Scalable Panoramic Wireless Video Streaming Relying on Optimal-Rate FEC-Coded Adaptive QAM

Yanqing Zhang, Jiankang Zhang, Senior Member, IEEE, Yongkai Huo,
Chao Xu, Senior Member, IEEE, Mohammed El-Hajjar, Senior Member, IEEE,
and Lajos Hanzo Fellow, IEEE

Abstract—The demanding bitrate requirements of supporting flawless streaming of immersive video remains a challenge. To minimize the bandwidth without degrading the user's experience, we may opt for partitioning the panoramic video frame into numerous tiles and only transmit those covered by the predicted Field of View (FoV), but naturally, this philosophy critically hinges on the accuracy of the Head Movement Prediction (HMP). However, since the HMP is never 100% accurate, rebuffering of the missing portions due to misprediction is likely to introduce video freezes or artefacts, which may significantly degrade the users' experience. Hence, instead of only transmitting the FoV tiles, Scalable Video Coding (SVC) comes to rescue. Explicitly, in SVC schemes, multiple layers having different importance provide a promising solution, where the basic quality of the entire panoramic video is supported by the Base Layer (BL) that only requires a low bitrate, while the Enhancement Layers (EL) are invoked for enhancing the quality of the predicted FoV. In this treatise, we propose coding rate adaptation assisted near-instantaneously Adaptive Quadrature Amplitude Modulation (AQAM) for layered panoramic video streaming. In our design, we categorize the video streaming into three priority classes according to the FoV and the SVC layer index, each of which is mapped to the most appropriate modulation mode determined by the instantaneous channel quality. Furthermore, we conceive an Evolutionary Algorithm (EA) assisted Forward Error Correction (FEC) coding rate optimization method for providing Unequal Error Protection (UEP) in order to maximize the uncoded source-rate according to the inter-frame and inter-layer decoding dependency for tile-based panoramic video streaming. The simulation results show that the proposed AQAM assisted UEP scheme configured by our EA assisted coding rate optimization algorithm significantly improves the overall video performance compared to its Equal Error Protection (EEP) counterpart, and provides perceptually pleasing video quality across a wide range of channel conditions by selecting the most appropriate modulation mode based on the instantaneous channel Signal-to-Noise Ratio (SNR).

Index Terms—Panoramic video streaming, SVC, unequal error protection, AQAM, adaptive protection, global optimization.

I. STATE-OF-THE-ART

Panoramic video, which offers a 360 degree immersive experience for the users, has gradually become a popular video format for presenting flawless ultra-high resolution contents. To reduce the demanding bitrate requirement of panoramic video streams, the Joint Video Expert Team (JVET) has been dedicated for years to the conception of the next-generation video coding standard, namely the H.266/Versatile Video Coding (VVC) standard [1], whose objective is to provide a significant improvement in compression performance over the existing H.265/High Efficiency Video Coding (HEVC) standard [2].

Furthermore, a method that spatially partitions the 360° video into multiple independent rectangular segments, also referred to as tiles, was proposed in [3], 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 [4], where only the specific portions of the scene observed by the users is transmitted [5], 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 [6]. However, since the success of the Field of View (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 [7].

Furthermore, to reduce the bitrate necessitated for panoramic video streaming, the concept of Scalable Video Coding (SVC) [8] that compresses video clips into multiple layers according to the requirement of the clients was conceived. He *et al.* [9] 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 Enhancement Layers (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 [10] for improving the coding efficiency of Virtual Reality (VR) applications, where the particular video frame portions outside the FoV were encoded as the Base Layer (BL) only, while the FoV portions were further enhanced by the ELs.

To elaborate a little further on a single-layer video stream, the Predicted (P)- and Bi-directionally predicted (B)-frames within a Group of Pictures (GoP) cannot be successfully decoded, if the Intra-coded (I)-frame is corrupted, even if the P and B frames were perfectly received. Similarly for SVC, in addition to the afore-mentioned inter-frame dependency, the ELs whose corresponding reference layers are corrupted have to be discarded as well. Therefore, it is important to allocate unequal protection to the different sources. The concept of Unequal Error Protection (UEP) was first proposed by Masnick and Wolf [11], which dedicates using stronger Forward Error Correction (FEC) to the more important I-frames and BLs, while weaker FEC is allocated to the less important P/B frames and ELs. UEP may also be achieved by allocating different transmission power to the individual bits according to their specific error sensitivity [12].

The UEP schemes used for video streaming can be categorized according to the OSI reference layers [13]. The authors of [14] proposed a novel scalable multicast system using application-layer FEC codes for transmitting scalable video streams over hostile channels exhibiting packet loss events, where the ELs conveyed parity information for protecting the more important BL. Hellge *et al.* [15] designed a Raptor coded cross-layer operation aided scheme for layered video transmission, which implants the bits of the BL into the ELs in order to recover the lost bits of the BL from the ELs.

Furthermore, a number of potent contributions employing UEP for improving the reconstructed video quality have been conceived for the physical layer, including sophisticated modulation schemes, transceiver techniques and channel coding, as seen in Table I [15]–[25]. Explicitly, observe from Table I that the majority of the contributions on UEP aided video streaming exploited the Channel State Information (CSI) knowledge. Furthermore, the authors of [16], [18], [21], [23] designed UEP relying on variable FEC code-rates. For example, Marx and Farah [16] minimized the mean video distortion for transmission over wireless networks using different punctured Turbo Codes (TC)s, where the FEC redundancy was non-uniformly shared across the consecutive video frames. Huo *et al.* [23] proposed a Recursive Systematic Convolutional (RSC) coded Inter-Layer operation-aided FEC (IL-FEC) technique for SVC streaming. They implanted the most important bits of the BL into the ELs by taking their modulo-2 addition and at the receiver they iteratively exchanged extrinsic information between them for the sake of improving the overall performance. The IL-FEC philosophy was then enhanced in [25] with the aid of a novel multifunctional transceiver relying on Multi-Set Space-Time Shift Keying (STSK) (MS-STSK) [26] that has the innate potential of providing UEP, even without varying the FEC code-rates. Additionally, UEP schemes were also designed by using sophisticated bit mapping schemes in [17], [20], [25]. Specifically, Chang *et al.* [17] mapped the important source bits to the specific Hierarchical Quadrature Amplitude Modulation (QAM) (HQAM) bits exhibiting the highest Euclidean distance for achieving UEP. As a further compelling feature of our proposed system, the optimal FEC code-rate allocation assisted UEP scheme also takes into account the unique characteristics of panoramic video, such as the specific importance of the different video tiles.

However, it is challenging to find the optimal coding rates, whilst relying on low-complexity implementations, especially when taking into account both the inter-frame and inter-layer coding dependency. Gratifyingly, the development of global optimization algorithms have found ever-increasing applications in communication and signal processing for creating globally optimal designs at affordable computational costs [27]. Zhang *et al.* [28] investigated the capability of powerful bio-inspired learning techniques, including the family of Genetic Algorithm (GA) [29], of Repeated Weighted Boosting Search (RWBS) [30], [31], of Particle Swarm Optimization (PSO) [32] and of Differential Evolutionary Algorithm (DEA) [33]. The design objective was to find optimal solutions for joint channel estimation and turbo multi-user detection/decoding in the context of Orthogonal Frequency Division Multiplexing/Space-Division Multiple-Access (OFDM/SDMA) [34], [35] systems. The GA was found to be capable of promptly determining the Reed-Solomon (RS) code allocation pattern for SVC streaming in [18], which was achieved by dynamically allocating the coding rate to different frames according to the specific layer and the GoP index. Additionally, the authors of [36] designed GA assisted optimal code rate allocation for layered video streaming, where the optimal coding rates were allocated to different layers per video segment before packing the video segment into different-resolution multiple descriptions protected by embedded FEC. Furthermore, the authors of [37] employed the powerful PSO algorithm to find the optimal resource allocation policy that is formulated as a constrained non-linear optimization problem for the sake of achieving the best reconstructed video quality.

Focusing our attention on the associated streaming techniques, numerous near-instantaneously adaptive schemes have been conceived for improving the robustness of the wireless system [38] by providing users with the best possible trade-off amongst a number of contradicting design factors [39]. In order to allow the transceiver to cope with the time-variant channel quality fluctuations of narrowband fading channels, the concept of Adaptive QAM (AQAM) was proposed by Steele and Webb, which provides the flexibility to adjust both the Bit Error Ratio (BER) and the bitrate to suit a particular application [40]. AQAM-aided wireless video transmission was conceived for example in [39], [41], [42], where the Burst-by-Burst (BbB) AQAM assisted system provided smooth Peak Signal-to-Noise Ratio (SNR) (PSNR) fluctuation, even when the channel SNR varied erratically.

Against this background, we conceive a UEP assisted AQAM scheme for tile-based panoramic video streaming over wireless channels, where we jointly consider both the video tile importance, as well as the inter-layer and inter-frame video coding dependency, as indicated by the tick-marks in Table I. Firstly, we employ twin-layer SVC-based panoramic video

Table I: Contrasting our contributions to the state-of-the-art.

Contributions	[15]	[16]	[17]	[18]	[19]	[20]	[21]	[22]	[23]	[24]	[25]	Ours
Variable FEC code-rate		✓		✓			✓		✓		✓	✓
IL-FEC	✓								✓		✓	
Content weight												✓
Bit mapping			✓			✓					✓	
Inter-layer dependency	✓			✓	✓			✓		✓	✓	✓
Inter-frame dependency		✓	✓	✓		✓	✓		✓			✓
Layer extraction												✓
CSI knowledge			✓	✓	✓	✓		✓	✓	✓	✓	✓
Antenna selection					✓							
FoV based												✓
Power control						✓		✓		✓		
Adaptive modulation												✓

coding, where the EL is invoked to enhance the FoV tiles of the user for the sake of improving the quality of experience. Then, the twin-layer video stream is further categorized into three classes according to their contributions to the users' experience, where their importance is identified using specific weighting coefficients. Additionally, the FEC coding rates are judiciously adjusted with the aid of an Evolutionary Algorithm (EA)[1] according to the specific channel SNR, with the objective of assigning stronger protection to more important video bits, such as the I-frame and the BL of SVC. Finally, a particular modulation switching criterion is proposed for selecting the most appropriate UEP assisted modulation mode exhibiting the highest uncoded source-rate. The novelty of this paper can be summarized as follows:

1) We design a variable-rate FEC assisted AQAM system for tile-based panoramic video streaming over Rayleigh fading channels. The panoramic video is compressed into two sub-layers with the aid of the Scalability extension of HEVC (SHVC) of [8], namely the BL and the EL, where only the FoV is included in the EL for the sake of further enhancing the FoV quality.

2) We exploit the different importance of the SVC aided panoramic video streaming classes, where three modulation modes exhibiting different robustness and throughput may be selected. More explicitly, the most robust modulation mode exhibiting the lowest throughput is utilized for transmitting the most important BL information, while the high-throughput modulation modes are activated for streaming the EL of the FoV.

3) We propose a FEC code-rate optimization solution using a novel EA for maximizing the weighted uncoded source-rate for each modulation mode in order to support robust panoramic video streaming, which is designed by taking into account both the inter-frame and inter-layer video coding dependencies, as well as the specific importance of the video tiles.

4) The proposed AQAM system activates the most ap-propriate modulation mode that exhibits the highest uncoded source-rate, whilst ensuring the flawless decod-ability[2] of the most important video information.

The rest of the paper is organized as follows. The basic concepts of panoramic video are introduced in Section II. In Section III, we briefly present the schematic diagram of our adaptive system proposed for transmitting the panoramic video over Rayleigh channels. Our coding rate optimization algorithm is detailed in Section IV, while the adaptive mode-switching operation is illustrated in Section V. Furthermore, in Section VI we compare our proposed system to its Equal Error Protection (EEP) counterpart in terms of both adaptive and fixed modes. Finally, we conclude in Section VII.

II. PANORAMIC VIDEO

In this section, we portray the state-of-the-art in panoramic video, including the projection methods that convert 3D spherical video representation into the conventional 2D plane for encoding and the recent Weighted-to-Spherically-uniform PSNR (WS-PSNR) algorithm that accurately estimates the subjective video quality.

A. ERP

Since the existing video standard is incapable of directly encoding spherical videos, the panoramic video is first con-verted to the two-dimensional plane using different projection methods, such as Icosahedral Projection (ISP) [43], Equal-Area Projection (EAP) [44], ERP [45] and Cubemap Projec-tion (CMP) [46] techniques.

Fig. 1 shows a conversion example from the spherical format of Fig. 1(a) to the ERP projection of Fig. 1(b). Further-more, Fig. 1(b) represents the tile-based compression method for ERP projection, where the ERP projection is partitioned into 32 tiles with each encoded independently. Fig. 2 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:

$$W = 2\pi\rho$$
$$H = \pi\rho. \tag{1}$$

---

[1]We employ the global optimization to find the empirical upper bound of the gain introduced by the proposed system, which also helps the employment of the suboptimal solutions to strike the trade-off between performance gain and complexity.

[2]We define decodability as the successful video frame decoding probability.

(a) Sphere format.

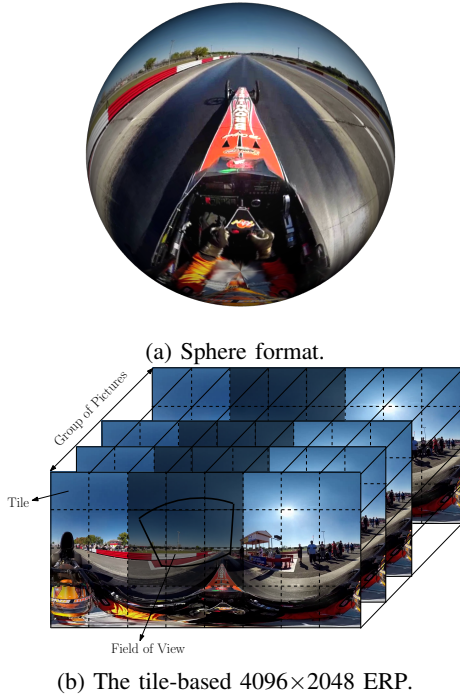
(b) The tile-based 4096×2048 ERP.
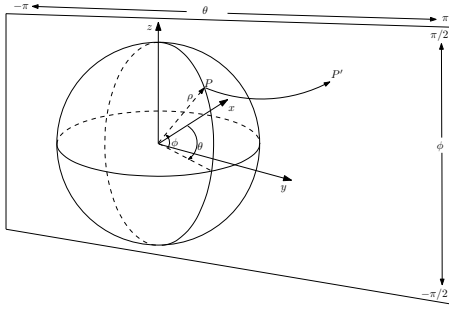
Fig. 1. Panoramic video projections.

Fig. 2. The ERP mapping procedure.

Assuming that $P(x_p, y_p, z_p)$ denotes an arbitrary point on the sphere of Fig. 2, the latitude ϕ can be expressed as:

$$\phi = \arcsin \frac{z_p}{\rho}, \quad -\rho \leq z_p \leq \rho, \qquad (2)$$

while the longitude θ is expressed as:

$$\theta = \arctan 2(y_p, x_p), \quad -\rho \leq x_p, y_p \leq \rho, \qquad (3)$$

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:

$$\begin{aligned} w_p &= \theta \cdot \rho \\ h_p &= \phi \cdot \rho. \end{aligned} \qquad (4)$$

B. WS-PSNR

Mapping from the representation space, such as the two-dimensional planes of Fig. 1(b), to the spherical surface of Fig. 1(a) 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 sphere may be repeatedly used in the 2D ERP projected format, which stretches the pole pixels and therefore introduces increased 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.* [47], where the distortion of the samples in the representation space is further weighted by the corresponding projection area in the observation space.

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 by [47]:

$$\vartheta(w, h) = \cos \frac{(h + 0.5 - H/2)\pi}{H}, \qquad (5)$$

which is depicted in Fig. 3, where the white colour represents higher importance, while the dark region represents lower importance. The Weighted Mean Square Error (WMSE) is therefore given by:

$$WMSE = \frac{\sum_{w=1}^{W} \sum_{h=1}^{H} \left[y(w, h) - y'(w, h) \right]^2 \vartheta(w, h)}{\sum_{w=1}^{W} \sum_{h=1}^{H} \vartheta(w, h)}, \qquad (6)$$

where $y(w, h)$ and $y'(w, h)$ are the original and received pixel values at the position (w, h) of the ERP. Hence, given the maximum possible intensity level of the image MAX_I, the WS-PSNR of panoramic video can be expressed as [47]:

$$WS - PSNR = 10 \log_{10} \frac{MAX_I^2}{WMSE}. \qquad (7)$$

III. SYSTEM MODEL

In this section, we introduce the proposed adaptive scheme conceived for layered panoramic video communications over wireless channels. Before elaborating on the system model, let us briefly introduce the importance classes of the different video tiles using Fig. 1(b), 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 Fig. 1(b) 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:

Fig. 3. Weighting coefficient of ERP.

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 code-rate optimization algorithms, which are part of the Rate Optimization block of Fig. 4 and will be detailed in Section IV.

A. Transmitter

The transmitter model of the proposed design is depicted in Fig. 4. 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 Fig. 1(b) 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 Fig. 4 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 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 $\{r_{i,j}^1\}_{i \in FoV}$, $\{r_{i,j}^1\}_{i \notin FoV}$ and $\{r_{i,j}^2\}_{i \in FoV}$, respectively, as depicted in Fig. 4, which are optimized by the Coding Rate Allocation block of Fig. 4 depending on the instantaneous channel SNR Γ.

Furthermore, the AQAM Space-Time Block Code (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 $\{f_{i,j}^1\}_{i \in FoV}$ only, while 16QAM transmits the C1 and C2 BL of all tiles, $\{f_{i,j}^1\}_{i \in FoV} \cup \{f_{i,j}^1\}_{i \notin FoV}$. Finally, all the C1 and C2 BLs and the C3 EL of the tiles within the FoV, namely $\{f_{i,j}^1\}_{i \in FoV} \cup \{f_{i,j}^1\}_{i \notin FoV} \cup \{f_{i,j}^2\}_{i \in FoV}$, 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 Fig. 4 explores the most appropriate coding rate allocation for $\{f_{i,j}^1\}_{i \in FoV}$ 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 Multiple-Input Multiple-Output (MIMO) techniques, such as Spatial Modulation (SM) [48], STSK [49] and MS-STSK [26].

B. Receiver

The processing of the received panoramic signals is portrayed in Fig. 4, which are first demapped by the AQAM STBC decoder block of Fig. 4 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 Fig. 4, 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.

IV. Evolutionary Algorithm Assisted FEC Code-Rate Optimization

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 Fig. 1(b) 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 Fig. 3, 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 IV-C.

The rest of this section is organized as follows. Our code-rate optimization problem is formulated in Section IV-A 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 IV-B, 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 IV-C, while the optimal FEC code-rates are discussed in Section IV-D. To clearly present the code-rate optimization algorithm, the most complex 128QAM conveying all the three
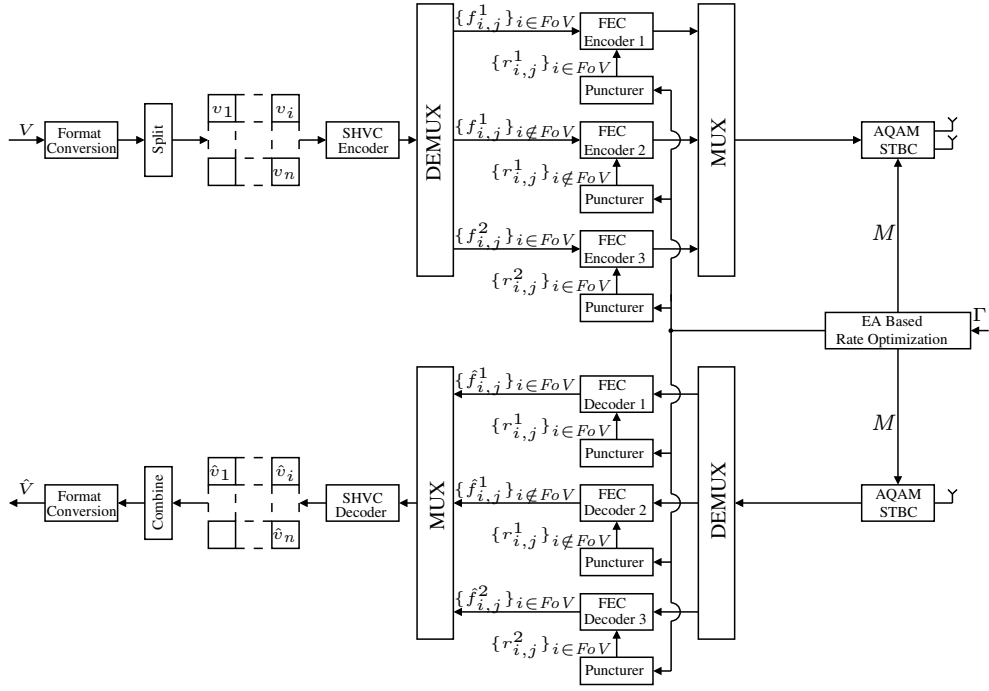
Fig. 4. Block diagram of the proposed system.

classes is exemplified in the following sections, but the CF can be readily modified for the other modulation modes.

A. 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 Fig. 3, 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:

$$\max \quad T_\omega\big(\{r_{i,j}^l\},\Gamma\big), \tag{8}$$

$$s.t.: \quad \sum_{i=1}^{N}\sum_{j=1}^{G}\frac{|f_{i,j}^1|}{r_{i,j}^1} + \sum_{i\in FoV}\sum_{j=1}^{G}\frac{|f_{i,j}^2|}{r_{i,j}^2} \leq T, \tag{9}$$

$$r_{i,j}^l \in \mathcal{R}. \tag{10}$$

(9) 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, (10) represents the constraint that the FEC code-rate has to be selected from the legitimate code-rate set $\mathcal{R}$. (8) can be reformulated as the weighted uncoded source-rate summation of the three classes, as given by:

$$T_\omega\big(\{r_{i,j}^l\},\Gamma\big) = \sum_{c=1}^{3} T_\omega^c\big(\{r_{i,j}^l\},\Gamma\big), \tag{11}$$

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

$$T_\omega^c\big(\{r_{i,j}^l\},\Gamma\big) = \begin{cases} \beta_1 \sum\limits_{i\in FoV} \alpha_i \sum\limits_{j=1}^{G} T_{i,j}^1\big(r_{i,j}^1,\Gamma\big), & c=1, \\ \beta_2 \sum\limits_{i\notin FoV} \alpha_i \sum\limits_{j=1}^{G} T_{i,j}^1\big(r_{i,j}^1,\Gamma\big), & c=2, \\ \beta_3 \sum\limits_{i\in FoV} \alpha_i \sum\limits_{j=1}^{G} T_{i,j}^2\big(r_{i,j}^2,\Gamma\big), & c=3. \end{cases} \tag{12}$$

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 Fig. 3, both of which will be detailed in Section IV-C. Furthermore, $T_{i,j}^l\big(\{r_{i,j}^l\},\Gamma\big)$ 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:

$$T_{i,j}^l\big(r_{i,j}^l,\Gamma\big) = |f_{i,j}^l| \prod_{j=1}^{j}\prod_{\ell=1}^{l} \psi(|f_{i,j}^\ell|, r_{i,j}^\ell,\Gamma). \tag{13}$$

In (13), $\psi\big(|f_{i,j}^\ell|, r_{i,j}^\ell,\Gamma\big)$ represents the decodability of video frame $f_{i,j}^\ell$ and $T_{i,j}^l\big(r_{i,j}^l,\Gamma\big)$ is obtained by taking into account the inter-layer and inter-frame coding dependencies. Substituting (12) and (13) into (11), we can have the expression of the weighted uncoded source-rate, as given by:

$$T_\omega\big(\{r_{i,j}^l\},\Gamma\big) = \beta_1 \sum_{i\in FoV} \alpha_i \sum_{j=1}^{G} |f_{i,j}^1| \prod_{j=1}^{j} \psi\big(|f_{i,j}^1|, r_{i,j}^1,\Gamma\big) \tag{14}$$

$$+ \beta_2 \sum_{i\notin FoV} \alpha_i \sum_{j=1}^{G} |f_{i,j}^1| \prod_{j=1}^{j} \psi\big(|f_{i,j}^1|, r_{i,j}^1,\Gamma\big)$$

$$+ \beta_3 \sum_{i \in FoV} \alpha_i \sum_{j=1}^{G} |f_{i,j}^2| \prod_{j=1}^{j} \prod_{\ell=1}^{2} \psi(|f_{i,j}^\ell|, r_{i,j}^\ell, \Gamma).$$

B. Successful Video Frame Decoding Probability

In this subsection, we embark on mathematically modelling the decodability of the variable-length video frame for the three modulation modes. We note that the FEC performance varies with the input packet length, regardless of the specific choice of FEC configuration [50]. Additionally, the video frames are usually compressed to bit streams exhibiting different frame lengths in terms of the number of bits. This represents a further challenge in terms of predicting the Frame Error Rate (FER) of the variable-length video frame. Fortunately, the authors of [12] [51] proposed a suitable FER estimation method for video frames resulting in different number of encoded bits, 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 Packet Error Ratio (PER) of a packet containing λ bits is p. Then the decodability of an $|f_{i,j}^l|$-bit video frame, where we have $|f_{i,j}^l| = n \times \lambda$, can be obtained by:

$$\psi = \left[1 - p\right]^n, \tag{15}$$

which is depicted in Fig. 5. Similarly, the decodability of an arbitrary video frame containing different number of bits can be found by exploiting the PER of a packet having a fixed length of λ, which is given by:

$$\psi(|f_{i,j}^\ell|, r_{i,j}^\ell, \Gamma) = \left[1 - p(r_{i,j}^\ell, \Gamma)\right]^{\frac{|f_{i,j}^\ell|}{\lambda}}, \tag{16}$$

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

$$p(r_{i,j}^\ell, \Gamma) = h(r, \Gamma) \approx \hbar(r, \Gamma) = \tag{17}$$
$$\frac{1}{1 + e^{-(a_4\Gamma^4 + a_3\Gamma^3 + a_2\Gamma^2 + a_1\Gamma + b_4r^4 + b_3r^3 + b_2r^2 + b_1r) + c}},$$

as exemplified in Fig. 6, where $\hbar(r, \Gamma)$ represents the fitted mathematical model. Specifically, Fig. 6(a) compares the scatter-diagrams of our Monte-Carlo simulations and the mathematically fitted curves, while Fig. 6(b) depicts the absolute
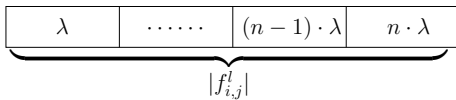

Fig. 5. A video frame $f_{i,j}^l$ having a frame length of $|f_{i,j}^l| = n \times \lambda$ bits may be partitioned into n shorter packets with a length of λ bits [23].

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 II. It is worth noting that both (17) and the values included in Table II have been evaluated empirically.

C. Estimation of the Weighting Factors

The weighting coefficients α and β in (14) 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 (14). Explicitly, the weighting coefficient α tends to guarantee that within the same class the tile exhibiting the lowest transformation distortion of Fig. 3 is strongly protected, while β is used for ensuring the robustness of the bit streams of the more important classes.

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 Fig. 3. In this section, we elaborate on calculating the weighting factor α_i for each tile v_i. Our methodology of generating α_i is to first partition the weighting coefficients obtained in (5) of Fig. 3 into the tiles of Fig. 1(b), yielding $\vartheta_i(w, h)$. Then, the weighting factor α_i of tile v_i can be obtained by averaging all the weighting coefficients of the partitioned $\vartheta_i(w, h)$, which is formulated as:

$$\bar{\alpha}_i = \frac{\sum_{w=1}^{W_i} \sum_{h=1}^{H_i} \vartheta_i(w, h)}{W_i \times H_i}, \tag{18}$$

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:

$$\{\alpha_i\}_{i=1}^{N} = \frac{\bar{\alpha}_i}{\sum_{i=1}^{N} \bar{\alpha}_i}. \tag{19}$$

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 model the weighted uncoded source-rate of (14), 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 $\min_{i \in FoV}\{|f_{i,G}^1|\}$, always remains higher than that of the most

Table II: Parameters values of (17) 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
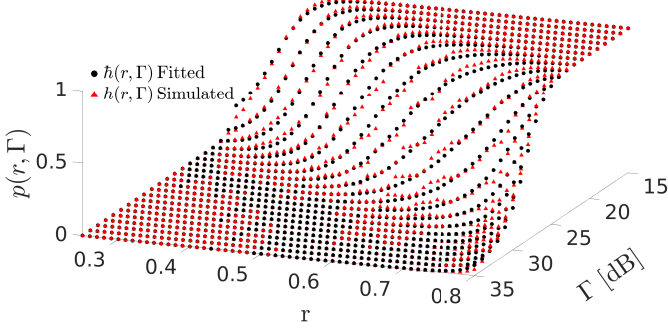
(a) Simulated scatter-diagram and its fitted mathematical counterpart for 128QAM, where the parameter are given in Table II and $\lambda = 1000$ bits.
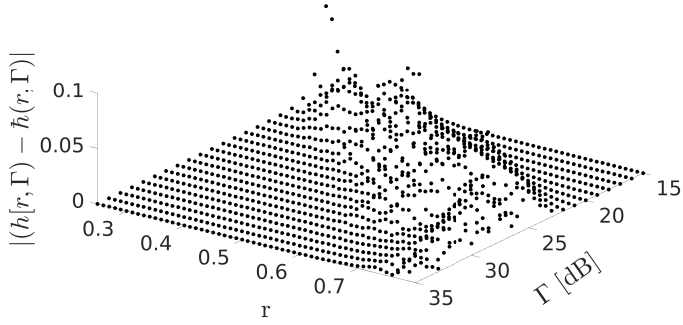
(b) Absolute difference between the Simulated and Fitted scatter points of Fig. 6(a).

Fig. 6. Model comparison between the simulated scatter-diagram and its fitted mathematical counterpart for 128QAM.

important uncoded source bits in C2, namely $\max\limits_{i \notin FoV}\{|f_{i,1}^1|\}$, which is formulated as:

$$\beta_1' \min\limits_{\forall i \in FoV}\{|f_{i,G}^1|\} \geq \beta_2' \max\limits_{\forall i \notin FoV}\{|f_{i,1}^1|\}. \tag{20}$$

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. $\min\limits_{\forall i \notin FoV}\{|f_{i,G}^1|\}$, are better protected than the most important video bits of C3, namely $\max\limits_{\forall i \in FoV}\{|f_{i,1}^2|\}$, which is formulated as:

$$\beta_2' \min\limits_{\forall i \notin FoV}\{|f_{i,G}^1|\} \geq \beta_3' \max\limits_{\forall i \in FoV}\{|f_{i,1}^2|\}. \tag{21}$$

Finally, setting β_3' representing the weight of the least important information to 1.0, the class weighting factors are

expressed as:

$$\beta_1' = \frac{\max\limits_{\forall i \notin FoV}\{|f_{i,1}^1|\}}{\min\limits_{\forall i \in FoV}\{|f_{i,G}^1|\}} \cdot \frac{\max\limits_{\forall i \in FoV}\{|f_{i,1}^2|\}}{\min\limits_{\forall i \notin FoV}\{|f_{i,G}^1|\}},$$

$$\beta_2' = \frac{\max\limits_{\forall i \in FoV}\{|f_{i,1}^2|\}}{\min\limits_{\forall i \notin FoV}\{|f_{i,G}^1|\}}, \tag{22}$$

$$\beta_3' = 1.0,$$

which are then normalized according to:

$$\beta_1 = \frac{\beta_1'}{\beta_1' + \beta_2' + \beta_3'},$$

$$\beta_2 = \frac{\beta_2'}{\beta_1' + \beta_2' + \beta_3'}, \tag{23}$$

$$\beta_3 = \frac{\beta_3'}{\beta_1' + \beta_2' + \beta_3'}.$$

D. Coding Rate Optimization

By substituting (13), (16), (17), (19) and (23), into (14), the weighted uncoded source-rate can be further detailed as:

$$T_\omega\big(\{r_{i,j}^l\}, \Gamma\big) = \beta_1 \sum_{i \in FoV} \alpha_i \sum_{j=1}^G |f_{i,j}^1| \prod_{j=1}^j \big[1 - \hbar(r_{i,j}^1, \Gamma)\big]^{\frac{|f_{i,j}^1|}{\lambda}}$$

$$+ \beta_2 \sum_{i \notin FoV} \alpha_i \sum_{j=1}^G |f_{i,j}^1| \prod_{j=1}^j \big[1 - \hbar(r_{i,j}^1, \Gamma)\big]^{\frac{|f_{i,j}^1|}{\lambda}}$$

$$+ \beta_3 \sum_{i \in FoV} \alpha_i \sum_{j=1}^G |f_{i,j}^2| \prod_{j=1}^j \prod_{\ell=1}^2 \big[1 - \hbar(r_{i,j}^\ell, \Gamma)\big]^{\frac{|f_{i,j}^\ell|}{\lambda}}, \tag{24}$$

where in addition to accounting for the inter-frame coding dependency, the third item of (24) 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 Fig. 4, the maximization of the weighted uncoded source-rate of (24) 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:

$$\{r_{i,j}^l\}_{opt} = \arg\max\limits_{\{r_{i,j}^l\}_{\forall i,j,l} \in \mathcal{R}} \Big\{T_\omega\big(\{r_{i,j}^l\}\big)\Big\}. \tag{29}$$

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 (10), 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.

Algorithm 1 Coding Rate Allocation Algorithm

1: **Initialization**: Set the number of generations to $g = 1$ and randomly generate the initial coding rate set $\{\hat{\mathbf{r}}_{g,p_s}\}_{p_s=1}^{P_s}$, 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 (9). 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 (24) for each $\hat{\mathbf{r}}_{g,p_s}$, where $1 \le p_s \le P_s$. 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:

$$\tilde{\mathbf{r}}_{g,p_s} = \hat{\mathbf{r}}_{g,p_s} + \lambda_{p_s}(\hat{\mathbf{r}}_{g,p_s,best} - \hat{\mathbf{r}}_{p_s}) \quad (25)$$
$$+ \lambda_{p_s}(\hat{\mathbf{r}}_{g,p_s,c_1} - \hat{\mathbf{r}}_{g,p_s,c_2}),$$

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:

$$\check{r}_{i,j,g,p_s}^l = \begin{cases} \tilde{r}_{i,j,g,p_s}^l, & rand(0,1) \le C_{r_{p_s}} \\ \hat{r}_{i,j,g,p_s}^l, & \text{otherwise}, \end{cases} \quad (26)$$

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 $T_\omega(\check{\mathbf{r}}_{g,p_s}) \ge T_\omega(\hat{\mathbf{r}}_{g,p_s})$, 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 [52]:

$$\mu_{C_r} = (1-c) \cdot \mu_{C_r} + c \cdot mean_A(S_{C_r}) \quad (27)$$
$$\mu_\lambda = (1-c) \cdot \mu_\lambda + c \cdot mean_L(S_\lambda), \quad (28)$$

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 [52], 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.

The EA assisted coding rate optimization is detailed in Algorithm 1, which aims for maximizing the weighted uncoded source-rate based CF of (24), whilst maintaining the throughput below the upper bound constraint of (9). It is worth noting that for different modulation modes the CF of (24) and the upper bound constraint of (9) 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 1 is simplified to the first term at the right hand side of (24).

V. ADAPTIVE QAM

Again, the proposed AQAM system of Fig. 4 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 [40], which has also been employed for video streaming [39], [41], [42]. 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:

$$M = \underset{M \in \{BPSK, 16QAM, 128QAM\}}{\arg\max}\{T_M\}, \quad (30)$$

where T_{BPSK}, T_{16QAM} and T_{128QAM} represent the uncoded source-rate provided by BPSK, 16QAM and 128QAM, respectively, which can be predicted with the aid of the corresponding optimal FEC code-rates. For example, the uncoded

source-rate for 128QAM is expressed as:

$$T_{128QAM} = \sum_{i \in FoV} \sum_{j=1}^{G} |f_{i,j}^1| \prod_{j=1}^{j} \left[1 - \hbar(r_{i,j}^1, \Gamma)\right]^{\frac{|f_{i,j}^1|}{\lambda}}$$
$$+ \sum_{i \notin FoV} \sum_{j=1}^{G} |f_{i,j}^1| \prod_{j=1}^{j} \left[1 - \hbar(r_{i,j}^1, \Gamma)\right]^{\frac{|f_{i,j}^1|}{\lambda}}$$
$$+ \sum_{i \in FoV} \sum_{j=1}^{G} |f_{i,j}^2| \prod_{j=1}^{j} \prod_{\ell=1}^{2} \left[1 - \hbar(r_{i,j}^\ell, \Gamma)\right]^{\frac{|f_{i,j}^\ell|}{\lambda}},$$
$$\tag{31}$$

where $p(r_{i,j}^\ell, \Gamma)$ can be found in (17) and Table II. It can be seen in (31) that obtaining the uncoded source-rate requires the knowledge of the optimal FEC code-rate assignment, which is exemplified in Section IV. Hence, the modulation selection (30) 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:

$$R = \frac{\sum_{i=1}^{N} \sum_{j=1}^{G} |f_{i,j}^1| + \sum_{i \in FoV} \sum_{j=1}^{G} |f_{i,j}^2|}{T}, \tag{32}$$

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

VI. PERFORMANCE RESULTS

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 III. The ERP having a resolution of 4096×2048 pixels, as depicted in Fig. 1(b), 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 Frame Per Second (FPS), each having Cyclic Redundancy Check (CRC) bits attached to its tail to check the error-free decodability 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 Instantaneous Decoding Refresh (IDR)/Clean Random Access (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
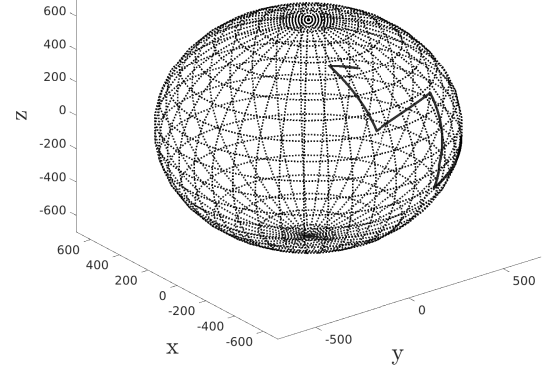

Fig. 7. FoV track prediction.

a consequence, the video sequence in our simulations only consists of I- and P-frames.

In addition to the video parameters, Table III 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 bitrates 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

Table III: 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	4×8=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

track is exemplified in Fig. 7 with the aid of the ERP mapping illustrated in Fig. 2, which can be estimated using the methods of [53] [54].

A. Complexity Analysis

Recall from Section IV that the optimization is carried out across the GoP period, exhibiting an optimization period of $\frac{G}{F}$, where F represents the FPS repetition rate of the video sequence. More particularly, in the example of Table III, where $G = 8$ and $F = 30$ represent the modulation mode and its associated FEC code-rate are optimized every $\frac{8}{30} \approx 0.267$ second. The optimization frequency is reduced upon extending the GoP period at the specific F. For example, the optimization period increases to 0.533 second when doubling G. However, this is achieved at the expense of an increased optimization complexity, since the increased number of elements to be optimized across a GoP period requires more generations for convergence 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 computational complexity order imposed. For a specific population size P_s and the number of generations G_{max}, the value of K employed by the EA for finding the optimal FEC code-rate set is equal to $P_s G_{max}$. By contrast, the value of K using exhaustive search for the modulation mode M is equivalent to $(N_M G)^{|\mathcal{R}|}$, where N_M represents the number of tiles conveyed when the modulation mode M is activated, and $|\mathcal{R}|$ denotes the number of FEC code-rates in the candidate set $\mathcal{R}$. Hence, the complexity imposed by the EA is given by a $\frac{P_s G_{max}}{(N_M G)^{|\mathcal{R}|}}$ 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 higher G_{max} has to be employed in order to achieve convergence. However, since in this treatise we aim for proposing an EA assisted FEC code-rate allocation for layered panoramic video streaming, further investigating the trade-off among the parameters N_M, G, $|\mathcal{R}|$, and the convergence performance G_{max} is left for our future work.

B. Performance Analysis

Fig. 8(a) depicts the Probability Density Function (PDF) of the three modulation modes versus the average channel SNR according to the mode-switching criterion detailed in Section V. 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, Fig. 8(b) 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 criterion of Section V.

Fig. 9 compares the simulation results of both our proposed UEP assisted AQAM and of its EEP assisted counterparts. To elaborate, Fig. 9(a)-(d) portray the PSNR of the FoV, the FER of the BL of the FoV, the WS-PSNR of the ERP projection
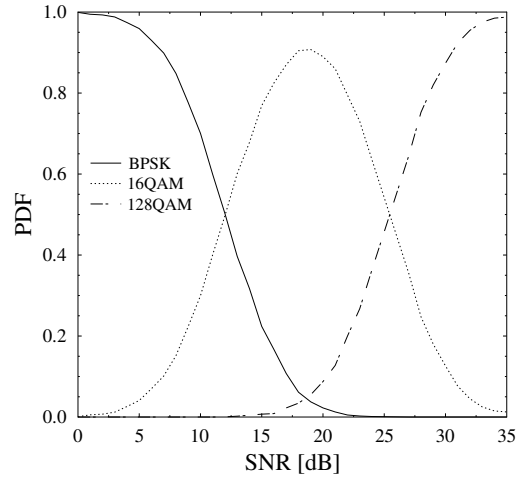
(a) PDF of proposed UEP assisted adaptive modem modes versus the channel SNR.
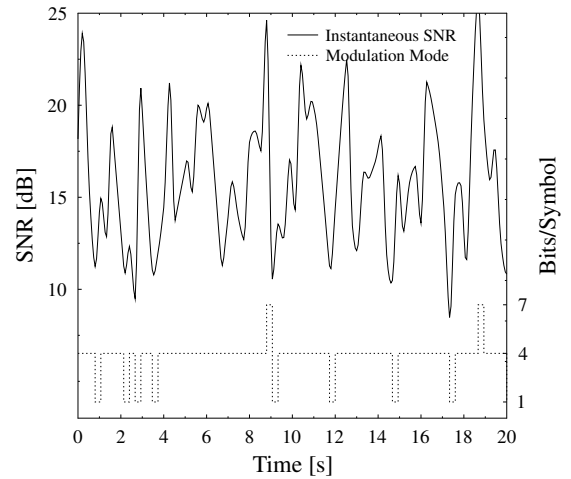
(b) Modulation mode switching at an average channel SNR of 16 dB.

Fig. 8. Modulation modes switching operation versus channel quality.

and the FER of the BL of the ERP projection, respectively. Observe from Fig. 9 that the EA assisted UEP designs on average outperform their EEP counterparts. The BPSK mode that only transmits the FoV bits is characterized in Fig. 9(a) and (b), where 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 the CF of (24) 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 Fig. 9(b). Furthermore, Fig. 9(a) shows that the PSNR of the FoV exhibited by
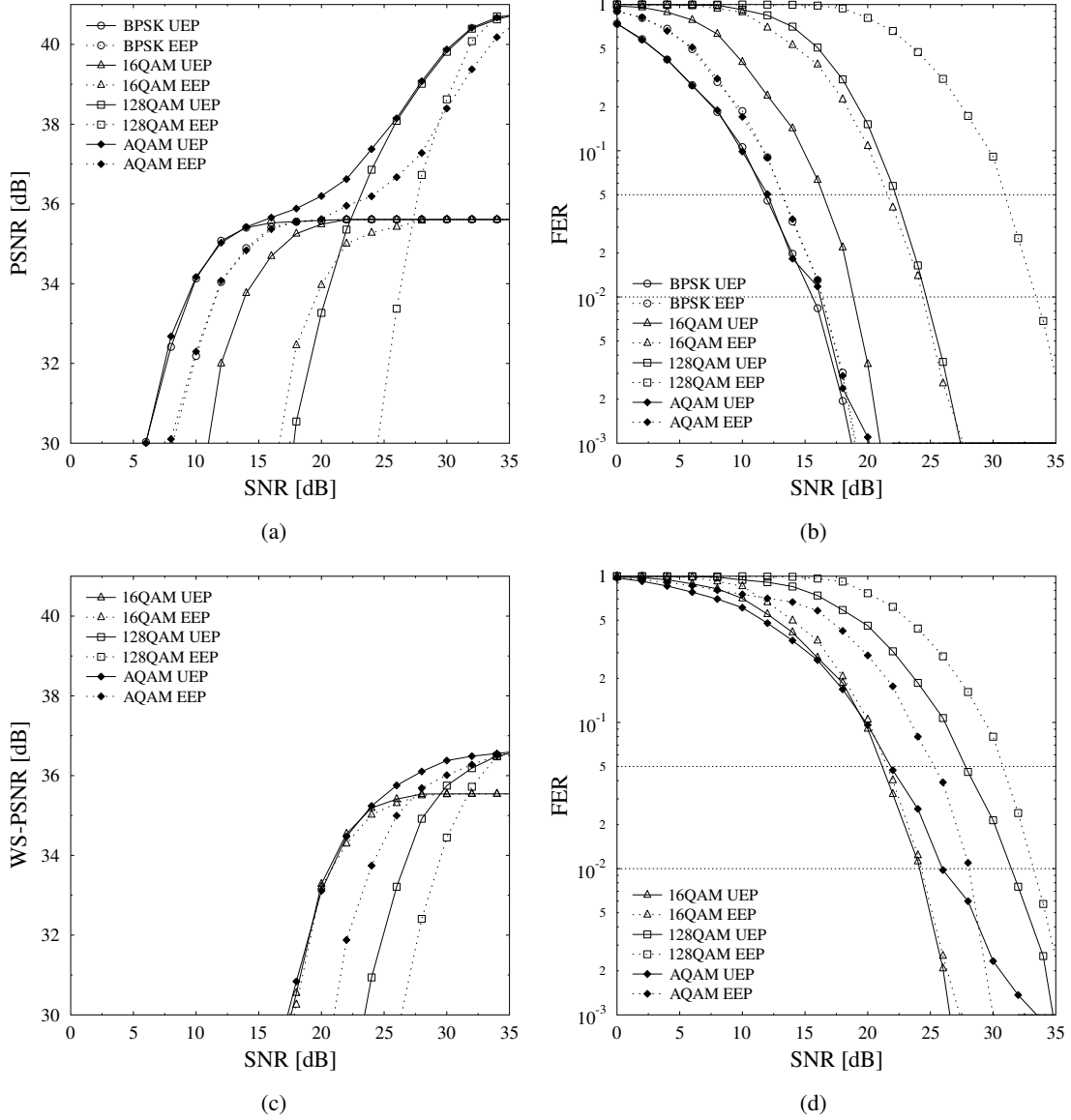
Fig. 9. The comparisons between the proposed systems and their EEP counterparts for the *RaceVR* test sequence against channel SNR, which are (a) the PSNR of the FoV, (b) the FER of the FoV, (c) the WS-PSNR of the ERP and (d) the FER of the BL of the ERP, respectively.

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 (24), the proposed UEP assisted 128QAM mode outperforms its EEP assisted counterpart by approximately 9 dB channel SNR at a 5% FER value of the FoV, as seen in Fig. 9(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 Fig. 9, 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 Fig. 10 at a channel SNR of 12 dB. The UEP assisted BPSK of Fig. 10(a) and the EEP assisted BPSK of Fig. 10(c) as well as UEP assisted 16QAM of Fig. 10(e) are capable of reconstructing the image, while the UEP assisted BPSK in Fig. 10(a) exhibits the smallest distortion. To make the

visual comparison more explicit, in Fig. 10(b), (d) and (f) we portrayed the error between the original and received video frames, where a darker region represents a higher error.

Finally, Fig. 11 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.

VII. CONCLUSIONS

In this treatise, we proposed a FEC code-rate optimization aided AQAM system for layered panoramic video streaming. We first devised a CF by carefully exploiting the weights of the

(a) BPSK UEP (35.1 dB) (b) BPSK UEP (c) BPSK EEP (34.0 dB) (d) BPSK EEP (e) 16QAM UEP (32.1 dB) (f) 16QAM UEP
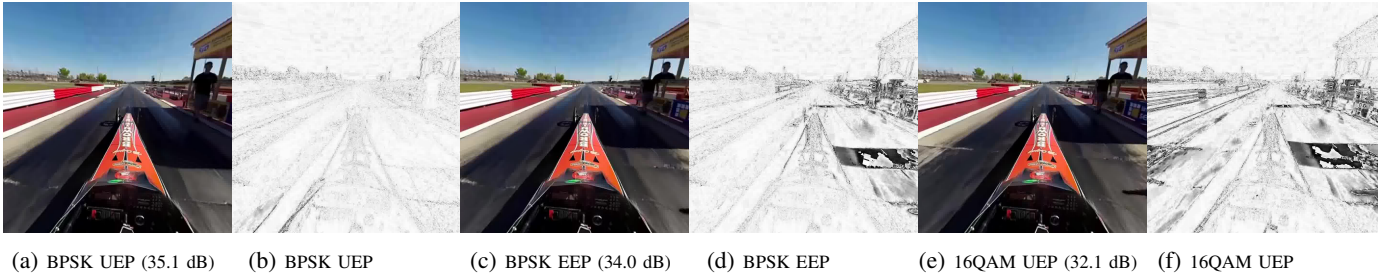
Fig. 10. Reconstructed FoV frame comparison of different modulation modes at the channel SNR of 12 dB, where (b), (d) and (f) portray the error between the original and received video of (a), (c) and (e), respectively.
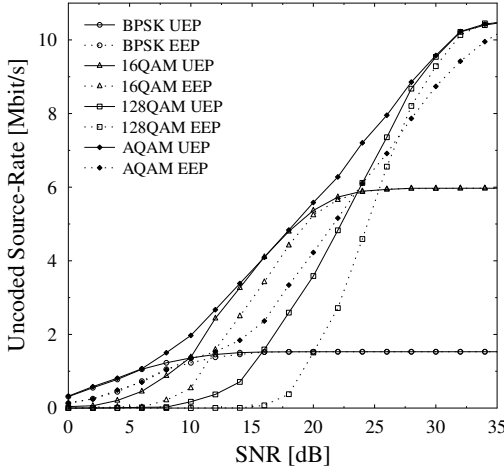

Fig. 11. Uncoded source-rate versus channel SNR comparison of the three fixed modulation modes and the adaptive AQAM in terms of both the UEP and EEP assisted schemes.

TCM (TTCM), Bit-Interleaved Coded Modulation (BICM) and BICM-Iterative Decoding (BICM-ID) [50], [55] in our future research, where numerous optimization algorithms, such as Monotonic Optimization (MO) [56], may be employed to find the globally optimal solution.

Finally, there is a radically new but extremely challenging trend, aiming for extending the classic single-component optimization to multi-component Pareto optimization by determining all optimal points of an optimization problem [57]. Explicitly, Pareto-optimization would determine all optimal points of the associated multi-component optimization problem, guaranteeing that none of the above-mentioned metric can be improved without degrading at least one of the other metrics.

tiles and different classes. An EA assisted FEC code-rate optimization algorithm was then invoked for transmitting the video data conveying different importance classes by taking into account their coding dependency, which dynamically assigns the lowest coding rates to the most important information.

Additionally, an UEP assisted AQAM adaptation scheme was conceived based on EA optimization algorithms, which maximized the uncoded source-rate when guaranteeing the FER value of the most important C1 information. Our simulation results show that both the UEP and EEP assisted AQAM schemes are capable of guaranteeing a FER below 5% for the most important information, while our proposed UEP assisted AQAM system progressively protects the less important information utilizing the remaining resources and hence maximizes the effective source-rate of the system.

Our EA assisted optimal FEC code-rate allocation aims for achieving the best possible improvement of the proposed system, albeit naturally at the expense of an increased computational complexity. In our future work, we may consider a range of lower-complexity near-optimal solutions. Moreover, we may also treat AQAM aided optimal FEC code-rate allocation as a joint optimization problem in joint coded modulation schemes such as Trellis Coded Modulation (TCM), Turbo

REFERENCES

[1] B. Bross, K. Andersson, M. Blåser, 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 Trans. Circuits Syst. Video Technol.*, vol. 30, no. 5, pp. 1226–1240, 2020.
[2] G. J. Sullivan, J. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, 2012.
[3] P. R. Alface, J.-F. Macq, and N. Verzijp, "Interactive omnidirectional video delivery: A bandwidth-effective approach," *Bell Labs Tech. J.*, vol. 16, no. 4, pp. 135–147, 2012.
[4] V. R. Gaddam, M. Riegler, R. Eg, C. Griwodz, and P. Halvorsen, "Tiling in interactive panoramic video: Approaches and evaluation," *IEEE Trans. Multimedia*, vol. 18, no. 9, pp. 1819–1831, 2016.
[5] X. Liu, Q. Xiao, V. Gopalakrishnan, B. Han, F. Qian, and M. Varvello, "360 Innovations for Panoramic Video Streaming," in *Proc. 16th ACM Workshop Hot Topics Netw.*, pp. 50–56, ACM, 2017.
[6] 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 Conf. Multimedia Conf.*, pp. 1190–1198, ACM, 2018.
[7] 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.
[8] J. M. Boyce, Y. Ye, J. Chen, and A. K. Ramasubramanian, "Overview of SHVC: scalable extensions of the high efficiency video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 1, pp. 20–34, 2016.
[9] G. He, J. Hu, H. Jiang, and Y. Li, "Scalable video coding based on user's view for real-time virtual reality applications," *IEEE Commun. Lett.*, vol. 22, no. 1, pp. 25–28, 2018.
[10] 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.
[11] B. Masnick and J. Wolf, "On linear unequal error protection codes," *IEEE Trans. Inf. Theory*, vol. 13, no. 4, pp. 600–607, 1967.

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

[13] H. Zimmermann, "OSI reference model-the ISO model of architecture for open systems interconnection," *IEEE Trans. Commun.*, vol. 28, no. 4, pp. 425–432, 1980.

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

[15] C. Hellge, D. Gomez-Barquero, T. Schierl, and T. Wiegand, "Layer-aware forward error correction for mobile broadcast of layered media," *IEEE Trans. Multimedia*, vol. 13, no. 3, pp. 551–562, 2011.

[16] F. Marx and J. Farah, "A novel approach to achieve unequal error protection for video transmission over 3G wireless networks," *Signal Process., Image Commun.*, vol. 19, no. 4, pp. 313–323, 2004.

[17] 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 Trans. Consum. Electron.*, vol. 52, no. 4, pp. 1153–1158, 2006.

[18] T. Fang and L.-P. Chau, "GOP-based channel rate allocation using genetic algorithm for scalable video streaming over error-prone networks," *IEEE Trans. Image Process.*, vol. 15, no. 6, pp. 1323–1330, 2006.

[19] 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 Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1218–1226, 2007.

[20] 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 Trans. Consum. Electron.*, vol. 55, no. 3, pp. 1089–1097, 2009.

[21] Nasruminallah and L. Hanzo, "Near-capacity H.264 multimedia communications using iterative joint source-channel decoding," vol. 14, no. 2, pp. 538–564, 2012.

[22] M. Li, Z. Chen, and Y. Tan, "Scalable resource allocation for SVC video streaming over multiuser MIMO-OFDM networks," *IEEE Trans. Multimedia*, vol. 15, no. 7, pp. 1519–1531, 2013.

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

[24] S. Cicalo and V. Tralli, "Distortion-fair cross-layer resource allocation for scalable video transmission in OFDMA wireless networks," *IEEE Trans. Multimedia*, vol. 16, no. 3, pp. 848–863, 2014.

[25] 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, 2018.

[26] 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.

[27] J. Zhang, S. Chen, X. Mu, and L. Hanzo, "Turbo multi-user detection for OFDM/SDMA systems relying on differential evolution aided iterative channel estimation," *IEEE Trans. Commun.*, vol. 60, no. 6, pp. 1621–1633, 2012.

[28] J. Zhang, S. Chen, X. Mu, and L. Hanzo, "Evolutionary-algorithm-assisted joint channel estimation and turbo multiuser detection/decoding for OFDM/SDMA," *IEEE Trans. Veh. Technol.*, vol. 63, no. 3, pp. 1204–1222, 2014.

[29] D. E. Goldberg and J. H. Holland, "Genetic algorithms and machine learning," *Machine learning*, vol. 3, no. 2, pp. 95–99, 1988.

[30] S. Chen, X. Wang, and C. J. Harris, "Experiments with repeating weighted boosting search for optimization signal processing applications," *IEEE Trans. Syst., Man, Cybern., Part B (Cybern.)*, vol. 35, no. 4, pp. 682–693, 2005.

[31] J. Zhang, S. Chen, X. Mu, and L. Hanzo, "Joint channel estimation and multiuser detection for SDMA/OFDM based on dual repeated weighted boosting search," *IEEE Trans. Veh. Technol.*, vol. 60, no. 7, pp. 3265–3275, 2011.

[32] J. Kennedy, "Particle swarm optimization," *Encyclopedia of machine learning*, pp. 760–766, 2010.

[33] K. Price, R. M. Storn, and J. A. Lampinen, *Differential evolution: a practical approach to global optimization*. Springer Science & Business Media, 2006.

[34] L. Hanzo, J. Akhtman, L. Wang, and M. Jiang, *MIMO-OFDM for LTE, WiFi and WiMAX: Coherent versus non-coherent and cooperative turbo transceivers*. John Wiley & Sons, 2010.

[35] L. Hanzo, M. Münster, B. Choi, and T. Keller, *OFDM and MC-CDMA for broadband multi-user communications, WLANs and broadcasting*. John Wiley & Sons, 2005.

[36] N.-S. Vo, T. Q. Duong, H. D. Tuan, and A. Kortun, "Optimal video streaming in dense 5G networks with D2D communications," *IEEE Access*, vol. 6, pp. 209–223, 2018.

[37] Y. Zhang, S. Qin, and Z. He, "Transmission distortion-optimized unequal loss protection for video transmission over packet erasure channels," in *2011 IEEE Int. Conf. Multimedia Expo*, pp. 1–6, IEEE, 2011.

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

[39] L. Hanzo, "Bandwidth-efficient wireless multimedia communications," *Proc. IEEE*, vol. 86, no. 7, 1998.

[40] W. Webb and R. Steele, "Variable Rate QAM for mobile radio," *IEEE Trans. Commun.*, vol. 43, no. 7, pp. 2223–2230, 1995.

[41] 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 Trans. Circuits Syst. Video Technol.*, vol. 10, no. 8, pp. 1355–1363, 2000.

[42] 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 Trans. Circuits Syst. Video Technol.*, vol. 16, no. 3, pp. 363–374, 2006.

[43] 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.

[44] Y. He, B. Vishwanath, X. Xiu, and Y. Ye, "AHG8: InterDigital's projection format conversion tool," *Document JVET-D0021, Chengdu, China*, 2016.

[45] J. P. Snyder, *Flattening the earth: two thousand years of map projections*. Univ. Chicago Press, 1997.

[46] M. Zhou, "AHG8: A study on compression efficiency of cube projection," *Document JVET-D0022, Chengdu, China*, 2016.

[47] Y. Sun, A. Lu, and L. Yu, "Weighted-to-spherically-uniform quality evaluation for omnidirectional video," *IEEE signal process. Lett.*, vol. 24, no. 9, pp. 1408–1412, 2017.

[48] R. Y. Mesleh, H. Haas, S. Sinanovic, C. W. Ahn, and S. Yun, "Spatial modulation," *IEEE Trans. Veh. Technol.*, vol. 57, no. 4, pp. 2228–2241, 2008.

[49] S. Sugiura, S. Chen, and L. Hanzo, "Coherent and differential space-time shift keying: A dispersion matrix approach," *IEEE Trans. Commun.*, vol. 58, no. 11, pp. 3219–3230, 2010.

[50] 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.

[51] 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.

[52] A. K. Qin, V. L. Huang, and P. N. Suganthan, "Differential evolution algorithm with strategy adaptation for global numerical optimization," *IEEE Trans. Evol. Comput.*, vol. 13, no. 2, pp. 398–417, 2009.

[53] H.-N. Hu, Y.-C. Lin, M.-Y. Liu, H.-T. Cheng, Y.-J. Chang, and M. Sun, "Deep 360 pilot: Learning a deep agent for piloting through 360 sports videos," in *Proc. CVPR*, pp. 1396–1405, 2017.

[54] 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 Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 11, pp. 2693 – 2708, 2018.

[55] L. Hanzo, C. H. Wong, and M.-S. Yee, *Adaptive wireless transceivers: turbo-coded, turbo-equalised and space-time coded TDMA, CDMA and OFDM systems*. John Wiley & Sons, 2002.

[56] L. P. Qian and Y. J. Zhang, "S-MAPEL: Monotonic optimization for non-convex joint power control and scheduling problems," *IEEE Transactions on Wireless Communications*, vol. 9, no. 5, pp. 1708–1719, 2010.

[57] J. Wang, C. Jiang, H. Zhang, Y. Ren, K. Chen, and L. Hanzo, "Thirty years of machine learning: The road to pareto-optimal wireless networks," *IEEE Commun. Surveys Tut.*, pp. 1–1, 2020.