Near-Instantaneously Adaptive Multi-Set Space-Time Shift Keying for UAV-Aided Video Surveillance

Yanqing Zhang, Chao Xu, Senior Member, IEEE, Ibrahim A. Hemadeh, Member, IEEE,
Mohammed El-Hajjar, Senior Member, IEEE, and Lajos Hanzo, Fellow, IEEE

Abstract—**Unmanned Aerial Vehicles (UAV) are more suitable for surveillance systems than their traditional stationary counterparts. To meet the associated demanding video compression requirements, the H.265/High Efficiency Video Coding (HEVC) codec is invoked. Its scalable video coding extension is capable of supporting diverse video resolutions by encoding the video clips captured into multiple layers, namely the Base Layer (BL) and several Enhancement Layers (EL). In this treatise, we propose a multi-functional Multiple-Input Multiple-Output (MIMO) transceiver relying on Multi-Set Space-Time Shift Keying (MS-STSK) for UAV-aided near-instantaneously adaptive layered video streaming, which is capable of promptly varying the throughput as a function of both the estimated channel quality as well as of the UAV mobility, while maintaining near-flawless video quality. Additionally, we invoke EXtrinsic Information Transfer (EXIT) charts for characterizing our Unequal Error Protection (UEP) based video transceiver with the aid of near-capacity channel coding. Some of the perceptually less important ELs may have to be dropped during instances of low channel quality to retain the robustness of the transmitted video stream when a low-throughput MIMO configuration is activated. The simulation results show that our proposed UEP assisted adaptive system is capable of attaining the best Y-Peak SNR (PSNR) performance in comparison to both its non-adaptive and Equal Error Protection (EEP) counterparts.**

Index Terms: Multilayer video transmission, index modulation, unequal error protection, MIMO, adaptive protection, iterative decoding, UAV surveillance.

I. Introduction

The beneficial attributes of Unmanned Aerial Vehicles (UAV) have fuelled their proliferation in a wide range of wireless applications [1]. UAVs can be deployed as aerial base stations to provide additional capacity and reliability for the existing cellular systems in emergency and public safety situations [2]. On the other hand, the UAVs are also capable of playing the role of aerial mobile terminals, while supporting real-time video surveillance and acting as communication relays [3]. UAV-aided surveillance is also capable of gathering tactical information about specific targets. However, in UAVs, the video data transmission characteristics substantially differ from those of manned vehicles. Since UAVs may fly hundreds or even thousands of kilometers away [4], the robustness of the video link becomes crucial for the safety and success of the mission.

Hench *et al.* [5] tackled numerous challenges related to channel estimation, rate adaptation, latency and scalable video streaming for demonstrating the feasibility of UAV-aided real-time channel-adaptive video communications. An adaptive multiscale optimization scheme was advocated by Reyes *et al.* [6] for UAV surveillance, relying on a resource allocation management scheme for dynamically activating the most appropriate mode of operation for the sake of improving the overall system performance. Zhao *et al.* proposed a secure transmission scheme for scalable video streaming [7], where the UAVs are deployed as small-cell base stations in order to improve the coverage and to enhance the bit rate supported.

However, the employment UAVs imposes problems on the existing communication networks, which have not been devised for high-Doppler airborne situations. The 3rd Generation Partnership Porject (3GPP) is investigating the benefits of UAV, both in the Long-Term Evolution (LTE) and in the 5th Generation (5G) networks, since the operational LTE systems are only capable of serving users travelling at a velocity of 500 km/h, but aircraft may reach a velocity of 1000 km/h [8]. More specifically, the conventional training-based Channel State Information (CSI) estimation used for classic coherent detection that assumes slowly fluctuating CSI may suffer from excessive estimation errors in the face of a high Doppler frequency [9]. To circumvent the problem imposed by high mobility, a number of coherent and non-coherent detection techniques have been investigated [10]. For coherent detection, Pilot Symbol Assisted Modulation (PSAM) becomes a favourable choice, where the pilot symbols known to both the transmitter and the receiver are periodically transmitted in order to sample the channel's complex-valued envelope [11] with the aid of numerous interpolation techniques, such as Least Square (LS), or Minimum Mean Square Error (MSE) (MMSE) based interpolation [12]. The PSAM-assisted channel estimation is then extended for estimating multi-antenna scenarios, including Single-Input Multiple-Output (SIMO) [13] and Multiple-Input Multiple-Output (MIMO) [14] schemes. On the other hand, for differentially encoded non-coherent schemes, high-complexity Multiple-Symbol Differential Detection (MSDD) [15] may have to be employed in order to attain an adequate performance.

Furthermore, given the relatively high complexity of video compression and the stringent power supply constraint of UAVs, the processing of collected data is a challenging task [16]. Additionally, as depicted in Figure 1, instead of having a single destination, there may be numerous destinations having different requirements concerning
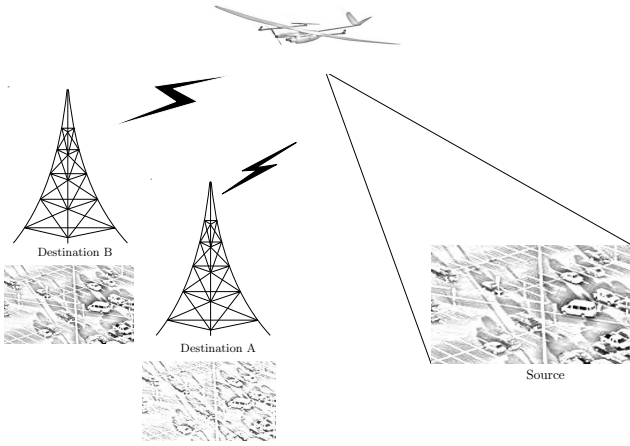
Figure 1: Layered video transmission for UAV surveillance.

the video resolution captured. Satisfying a broad range of technical specifications and user requirements imposes significant challenges on video acquisition, compression and transmission [17].

Over the past few decades, the family of video compression techniques has gradually evolved from H.261 [18] to H.265/High Efficiency Video Coding (HEVC) [19] and ratification of H.266/Versatile Video Coding (VVC) [20] is imminent. Scalable Video Coding (SVC) has attracted particular attention as a potential solution for enhancing the perceptual video quality ever since the H.262 standard was conceived [21]. However, this scalable construction has remained an open challenge due to its limited coding efficiency as well as owing to its considerable complexity, until the H.264 standard was conceived. Explicitly, H.264 finally succeeded in significantly improving the video compression capability attained without dramatically increasing the coding complexity [21]. This scalable technique has been extended to the HEVC standard, also referred to as the Scalability extension of HEVC (SHVC) [22].

Since the decoding of the Enhancement Layers (EL) in SVC hinges on the flawless decoding of its reference layers, this layered coding structure benefits from the concept of Unequal Error Protection (UEP), which allocates different resources to various data sources according to their specific importances [21]. The UEP technique was first proposed by Masnick and Wolf [23], who assigned stronger Forward Error Correction (FEC) to the more important information, while weaker FEC is allocated to the less important data. Since then, numerous UEP schemes have been constructed [24]–[35], as listed in Table I.

Observe in Table I that the majority of UEP schemes for layered video streaming is designed according to layer importance, which assign stronger protection to the more important layer. Furthermore, potent UEP schemes have been conceived with the aid of FEC techniques [24], [26], [30], [32], [34] by allocating more robust protection to the more important video bits, such as those of the picture-start code. For example, Marx and Farah [24] minimized the mean video distortion using different-rate punctured Turbo Codes (TCs), where the redundancy imposed on the compressed stream is non-uniformly distributed across the consecutive video frames. Fang *et al.* conceived a Genetic Algorithm (GA) assisted optimal FEC code-rate allocation scheme for SVC streaming. Moreover, Huo *et al.* [32] proposed a Recursive Systematic Convolutional (RSC) coded interlayer (IL) operation-aided FEC (IL-FEC) technique that implants the bits of the Base Layer (BL) into the ELs with the aid of taking their modulo-2 addition, where iterative decoding is invoked for exchanging extrinsic information between the two layers, when the BL is not successfully decoded. This IL-FEC was then enhanced by the authors of [35] with the aid of providing UEP for the EL. Additionally, UEP schemes were also designed by using bit mapping scheme [25], [28], [35]. Specifically, Chang *et al.* [25] mapped the important source bits to the specific Hierarchical Quadrature Amplitude Modulation (QAM) (HQAM) bits exhibiting the highest Euclidean distance for UEP.

Let us now briefly focus our attention on the evolution of advanced wireless solutions, with an emphasis on the development of MIMO techniques conceived for establishing reliable high-rate links [36]. More explicitly, MIMO techniques are capable of enhancing the multiplexing gain by invoking for example the Bell Laboratories Layered Space-Time (BLAST) architecture [37], or improving the diversity gain by employing Space-Time Block Codes (STBC) [38]. A combination of both gains can be attained by employing multi-functional MIMOs, such as Layered Steered Space-Time Coding (LSSTC) [39] that combines the benefits of Vertical BLAST (V-BLAST) and STBC techniques. The Spatial Modulation (SM) concept advocated in [40] is capable of providing a high normalized throughput at the expense of low complexity. Since only a single antenna is activated, which is selected from multiple antennas, only a single Radio Frequency (RF) chain is required. A multi-functional MIMO arrangement referred to as Space-Time Shift Keying (STSK) was proposed in [41], 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. Recently, as a scalable extension of STSK, the novel concept of Mult-Set Space-Time Shift-Keying (MS-STSK) was proposed [42], which achieves both higher data rates and a lower Bit Error Ratio (BER) than its conventional STSK counterpart. However, this is achieved at the expense of needing more Transmit Antennas (TA)s and by imposing additional complexity during the antenna-switching operation.

Against this background, in this paper we conceive a radically new MS-STSK-assisted UEP adaptive system for transmitting the SVC bit streams of the family of fixed-wing UAV-aided surveillance systems. The MS-STSK transceiver is configured to be near-instantaneously adaptive in order to transmit different number of layered video streams according to the prevalent channel Signal-to-Noise Ratio (SNR) as well as to the mobility, while taking into account both the effect of shadowing and of the Doppler spread. The video bits of different sub-layers are mapped to the different-integrity MS-STSK

Table I: Major contributions on UEP for video communications.

Contributions	[24]	[25]	[26]	[27]	[28]	[29]	[30]	[31]	[32]	[33]	[34]	[35]	Ours
Variable FEC rate	✓		✓				✓		✓		✓		
IL-FEC						✓			✓			✓	
Bit mapping		✓			✓							✓	✓
Iterative decoding							✓		✓				✓
Layer based				✓		✓		✓		✓	✓	✓	✓
Frame based	✓	✓	✓		✓		✓		✓				
CSI estimation				✓									✓
Antenna selection				✓							✓		
Layer extraction											✓		✓
Power control					✓			✓		✓			
Mobility effect				✓		✓				✓			✓

subchannels for achieving carefully matched UEP. The novel contributions of this paper are:

1) A powerful EXtrinsic Information Transfer (EXIT)-chart-aided multimedia design procedure is conceived and its benefits are demonstrated in the context of a sophisticated SVC-MS-STSK-aided UAV assisted adaptive video streaming system. Our design procedure is however applicable to a broad class next-generation of multimedia systems.

2) **More explicitly, the proposed system is specifically designed taking into account various UAV mobility scenarios, including different Doppler spreads and shadowing effects, which determine the prerecorded thresholds used for the mode-switching operation.**

3) **We design the source-bit mapping scheme for our MS-STSK transceiver, where the video bits exhibiting different importance are mapped to different-integrity MS-STSK subchannels in order to better protect the more important BL and ELs.**

4) **As an evolution from [35], the proposed video system exploits the near-capacity UEP capability of the sophisticated MS-STSK scheme with the aid of EXIT charts. Additionally, a flexible layer extraction scheme is conceived, where the EL may have to be dropped during instances of low channel quality, whilst maintaining error-free but reduced-PSNR video quality, when a low-throughput MIMO configuration is activated.**

5) Explicitly, we conceive an SVC streaming scheme by jointly designing the video bit mapping, adaptive layer extraction and iterative decoding schemes, as shown in Table I, intrinsically amalgamated with adaptive FEC.

The proposed system is also applicable for terrestrial mobile scenario, where the system parameters may be near-instantaneously updated. The rest of the paper is organized as follows. The system structure is introduced in Section II, where we briefly discuss the SVC-based MS-STSK video transceiver relying on near-instantaneous adaptivity. Our PSAM-assisted MS-STSK technique is portrayed in Section III. In Section IV, we quantify the iterative decoding gain of the MS-STSK transceiver with the aid of EXIT charts and conceive a sophisticated UEP-aided MS-STSK-assisted layered video streaming arrangement. Our simulation results are provided in Section V, while Section VI offers our conclusions.

II. SYSTEM MODEL

We propose an MS-STSK-assisted near-instantaneously adaptive video streaming system for UAV-aided surveillance over an Air-to-Ground (AG) channel. The video captured is transmitted to the destination taking into account the video resolution required, the estimated CSI and the UAV's 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 bit rates. In the following, we will commence by introducing the MS-STSK transceiver, followed by the transmitter and receiver models.

A. MS-STSK

The structure of MS-STSK is depicted in Figure 2, where it can be seen that the MS-STSK transceiver consists of three components, namely the Antenna Selection Unit (ASU), the classic L-QAM/PSK modulator and the dispersion matrix generator. We will show later in Figure 4 of this section that the different MS-STSK subchannels of Figure 2 are capable of yielding different BER performances, hence they have the inherent potential to provide UEP. The MS-STSK system [42] is denoted as $MS\text{-}STSK(N_t, N_r, M, T, Q, L)|_{PSK/QAM}$, which indicates that there are N_t TAs, N_r Receive Antennas (RA), M RF chains, T time slots, Q dispersion matrices and finally, an L-QAM/PSK modulator. Let us first consider the STSK block of Figure 2, which consists of a dispersion matrix generator that generates Q dispersion matrices and the

Figure 2: Structure of MS-STSK [42].

classic L-QAM/PSK modulator, outputting $\mathbf{A}_q \in \mathbb{C}^{M \times T}$ selected from the set $\{\mathbf{A}_q\}_{q=1}^{Q}$ satisfying the power constraint of $\{tr(\mathbf{A}_q^H \mathbf{A}_q) = T\}_{\forall q}$ and $s_M \in S_l$, respectively. Hence, the output of the STSK block can be expressed as:

$$\{\mathbf{X} \in \mathbb{C}^{M \times T}\} = \mathbf{A}_q s_M = [\mathbf{x}_1 ... \mathbf{x}_m ... \mathbf{x}_M]^T, \qquad (1)$$

where $\mathbf{x}_m \in \mathbb{C}^{1 \times T}$ is the m-th row of $\mathbf{X}$. The ASU of Figure 2 then selects M out of a total of N_t TAs for conveying the associated STSK-encoded data matrix $\mathbf{X}$ according to the input bits fed into the ASU block of Figure 2, finally forming the MS-STSK symbol block of $\tilde{\mathbf{X}} \in \mathbb{C}^{N_t \times T}$.
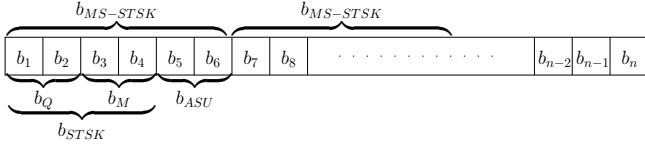
Figure 3: Bit structure of MS-STSK [35].

Figure 3 illustrates the specific mapping of the bits to the MS-STSK codeword. In the example of Figure 3, the input bit stream is partitioned into $b_{MS\text{-}STSK} = 6$-Bit Per Channel Use (BPCU) that contain $b_Q = 2$ bits, $b_M = 2$ bits and $b_{ASU} = 2$ bits, in which b_Q and b_M form b_{STSK}. In Figure 4, we plot the BER of different blocks of the MS-STSK transceiver for transmission over uncorrelated non-dispersive Rayleigh fading channels using the configurations of $MS\text{-}STSK(16,2,2,2,8,8)|_{QPSK}$, $MS\text{-}STSK(4,2,2,2,8,4)|_{QAM}$ and $MS\text{-}STSK(8,2,2,2,4,4)|_{QAM}$. Obverse in Figure 4 that the BER performance of the different bits conveyed by the three MS-STSK subchannels is different [35]. Generally, the overall performance of b_{ASU}s is better than that of b_Ms, while that of b_Qs exhibits the worst BER performance. Hence, Figure 4 demonstrates that the MS-STSK transceiver is capable of providing UEP by mapping the different-sensitivity video bits into different MS-STSK blocks representing the subchannels, as it will be further discussed in Section IV.
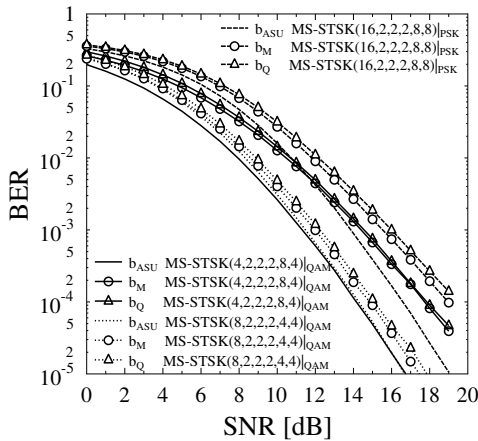
Figure 4: BER performances of different blocks of MS-STSK over uncorrelated non-dispersive Rayleigh channel [35].

B. Transmitter Model

Based on the above different-integrity subchannels, let us now detail the transmitter of our proposed adaptive system, where the UAV captures and transmits the video signals to the ground stations. Three scalable video layers are taken into account in our design example, but the number of the video sub-layers can be readily customized according to the specific requirements of the system.
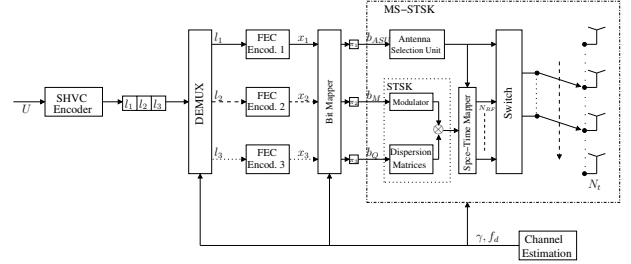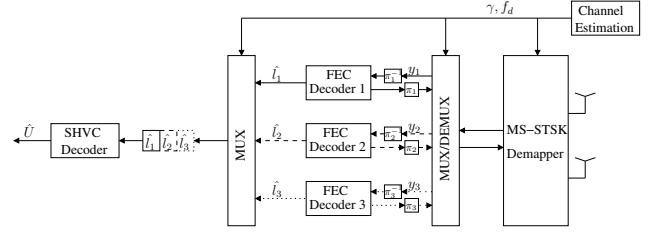

Figure 5: The transmitter block diagram of the proposed MS-STSK aided adaptive system for scalable video streaming.

Figure 5 depicts the transmitter architecture of our proposed MS-STSK-assisted adaptive system conceived for layered video streaming. It can be seen from Figure 5 that the captured video U 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:

$$Mode = \begin{cases} Mode\ 1 & \gamma_{f_d} \leq t_{1,f_d} \\ Mode\ 2 & t_{1,f_d} < \gamma_{f_d} \leq t_{2,f_d} \\ Mode\ 3 & t_{2,f_d} < \gamma_{f_d}, \end{cases} \qquad (2)$$

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 Packet Loss Ratio (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:

$$\begin{cases} P(l_1) & i = 1 \\ \sum\limits_{m=2}^{i} P(l_m) \prod\limits_{n=1}^{m-1} [1 - P(l_n)] + P(l_1) & i > 1. \end{cases} \quad (3)$$

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% [43], 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 $b_{MS-STSK} = 2$ 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 5 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 5 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 $b_{MS-STSK} = 5$ bits for transmitting the additional bit stream of the first EL l_2. In this mode, the FEC Encoders 1 and 2 of Figure 5 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 5 utilizes the bits that are capable of exhibiting the best BER performance in the MS-STSK codeword of Figure 3 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 $b_{MS-STSK} = 10$ 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 5 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 5 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 IV. The modulated bits are then transmitted by the MS-STSK transceiver over the AG channel to be detailed in Section III-A. 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.

C. Receiver Model

Figure 6: The receiver block diagram of the proposed MS-STSK aided adaptive system for scalable video streaming.

Figure 6 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 6 for generating the soft Log-Likelihood Ratio (LLR) output of the MS-STSK scheme. Then, the MUX/DEMUX of Figure 6 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 6.

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 6 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 6, which are then forwarded to the SHVC Decoder of Figure 6 in order to reconstruct the video.

III. PSAM FOR MS-STSK

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.

A. Air-to-Ground Channel Model

The 3GPP is currently developing both the LTE Advanced Pro and the 5G networks, making them suitable for supporting Vehicle-to-Everything (V2X) [44] and for UAV scenarios [45], including the UAV and Base Station (BS) heights, Angle of Arrival (AoA), Angle of Departure (AoD), Spread of Arrival (ASA) and 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 [45] 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 [46]:

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 [47]:

$$\mathbf{H}_n = \mathbf{H}_n^D + \mathbf{H}_n^S, \tag{4}$$

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 $\Delta f_{LOS} = f_d \cos(\Delta f_{offset}) \leq f_d$ [46]. 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:

$$\mathbf{a}_t = [1, e^{j2\pi dcos(\phi_t)}, ..., e^{j2\pi d(N_t-1)cos(\phi_t)}]^T, \tag{5}$$

and

$$\mathbf{a}_r = [1, e^{j2\pi dcos(\phi_r)}, ..., e^{j2\pi d(N_r-1)cos(\phi_r)}]^T, \tag{6}$$

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 $E\{\mathbf{H}_{n+k}^S(\mathbf{H}_n^S)^H\} = \frac{N}{K+1}J_0(2\pi f_d k)\mathbf{I}_M$ [46]. This autocorrelation matrix is eminently suitable for PSAM-assisted channel estimation, as described in the next section.

B. PSAM for MS-STSK

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.

1) Preliminaries of PSAM: The PSAM frame-structure proposed by Cavers in [11] is portrayed in Figure 7, which is suitable for a Single-In Single-Out (SISO) fading channel [46]. We denote the j-th data symbol of the i-th PSAM frame by $\{s_{i,j}\}_{j=1}^{N_{PS-1}}$, which is identical to the

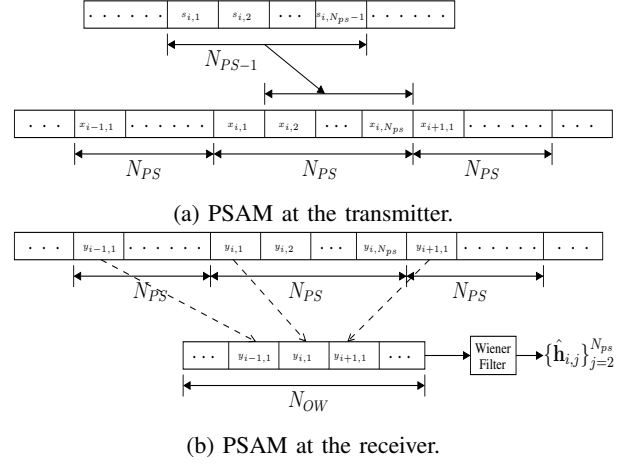(a) PSAM at the transmitter.

(b) PSAM at the receiver.

Figure 7: The frame structure of PSAM.

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 7(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:

$$y_{i,j} = x_{i,j}h_{i,j} + v_{i,j}, \tag{7}$$

with v representing the Additive White Gaussian Nosie (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 7(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 the channel. The multiplicative channel gain of each detected symbol is derived by interpolation, yielding:

$$\{\hat{h}_{i,j}\}_{j=2}^{N_{PS}} = \sum_{\bar{i}=-N_{OW}^a}^{N_{OW}^b} w_{\bar{i},j}y_{i+\bar{i},1}/x_{i+\bar{i},1}$$

$$= \sum_{\bar{i}=1}^{N_{OW}} w_{\bar{i},j}y_{i-N_{OW}^a-1+\bar{i},1}/x_{i-N_{OW}^a-1+\bar{i},1}$$

$$= \mathbf{w}_j^T(\mathbf{X}_i^P)^H\mathbf{y}_i^P, \tag{8}$$

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 $h_{i,j}$ [46]. The corresponding MSE cost function may be formulated as:

$$\sigma_{MSE}^2 = E\{\|h_{i,j} - \hat{h}_{i,j}\|^2\} = 1 - 2\mathbf{w}_j^T\tilde{\mathbf{e}} + \mathbf{w}_j^T\tilde{\mathbf{C}}\mathbf{w}_j^*. \tag{9}$$

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

$$\tilde{\mathbf{C}} = \begin{bmatrix} \psi[0]+N_0 & \psi[-1] & ... & \psi[-(N_{OW}-1)] \\ \psi[1] & \psi[0]+N_0 & ... & \psi[-(N_{OW}-2)] \\ \psi[2] & \psi[1] & ... & \psi[-(N_{OW}-3)] \\ \vdots & \vdots & \ddots & \vdots \\ \psi[-(N_{OW}-1)] & \psi[-(N_{OW}-2)] & ... & \psi[0]+N_0 \end{bmatrix}, \quad (10)$$

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

$$\tilde{\mathbf{e}} = [\psi[-N_{OW}^a N_{PS} - j + 1], \psi[-(N_{OW}^a - 1)N_{PS} - j + 1], ...,$$
$$\psi[-j + 1], ..., \psi[N_{OW}^b N_{PS} - j + 1]]^T. \quad (11)$$

Therefore, by invoking the 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:

$$\mathbf{w}_j^* = \tilde{\mathbf{C}}^{-1}\tilde{\mathbf{e}}^T. \quad (12)$$

Hence, the channel estimates $\{\hat{h}_{i,j}\}_{j=2}^{N_{PS}}$ can be acquired by invoking Eq. (8).

2) 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 [46] [48] [49]. 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:

$$\mathbf{Y}_{i,j} = \mathbf{H}_{i,j}\bar{\mathbf{X}}_{i,j} + \mathbf{V}_{i,j}, \quad (13)$$

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:

$$\{\hat{\mathbf{H}}_{i,j}\}_{j=2}^{N_{PS}} = \sum_{\bar{i}=-N_{OW}^a}^{N_{OW}^b} \mathbf{W}_{\bar{i},j}\bar{\mathbf{X}}_{i+\bar{i}}^H\mathbf{Y}_{i+\bar{i},1} = \mathbf{W}_j^T(\bar{\mathbf{X}}_i^P)^H\mathbf{Y}_i^P.$$
$$(14)$$

The filter taps $\{\mathbf{W}_{\bar{i},j}\}_{\bar{i}=1}^{N_{ow}} = [\mathbf{W}_{1,j}^T, \mathbf{W}_{2,j}^T, ..., \mathbf{W}_{N_{ow},j}^T]$ then become $(N_t \times N_t)$-element matrices. The pilots of the MS-STSK transceiver become $\bar{\mathbf{X}}_i^P = \text{diag}\{\bar{\mathbf{X}}_{i-N_{OW}^a,1}, ..., \bar{\mathbf{X}}_{i,1}, ..., \bar{\mathbf{X}}_{i+N_{OW}^b,1}\}$, 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:

$$\sigma_{MSE}^2 = E\{\|\mathbf{H}_{i,j} - \hat{\mathbf{H}}_{i,j}\|^2\}/(N_t N_r)$$
$$= 1 - 2\text{tr}(\mathbf{W}_j^T\tilde{\mathbf{e}}) + \text{tr}(\mathbf{W}_j^T\tilde{\mathbf{C}}\mathbf{W}_j^*). \quad (15)$$

Substituting $\psi[k]$ of Eq. (10) by the $(N_t \times N_t)$-element matrix $\{\Psi[k] = \frac{K}{N_t N_r(K+1)}e^{(j2\pi\Delta_{LOS}kN_{PS})}\mathbf{R}_{AA} +$

$\frac{1}{N_t(K+1)}J_0(|2\pi f_d k N_{PS}|)\mathbf{I}_{N_t}\}_{k=-(N_{ow}-1)}^{N_{ow}-1}$, where $J_0(\cdot)$ is the zero-order Bessel function of the first kind, the resultant MMSE solution is given by [46]:

$$\mathbf{W}_t^* = \tilde{\mathbf{C}}^{-1}\tilde{\mathbf{e}}. \quad (16)$$

Therefore, the channel matrices $\{\hat{\mathbf{H}}_{i,j}\}_{j=2}^{N_{PS}}$ of the MS-STSK symbol blocks can be obtained by substituting Eq. (16) into Eq. (14).

IV. EXIT-CHART ASSISTED UEP FOR CODED MS-STSK

The BER performance of the uncoded MS-STSK system was shown in Figure 4 for three different MS-STSK configurations, which were distinguished by continuous, dashed and dotted lines, respectively. Within each of these three categories, we have the ASU bits, the dispersion-matrix index bits and the classic QAM bits. All the ASU index bits exhibit the best error-sensitivity. By contrast, the b_Q-bits and b_M-bits of the Q dispersion matrix indices and of the QAM-bits defined in Figure 5 tend to have different error-sensitivity. Hence, we have to deal with the bit-sensitivities on a bit-level basis by providing stronger protection for the more important video bits.

This UEP regime could be designed by simply running BER versus SNR simulations for each individual bit using channel coding. Naturally, the BER versus SNR curves would be different for each FEC code. Furthermore, we have to match the number of bits in each EL to the throughput of each MS-STSK bit stream. We solve this problem by assuming the employment of near-capacity channel-coding, so that our design becomes valid regardless of the specific choice of the channel coding and determine the mapping of video bits to the MS-STSK bits using EXIT chart analysis.
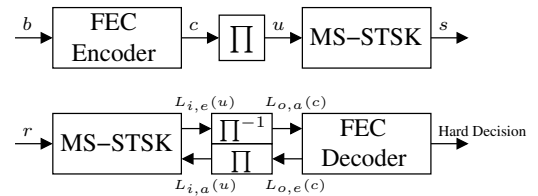

Figure 8: Block diagram of a two-stage iteratively decoded system using MS-STSK and FEC.

Figure 8 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 8 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 8 indicate *a priori* and *a posteriori* information, respectively. The received signal stream r of Figure 8 is demapped first by the MS-STSK decoder with the aid of the Logarithmic Maximum *A Posteriori* (Log-MAP) algorithm to its LLR representation,
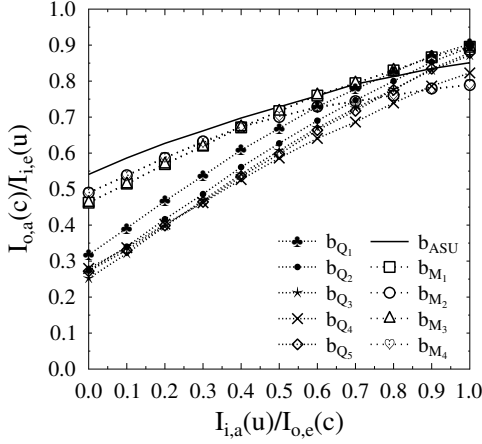
Figure 9: The *MS-STSK(4,2,2,2,32,16)|$_{QAM}$* scheme's MI curve for transmission over Ricean fading channels ($K = 0$ dB) at the channel SNR of 6 dB.

namely to $L_{i,e}(u)$, which is then deinterleaved by the soft-bit deinterleaver block Π^{-1} of Figure 8 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 8, 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 MI I is used for quantifying the relationship between two simultaneously-sampled random variables [50].

Briefly, EXIT charts constitute powerful tools of visualising the exchange of extrinsic information between the concatenated decoders of Figure 8, which are capable of accurately predicting the convergence behaviour of the iterative receiver based on the MI exchanged between its components [51] [52]. Let us now embark on designing EXIT-chart-aided the source-bit to MS-STSK-bit allocation of the Bit Mapper unit of Figure 5 for *Mode* 2 and *Mode* 3 of Section II-B 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.

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

(b) *MS-STSK(4,2,2,2,32,16)|$_{QAM}$*.
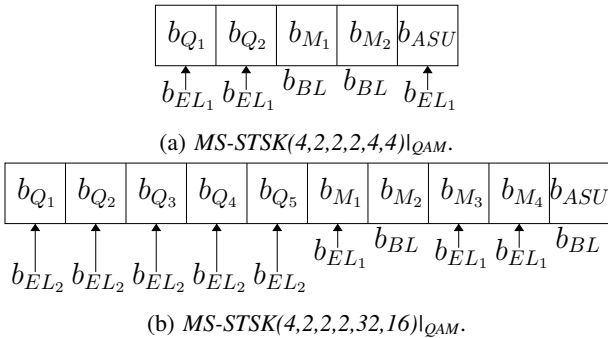
Figure 10: Bit mapping allocation.

2) Set up the system model and testify the system performance via Monte-Carlo experiments, as illustrated in Figure 9.
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 5 according to the estimated $\mathcal{A}$ of MS-STSK bits, as shown in Figure 10.

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 [53] [54]. 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 Figure 5 for *Mode* 2 and *Mode* 3 are given in Figure 10(a) and 10(b), respectively.

V. 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.

The parameters of the video clip are listed in Table II. A video clip represented in (352×288)-pixel Common Intermediate Format (CIF)[1] 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 Frame Per Second (FPS) and the Group Of Pictures (GOP) interval was set to 8 for all video simulations, which means that the Instantaneous Decoding Refresh (IDR)/Clean Random Access (CRA) [19] frames are inserted every 8 frames. No Bi-directionally predicted (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 Intra-coded (I) frames and Predicted (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 decodability 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 II 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 II, which is punctured to half rate. The number of inner iterations of the TCs between the inner

---

[1]In this treatise, we focus our attention on designing the MS-STSK assisted UEP system aiming for find the most appropriate MS-STSK bit mapping scheme for robust video streaming, regardless of the video-resolution. Naturally, our specific mapping scheme only relies on the MS-STSK configurations, which is applicable to a variety of video-resolutions.

Table II: 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-Peak Signal-to-Noise Ratio (PSNR) (dB)	33.45, 36.27, 39.6
Bitrate (kbps)	239, 332, 654
Error Concealment	Frame-Copy
Channel Coding	TCs [111 101]
TCs Inner Iteration	8
Outer Iteration	2
Channel	Ricean (K = 0 dB)
Simulation Repeated	200

RSCs is fixed to 8. Recall that the channel model was given in Section III-A, 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 III. 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.

A. Complexity Analysis

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 5. We partition the bit-mapping procedure into offline preparation and online operation. As part of offline design, we illustrate the preparation required for designing the mapping scheme, while the online operation carries out the bit-mapping for achieving UEP.
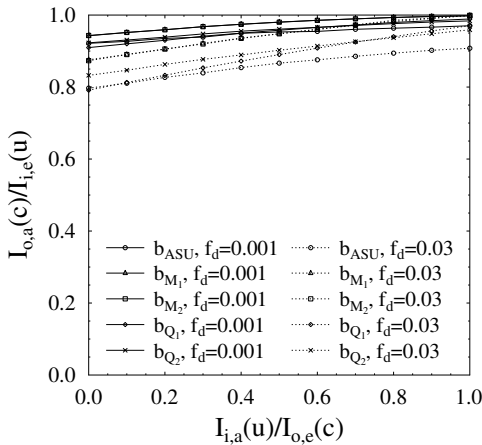
Figure 11: MI performance of our *MS-STSK(4,2,2,2,4,4)|$_{QAM}$* under two different normalized Doppler frequencies, when $K = 0$ dB at a channel SNR of 7 dB.
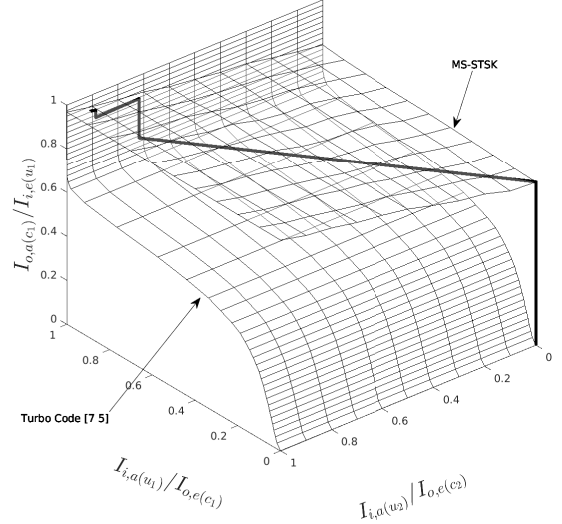

Figure 12: The trajectory of the BL between the MS-STSK configured with *MS-STSK(4,2,2,2,4,4)|$_{QAM}$* and the half rate TCs [7 5] for $Mode$ 2 at the channel SNR of 3 dB.

To evaluate the BER of the individual MS-STSK bits of Figure 3, an empirical Monte Carlo simulation is employed using the parameters of Table III, used in our simulations before carrying on the simulations. Then, the MI of all the MS-STSK bits of Figure 3 is plotted in Figure 9, which is then exploited for designing the bit-mapping scheme of Figure 10. Specifically, Section IV 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 TCs and Low Density Parity Check (LDPC) codes.

The online complexity imposed by our proposed mapping design is dominated by the Bit Mapper block of Figure 5. The FEC encoded bits are fed into the Bit Mapper block, which rearranges the bits according to our mapping guideline of Figure 10, imposing negligible online complexity. Explicitly, the complexity of this block is similar to that of the classic bit mapping scheme of HQAM shown in [25].

B. Performance Analysis

Based on the AG model of Section III-A, 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 $\phi = \phi_0 = \phi_r = \phi_t \in [-\pi, \pi]$ [46]. These values remain unchanged for $N_{PS}N_{OW}$ number of MS-STSK symbol durations. Figure 11 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 per 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
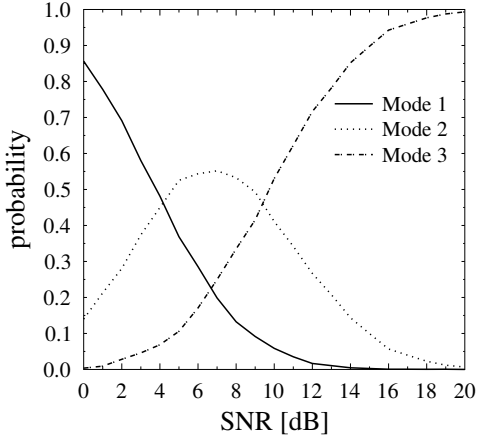
(a) PDF of the three operating modes versus channel SNR for the mobility Scenario 1 of Table IV.
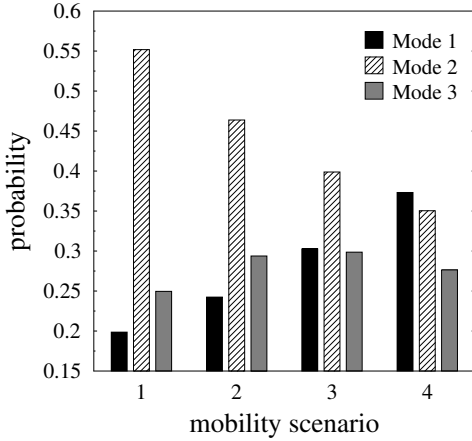
(b) PDF of the three operating modes versus the mobility Scenarios at SNR of 7 dB.

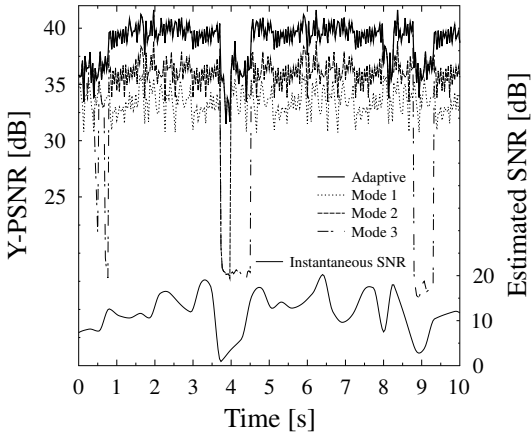Figure 13: PDF of the three fixed modes of operation.

Figure 14: 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.

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 12 portrays the 3D EXIT chart [55] based iterative decoding trajectory of the l_1 between the MS-STSK transceiver and the half rate TCs when $Mode$ 2 is selected, where only two video sub-layers, namely l_1 and l_2, are transmitted, as shown in Figure 10(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 10(a), respectively. The MS-STSK plane of $I_{i,e(u_1)}$ in Figure 12 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 12 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 12 exhibits the MI transfer between the TCs decoder and the MS-STSK transceiver for l_1. In Figure 12, 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 TCs 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 TCs 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 12 that the improvement after the second outer iteration becomes fairly minor. Therefore, the number of outer iterations between the TCs and the MS-STSK transceiver is set to 2, based on our observation of Figure 12, namely that the improvements gleaned by more outer iterative decoding operations become negligible.

Table IV 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 [3] [56]. 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 ac-

Table III: 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)$|_{PSK}$	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

Table IV: Parameter sets for different mobilities.

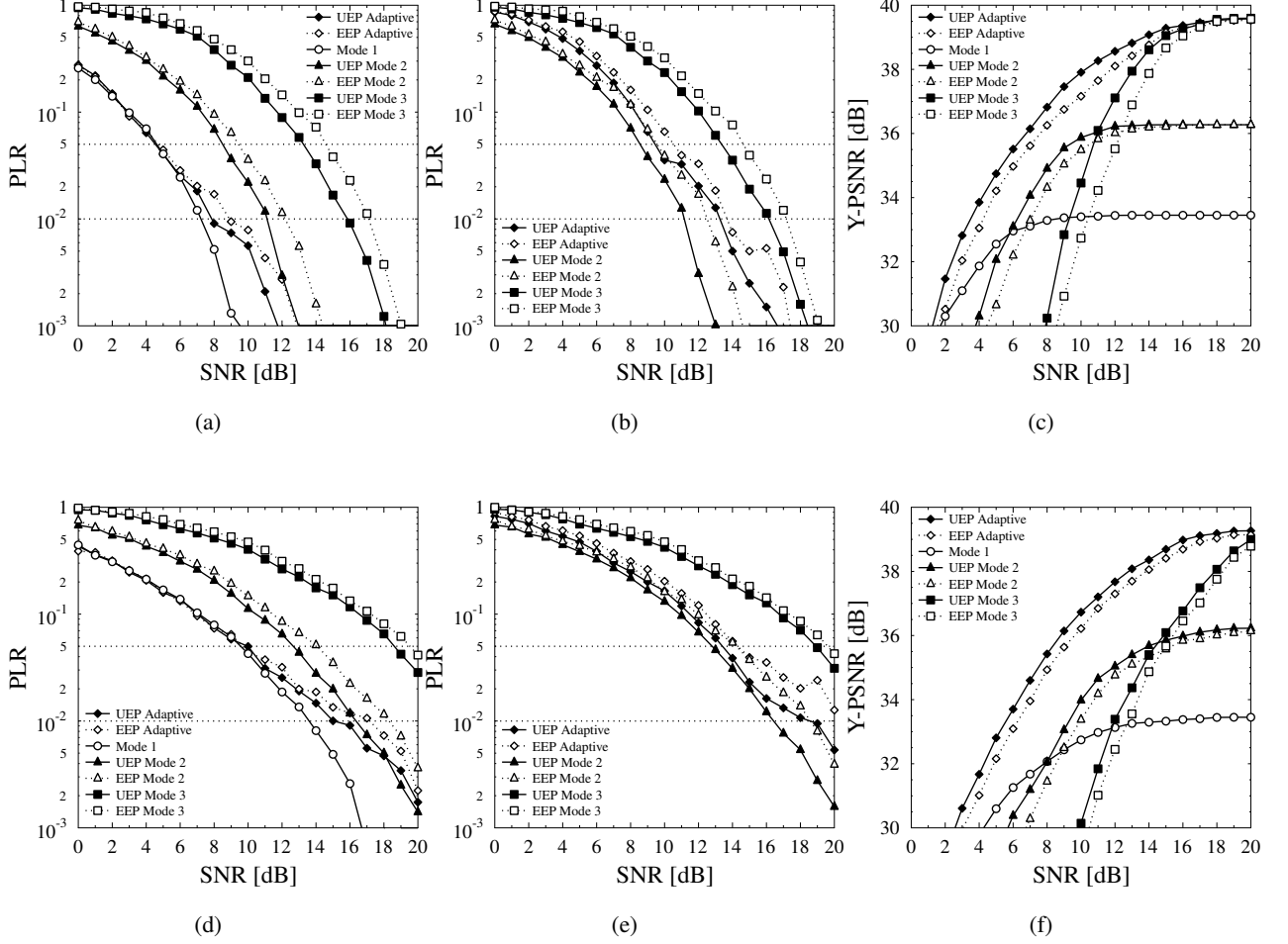| Mobility Scenario	1	2	3	4
f_d	0.001	0.01	0.02	0.03
σ_{sf} (dB)	4	5	6	7
$t_{1_{UEP}}$ (dB)	3.3	3.5	3.9	4.5
$t_{1_{EEP}}$ (dB)	4	4.3	4.8	5.2
$t_{2_{UEP}}$ (dB)	9.2	9.5	9.8	10.3
$t_{2_{EEP}}$ (dB)	10.2	10.5	11	11.4

Figure 15: The comparison between the three fixed modes and the adaptive system, where the first, second and third column represent the PLR value of the l_1, the e-PLR value of l_2 and the image quality (PSNR), respectively, for mobility Scenario 1 (first row) and 3 (second row).

curacy 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 III 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 Equal Error Protection (EEP) based adaptive MS-STSK counterpart is considered as the benchmarker, where the bit-to-MS-STSK subchannel mapping of Figure 10 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 IV, requiring an additional 1.1 dB of channel SNR for switching to the high-throughput video modes.

Figure 13(a) depicts the Probability Density Function (PDF) of the three fixed modes versus the channel SNR for the thresholds listed in Table IV. Observe in Figure 13(a) that for the mobility Scenario 1 described in Table IV,

$Mode$ 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 frequent mode. To elaborate a little further, Figure 13(b) shows that upon increasing the mobility, the robust $Mode$ 1 becomes more dominant, when aiming for guaranteeing the best possible performance of the system.

In the time interval shown in Figure 14, 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 15 in more detail.

To elaborate, Figures 15(a), (b), (d) and (e) show the simulation results in terms of the PLR, while Figure 15(c) 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
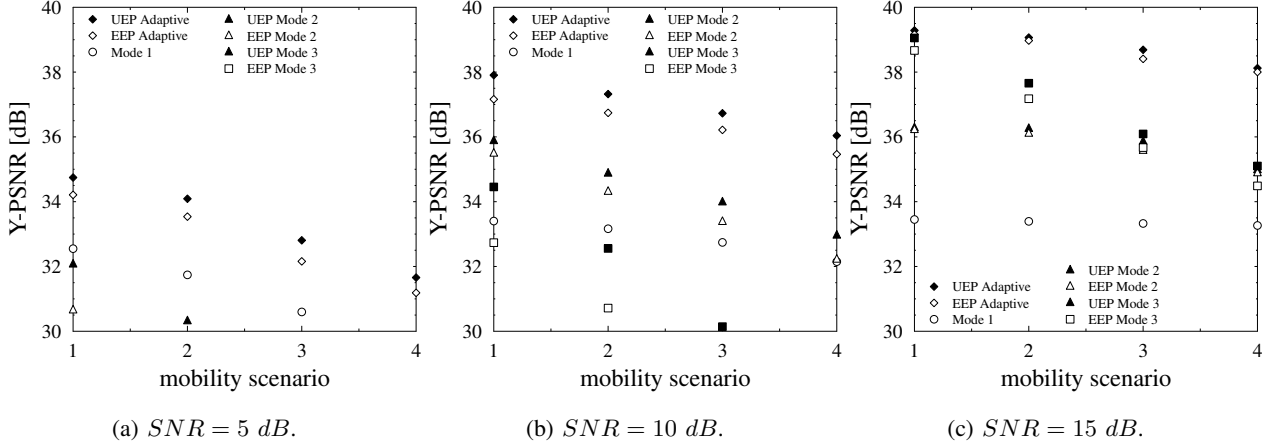
(a) $SNR = 5\ dB$. (b) $SNR = 10\ dB$. (c) $SNR = 15\ dB$.

Figure 16: Image quality (Y-PSNR) versus mobility scenarios for the EEP and UEP schemes.

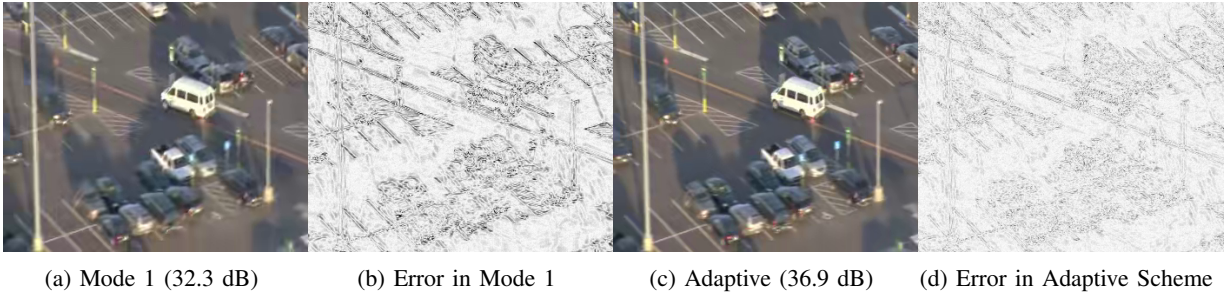(a) Mode 1 (32.3 dB) (b) Error in Mode 1 (c) Adaptive (36.9 dB) (d) Error in Adaptive Scheme

Figure 17: The subjective image comparison at the instantaneous channel SNR of 4 dB

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)(b)(c)) and the second row ((d)(e)(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 15 that the system performance of mobility Scenario 3, shown in the second row of Figure 15 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 significantly degrading the overall system performance.

Figure 15 shows that our proposed UEP allocation methods of Figure 10 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 15(a) and (d) 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 13. 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 15(b) and (e), which represents the equivalent PLR of l_2 by taking into account its reference layer, namely l_1, as well. Observe in Figures 15(b) and (e) 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 15(b) and (e) 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 15(c) 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 15(c) 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.

Additionally, Figure 15 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 Figure 15(c) and (f). The gain of the image quality (Y-PSNR) improved by the adaptive system becomes more distinct in the mo-

bility Scenario 3, where the instantaneous channel quality fluctuates more dramatically due to the higher shadow fading.

Figure 16 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 16 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 10 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 17. Explicitly, observe in Figure 17 that the adaptive system is capable of switching to the higher-throughput, higher-PSNR modes, hence exhibiting a more detailed video texture in Figure 17(c) than that of its Mode 1 counterpart in Figure 17(a). To make the visual comparison more explicit, in Figure 17(b) and (d) we portrayed the error between the original and received video frames, where the latter exhibits a lower error.

VI. CONCLUSIONS

In this treatise, we designed an adaptive MS-STSK system for surveillance UAVs operating under various mobility scenarios, where the PSAM technique is employed by the MS-STSK transceiver for estimating the instantaneous channel SNR. We devised the video-to-MS-STSK mapping of Figure 10 with the aid of EXIT charts for improving both the attainable PLR and image quality (Y-PSNR) performances. The simulation results showed that the proposed UEP-assisted adaptive scheme is always capable of providing better reconstructed video quality than any of its counterparts, where the improvements introduced by the adaptive system become distinct in high-mobility scenarios compared to the corresponding fixed counterparts.

REFERENCES

[1] E. W. Frew and T. X. Brown, "Airborne communication networks for small unmanned aircraft systems," *Proc. IEEE*, vol. 96, no. 12, pp. 2008–2027, 2008.

[2] Y. Zeng, R. Zhang, and T. J. Lim, "Wireless communications with unmanned aerial vehicles: opportunities and challenges," *arXiv preprint arXiv:1602.03602*, vol. 54, no. 5, pp. 36–42, 2016.

[3] 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 Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2361–2391, 2019.

[4] M. Mozaffari, W. Saad, M. Bennis, Y.-H. Nam, and M. Debbah, "A tutorial on UAVs for wireless networks: Applications, challenges, and open problems," *arXiv preprint arXiv:1803.00680*, 2018.

[5] D. L. Hench, P. N. Topiwala, and Z. Xiong, "Channel adaptive video compression for unmanned aerial vehicles (UAVs)," in *Applications of Digital Image Process. XXVII*, vol. 5558, pp. 475–485, International Society for Optics and Photonics, 2004.

[6] I. O. Reyes, P. A. Beling, and B. M. Horowitz, "Adaptive multiscale optimization: Concept and case study on simulated UAV surveillance operations," *IEEE Syst. J.*, vol. 11, no. 4, pp. 1947–1958, 2017.

[7] N. Zhao, F. Cheng, F. R. Yu, J. Tang, Y. Chen, G. Gui, and H. Sari, "Caching UAV assisted secure transmission in hyper-dense networks based on interference alignment," *IEEE Trans. Commun.*, vol. 66, no. 5, pp. 2281–2294, 2018.

[8] J. Wu and P. Fan, "A survey on high mobility wireless communications: Challenges, opportunities and solutions," *IEEE Access*, vol. 4, pp. 450–476, 2016.

[9] 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.

[10] M. El-Hajjar and L. Hanzo, "Dispensing with Channel Estimation...," *IEEE Veh. Technol. Mag.*, vol. 5, no. 2, pp. 42–48, 2010.

[11] J. K. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels (mobile radio)," *IEEE Trans. Veh. Technol.*, vol. 40, no. 4, pp. 686–693, 1991.

[12] C. Xu, X. Zuo, S. X. Ng, R. G. Maunder, and L. Hanzo, "Reduced-complexity soft-decision multiple-symbol differential sphere detection," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3275–3289, 2015.

[13] G. E. Oien, H. Holm, and K. J. Hole, "Impact of channel prediction on adaptive coded modulation performance in Rayleigh fading," *IEEE Trans. Veh. Technol.*, vol. 53, no. 3, pp. 758–769, 2004.

[14] S. Zhou and G. B. Giannakis, "How accurate channel prediction needs to be for transmit-beamforming with adaptive modulation over Rayleigh MIMO channels?," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1285–1294, 2004.

[15] D. Divsalar and M. K. Simon, "Multiple-symbol differential detection of MPSK," *IEEE Trans. Commun.*, vol. 38, no. 3, pp. 300–308, 1990.

[16] N. H. Motlagh, M. Bagaa, and T. Taleb, "UAV-based IoT platform: A crowd surveillance use case," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 128–134, 2017.

[17] R. Beard, D. Kingston, T. W. McLain, and D. Nelson, "Decentralized cooperative aerial surveillance using fixed-wing miniature UAVs," *Proc. IEEE*, vol. 94, no. 7, pp. 1306–1324, 2006.

[18] "Video Codec for Audiovisual Services at p×64 kbit/s ITU-T Recommendation H.261, Version 1," *ITU-T, ITU-T Recommendation H.261 Version 1*, 1990.

[19] 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.

[20] 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.

[21] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1103–1120, 2007.

[22] J. M. Boyce, Y. Ye, J. Chen, and A. K. Ramasubramonian, "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.

[23] B. Masnick and J. Wolf, "On linear unequal error protection codes," *IEEE Trans. Inf. Theory*, vol. 13, no. 4, pp. 600–607, 1967.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

[28] Chang, Yoong Choon and Lee, Sze Wei and Komiya, Ryoichi, "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.

[29] 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.

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

[31] 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.

[32] 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.

[33] 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.

[34] C. Zhou, C. Lin, X. Zhang, and Z. Guo, "A novel JSCC scheme for UEP-based scalable video transmission over MIMO systems," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 6, pp. 1002–1015, 2015.

[35] 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.

[36] L. Hanzo, H. Haas, S. Imre, D. O'Brien, M. Rupp, and L. Gyongyosi, "Wireless myths, realities, and futures: from 3G/4G to optical and quantum wireless," *Proc. IEEE*, vol. 100, no. Special Centennial Issue, pp. 1853–1888, 2012.

[37] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell labs technical journal*, vol. 1, no. 2, pp. 41–59, 1996.

[38] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1456–1467, 1999.

[39] M. El-Hajjar, O. Alamri, J. Wang, S. Zummo, and L. Hanzo, "Layered steered space-time codes using multi-dimensional sphere packing modulation," *IEEE Trans. Wireless Commun.*, vol. 8, no. 7, pp. 3335 – 3340, 2009.

[40] 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.

[41] 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.

[42] 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.

[43] 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.

[44] 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 Commun. Standards Mag.*, vol. 1, no. 2, pp. 70–76, 2017.

[45] *Technical Specification Group Radio Access Network; Study on Enhanced LTE Support for Aerial Vehicles (Release 15)*, document 3GPP 36.777, Jan. 2018.

[46] C. Xu, J. Zhang, T. Bai, 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 Trans. Commun.*, vol. 67, no. 2, pp. 1099–1116, 2018.

[47] E. Haas, "Aeronautical channel modeling," *IEEE Trans. Veh. Technol.*, vol. 51, pp. 254–264, March 2002.

[48] S. Zhou and G. B. Giannakis, "How accurate channel prediction needs to be for transmit-beamforming with adaptive modulation over Rayleigh MIMO channels?," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1285–1294, 2004.

[49] 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 Commun.*, vol. 26, no. 4, pp. 170–177, 2019.

[50] Z. Ghahramani, "Information theory," *Encyclopedia of Cognitive Science*, 2006.

[51] S. Ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. Commun.*, vol. 49, no. 10, pp. 1727–1737, 2001.

[52] 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.

[53] M. El-Hajjar and L. Hanzo, "EXIT charts for system design and analysis," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 1, pp. 127–153, 2014.

[54] 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.

[55] J. Kliewer, S. X. Ng, and L. Hanzo, "Efficient computation of EXIT functions for nonbinary iterative decoding," *IEEE Trans. Commun.*, vol. 54, no. 12, pp. 2133–2136, 2006.

[56] W. Khawaja, I. Guvenc, and D. Matolak, "UWB channel sounding and modeling for UAV air-to-ground propagation channels," in *Global Commun. Conference (GLOBECOM), 2016 IEEE*, pp. 1–7, IEEE, 2016.