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

Chuan Zhu, Yongkai Huo, Bo Zhang, Rong Zhang, Mohammed El-Hajjar and Lajos Hanzo, *Fellow, IEEE*
School of ECS, University of Southampton, UK.
Email: {cz12g09, yh3g09, rz, bz2g10, meh, lh}@ecs.soton.ac.uk, http://www-mobile.ecs.soton.ac.uk

***Abstract*— Inter-layer forward error correction coding (IL-FEC) constitutes an effective unequal error protection (UEP) scheme conceived for transmitting layered video over wireless channels. In contrast to traditional UEP schemes, it operates by embedding extra information concerning the base layer (BL) into the enhancement layers (ELs) without requiring extra transmission resources. The received EL packets assist in the decoding of the BL in order to reduce the distortion of the reconstructed video. The optimum scheduling of the IL-FEC coded layered video streaming in a truncated HARQ (THARQ) aided system is an open problem. Hence in this treatise, we conceive an adaptive THARQ (ATHARQ) algorithm for finding the most appropriate scheduling of the IL-FEC coded layered video packets for the sake of minimizing the video distortion under the constraint of a given total number of transmission time slots. Furthermore, we develop a method of on-line coding-rate optimization algorithm for our IL-ATHARQ transmission scheme, in order to find the best FEC code rate distribution among the video layers that results in the lowest possible video distortion. When using a recursive systematic convolutional (RSC) code, our simulation results show that the proposed rate-optimized IL-ATHARQ system outperforms the traditional THARQ transmission scheme by about 5.3 dB of E_b/N_0 at a peak signal-to-noise ratio (PSNR) of 38.5 dB. Viewing the improvements in terms of the video quality, 2.5 dB of PSNR improvement is attained at an E_b/N_0 of 15 dB.**

I. Introduction

The wireless channel is subject to impairments imposed both by the noise and fading. Thus the task of transmitting video contents to users equipped with various terminals is a challenging one both in terms of source coding and transmission techniques [1]. Layered video coding [2] is a widely used scheme conceived for handling this heterogeneous networking problem. By providing multiple layers of different importance, layered video coding is capable of supporting progressive reception of video steams, depending both on the prevalent channel conditions and on the hardware requirements of the individual users. More specifically, the most important layer is referred to as the base layer (BL), while the enhancement layers (ELs) are capable of providing additional video quality refinements during instances of higher channel qualities. Hence the popular video standards [3]–[7] are capable of supporting layered video coding. For example, H.264 provides partitioned video coding [6] for generating multiple layers (or partitions) of different error-sensitivity. The multiview profile (MVP) [4] developed by the moving picture expert group (MPEG) generates different encoded views as different layers. The scalable compression based extension of the H.264/AVC standard [6] is referred to as scalable video coding (SVC) [5], [6], which generates an encoded stream containing multiple interdependent layers, where some of the layers can be discarded in case of network-congestion for example, in order to tailor the bit-rate according to the specific user-requirements and/or channel quality. At the time of writing, the high efficiency video coding (HEVC) scheme, also known as the H.265 standard [3], is being further developed to create an extension referred to as scalable high-efficiency video coding (SHVC) [8], [9] in order to support scalability.

Hybrid automatic retransmission request (HARQ) aided systems rely on the combination of two error correction mechanisms that are capable of improving the reliability of transmissions: automatic retransmission request (ARQ) and forward error correction (FEC), where the original signals are retransmitted upon requests, when the signals cannot be flawlessly decoded by the FEC decoders. In Type-I HARQ, the transmitter retransmsits the original packet upon reception of a negative acknowledgment (NACK) feedback. In order to provide a more reliable decision concerning the original packet and to achieve a diversity gain, the best approach at the receiver is to combine the various corrupted retransmitted signals according to the maximal ratio combining (MRC) principle, which is carried out by adding the Log-Likelihood Ratios (LLRs) of several packet replicas. This approach is also referred to as Type-I HARQ relying on Chase Combining (CC) [10]. In Type_II HARQ, incremental redundancy (IR) generated from the original packet in form of additional parity bits is transmitted instead of the original packet upon receiving a NACK feedback. Finally, all the information is appropriately combined at the receiver. This scheme is often referred to as Type_II HARQ with IR.

Due to the delay-constraints of near-real-time video transmission systems, only the employment of truncated HARQ (THARQ), relying on a limited number of retransmissions is realistic. The energy efficiency of THARQ protocols designed for a single-user link or assisted by relay stations was considered in [11]. The closed-form analytical expressions of the achievable throughput, of the average packet delay and of the packet loss rate was provided in [12], where the maximization of the system throughput was also carried out. The performance analysis of a wireless network using adaptive modulation and coding combined with THARQ-CC at the data link was presented in [13]. The transmission of control messages using adaptive modulation and coding was considered in [14] in the scenario of voice over Internet Protocol (VoIP) services supported by THARQ. However, the associated video characteristics had not been addressed in the aforementioned THARQ schemes. As a further advance, a

The financial support of the EU's Concerto project, of the EPSRC under the auspices of the India-UK Advanced Technology Centre (IU-ATC) and that of the ERC's Advanced Fellow Grant is gratefully acknowledged.

video transmission system was proposed and analyzed in [15], which relied both on THARQ and selective combining, as well as on rate-compatible punctured convolutional (RCPC) codes for transmission over fading channels. A finite-state Markov model was used for representing the Rayleigh fading channels. An improved video quality was achieved by the proposed scheme at a limited delay. Layered video has been considered for transmission using HARQ schemes in either unicast or multicast scenarios [16]–[20]. The authors of [16] presented a theoretical analysis as well as rich experimental results for characterizing both unicast and multicast scenarios for transmission over packet-erasure channels, while the authors of [17]–[20] provided solutions for multicast systems transmitting layered video using various HARQ schemes.

The transmission of layered video can be protected by unequal error protection (UEP) [21]. In [22], the cross-layer design of FEC schemes is investigated by using UEP Raptor codes at the application layer (AL), and UEP RCPC codes at the physical layer (PHY) for the prioritized video packets, which are prioritized based on their contribution to the received video quality. The authors of [23] introduced an APP/MAC/PHY cross-layer architecture that improves the perceptual quality of delay-constrained scalable video transmission. Furthermore, an online QoS-to-QoE mapping technique is proposed in [23] for quantifying the QoE reduction imposed by each video layer using both the ACK history and a variety of perceptual metrics. The authors of [24] studied the channel-dependent adaptation capability of SVC by conceiving a solution for transmission over an orthogonal frequency division multiplexing (OFDM) based broadband network relying on cross-layer optimization. The FEC protected UEP may be classified into two categories, namely the packet-level schemes [25]–[32] and bit-level schemes [33]–[40]. The packet-level contributions [25]–[32] usually employ hard decoded FEC codes for mitigating the packet loss events at the application layer [41], while the bit-level ones operate at the physical layer and rely on soft decoded FEC codes for correcting bit-errors in wireless scenarios [42]. Traditional UEP schemes designed for layered video transmission only handle the different importance of separate video layers by assigning different-rate FEC codes to them. By contrast, the recent contributions [28], [31], [32], [38], [40] explored the dependencies amongst the layers and conceived UEP schemes by exploiting this sophisticated feature. Specifically, the unsuccessful decoding of the BL will instruct the video decoder to discard all the ELs depending on it, regardless whether they have or have not been successfully decoded. Naturally, this course of action wasted the transmit power assigned to the dependent layers. Thus we proposed in our previous work [40] a bit-level inter-layer coded FEC (IL-FEC) scheme that embeds the BL into the FEC coded ELs, so that the reception of the BL can be improved with the aid of the ELs using soft decoding. In our subsequent work [43] we conceived a sophisticated on-line real-time video distortion estimation technique, which is suitable for diverse channel conditions and system configurations. More explicitly, in [43] we proposed an on-line code rate optimization method for minimizing the video distortion.

A range of UEP schemes have been conceived for HARQ [44]–[52] in order to improve the video quality of layered videos. The authors of [44] proposed UEP by appropriately sharing the bitrate budget between the source and channel encoders based on either the minimum-distortion or on the minimum-power consumption criterion. In [50], [51], UEP was achieved by assigning each video layer a different retransmission limit. Another stream of contributions [44], [46]–[49] adopted the so-called limited-retransmission based priority encoding transmission (PET) scheme [53], where UEP is achieved by varying the source block-length across the different source layers, while keeping the FEC-dcoding block-length fixed. This allows the PET to have a packetization scheme that ensures that the source layers of an FEC-coded block are dropped according to their significance, commencing by dropping the least significant one first.

Against this background, in this treatise, we conceive an adaptive THARQ (ATHARQ) transmission scheme in support of IL-FEC coded layered video for minimizing the video distortion under the constraint of a given total number of transmission time slots. In our previous work [40], the transmission environment of THARQ was not considered. Furthermore, the merits of IL-FEC schemes have not been investigated in the context of THARQ transmission schemes. However, the packet scheduling schemes should be carefully designed by ensuring that instead of the sequential packet transmissions assumed in [40], we have to exploit the specific characteristics of each IL-FEC coded packet. Furthermore, we develop a method of on-line optimization for our IL-ATHARQ transmission scheme, in order to find the most appropriate FEC code rate distribution among the video layers that reduces the video distortion. Type-I HARQ relying on Convolutional Codes (CC) is used for simplicity, because our focus is on the design of the scheduling schemes. Our proposed technique is significantly different from the existing contributions, such as the PET framework of [53], as detailed below. Firstly, our transmission scheme is proposed for wireless channels, while most contributions on PET [53] operate at the packet-level and consider the Binary Erasure Channel (BEC). Secondly, IL-FEC typically relies on bit-level FEC decoders using soft decoding, such as a Recursive Systematic Convolutional (RSC) code, while PET employs hard-decoded codes, such as the family of (N, k) block codes.

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

1) We intrinsically amalgamated IL-FEC coding with the THARQ-aided transmission of layered video. We conceived an ATHARQ transmission scheme for adaptively scheduling the IL-FEC coded video layer packets for minimizing the video distortion under the constraint of a certain total number of transmission time slots.
2) We develop a method of on-line optimization for our IL-ATHARQ transmission scheme, in order to find the optimal FEC code rate distribution, sharing among the video layers that results in a reduced video distortion. Quantitatively, about 2.5 dB of PSNR video quality improvement may be observed at an E_b/N_0 of 15 dB, over the traditional THARQ benchmaker. Alternatively, we will demonstrate that an E_b/N_0 reduction of about

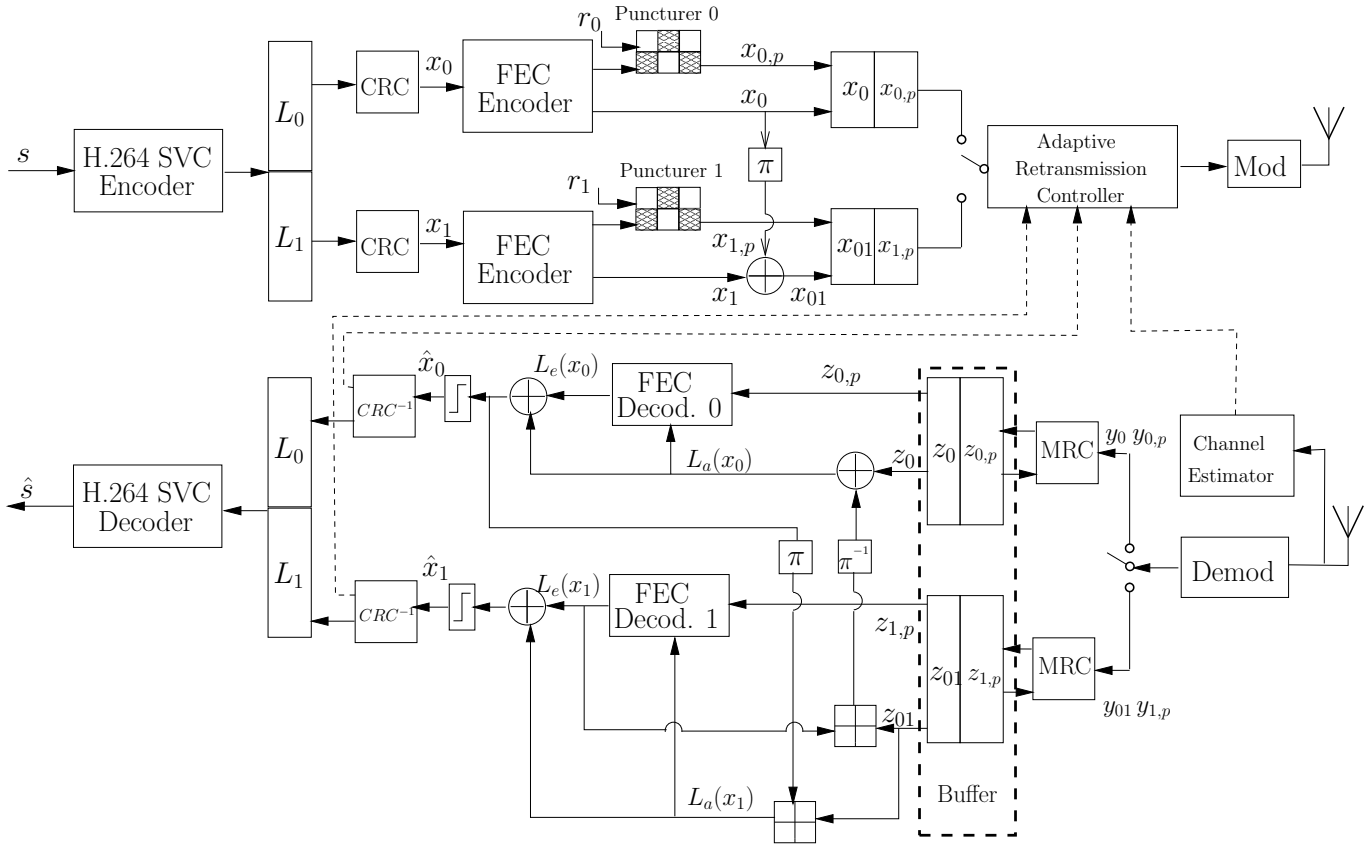


Fig. 1: Block diagram of the proposed ATHARQ-IL-FEC coded SVC H.264/AVC coded video system, where r_0 and r_1 represent the code rates for FEC encoder 0 and 1, respectively.

5.3 dB at a PSNR of 38.5 dB can be achieved.

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

II. System Overview

Here we introduce the IL-ATHARQ-aided inter-layer video transceiver shown in Fig. 1. The system consists of two major parts: the inter-layer FEC (IL-FEC) protected video codec and the retransmission control protocol. Firstly, the former one is introduced based on [40], where the inter-layer FEC architecture is described in detail. We will briefly describe both the IL-FEC transceiver architecture, as well as the IL-ATHARQ protocol in this section.

A. Transmitter Model

The original video sequence is firstly encoded into a scalable video stream by invoking the SVC extension of H.264 [6]. The compressed video stream consists of the layers L_0, L_1,... L_n with the dependency of $L_0 \Leftarrow L_1 \Leftarrow \ldots \Leftarrow L_n$, where each item on the right of the $\Leftarrow$ symbol depends on all the items to the left of it. To utilize the n-th layer for successful decoding, the decoder has to invoke the information from all the previous $(n-1)$ layers. For simplicity of illustration, only the pair of layers L_0 and L_1 are used in our description of IL-FEC, where L_0 is the BL and L_1 is the EL.

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

As part of the IL-FEC mechanism, the systematic part of the encoded layer L_0, namely x_0, is interleaved and then embedded into the systematic part x_1 of L_1, using the bit-wise XOR operation, producing the bit sequence x_{01}. In the case

that L_0 and L_1 are different in length, the solution detailed in [40] may be invoked. Then the systematic bits and the parity bits of the BL are concatenated. Similarly, the EL, which contains the systematic bits and parity bits of the original EL are also concatenated. For each time slot, the adaptive retransmission controller picks the packets from one of the two layers and transmits them using BPSK over the wireless channel, which is modeled as an uncorrelated Rayleigh-faded channel.

B. Receiver Model

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

After updating the buffers, the decoder carries out the IL-FEC decoding process. The pair of FEC decoders shown in Fig. 1 invokes the BCJR algorithm [54] to produce the extrinsic information for x_0 and x_1, given the *a priori* information of their systematic bits and parity bits.

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

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

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

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

C. Major Assumptions and Transmission Protocol

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

The traditional THARQ transmission protocol conceived for the FEC coded video layers is shown in Fig. 2(a). The BL is transmitted first, followed by the ELs. Each layer is transmitted a maximum number of n times, regardless, whether or not it is correctly received. However, according to the dependency between the video layers, there is no need to transmit the ELs, if the BL is lost. Therefore it is sub-optimum to assign the same retransmission limit to each layer. Thus we adapt the traditional THARQ by defining a total maximum retransmission limit for a specific video slice. Explicitly, for a total of n transmissions, the BL is allowed to have a higher number of transmissions than the EL. This plausible prioritization principle may be readily extended to an arbitrary number of layers, where the less dependent layers are granted more transmission opportunities than the more dependent layers.

With the introduction of IL-FEC coding into the THARQ aided SVC coded video steam, the constraints imposed may be relaxed, because the IL-FEC coded layers of higher dependencies may have sufficient information concerning the layers of lower dependencies and hence they may be capable of recovering them, even they were incorrectly recovered during the previous transmissions. As a result, it may in fact become wasteful to complete the recovery of the BL before transmitting the ELs. Hence we have to carefully consider the choice of transmission limits for each layer. The total number of transmissions dedicated to a specific video slice remains the same as defined previously for fair comparison. The philosophy of this scheme is illustrated by a specific example in Fig. 2(b).

We set out to improve the THARQ regime introduced above, which relies on the CRC check result of the decoded layers. In the traditional regime, the transmitter only knows whether the current layer has or has not been successfully recovered. However, it has no quantitative knowledge about the specific

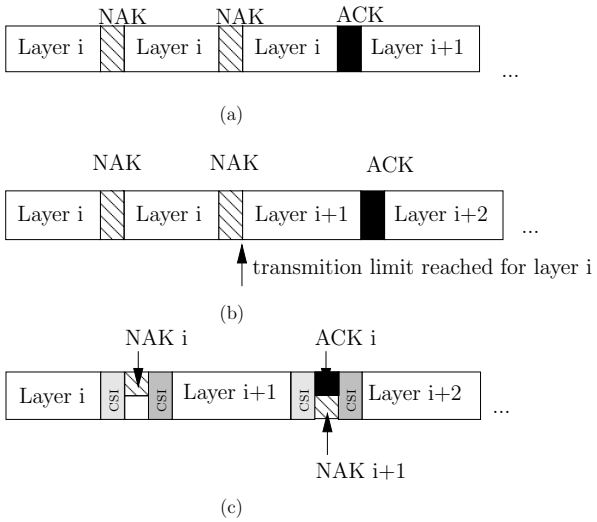


Fig. 2: THARQ schemes conceived for scalable video transmission: (a) pure THARQ (b) IL-FEC aided THARQ where each video layer has a different retransmission limit (c) proposed IL-FEC aided adaptive THARQ()

grade of degradation imposed on the unsuccessfully decoded layers in the buffer, if any. Similarly, the transmitter has no knowledge of the next transmission's contribution towards the successful decoding of the video slice. With the goal of improving the performance, we set out to estimate both and hence to make better-informed decisions. Therefore in our new regime, the receiver has to provide a feedback for the transmitter concerning the CRC result, as well as additionally has to feed back the channel state information (CSI) of both the most recent transmission and of the next transmission.

In reciprocal channels typically encountered in Time Division Duplex (TDD) systems, the CSI of the next transmission can be acquired by appending pilot symbols to the reverse-direction ACK/NAK feedback, which allows us to estimate the CSI and assuming that the coherence-time is less than 30 ms [55], use it for predicting the channel of the next transmission. Owing to using this low-complexity zero-order prediction, a prediction error may be introduced at this stage. At the reception of a packet, the receiver sends a feedback message to the transmitter, which includes the CRC results of the layers of interest, plus the estimated CSI of this specific transmission. The latter one assists the transmitter in rectifying the previous CSI prediction error at the transmitter, and as a benefit, this measure prevents error propagation in the subsequent prediction process. As shown in Fig. 2(c), the transmitter becomes capable of estimating the contribution of each possible transmission at the receiver, hence intelligently controlling the retransmission process by maximizing the video quality after the next transmission attempt.

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

As mentioned, each video layer is packaged into a single NALU and can be transmitted over a single channel instance. By contrast, when the video layers of high-resolution sequences are represented by more bits, each video layer may be packaged into several NALUs and transmitted over different channel instances. In that case, advanced channel estimation techniques [56], [57] may be adopted for acquiring the channel estimates for the sake of predicting the video qualities, which is beyond the scope of this treatise.

III. Adaptive Truncated HARQ Transmission

In this section, we will describe our adaptive truncated HARQ aided IL-FEC coded video streaming scheme, which is used in the "Adaptive Retransmission Controller" block of Fig. 1.

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

In order to characterize the behavior of the receiver seen in Fig. 1 relying on the proposed algorithm, a classic RSC codec is used as the FEC code. However, the employment of our proposed techniques is not limited to the RSC codec.

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

- N_T: limit of the total number of transmission time slots (TS);
- N_L: total number of video layers;

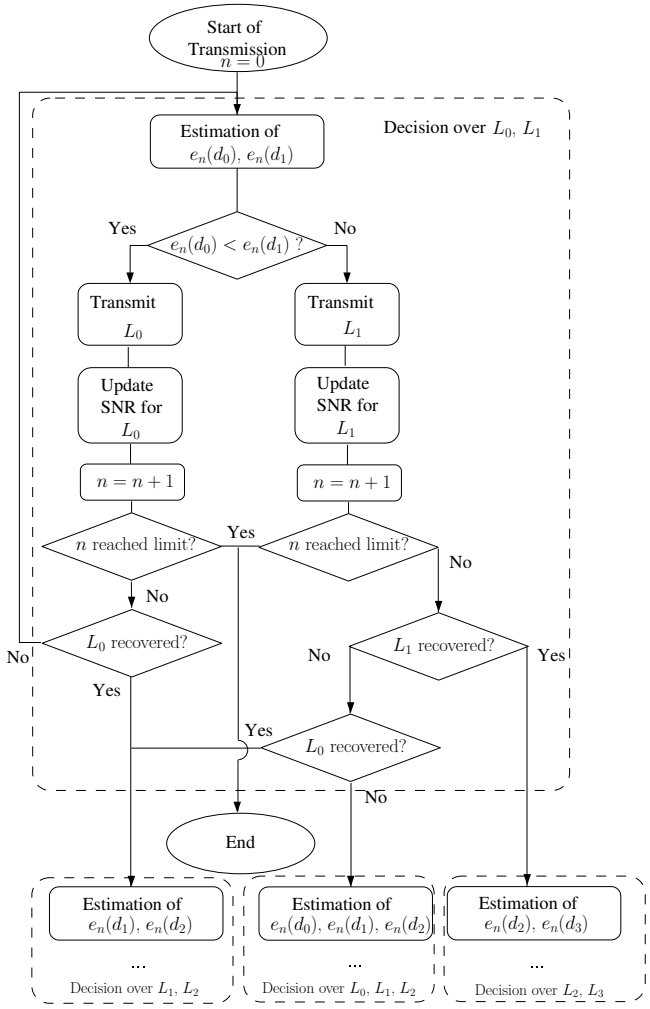


Fig. 3: Adaptive THARQ algorithm for the Adaptive Retransmission Controller

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

The scheduling procedure of the adaptive retransmission controller is shown in Fig. 3. Each time the algorithm considers a number of video layers, for making decisions, as shown in the dashed boxes of Fig. 3.

In order to generalize the scheduling process, the concept of a decision window can be introduced, which contains the layers to be chosen by our algorithm, as shown in Fig. 4. This decision window always contains one more layer than the layers transmitted during the most recent history, which is layer $(i + j + 1)$ in the example shown in Fig. 4. Upon reception of an ACK for a successfully recovered layer, the layer is removed from the window. At the beginning of the entire transmission, the decision window only contains L_0 and L_1.

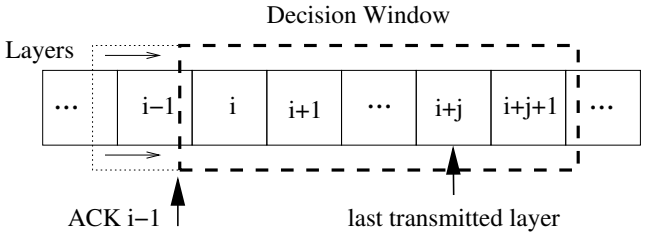


Fig. 4: Decision Window of the Adaptive THARQ algorithm

Now we set out to describe the algorithm in each of the dashed boxes of Fig. 3. As described in Section II-C, the prediction of the forthcoming channel amplitude is given by $\left|\widetilde{h}_n\right|$. Furthermore, when using a Maximum Ratio Combining (MRC) receiver, given the E_b/N_0 value, the combined signal's SNR in the receiver's buffer can be predicted as:

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

which means that the SNR value of the buffered signal corresponding to the k-th layer will increase by the value of $\frac{E_b}{N_0} R \left|\widetilde{h}_n\right|^2$, if the decision d_k is made. The channel prediction $\widetilde{h}_n$ can be expressed as [58]

$$\widetilde{h}_n = h_n + \zeta_n, \tag{3}$$

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

To obtain the estimate of the distortion, each layer's PER $p_{n,i}(d_k)$ and its distortion ϵ_i should be acquired. The latter can be obtained by the so-called offline removal-decoding test, which was advocated in [59]. Explicitly, ϵ_i is measured by

comparing the PSNRs of the reconstructed video with the bit stream of the i-th video layer removed, and the one relying on the intact bit stream for the i-th video layer. Again, this measurement is carried out offline, before the transmissions begin.

Let us now consider the "conditional" PER of layer i, which refers to the PER of the i-th video layer at TS n corresponding to decision d_k, given that layer $(i-1)$ has been successfully recovered, which is denoted by $p_{n,i}(d_k)$. In order to formulate the PER $p_{n,i}(d_k)$, here we introduce the function $f_i(.)$ which was defined by Eq. (17), (18) in [43]:

$$f_i(SNR_{n,i}, \ell_i, r_i) = 1 - [1 - T_p[SNR_{n,i}, I_a, r_i]]^{\ell_i/\ell}. \quad (4)$$

Given the $SNR_{n,i}$ of the buffered signal corresponding to layer i, the systematic bit-length ℓ_i, and the FEC coding rate r_i, the $f_i(.)$ function gives the PER estimate $p_{n,i}$ of layer i. At the right side of Eq. 4, the pre-generated look-up table (LUT) T_p is used for obtaining the PER assuming a fixed systematic bit-length of ℓ. Furthermore, T_p is a 4-dimensional LUT that has three input parameters to index the specific PER needed. Apart from the aforementioned $SNR_{n,i}$ and ℓ_i, the mutual information (MI) I_a gleaned from the estimation of the decoding output of layer $i+1$ is needed. Further details concerning the estimate of I_a can be found in [43]. Therefore $p_{n,i}(d_k)$ can be readily formulated as

$$\begin{aligned} p_{n,i}(d_k) &\doteq p_{n,i}\left(d_k, \left|\widetilde{h}_n\right|, SNR_{n-1}\right) \\ &= f_i(SNR_n(d_k), \ell_i, \ldots, \ell_{N_L-1}, \\ &\quad r_i, \ldots, r_{N_L-1}). \end{aligned} \quad (5)$$

Firstly, the $SNR_n(d_k)$, which is required for the estimation of the PER $p_{n,i}$ can be obtained from Eq. 2. As for the decoding process, the IL decoder commences its operation from the specific layer having the highest grade of dependency, which is layer (N_L-1). Then it exchanges information between the decoding of two consecutive video layers during each iteration, as illustrated in Section II-B. Given the intact layer $(i-1)$, the successful decoding of layer i depends on the assistance of layer $(i+1)$, which in turn depends on layer $(i+2)$, etc. Therefore the estimation of the "conditional" PER in Eq. 5 depends on the properties of all the layers spanning from i to (N_L-1), which includes both the lengths of their coded blocks and their coding rates. Specifically, the layer (N_L-1) associated with the highest grade of dependency but receiving no extra protection from the other layers has the "conditional" PER that only depends on the layer (N_L-1) itself, which is formulated as

$$p_{n,N_L-1}(d_k) = f_{N_L-1}[SNR_n(d_k), \ell_{N_L-1}, r_{N_L-1}]. \quad (6)$$

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

$$\begin{aligned} e_n(d_k) &\doteq e_n\left(d_k, SNR_{n-1}, |h_n|^2\right) \\ &= p_{n,0}(d_k) \cdot \epsilon_0 + \\ &\quad [1 - p_{n,0}(d_k)] \cdot p_{n,1}(d_k) \cdot \epsilon_1 + \ldots \\ &= \sum_{i=0}^{N_L-1} \epsilon_i \cdot p_{n,i}(d_k) \cdot \prod_{j=0}^{i-1} [1 - p_{n,j}(d_k)] \end{aligned}, \quad (7)$$

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

The retransmission controller of Fig. 3 opts for transmitting the specific video layer that ends up with the minimum distortion of the decoded video. Hence the final decision carried out by the controller is

$$D_n\left(SNR_{n-1}, |h_n|^2\right) = \underset{d_k \in \mathcal{D}}{\arg\min}\{e_n(d_k)\}, \quad (8)$$

which may be compactly expressed as D_n. At the commencement of transmissions, the module estimates the distortions that two different scheduling decisions would impose, namely when transmitting L_0 or transmitting L_1, which may be denoted by d_0/d_1. As shown in Fig. 3, the receiver buffer is empty at the beginning of a transmission session, and both d_0 as well as d_1 are compared by the retransmission controller, as the potentially available choices. If we have $D_1 = d_0$, which means that the controller of Fig. 3 decides to transmit L_0, and L_0 is successfully recovered, then the controller will consider both d_1 and d_2, provided that more transmission TSs are available. Otherwise, if the receiver failed to decode L_0, then both d_0 and d_1 will be reconsidered as retransmission candidates for the next retransmission attempt. On the other hand, if we have $D_1 = d_1$, which implies that L_1 is selected for transmission, and L_1 is successfully recovered, the controller of Fig. 3 will naturally move on to consider both d_2 and d_3, if possible. However, if L_1 was unsuccessfully decoded, the coded packet of L_1 will be stored in the receiver's buffer, and hence we have to further consider the decoding outcome of L_0. If the decoding of L_0 turned out to be successful, the transmitter will consider retransmitting L_1 or transmitting the new L_2 packet. But if L_0 also failed, with d_0 and d_1 on the table, d_2 should also be considered. Since the IL-encoded L_2 includes the redundancy protecting L_1, it may be capable of improving the decoding of L_1. In turn, the improved decoding of L_1 may become capable of providing beneficial information for the BL L_0.

After each transmission, the transmitter will receive the updated version of the channel's amplitude $|h_n|$, in order to replace the predicted version $\left|\widetilde{h}_n\right|$, as described by our protocol in Section II-C. Here the updated version may be considered accurate, since it exactly equals $|h_n|$. The retransmission controller of Fig. 3 will then update the estimation of the SNRs of the signals in the receiver buffer, using both the E_b/N_0 value, as well as the past transmission decision records and channel amplitudes, as formulated in Eq. 2. The SNR estimate

of SNR_{n-1} can be expressed as

$$SNR_{n-1} = \frac{E_b}{N_0} R \cdot \sum_{t=1}^{n-1} D_t \left|h_t\right|^2, \tag{9}$$

and upon substituting it into Eq. 2, we arrive at:

$$SNR_n\left(d_k\right) = \frac{E_b}{N_0} R \cdot \left(d_k \left|\widetilde{h}_n\right|^2 + \sum_{t=1}^{n-1} D_t \left|h_t\right|^2 \right). \tag{10}$$

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

IV. FEC Coding Rate Optimization

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

According to Eq. 8, the distortion of the reconstructed video frame after the n-th transmission is given by:

$$\mathcal{E}_n\left(SNR_{n-1}, \left|h_n\right|^2\right) = \min_{d_k \in \mathcal{D}} \left\{e_n\left(d_k\right)\right\}. \tag{11}$$

Given the video distortion definition in Eq. 7, by substituting SNR_{n-1} from Eq. 9 into Eq. 11, we get:

$$\begin{aligned} &\mathcal{E}_n\left(\left|h_1\right|^2, \ldots, \left|h_n\right|^2\right) \\ &\quad = \mathcal{E}_n\left(A_1, \ldots, A_n\right) \\ &\quad = \min_{d_k \in \mathcal{D}} \left\{ e_n\left(d_k, \frac{E_b}{N_0} R \cdot \sum_{t=1}^{n-1} D_t A_t, A_n\right) \right\}. \end{aligned} \tag{12}$$

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

$$\begin{aligned} E\left(\mathcal{E}_n\right) = \int f\left(A_1\right) dA_1 \int f\left(A_2\right) dA_2 \cdots \\ \int \mathcal{E}_n\left(A_1, \ldots, A_n\right) f\left(A_n\right) dA_n \end{aligned}, \tag{13}$$

where $f\left(\cdot\right)$ is the probability density function (PDF) of the fading channel. For Rayleigh-faded channels $f\left(\cdot\right)$ is given by the PDF of the Gamma distribution. Since Eq. 13 is difficult to evaluate in a closed form, we carried out Monte-Carlo simulations using Eq. 12. It is worth noting that each experiment is based on low-complexity table-look-up operations without any actual encoding or decoding operations, therefore imposing an affordable complexity. A numerical example is provided in Fig. 5 for the Football sequence in terms of the video distortion vs coding rates.

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

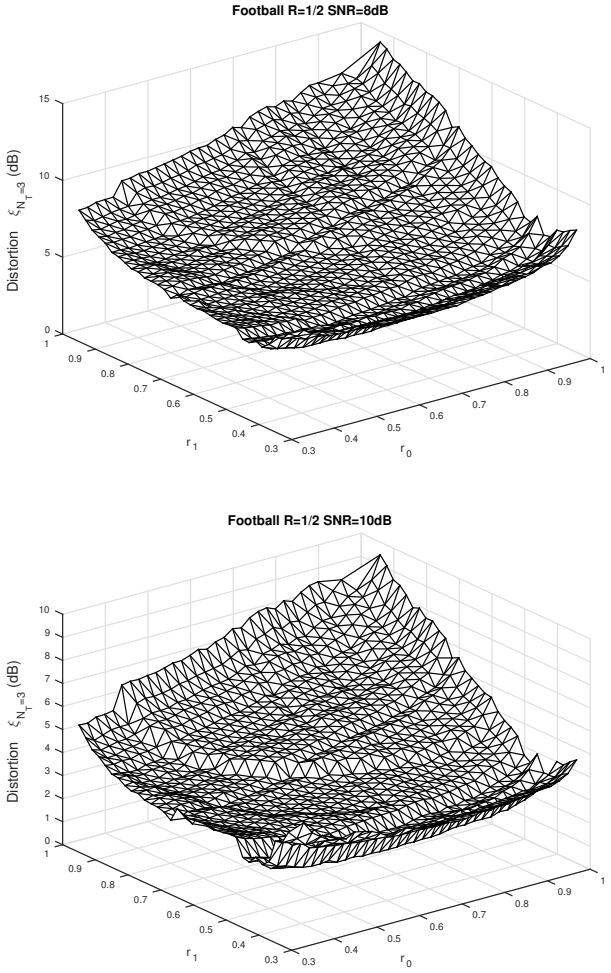


Fig. 5: Video distortion versus FEC coding rate performance according to Eq. 12

$$\arg\min_{\boldsymbol{r} \in \Gamma} \left\{E\left(\mathcal{E}_{N_T}\right)\right\}, \tag{14}$$

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

$$\Gamma = \left\{ \boldsymbol{r} \left| \sum_{i=0}^{N_L-1} \frac{\ell_i}{r_i} = \frac{\sum_{i=0}^{N_L-1} \ell_i}{R} \right. \right\}. \tag{15}$$

Naturally, the system performance formulated in Eq. 14 may be affected by excessive video distortion estimation errors at lower E_b/N_0 values. Therefore an amended version of Eq. 14 can be formulated as

$$\boldsymbol{r}_{op} = \begin{cases} \arg\min_{\boldsymbol{r} \in \Gamma} \left\{E\left(\mathcal{E}_{N_T}\right)\right\} & \min\left\{E\left(\mathcal{E}_{N_T}\right)\right\} < \\ & \underset{\boldsymbol{r}=\boldsymbol{r}_{as}}{E}\left(\mathcal{E}_{N_T}\right) + \delta \\ \boldsymbol{r}_{as} & otherwise \end{cases}, \tag{16}$$

where $\boldsymbol{r}_{as} \in \Gamma$ represents the code rates in ascending order, and the BL has the lowest FEC coding rate, i.e. the highest

protection. Still referring to Eq. 16, δ is the estimation error tolerance threshold, which is found experimentally.

To evaluate the effect of the value of δ on the final PSNR performance of the system, we carried out simulations for various settings of δ, and the corresponding PSNR results are shown in Fig. 6. We can observe from Fig. 6a for $N_T = 3$ using the Football sequence that the PSNR performance is not very sensitive to the δ values at high E_b/N_0 values, say for 13~16 dB. However, a slight improvement of PSNR can be observed for lower E_b/N_0 values around the δ values of 0.5 dB. Similar trends can also be observed for $N_T = 4$ in Fig. 6b, except that a marginal PSNR reduction is encountered upon increasing δ at E_b/N_0 values above 10 dB. Since the estimation error $\mathcal{E}_{N_T}$ is difficult to model analytically, we found the optimum value of δ experimentally. In order to improve the PSNR performance at lower E_b/N_0 values, we set $\delta = 0.6$ for our simulations, which was also found to be beneficial for the other video sequences investigated in Section. V.

V. System Performance

In this section, we will quantify the attainable performance gain of our proposed ATHARQ-IL transmission scheme, as well as the additional performance gain of our rate-optimized ATHARQ-IL scheme. Furthermore, we will characterize both the delay performance and the robustness of the aforementioned systems against channel prediction errors. The main system parameters are listed in Table I. Three 4:2:0 YUV format video sequences were chosen for transmissions, namely the Football, the Soccer and the Crew video clips. The 15-frame Football sequence is in the (176×144)-pixel quarter common intermediate format (QCIF) and has a frame rate of 15 frames/second (FPS). The other two 60-frame sequences are in the (704×576)-pixel 4CIF format, and were recorded at 60 FPS.

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

TABLE I: The parameters of the testing video sequences

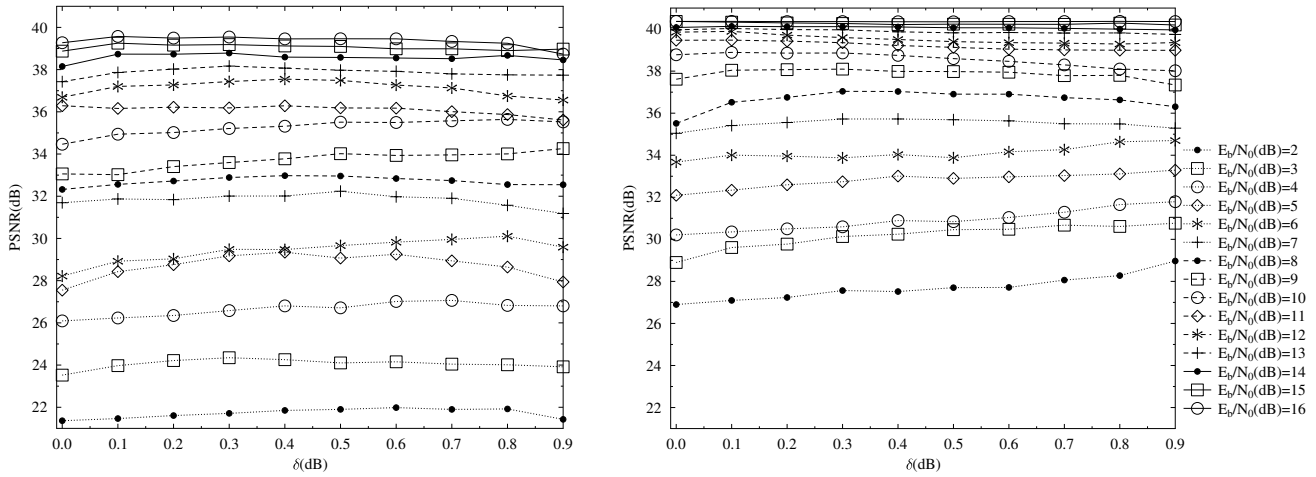
Regime	Coding Rate (r_0, r_1)	Other Parameters
THARQ	$(1/2, 1/2)$, $R = 1/2$	
IL-THARQ	$(1/2, 1/2)$, $R = 1/2$	(n_0, n_1): limit of transmission times for the video layers
IL-ATHARQ	$(1/2, 1/2)$, $R = 1/2$	
	$(1/3, 1/2)$, $R = 1/2$	
	$(1/3, 1/3)$, $R = 1/2$	
RO-IL-ATHARQ	$R = 1/2$	non-modified
	$R = 1/2$	$\delta = 0.6$ for all the sequences

TABLE II: The regimes with their settings characterized in Section V, where R stands for the overall coding rate of the system.

We use the JSVM H.264/AVC reference video codec as the SVC codec. The video encoder relies on a group of pictures (GOP) duration of 15 frames and the bi-directionally predicted (B) frames are disabled. We enabled the Medium Grain Scalability (MGS) [5], [60] feature for encoding the video sequences into three layers with the aid of the standardized quantization parameters (QP) of 40, 32 and 24, respectively. The average PSNRs achieved by the decoder for different sequences are 40.46 dB, 42.62 dB and 42.82 dB, respectively.

Based on our configuration of the SVC encoder, each slice is encoded into three layers and each layer is encapsulated into a network abstraction layer unit (NALU) [6]. The NALUs are transmitted sequentially using our proposed system. Should the CRC check of a certain NALU indicate a decoding failure, these NALUs are discarded. The SVC decoder uses the low-complexity error concealment method of frame-copying in order to compensate for the lost frames.

The RSC code having a code-rate of $1/3$ and the generator polynomials of $[1011, 1101, 1111]$ is employed as the FEC code in our system. The reconfigurable puncturers employed are capable of adjusting the FEC code rate on a fine scale, ranging from its original $1/3$ to 1, thus providing a wide range of design options. The FEC encoded signals are BPSK modulated *and transmitted through a block-fading non-dispersive uncorrelated Rayleigh channel*. The total coding rate of the system is assumed to be $1/2$. The channel is static for each FEC encoded NALU, but it is faded independently between NALUs. As we are considering delay-constrained systems, we characterize the attainable performance of the proposed scheme using two scenarios, where either $N_T = 3$ or $N_T = 4$ transmissions are allowed in total, respectively. The regimes and their settings characterized in this section are listed in Table II.



(a) PSNR vs δ with RO-IL-ATHARQ(δ) for *Football,* $N_T = 3$

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

Fig. 6: PSNR versus δ performance with RO-IL-ATHARQ(δ) systems. The *Football* sequence was transmitted over block-fading non-dispersive uncorrelated Rayleigh channels

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

TABLE III: Example of the LUT $T_p(SNR, I_s, r)$.

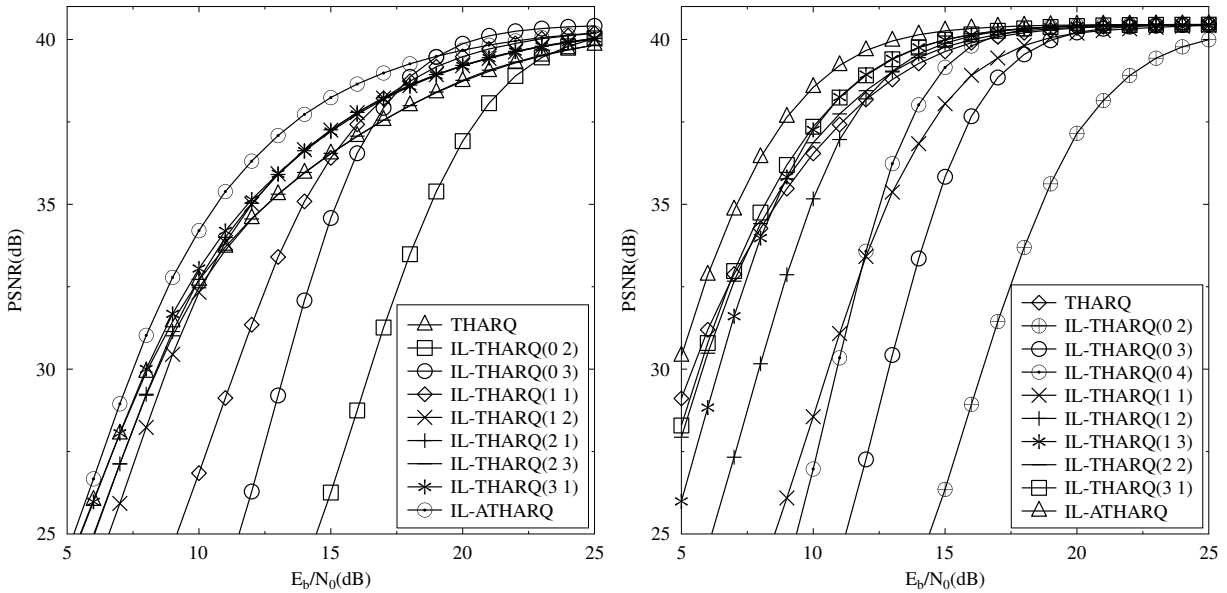
A. Off-line LUTs Generation

As described in Eq. 4 of Section III, the estimation of the PER relies on the LUT T_p. Here we describe the implementation of the LUT T_p that is used in our experiments. As mentioned in Section III, the LUT T_p is indexed by three parameters, namely SNR, I_a, r. To generate T_p, we fix the block-length ℓ of the FEC and obtain the outputs, namely the extrinsic information I_e and the PER $p(\ell)$ of the component FEC by scanning the practical coding parameter ranges of SNR, I_a, r at certain intervals. Specifically, the SNR is considered over the range of $[0, 25]$ dB, using a step-size of 0.2 dB, I_a is scanned over the range of $[0, 1]$ at intervals of 0.01, and finally r is scanned across the range of $[0.33, 1]$ at intervals of 0.02. This makes the total number of legitimate settings $n_{Tp} = n_{snr} n_I n_r = 126 \times 101 \times 33$, which is 419,958. All 5 items corresponding to each setting can be individually stored as floats in 8 bytes. Thus the total size of the LUT T_p is 16 MB. In Table III, we show an example of the LUT T_p that is used in our simulations.

B. Performance of the Adaptive Rate Controller

In order to demonstrate the attainable performance gain of our IL-ATHARQ algorithm, we compare its PSNR performance to that of the aforementioned traditional THARQ as well as to that of the IL-THARQ scheme relying on fixed transmission limits, using the Football sequence, as listed in Table I. For the IL-THARQ scheme, we use the compact form of IL-THARQ(n_0, n_1) to represent different configurations, where n_0 and n_1 denote the number of transmission times allowed for L_0 and L_1, respectively. Since the total number of transmission is fixed to N_T, L_2 is allowed to transmit as long as L_0 and L_1 have completed their transmission, provided that the total transmission attempts N_T has not been exceeded. At this stage we assume that all the three video layers are encoded using the same FEC coding rate of $1/2$, which is the total coding rate of the system. The corresponding results are shown in Fig. 7.

Observe in Fig. 7 that the PSNR versus E_b/N_0 performances of our proposed IL-ATHARQ system, relying on $N_T = 3$ or $N_T = 4$ transmissions are portrayed separately in Fig. 7a and Fig. 7b. Observe in Fig. 7a that the IL-THARQ(n_0, n_1) schemes perform differently for the different configurations of n_0 and n_1, given $N_T = 3$. We opted for characterizing the most typical combinations of n_0 and n_1, noting that others have similar results, hence we limited the number of combinations to make the figure more readable. It is clear from Fig. 7a that the IL-THARQ(0, n_1) class of systems performs relatively poorly at low E_b/N_0 values, because the BL L_0 is never transmitted, and the recovery of L_0 solely depends on the accumulation of the MI provided by the information embedded in L_1, which is not necessarily beneficial, given the limited number of transmission slots. However, this drawback turns into a benefit, when the E_b/N_0 value reaches higher levels, where L_0 can be readily recovered with the aid of L_1 and the remaining TSs can be saved for transmitting other layers for the sake of improving the



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

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

Fig. 7: PSNR versus E_b/N_0 performance of our proposed IL-ATHARQ system versus both the traditional THARQ transmission and the IL-HARQ scheme in conjunction with fixed transmission limits (IL-THARQ(n_0, n_1)) as benchmarks. The Football sequence is used for transmission over block-faded non-dispersive uncorrelated Rayleigh channels

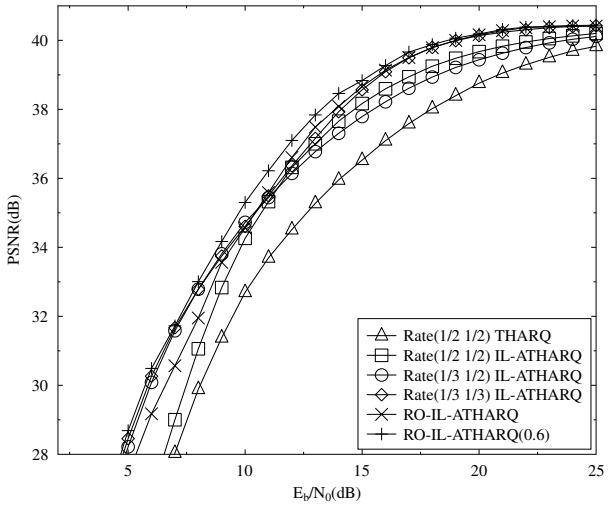
video quality. The traditional THARQ scheme performs better than the IL-THARQ(n_0, n_1) schemes at low levels of E_b/N_0 because this scheme prioritizes the transmission of L_0 and indeed, recovers L_0 with a high probabiliry. However, this scheme is not so efficient at high E_b/N_0 values, because each layer is transmitted at least once, which is not always necessary in IL-based schemes, since the skipped layer can be recovered later using the information embedded into the other layers. Compared to the benchmarks, our proposed IL-ATHARQ scheme results in an improved performance by virtue of its adaptive nature. As observed in Fig. 7a, the IL-ATHARQ scheme outperforms the traditional THARQ scheme all the way and achieves an E_b/N_0 reduction of about 3.8 dB at a PSNR of 38.5 dB. Alternatively, 1.8 dB of PSNR video quality improvement may be observed at an E_b/N_0 of 15 dB. Furthermore, IL-ATHARQ also outperforms most IL-THARQ(n_0, n_1) schemes, except for the IL-THARQ(0, 3) scheme, which shows an exceptionally good performance at sufficiently high E_b/N_0 values and slightly outperforms IL-ATHARQ. This may be due to the inaccuracy of the distortion estimation function invoked by IL-ATHARQ. Nonetheless, an approximately 1.9 dB of power reduction is achieved by the IL-ATHARQ arrangement compared to the IL-THARQ(0, 3) scheme at a PSNR of 38.5 dB. Alternatively, about 1.1 dB of PSNR video quality improvement may be observed at an E_b/N_0 of 15 dB, compared to the IL-THARQ(2, 1) scheme, which is the best performer amongst the IL-THARQ(n_0, n_1) schemes at an E_b/N_0 of 15 dB.

Similarly, observe from Fig. 7b that given $N_T = 4$, the IL-ATHARQ scheme outperforms the traditional THARQ arrangement and achieves an E_b/N_0 reduction of about 2.7 dB at a PSNR of 38.5 dB. Alternatively, about 1.9 dB of PSNR video quality improvement may be observed at an E_b/N_0 of 11 dB. IL-ATHARQ outperforms all of the IL-THARQ(n_0, n_1) schemes at all the E_b/N_0 values considered. More specifically, about 1.5 dB of power reduction is achieved by the IL-ATHARQ scheme compared to the IL-THARQ(1, 3) scheme at a PSNR of 38.5 dB. Alternatively, about 1.0 dB of PSNR video quality improvement may be observed at an E_b/N_0 of 11 dB compared to the IL-THARQ(1, 3) scheme, which is the best performance amongst the IL-THARQ(n_0, n_1) schemes at the E_b/N_0 of 11 dB. Generally speaking, the video performance gain becomes relatively modest upon increasing $N_T = 3$ to $N_T = 4$. This is because the adaptive scheduling of the layers becomes less important, when there are sufficient TSs.

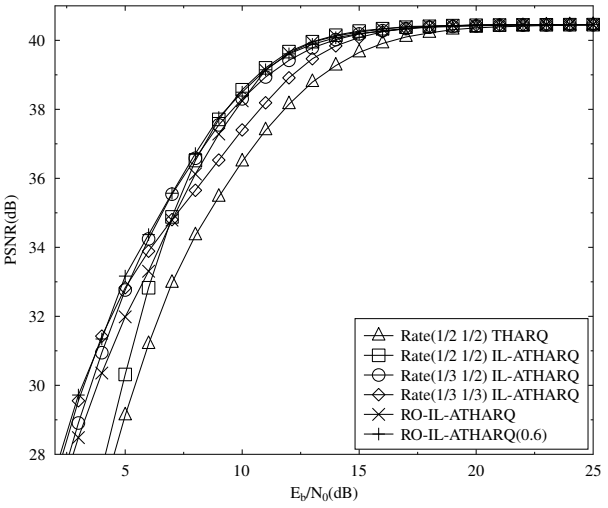
C. Optimized Coding Rates

In order to characterize the PSNR versus E_b/N_0 performance both of our proposed RO-IL-ATHARQ system and of the modified RO-IL-ATHARQ scheme, we compare them to two benchmarks, namely to a fixed-rate IL-ATHARQ scheme and to traditional THARQ transmission over block-fading non-dispersive uncorrelated Rayleigh channels, as listed in Table II. These comparisons are shown in Fig. 8, which were carried out using three different video sequences, namely the Football, the Soccer and the Crew sequences, as listed in Table I. Three different IL-ATHARQ transmission schemes denoted by $Rate\,(r_0, r_1)$ were simulated, namely the $Rate\,(1/2, 1/2)$, the $Rate\,(1/3, 1/2)$ and the $Rate\,(1/3, 1/3)$ schemes, where the $Rate\,(1/2, 1/2)$ scheme is the same as the IL-ATHARQ scheme we used in Section V-B.

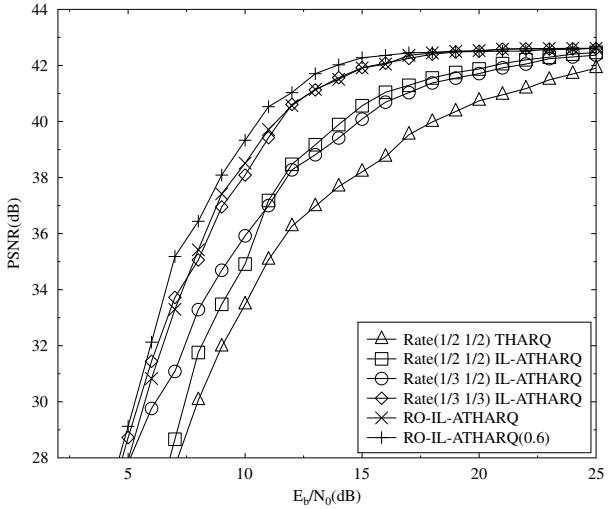
The results recorded for the Football sequence with the aid of three transmission TSs are shown in Fig. 8a. It can



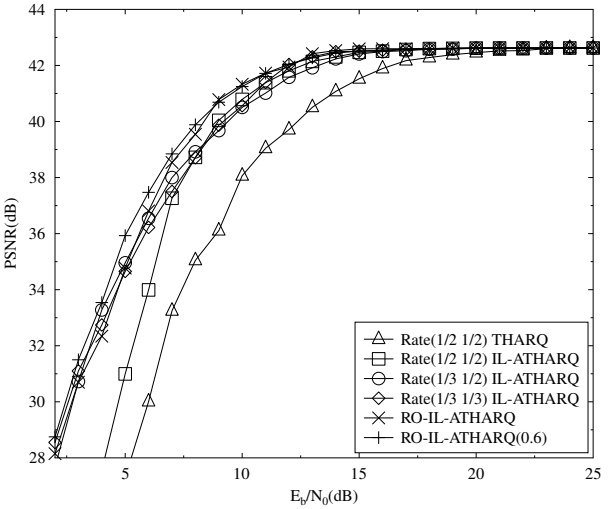
(a) PSNR vs E_b/N_0 for *Football with* $N_T = 3$



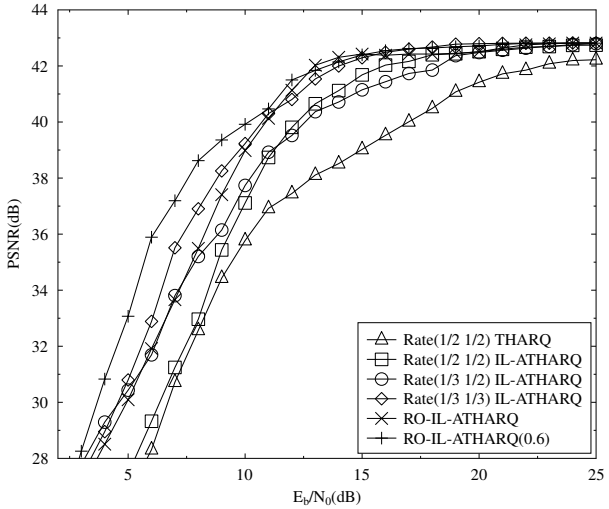
(b) PSNR vs E_b/N_0 for *Football with* $N_T = 4$



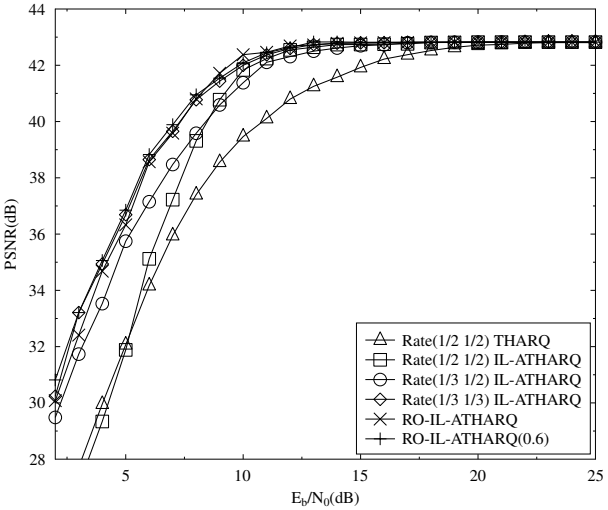
(c) PSNR vs E_b/N_0 for *Soccer with* $N_T = 3$



(d) PSNR vs E_b/N_0 for *Soccer with* $N_T = 4$



(e) PSNR vs E_b/N_0 for *Crew with* $N_T = 3$



(f) PSNR vs E_b/N_0 for *Crew with* $N_T = 4$

Fig. 8: PSNR versus E_b/N_0 performance of our proposed RO-IL-ATHARQ system and of the modified RO-IL-ATHARQ scheme in comparison to both the IL-ATHARQ transmission and to the traditional THARQ transmission as benchmarks. The Football, Soccer and Crew sequences are used for transmission over block-fading non-dispersive uncorrelated Rayleigh channels

Fig. 9: Comparison of decoded frames of the 26-th frame at E_b/N_0 of 10 dB for the *Soccer* sequences and $N_T = 3$. The upper five columns (from left to right) indicate frames of the original video, the rate($1/2, 1/2$) IL-THARQ(1 1) scheme, rate($1/2, 1/2$) THARQ scheme, the rate($1/2, 1/2$) IL-ATHARQ scheme and the RO-IL-ATHARQ(0.6) scheme, respectively. The lower row correspond to the difference frames between the top ones and the original video frame.

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

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

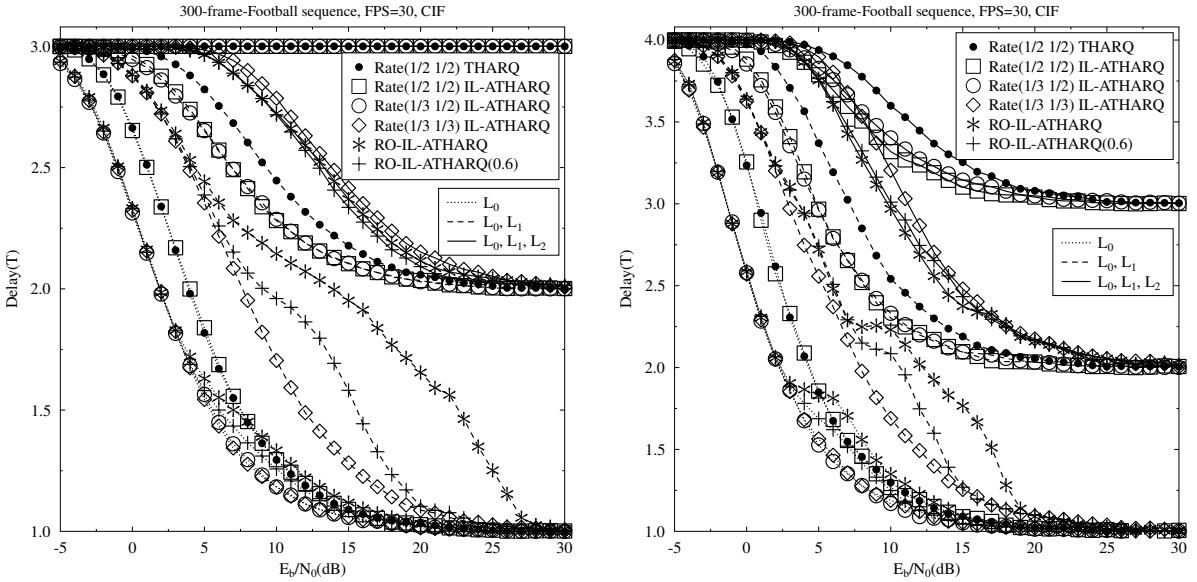
Similar trends can be observed, when the Soccer or Crew sequences are used, as shown in Fig. 8c to Fig. 8f. We infer from these results that our RO-IL-ATHARQ scheme is applicable to video sequences of diverse natures, and it is capable of achieving a beneficial performance gain for both $N_T = 3$ and 4. The subjective comparison of the decoded videos associated with our different regimes is discussed in Section V-D.

D. Subjective Comparison

Explicitly, Fig. 9 shows the subjective comparison of the decoded video frames associated with our different regimes using the Soccer sequence and $N_T = 3$ at the E_b/N_0 value of 10 dB. The 26-th frame of the recovered videos of some of our schemes are shown in the top row of Fig. 9. The rate($1/2, 1/2$) IL-THARQ(1 1) scheme is more error-prone according to Section V-B, and in this regime all three layers of this frame failed to be recovered, and so did all their preceding frames. The difference frame, which is obtained by subtracting the recovered frame from the 26-th frame of the original video, has substantial non-zero values. Continuing from left to right, we can observe that the frames corresponding to the rate($1/2, 1/2$) THARQ scheme, the rate($1/2, 1/2$) IL-ATHARQ scheme and the RO-IL-ATHARQ(0.6) scheme are becoming sharper and containing more intricate video details, while the corresponding difference frames having less and less non-zero values, which indicates the improvement of the video quality.

E. Transmission Delay

In Fig. 10, the average number of TSs required for receiving the i-th layer employing various transmission schemes is displayed. When $N_T = 3$ is used, the average number of TSs versus the E_b/N_0 characteristics are shown in Fig. 10a. When the $Rate\,(1/2, 1/2)$ THARQ, $Rate\,(1/2, 1/2)$ IL-ATHARQ and $Rate\,(1/3, 1/2)$ IL-ATHARQ schemes are used along with $N_T = 3$, three TSs are occupied, regardless of the E_b/N_0 value. On the other hand, observe in Fig. 10a that the $Rate\,(1/3, 1/3)$ IL-ATHARQ, RO-IL-ATHARQ and RO-IL-ATHARQ(0.6) schemes only use two TSs on average, in order to successfully receive all transmissions at high E_b/N_0 values. Furthermore, as seen in Fig. 10a, both the RO-IL-ATHARQ(0.6) and RO-IL-ATHARQ schemes require



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

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

Fig. 10: Average transmission time slots versus E_b/N_0 performance comparison of the THARQ, IL-ATHARQ, RO-IL-ATHARQ, and modified RO-IL-ATHARQ schemes for the *Football* sequence, transmitted over the quasi-static non-dispersive uncorrelated Rayleigh fading wireless channels

slightly less transmission TSs than the $Rate\,(1/3, 1/3)$ IL-ATHARQ at the same E_b/N_0 value. It can also be observed in Fig. 10a that for the first two layers, namely for L_0 and L_1, the $Rate\,(1/2, 1/2)$ THARQ, $Rate\,(1/2, 1/2)$ IL-ATHARQ and $Rate\,(1/3, 1/2)$ IL-ATHARQ schemes require two TSs on average at sufficiently high E_b/N_0 values, while the $Rate\,(1/3, 1/3)$ IL-ATHARQ, RO-IL-ATHARQ and RO-IL-ATHARQ(0.6) schemes only need one TS. It can also be inferred from Fig. 10a that the rate-optimized schemes occupy more TSs by successfully conveying L_0 and L_1. The reason behind this phenomenon is that the optimization algorithm strikes a more balanced compromise instead of assigning all the resources for protecting L_0 and L_1, where a reasonable reduction of the protection of L_1 can be compensated by successfully decoding both L_1 and L_2 in a single reception, if the latter one is well protected and ends up possessing high MI values. Finally, if we consider the transmission of L_0, we find from Fig. 10a that the $Rate\,(1/3, 1/2)$ and $Rate\,(1/3, 1/3)$ IL-ATHARQ, as well as the RO-IL-ATHARQ and RO-IL-ATHARQ(0.6) generally necessitates less transmission TSs, than the $Rate\,(1/2, 1/2)$ THARQ and $Rate\,(1/2, 1/2)$ IL-ATHARQ.

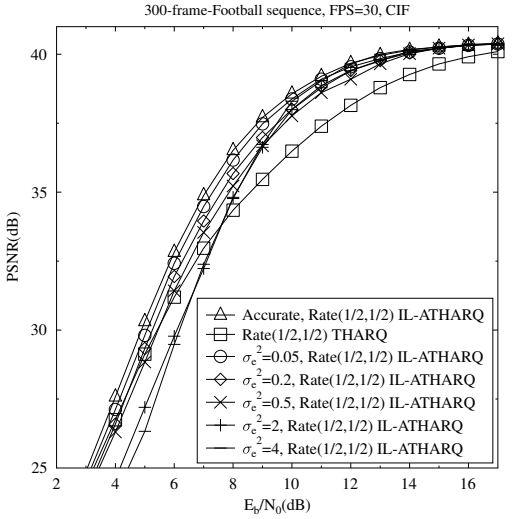
In Fig. 10b, the average number of TSs used versus the E_b/N_0 is portrayed for $N_T = 4$. Similar trends can be observed to those recorded in Fig. 10a for $N_T = 3$. Hence we conclude that the IL-ATHARQ is capable of efficiently reducing the number of TSs required for transmission, and the RO-IL-ATHARQ, although optimized for minimum distortion, additionally occupies less TSs.

F. Effect of Channel Prediction Errors

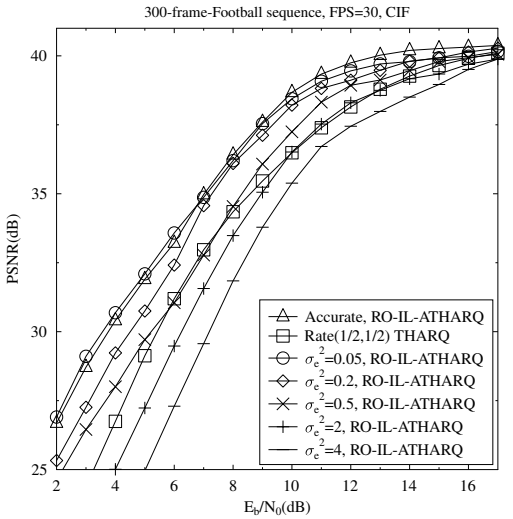
To demonstrate the effect of the channel prediction errors on the performance of our proposed system, we include the simulation results for both the IL-ATHARQ and RO-IL-ATHARQ(0.6) transmission schemes contaminated by channel prediction errors [58] in Fig. 11, which obeyed a Gaussian distribution.

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

As for the performance of the RO-IL-ATHARQ system, we can observe in Fig. 11b that the channel prediction error affected the systems more severely than for the IL-ATHARQ systems. At the E_b/N_0 of 6 dB, the PSNR performance associated with $\sigma_e^2 = 0.5$ is 2.4 dB worse than that of perfect channel prediction, while the system performance relying on $\sigma_e^2 = 4$ is 6.1 dB worse. At the E_b/N_0 of 11 dB, the PSNR performance of $\sigma_e^2 = 0.5$ is 1 dB lower than the one with perfect channel prediction, while that associated with $\sigma_e^2 = 4$ is 2.6 dB worse. The system's performance recorded for $\sigma_e^2 = 2$ or $\sigma_e^2 = 4$ is even worse than that of the THARQ benchmark



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



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

Fig. 11: PSNR versus E_b/N_0 performance in the presence of channel prediction errors, which affected the $Rate\,(1/2, 1/2)$ IL-ATHARQ, RO-IL-ATHARQ system for different σ_e^2 values against that of the THARQ system used as the benchmark.

system due to the severe error propagation imposed by the channel prediction. We may observe that in Fig. 11b the PSNR performance achieved with the aid of perfect channel prediction is not substantially better than the one associated with $\sigma_e^2 = 0.05$, when we have $E_b/N_0 \leq 6$ dB. Recall that the unmodified RO-IL-ATHARQ scheme of Section IV is subject to a certain level of PER estimation errors introduced by the algorithm itself and it is even more so in conjunction with larger N_T values because of the error propagation. The effect of the channel estimation error may be deemed comparable to that of the PER estimation errors, provided that it is not excessive. The interaction of these two types of errors may not be additive.

VI. Conclusions

We conceived an adaptive THARQ (ATHARQ) algorithm for IL-FEC coded layered video streaming for the sake of minimizing the video distortion under the constraint of a given total number of transmission TSs. The adaptive retransmission controller predicts the channel conditions and estimates the SNR values at the receiver for the sake of appropriately configuring the transmitter. The specific video layer, which would most effectively reduce the video distortion at the receiver is chosen for transmission. Furthermore, we developed an on-line optimization technique for our IL-ATHARQ transmission scheme, in order to find the most beneficial FEC code rate for each of the video layers that results in a reduced video distortion. A method of estimating the video distortions related to each code rate assignment was conceived for the IL-ATHARQ transmission.

Our simulation results demonstrated that the optimized IL-FEC system outperforms the traditional THARQ system by an E_b/N_0 value of about $5.3\ dB$ at a PSNR of $38.5\ dB$. Alternatively, about 2.5 dB of PSNR video quality improvement may be observed at an E_b/N_0 of 15 dB, when employing a RSC code.

In our future work, we will further develop our THARQ scheme for incremental redundancy aided schemes.

References

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

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

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

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

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

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

[7] A. Vetro, T. Wiegand, and G. Sullivan, "Overview of the stereo and multiview video coding extensions of the H.264/MPEG-4 AVC standard," *Proceedings of the IEEE*, vol. 99, pp. 626–642, April 2011.

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

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

[10] D. Chase, "A combined coding and modulation approach for communication over dispersive channels," *IEEE Transactions on Communications*, vol. 21, pp. 159–174, March 1973.

[11] I. Stanojev, O. Simeone, Y. Bar-Ness, and D. H. Kim, "Energy efficiency of non-collaborative and collaborative hybrid-ARQ protocols," *IEEE Transactions on Wireless Communications*, vol. 8, pp. 326–335, January 2009.

[12] J. Ramis and G. Femenias, "Cross-layer design of adaptive multirate wireless networks using truncated HARQ," *IEEE Transactions on Vehicular Technology*, vol. 60, pp. 944–954, March 2011.

[13] J. Ramis, G. Femenias, F. Riera-Palou, and L. Carrasco, "Cross-layer optimization of adaptive multi-rate wireless networks using truncated chase combining HARQ," in *IEEE Global Telecommunications Conference (GLOBECOM)*, pp. 1–6, December 2010.

[14] T. Kwon and D.-H. Cho, "Adaptive-modulation-and-coding-based transmission of control messages for resource allocation in mobile communication systems," *IEEE Transactions on Vehicular Technology*, vol. 58, pp. 2769–2782, July 2009.

[15] Q. Zhang and S. Kassam, "Hybrid ARQ with selective combining for fading channels," *IEEE Journal on Selected Areas in Communications*, vol. 17, pp. 867–880, May 1999.

[16] A. Majumda, D. Sachs, I. Kozintsev, K. Ramchandran, and M. Yeung, "Multicast and unicast real-time video streaming over wireless LANs," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, pp. 524–534, June 2002.

[17] Q. Zhang, Q. Guo, Q. Ni, W. Zhu, and Y.-Q. Zhang, "Sender-adaptive and receiver-driven layered multicast for scalable video over the internet," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, pp. 482–495, April 2005.

[18] Z. Liu, Z. Wu, P. Liu, H. Liu, and Y. Wang, "Layer bargaining: multicast layered video over wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 28, pp. 445–455, April 2010.

[19] M. Wu, S. Makharia, H. Liu, D. Li, and S. Mathur, "IPTV multicast over wireless LAN using merged hybrid ARQ with staggered adaptive FEC," *IEEE Transactions on Broadcasting*, vol. 55, pp. 363–374, June 2009.

[20] I. Bajic, "Efficient cross-layer error control for wireless video multicast," *IEEE Transactions on Broadcasting*, vol. 53, pp. 276–285, March 2007.

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

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

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

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

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

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

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

[28] D. Vukobratović, V. Stanković, D. Sejdinović, L. Stanković, and Z. Xiong, "Scalable video multicast using expanding window fountain codes," *IEEE Transactions on Multimedia*, vol. 11, pp. 1094–1104, October 2009.

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

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

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

[32] M. Halloush and H. Radha, "Network coding with multi-generation mixing: A generalized framework for practical network coding," *IEEE Transactions on Wireless Communications*, vol. 10, pp. 466–473, February 2011.

[33] F. Marx and J. Farah, "A novel approach to achieve unequal error protection for video transmission over 3G wireless networks," *Signal Processing: Image Communication*, vol. 19, pp. 313–323, April 2004.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

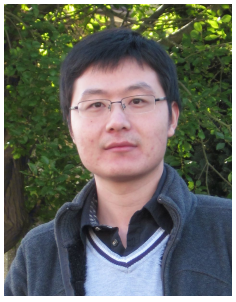
[53] A. Albanese, J. Blomer, J. Edmonds, M. Luby, and M. Sudan, "Priority encoding transmission," *IEEE Transactions on Information Theory*, vol. 42, pp. 1737–1744, November 1996.

[54] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Transactions on Information Theory*, vol. 20, pp. 284–287, March 1974.

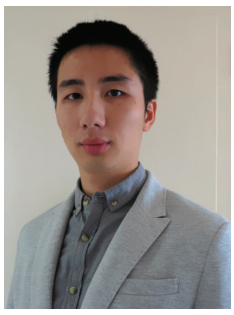
[55] L. Cheng, B. Henty, D. Stancil, F. Bai, and P. Mudalige, "Mobile vehicle-to-vehicle narrow-band channel measurement and characterization of the 5.9 ghz dedicated short range communication (dsrc) frequency band," *Selected Areas in Communications, IEEE Journal on*, vol. 25, pp. 1501–1516, October 2007.

[56] L. Hanzo, J. Blogh, and S. Ni, *3G, HSPA and FDD versus TDD networking : smart antennas and adaptive modulation*. Chichester: John Wiley & Sons [s.l.], 2008.
[57] L. L. Hanzo, Y. Akhtman, L. Wang, and M. Jiang, *MIMO-OFDM for LTE, WiFi and WiMAX: Coherent versus Non-coherent and Cooperative Turbo Transceivers*. John Wiley & Sons, IEEE press, October 2010.
[58] H. Gacanin, M. Salmela, and F. Adachi, "Performance analysis of analog network coding with imperfect channel estimation in a frequency-selective fading channel," *IEEE Transactions on Wireless Communications*, vol. 11, pp. 742–750, February 2012.
[59] H. Xiao, Q. Dai, X. Ji, and W. Zhu, "A novel JSCC framework with diversity-multiplexing-coding gain tradeoff for scalable video transmission over cooperative MIMO," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, pp. 994–1006, July 2010.
[60] A. Detti, G. Bianchi, C. Pisa, F. Proto, P. Loreti, W. Kellerer, S. Thakolsri, and J. Widmer, "SVEF: an open-source experimental evaluation framework for H.264 scalable video streaming," in *IEEE Symposium on Computers and Communications, ISCC*, (Sousse, Tunisia), pp. 36–41, July 2009.

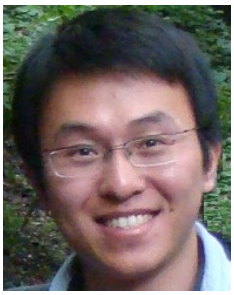
Chuan Zhu received the B.Eng. degree from Southeast University, Nanjing, China, and the M.Sc. degree (with distinction) in radio-frequency communication systems from the University of Southampton, Southampton, U.K., in 2010. He is currently working toward the Ph.D. degree with the Communications Research Group, School of Electronics and Computer Science, University of Southampton. His research interests include joint source–channel decoding, video compression and transmission, and extrinsic information transfer chart-aided turbo detection, as well as cooperative communications. Mr. Zhu received the Student Case Award toward his Ph.D. study from British Telecom, London, U.K.



Yongkai Huo received the B.Eng. degree with distinction in computer science and technology from Hefei University of Technology, Hefei, China, in 2006 and the M.Eng. degree in computer software and theory from University of Science and Technology of China, Hefei, China, in 2009. In 2014, he was awarded a Ph.D in Wireless Communications group, School of Electronics and Computer Science, University of Southampton, Southampton, UK, where he is currently working as a research fellow. He received a scholarship under the China-U.K. Scholarships for Excellence Programme. His research interests include distributed video coding, multiview video coding, robust wireless video streaming and joint source-channel decoding.



Bo Zhang received his B.Eng. degree in Information Engineering from National University of Defense Technology, China, in 2010. He received the Ph.D. degree in wireless communications from School of Electronics and Computer Science, University of Southampton, Southampton, U.K., in 2015. His research interests in wireless communications include design and analysis of cooperative communications, MIMO and network-coded systems.



Rong Zhang (M'09) received his PhD (Jun 09) from Southampton University, UK and his BSc (Jun 03) from Southeast University, China. Before doctorate, he was an engineer (Aug 03-July 04) at China Telecom and a research assistant (Jan 06-May 09) at Mobile Virtual Center of Excellence (MVCE), UK. After being a post-doctoral researcher (Aug 09-July 12) at Southampton University, he took industrial consulting leave (Aug 12-Jan 13) for Huawei Sweden R\& D as a system algorithms specialist. Since Feb 13, he has been appointed as a lecturer at CSPC group of ECS, Southampton University. He has 40+ journals in prestigious publication avenues (e.g. IEEE, OSA) and many more in major conference proceedings. He regularly serves as reviewer for IEEE transactions/journals and has been several times as TPC member/invited session chair of major conferences. He is the recipient of joint funding of MVCE and EPSRC and is also a visiting researcher under Worldwide University Network (WUN). More details can be found at http://www.ecs.soton.ac.uk/people/rz

Mohammed El-Hajjar is a lecturer in the Electronics and Computer Science in the University of Southampton. He received his BEng degree in Electrical Engineering from the American University of Beirut, Lebanon in 2004. He then received an MSc in Radio Frequency Communication Systems and PhD in Wireless Communications both from the University of Southampton, UK in 2005 and 2008, respectively. Following the PhD, he joined Imagination Technologies as a design engineer, where he worked on designing and developing Imagination's multi-standard communications platform, which resulted in three patents. In January 2012, he joined the Electronics and Computer Science in the University of Southampton as a lecturer in the Southampton Wireless research group. He is the recipient of several academic awards including the Dean's award for creative achievement, Dorothy Hodgkin postgraduate award and IEEE ICC 2010 Best paper award. He has published a Wiley-IEEE book and in excess of 50 journal and international conference papers. His research interests are mainly in the development of intelligent communications systems including energy-efficient transceiver design, cross-layer optimisation for large-scale networks, MIMO, millimetre wave communications and Radio over fibre systems.

Lajos Hanzo (http://www-mobile.ecs.soton.ac.uk) FREng, FIEEE, FIET, Fellow of EURASIP, DSc received his degree in electronics in 1976 and his doctorate in 1983. In 2009 he was awarded the honorary doctorate "Doctor Honoris Causa" by the Technical University of Budapest. During his 38-year career in telecommunications he has held various research and academic posts in Hungary, Germany and the UK. Since 1986 he has been with the School of Electronics and Computer Science, University of Southampton, UK, where he holds the chair in telecommunications. He has successfully supervised about 100 PhD students, co-authored 20 John Wiley/IEEE Press books on mobile radio communications totalling in excess of 10 000 pages, published 1400+ research entries at IEEE Xplore, acted both as TPC and General Chair of IEEE conferences, presented keynote lectures and has been awarded a number of distinctions. Currently he is directing a 100-strong academic research team, working on a range of research projects in the field of wireless multimedia communications sponsored by industry, the Engineering and Physical Sciences Research Council (EPSRC) UK, the European Research Council's Advanced Fellow Grant and the Royal Society's Wolfson Research Merit Award. He is an enthusiastic supporter of industrial and academic liaison and he offers a range of industrial courses. He is also a Governor of the IEEE VTS. During 2008 - 2012 he was the Editor-in-Chief of the IEEE Press and a Chaired Professor also at Tsinghua University, Beijing. His research is funded by the European Research Council's Senior Research Fellow Grant. For further information on research in progress and associated publications please refer to http://www-mobile.ecs.soton.ac.uk Lajos has 20 000+ citations.