364–4379, 2013.

[48] Y. Huo, C. Zhou, J. Jiang, and L. Hanzo, "Historical information aware unequal error protection of scalable HEVC/H. 265 streaming over free space optical channels," *IEEE Access*, vol. 4, pp. 5659–5672, 2016.

[49] X. Chen, J. Hwang, J. A. Ritcey, C. Lee, and F. Yeh, "Quality-Driven Joint Rate and Power Adaptation for Scalable Video Transmissions Over MIMO Systems," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 2, pp. 366–379, 2017.

[50] Y. Huo and X. Wang and P. Zhang and J. Jiang and L. Hanzo, "Unequal Error Protection Aided Region of Interest Aware Wireless Panoramic Video," *IEEE Access*, vol. 7, pp. 80262–80276, 2019.

[51] B. Paudel and S. Vafi, "Efficient Unequal Error Protection Techniques for Tile-Based Transmission of HEVC Videos," *IEEE Access*, vol. 8, pp. 128591–128601, 2020.

[52] Y. Huo, T. Wang, R. G. Maunder, and L. Hanzo, "Motion-Aware Mesh-Structured Trellis for Correlation Modelling Aided Distributed Multi-View Video Coding," *IEEE Transactions on Image Processing*, vol. 23, no. 1, pp. 319–331, 2014.

[53] Y. Huo, P. T. Kovcs, T. J. Naughton, and L. Hanzo, "Wireless Holographic Image Communications Relying on Unequal Error Protected Bitplanes," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 8, pp. 7136–7148, 2017.

[54] 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, no. 8, pp. 4306–4315, 2009.

[55] X. Wu, X. Li, J. Xu, and B. Chen, "A joint source-channel coding scheme for unequal protection of SVC video using LDPC codes," in *Computational Problem-Solving (ICCP), 2010 International Conference on*, pp. 370–373, IEEE, 2010.

[56] D. Vukobratovic, V. Stankovic, D. Sejdinovic, L. Stankovic, and Z. Xiong, "Scalable video multicast using expanding window fountain codes," *IEEE Transactions on Multimedia*, vol. 11, no. 6, pp. 1094–1104, 2009.

[57] K. Nguyen, T. Nguyen, and S.-C. Cheung, "Video streaming with network coding," *Journal of Signal Processing Systems*, vol. 59, no. 3, pp. 319–333, 2010.

[58] 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, no. 3, pp. 551–562, 2011.

[59] 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, 2010.

[60] 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, no. 9, pp. 1622–1634, 2013.

[61] C. Zhu, Y. Huo, B. Zhang, R. Zhang, M. El-Hajjar, and L. Hanzo, "Adaptive-Truncated-HARQ-Aided Layered Video Streaming Relying on Interlayer FEC Coding," *IEEE Transactions on Vehicular Technology*, vol. 65, pp. 1506–1521, March 2016.

[62] Toan Duc Bui, Phuc Chau, and Jitae Shin, "Content-aware raptorq," in *2017 International Conference on Information Networking (ICOIN)*, pp. 206–209, 2017.

[63] H. Chen, X. Zhang, Y. Xu, Z. Ma, and W. Zhang, "Efficient Mobile Video Streaming via Context-Aware RaptorQ-Based Unequal Error Protection," *IEEE Transactions on Multimedia*, vol. 22, no. 2, pp. 459–473, 2020.

[64] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "Self-concatenated trellis coded modulation with self-iterative decoding," in *IEEE GLOBECOM 1998 (Cat. NO. 98CH36250)*, vol. 1, pp. 585–591 vol.1, 1998.

[65] Y. Huo, M. El-Hajjar, R. G. Maunder, and L. Hanzo, "Layered wireless video relying on minimum-distortion inter-layer FEC coding," *IEEE Transactions on Multimedia*, vol. 16, no. 3, pp. 697–710, 2014.

[66] T. Cover, "Broadcast channels," *IEEE Transactions on Information Theory*, vol. 18, no. 1, pp. 2–14, 1972.

[67] P. Bergmans, "Random coding theorem for broadcast channels with degraded components," *IEEE Transactions on Information Theory*, vol. 19, no. 2, pp. 197–207, 1973.

[68] B. Barmada, M. M. Ghandi, E. V. Jones, and M. Ghanbari, "Prioritized transmission of data partitioned H. 264 video with hierarchical QAM," *IEEE Signal Processing Letters*, vol. 12, no. 8, pp. 577–580, 2005.

[69] 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, no. 4, pp. 1153–1158, 2006.

[70] M. M. Ghandi and M. Ghanbari, "Layered H. 264 video transmission with hierarchical QAM," *Journal of Visual Communication and Image Representation*, vol. 17, no. 2, pp. 451–466, 2006.

[71] 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.

[72] 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, no. 4, pp. 2741–2746, 2010.

[73] L. Cai, S. Xiang, Y. Luo, and J. Pan, "Scalable Modulation for Video Transmission in Wireless Networks," *IEEE Transactions on Vehicular Technology*, vol. 60, pp. 4314–4323, Nov 2011.

[74] 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, pp. 731–738, August 2012.

[75] S. Chang, M. Rim, P. C. Cosman, and L. B. Milstein, "Optimized Unequal Error Protection Using Multiplexed Hierarchical Modulation," *IEEE Transactions on Information Theory*, vol. 58, pp. 5816–5840, Sep. 2012.

[76] T. V. Nguyen, P. C. Cosman, and L. B. 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.

[77] J. She, J. Ho, Z. Chen, and P. Ho, "Logical Superposition Coded Video Multicast/Broadcast," *IEEE Transactions on Vehicular Technology*, vol. 66, pp. 1379–1392, Feb 2017.

[78] J. Lee, S. Park, S. Kwon, B. Lim, H. M. Kim, N. Hur, A. Pesin, J. Chevet, J. Llach, A. J. Stein, S. Jeon, and Y. Wu, "Efficient Transmission of Multiple Broadcasting Services Using LDM and SHVC," *IEEE Transactions on Broadcasting*, vol. 64, pp. 177–187, June 2018.

[79] D. Song and C. W. Chen, "Scalable H. 264/AVC video transmission over MIMO wireless systems with adaptive channel selection based on partial channel information," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1218–1226, 2007.

[80] 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.

[81] S. Chang, M. Rim, P. C. Cosman, and L. B. Milstein, "Superposition MIMO Coding for the Broadcast of Layered Sources," *IEEE Transactions on Communications*, vol. 59, pp. 3240–3248, December 2011.

[82] H. Kim, P. C. Cosman, and L. B. Milstein, "Motion-Compensated Scalable Video Transmission Over MIMO Wireless Channels," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, pp. 116–127, Jan 2013.

[83] M. Li, Z. Chen, and Y. Tan, "Scalable Resource Allocation for SVC Video Streaming Over Multiuser MIMO-OFDM Networks," *IEEE Transactions on Multimedia*, vol. 15, pp. 1519–1531, Nov 2013.

[84] S. Wang and W. Liao, "Cooperative Multicasting for Wireless Scalable Video Transmissions," *IEEE Transactions on Communications*, vol. 61, pp. 3980–3989, Sep. 2013.

[85] C. Zhou, X. Zhang, and Z. Guo, "Optimal adaptive channel scheduling for scalable video broadcasting over MIMO wireless networks," *Computer Networks*, vol. 57, no. 15, pp. 3039–3050, 2013.

[86] C. Zhou, C. Lin, X. Zhang, and Z. Guo, "A Novel JSCC Scheme for UEP-Based Scalable Video Transmission Over MIMO Systems," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, pp. 1002–1015, June 2015.

[87] H. Cui, C. Luo, C. W. Chen, and F. Wu, "Scalable Video Multicast for MU-MIMO Systems With Antenna Heterogeneity," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, pp. 992–1003, May 2016.

[88] Y. I. Choi and C. G. Kang, "Scalable Video Coding-based MIMO Broadcasting System With Optimal Power Control," *IEEE Transactions on Broadcasting*, vol. 63, pp. 350–360, June 2017.

[89] C. Guo, Y. Cui, D. W. K. Ng, and Z. Liu, "Multi-Quality Multicast Beamforming With Scalable Video Coding," *IEEE Transactions on Communications*, vol. 66, pp. 5662–5677, Nov 2018.

[90] A. Goldsmith, *Wireless communications*. Cambridge university press, 2005.

[91] S. Nanda, K. Balachandran, and S. Kumar, "Adaptation techniques in wireless packet data services," *IEEE Communications Magazine*, vol. 38, no. 1, pp. 54–64, 2000.

[92] L. Hanzo, "Bandwidth-efficient wireless multimedia communications," *Proceedings of the IEEE*, vol. 86, no. 7, pp. 1342–1382, 1998.

[93] P. Cherriman, C. H. Wong, and L. Hanzo, "Turbo-and BCH-coded wide-band burst-by-burst adaptive H. 263-assisted wireless video telephony," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, no. 8, pp. 1355–1363, 2000.

[94] S. X. Ng, J. Y. Chung, P. Cherriman, and L. Hanzo, “Burst-by-burst adaptive decision feedback equalized TCM, TTCM, and BICM for H. 263-assisted wireless video telephony,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 3, pp. 363–374, 2006.

[95] W. Webb and R. Steele, “Variable Rate QAM for mobile radio,” *IEEE Transactions on Communications*, vol. 43, no. 7, pp. 2223–2230, 1995.

[96] R. Y. Mesleh, H. Haas, S. Sinanovic, C. W. Ahn, and S. Yun, “Spatial modulation,” *IEEE Transactions on Vehicular Technology*, vol. 57, no. 4, pp. 2228–2241, 2008.

[97] S. Sugiura, S. Chen, and L. Hanzo, “Coherent and differential space-time shift keying: A dispersion matrix approach,” *IEEE Transactions on Communications*, vol. 58, no. 11, pp. 3219–3230, 2010.

[98] I. A. Hemadeh, M. El-Hajjar, S. Won, and L. Hanzo, “Multi-set space-time shift-keying with reduced detection complexity,” *IEEE Access*, vol. 4, pp. 4234–4246, 2016.

[99] Y. Zhang, I. A. Hemadeh, M. El-Hajjar, and L. Hanzo, “Multi-Set Space-Time Shift Keying Assisted Adaptive Inter-Layer FEC for Wireless Video Streaming,” *IEEE Access*, vol. 7, pp. 3592–3609, 2019.

[100] Y. Zhang, J. Zhang, Y. Huo, C. Xu, M. El-Hajjar, and L. Hanzo, “Scalable Panoramic Wireless Video streaming Relying on Optimal-Rate FEC-Coded Adaptive QAM,” *IEEE Transactions on Vehicular Technology*, pp. 1–1, 2020.

[101] Y. Zhang, c. xu, I. A. Hemadeh, M. El-Hajjar, and L. Hanzo, “Near-instantaneously adaptive multi-set space-time shift keying for uav-aided video surveillance,” *IEEE Transactions on Vehicular Technology*, pp. 1–1, 2020.

[102] 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, no. 3, pp. 1089–1097, 2009.

[103] Lie-Liang Yang and L. Hanzo, “A recursive algorithm for the error probability evaluation of M-QAM,” *IEEE Communications Letters*, vol. 4, no. 10, pp. 304–306, 2000.

[104] L. Hanzo, S. X. Ng, W. Webb, and T. Keller, *Quadrature amplitude modulation: From basics to adaptive trellis-coded, turbo-equalised and space-time coded OFDM, CDMA and MC-CDMA systems*. IEEE Press-John Wiley, 2004.

[105] P. Patel and J. Holtzman, “Analysis of a simple successive interference cancellation scheme in a DS/CDMA system,” *IEEE Journal on Selected Areas in Communications*, vol. 12, no. 5, pp. 796–807, 1994.

[106] L. Hanzo, T. Liew, B. Yeap, R. Tee, and S. X. Ng, *Turbo Coding, Turbo Equalisation and Space-Time Coding.: EXIT-Chart-Aided Near-Capacity Designs for Wireless Channels*, vol. 22. John Wiley & Sons, 2011.

[107] R. Gallager, "Low-density parity-check codes," *IRE Transactions on Information Theory*, vol. 8, no. 1, pp. 21–28, 1962.

[108] J. M. Meredith, "Study on downlink multiuser superposition transmission for LTE," in *TSG RAN Meeting*, vol. 67, 2015.

[109] L. Fay, L. Michael, D. Gómez-Barquero, N. Ammar, and M. W. Caldwell, "An overview of the ATSC 3.0 physical layer specification," *IEEE Transactions on Broadcasting*, vol. 62, no. 1, pp. 159–171, 2016.

[110] Y. I. Choi and C. G. Kang, "Scalable video coding-based MIMO broadcasting system with optimal power control," *IEEE Transactions on Broadcasting*, vol. 63, no. 2, pp. 350–360, 2016.

[111] H.-T. Kim, S. H. Lim, I. Lee, S. Kim, and S.-Y. Chung, "Code design for MIMO downlink with imperfect CSIT," *IEEE Transactions on Communications*, vol. 58, no. 1, pp. 89–94, 2010.

[112] R. E. Walpole, R. H. Myers, S. L. Myers, and K. Ye, *Probability and statistics for engineers and scientists*, vol. 5. Macmillan New York, 1993.

[113] J. Zhang, S. Chen, X. Mu, and L. Hanzo, "Evolutionary-algorithm-assisted joint channel estimation and turbo multiuser detection/decoding for OFDM/SDMA," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 3, pp. 1204–1222, 2014.

[114] R. Mesleh, H. Haas, C. W. Ahn, and S. Yun, "Spatial Modulation - A New Low Complexity Spectral Efficiency Enhancing Technique," in *2006 First International Conference on Communications and Networking in China*, pp. 1–5, 2006.

[115] J. Jeganathan, A. Ghrayeb, and L. Szczecinski, "Spatial modulation: optimal detection and performance analysis," *IEEE Communications Letters*, vol. 12, no. 8, pp. 545–547, 2008.

[116] B. Hassibi and B. M. Hochwald, "High-rate codes that are linear in space and time," *IEEE Transactions on Information Theory*, vol. 48, no. 7, pp. 1804–1824, 2002.

[117] Z. Ghahramani, "Information theory," *Encyclopedia of Cognitive Science*, 2006.

[118] S. Ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," vol. 49, no. 10, pp. 1727–1737, 2001.

[119] C. Xu, S. Sugiura, S. X. Ng, P. Zhang, L. Wang, and L. Hanzo, “Two decades of MIMO design tradeoffs and reduced-complexity MIMO detection in near-capacity systems,” *IEEE Access*, vol. 5, pp. 18564–18632, 2017.

[120] S. Chen, J. Hu, Y. Shi, Y. Peng, J. Fang, R. Zhao, and L. Zhao, “Vehicle-to-Everything (v2x) Services Supported by LTE-Based Systems and 5G,” *IEEE Communications Standards Magazine*, vol. 1, no. 2, pp. 70–76, 2017.

[121] *Technical Specification Group Radio Access Network; Study on Enhanced LTE Support for Aerial Vehicles (Release 15)*, document 3GPP 36.777, Jan. 2018.

[122] C. Xu, J. Zhang, T. Bai, S. P. Botsinis, R. G. Maunder, R. Zhang, and L. Hanzo, “Adaptive Coherent/Non-Coherent Single/Multiple-Antenna Aided Channel Coded Ground-to-Air Aeronautical Communication,” *IEEE Transactions on Communications*, 2018.

[123] E. Haas, “Aeronautical channel modeling,” *IEEE Transactions on Vehicular Technology*, vol. 51, pp. 254–264, March 2002.

[124] J. K. Cavers, “An analysis of pilot symbol assisted modulation for Rayleigh fading channels (mobile radio),” *IEEE Transactions on Vehicular Technology*, vol. 40, no. 4, pp. 686–693, 1991.

[125] S. Zhou and G. B. Giannakis, “How accurate channel prediction needs to be for transmit-beamforming with adaptive modulation over Rayleigh MIMO channels?,” *IEEE Transactions on Wireless Communications*, vol. 3, no. 4, pp. 1285–1294, 2004.

[126] C. Xu, T. Bai, J. Zhang, R. Rajashekar, R. G. Maunder, Z. Wang, and L. Hanzo, “Adaptive coherent/non-coherent spatial modulation aided unmanned aircraft systems,” *IEEE Wireless Communications*, vol. 26, no. 4, pp. 170–177, 2019.

[127] M. El-Hajjar and L. Hanzo, “EXIT charts for system design and analysis,” *IEEE Communications Surveys & Tutorials*, vol. 16, no. 1, pp. 127–153, 2014.

[128] J. Kliewer, S. X. Ng, and L. Hanzo, “Efficient computation of EXIT functions for nonbinary iterative decoding,” *IEEE Transactions on Communications*, vol. 54, no. 12, pp. 2133–2136, 2006.

[129] W. Khawaja, I. Guvenc, D. W. Matolak, U. Fiebig, and N. Schneckenburger, “A survey of air-to-ground propagation channel modeling for unmanned aerial vehicles,” *IEEE Communications Surveys & Tutorials*, vol. 21, no. 3, pp. 2361–2391, 2019.

[130] W. Khawaja, I. Guvenc, and D. Matolak, “UWB channel sounding and modeling for UAV air-to-ground propagation channels,” in *Global Communications Conference (GLOBECOM), 2016 IEEE*, pp. 1–7, IEEE, 2016.

[131] J. Hagenauer, “The EXIT chart-introduction to extrinsic information transfer in iterative processing,” in *Signal Processing Conference, 2004 12th European*, pp. 1541–1548, IEEE, 2004.

[132] J. Hagenauer, E. Offer, and L. Papke, “Iterative decoding of binary block and convolutional codes,” *IEEE Transactions on Information Theory*, vol. 42, no. 2, pp. 429–445, 1996.

[133] I. A. Hemadeh, P. Botsinis, M. El-Hajjar, S. H. Won, and L. Hanzo, “Reduced-RF-chain aided soft-decision multi-set steered space-time shift-keying for millimeter-wave communications,” *IEEE Access*, 2017.

[134] I. A. Hemadeh, M. El-Hajjar, S. Won, and L. Hanzo, “Multi-set space-time shift keying and space-frequency space-time shift keying for millimeter-wave communications,” *IEEE Access*, vol. 5, pp. 8324–8342, 2017.

[135] I. A. Hemadeh, M. El-Hajjar, S. Won, and L. Hanzo, “Multiuser steered multiset space-time shift keying for millimeter-wave communications,” *IEEE Transactions on Vehicular Technology*, vol. 66, no. 6, pp. 5491–5495.

[136] X. Liu, Q. Xiao, V. Gopalakrishnan, B. Han, F. Qian, and M. Varvello, “360 Innovations for Panoramic Video Streaming,” in *Proceedings of the 16th ACM Workshop on Hot Topics in Networks*, pp. 50–56, ACM, 2017.

[137] A. Nguyen, Z. Yan, and K. Nahrstedt, “Your Attention is Unique: Detecting 360-Degree Video Saliency in Head-Mounted Display for Head Movement Prediction,” in *2018 ACM Multimedia Conference on Multimedia Conference*, pp. 1190–1198, ACM, 2018.

[138] F. Qian, L. Ji, B. Han, and V. Gopalakrishnan, “Optimizing 360 video delivery over cellular networks,” in *Proceedings of the 5th Workshop on All Things Cellular: Operations, Applications and Challenges*, pp. 1–6, ACM, 2016.

[139] J. P. Snyder, *Flattening the earth: two thousand years of map projections*. University of Chicago Press, 1997.

[140] M. Zhou, “AHG8: A study on compression efficiency of cube projection,” *Document JVET-D0022, Chengdu, China*, 2016.

[141] Y. He, B. Vishwanath, X. Xiu, and Y. Ye, “AHG8: InterDigital’s projection format conversion tool,” *Document JVET-D0021, Chengdu, China*, 2016.

[142] V. Zakharchenko, E. Alshina, K. Choi, A. Singh, and A. Dsouza, “AHG8: Icosahedral projection for 360-degree video content,” *document ITU-T SG16 WP3, ISO/IEC JTC1/SC29/WG11, and JVETD0028, Joint Video Explor. Team, ITU-T, Geneva, Switzerland*, 2016.

[143] T. Engelhardt and C. Dachsbacher, "Octahedron Environment Maps.," in *VMV*, pp. 383–388, 2008.

[144] G. Van der Auwera, M. Coban, M. K. Hendry, and M. Karczewicz, "AHG8: Truncated square pyramid projection (tsp) for 360 video," in *Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JVET-D0071, 4th Meeting*, 2016.

[145] C. Zhang, Y. Lu, J. Li, and Z. Wen, "AHG8: Segmented Sphere Projection for 360-degree video," *Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JVET-E0025*, 2017.

[146] Y. Sun, A. Lu, and L. Yu, "Weighted-to-spherically-uniform quality evaluation for omnidirectional video," *IEEE Signal Processing Letters*, vol. 24, no. 9, pp. 1408–1412, 2017.

[147] H. Lin, C. Huang, C. Li, *et al.*, "AHG8: inter Digitals projection format conversion tool Joint Video Exploration Team of ITU-T SG16 WP3 and ISO," IEC JTC1/SC29/WG11, JVET-D0021. In: 5th Meeting, Geneva, 2017.

[148] W. Rudin *et al.*, *Principles of mathematical analysis*, vol. 3. McGraw-hill New York, 1964.

[149] P. R. Alface, J.-F. Macq, and N. Verzijp, "Interactive omnidirectional video delivery: A bandwidth-effective approach," *Bell Labs Technical Journal*, vol. 16, no. 4, pp. 135–147, 2012.

[150] V. R. Gaddam, M. Riegler, R. Eg, C. Griwodz, and P. Halvorsen, "Tiling in interactive panoramic video: Approaches and evaluation," *IEEE Transactions on Multimedia*, vol. 18, no. 9, pp. 1819–1831, 2016.

[151] Y. Bao, H. Wu, T. Zhang, A. A. Ramli, and X. Liu, "Shooting a moving target: Motion-prediction-based transmission for 360-degree videos," in *BigData*, pp. 1161–1170, 2016.

[152] W. Huang, L. Ding, H.-Y. Wei, J.-N. Hwang, Y. Xu, and W. Zhang, "Qoe-oriented resource allocation for 360-degree video transmission over heterogeneous networks," *arXiv preprint arXiv:1803.07789*, 2018.

[153] G. He, J. Hu, H. Jiang, and Y. Li, "Scalable video coding based on user's view for real-time virtual reality applications," *IEEE Communications Letters*, vol. 22, no. 1, pp. 25–28, 2018.

[154] D. Liu, P. An, R. Ma, W. Zhan, and L. Ai, "Scalable Omnidirectional Video Coding for Real-Time Virtual Reality Applications," *IEEE Access*, vol. 6, pp. 56323–56332, 2018.

[155] L. Hanzo, T. H. Liew, and B. L. Yeap, *Turbo coding, turbo equalisation, and space-time coding*. Wiley Online Library, 2002.

[156] A. K. Qin, V. L. Huang, and P. N. Suganthan, “Differential evolution algorithm with strategy adaptation for global numerical optimization,” *IEEE Transactions on Evolutionary Computation*, vol. 13, no. 2, pp. 398–417, 2009.

[157] H. Hu, Y. Lin, M. Liu, H. Cheng, Y. Chang, and M. Sun, “Deep 360 pilot: Learning a deep agent for piloting through 360 sports videos,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1396–1405, 2017.

[158] M. Xu, Y. Song, J. Wang, M. Qiao, L. Huo, and Z. Wang, “Predicting head movement in panoramic video: A deep reinforcement learning approach,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018.

Glossary

AC	Antenna Combination
ACS	Adaptive Channel Selection
AG	Air-to-Ground
AoA	Angle of Arrival
AoD	Angle of Departure
AQAM	Adaptive Quadrature Amplitude Modulation
ARQ	Automatic Repeat Request
AS	Antenna Set
ASA	Angle Spread of Arrival
ASD	Angle Spread of Departure
ASP	Advanced Simple Profile
ASU	Antenna Selection Unit
ATSC	Advanced Television System Committee
AVC	Advanced Video Coding
AVS	Audio Video Coding Standard
AWGN	Adaptive White Gaussian noise
BbB	Burst-by-Burst
BER	Bit Error Rate
BL	Base Layer
BLAST	Bell-Labs Layered Space-Time
BPCU	Bits Per Channel Use

BS	Base Station
BSC	Binary Symmetric Channel
B-frame	Bi-Directionally Predicted Frame
CC	Convolutional Code
CCMC	Continuous-Input Continuous-Output Memoryless Channel
CF	Cost Function
CGS	Coarse-Grain Scalable
CIF	Common Intermediate Format
CMP	Cube Map Projection
CND	Check Node Decoder
CPP-PSNR	Craster Parabolic Projection PSNR
CRA	Clean Random Access
CRC	Cyclic Redundancy Check
CSI	Channel State Information
DAC	Distinct AC
DCMC	Discrete-Input Continuous-Output Memoryless Channel
DPCM	Differential Pulse Code Modulation
DV	Digital Video
DVD	Digital Versatile Disc
EA	Evolutionary Algorithm
EEP	Equal Error Protection
EL	Enhancement Layer
ERP	Equirectangular Projection
EWF	Expanding Window Fountain
EXIT	EXtrinsic Information Transfer
e-PLR	Equivalent PLR
FEC	Forward Error Correction
FER	Frame Error Rate
FoV	Filed of View
FPS	Frame Per Second

GA	Genetic Algorithm
GoP	Group of Pictures
HARQ	Hybrid ARQ
HEVC	High Efficiency Video Coding
HMP	Head Movement Prediction
HQAM	Hierarchical QAM
IDR	Instantaneous Decoding Refresh
IL	Inter-layer
IL-FEC	IL operation-aided FEC
IM	Index Modulation
ISI	Inter-Symbol Interference
ISO	International Organization for Standardization
ITU	International Telecommunication Unit
ITU-T	ITU Telecommunication Standardization Sector
I-frame	Intra-Coded Frame
JCT-VC	Joint Collaborative Team on Video Coding
JPEG	Joint Photographic Experts Group
JVET	Joint Video Experts Team
JVT	Joint Video Team
LA-FEC	Layer-Aware FEC
LDC	Linear Dispersion Code
LDM	Layered Division Multiplexing
LDPC	Low Density Parity Check
LLR	Log-Likelihood Ratio
Log-MAP	Logarithmic Maximum *a posteriori*
LSSTC	Layered Steered Space-Time Coding
LTE	Long-Term Evolution
LUT	LookUp Table
MI	Mutual Information
MIMO	Multiple-Input Multiple-Output

ML	Maximum Likelihood
MMSE	Minimum Mean Square Error
MOS	Mean Opinion Score
MPEG	Moving Picture Experts Group
MUD	Multi-User Detection
MUST	Multi-User Superposition Transmission
MSE	Mean Square Error
MS-STSK	Multi-Set STSK
NALU	Network Abstraction Layer Unit
NOMA	Non-Orthogonal Multiplex Access
NTSC	National Television System Committee
OF	Objective Function
OFDM	Orthogonal Frequency Division Multiplexing
OMA	Orthogonal Multiplex Access
OSI	Open Systems Interconnection
PL	Path Loss
PDF	Probability Density Function
PLR	Packet Loss Ratio
PSAM	Pilot Symbol Assisted Modulation
PSK	Phase Shift Keying
PSNR	Peak SNR
P-frame	Predicted frame
QCIF	Quarter CIF
QoE	Quality of Experience
QP	Quantization Parameter
RA	Receive Antenna
RF	Radio Frequency
RoI	Region of Interest
RS	Reed-Solomon
RSC	Recursive Systematic Convolutional

SBC	Short Block Code
SC	Superposition Coding
SDM	Spatial Division Multiplexing
SECCC	Self-Concatenated Convolutional Code
SHVC	Scalability extension of HEVC
SIC	Successive Interference Cancellation
SISO	Single-In Single-out
SM	Spatial Modulation
SNR	Signal-to-Noise Ratio
SP	Simple Profile
SSK	Space-Shift Keying
STBCs	Space-Time Block Codes
STSK	Space-Time Shift Keying
SVC	Scalable Video Coding
S-PSNR-NN	Spherical PSNR using the Nearest Neighbor position without interpolation
TA	Transmit Antenna
TC	Turbo Code
TDM	Time Division Multiplexing
TV	TeleVision
UAV	Unmanned Aerial Vehicle
UEP	Unequal Error Protection
UHD	Ultra High Definition
VCEG	Video Coding Experts Group
VLC	Variable Length Coding
VND	Variable Node Decoder
VR	Virtual Reality
VVC	Versatile Video Coding
V2X	Vehicle-to-Everything
WS-PSNR	Weighted-to-Spherically-uniform PSNR
5G	5-th Generation