Adaptive Cross-Packet HARQ

Mohammed Jabi, Abdellatif Benyouss, Mael Le Treust, Etienne Pierre-Doray, Leszek Szczecinski

Adaptive Cross-Packet HARQ

Mohammed Jabi, *Student Member, IEEE*, Abdellatif Benyouss, Maël Le Treust, *Member, IEEE*, Étienne Pierre-Doray, and Leszek Szczecinski, *Senior Member, IEEE*

***Abstract*—In this work, we investigate a coding strategy devised to increase the throughput in hybrid ARQ (HARQ) transmission over block fading channel. In our approach, the transmitter jointly encodes a variable number of bits for each round of HARQ. The parameters (rates) of this joint coding can vary and may be based on the negative acknowledgment (NACK) provided by the receiver or, on the past (outdated) information about the channel states. These new degrees of freedom allow us to improve the match between the codebook and the channel states experienced by the receiver. The results indicate that gains obtained using the proposed cross-packet coding strategy are particularly notable for large values of the throughput. In this region, the conventional HARQ fails to offer throughput improvement even if the number of transmission rounds is increased. We implement the proposed cross-packet HARQ using turbo codes where we show that the theoretically predicted throughput gains materialize in practice; the implementation challenges are also discussed.**

***Index terms*—** Block fading channels, Energy efficiency, Hybrid automatic repeat request, HARQ, Markov decision process, MDP, Rate adaptation, Throughput.

I. Introduction

This work is concerned with HARQ transmission over equal-length, block-fading channel when the instantaneous channel state information (CSI) is not available at the transmitter. Our goal is to propose a general coding strategy which will allow us to maximize the throughput in the operational range where the conventional HARQ fails to do so. To this end, we propose a joint coding of multiple information packets into the same channel block and develop methods to optimize the corresponding coding rates.

A. HARQ and the throughput

HARQ is used in communication systems to deal with unavoidable transmission errors caused by unpredictable changes in the channel (e.g., due to fading, in wireless transmission), or by the distortion of the transmitted signals (due to the noise or the interference). It relies on the feedback/acknowledgement channel, which is used by the receiver to inform the transmitter about the decoding errors (via NACK) and about the decoding success, via positive acknowledgment (ACK). After NACK, the transmitter makes a new transmission *round* which conveys additional information necessary to decode the packet. In *persistent* HARQ, this continues till ACK is received and then a new HARQ *cycle* starts again for another information packet. For *truncated* HARQ, the cycle stops also if a predefined number of rounds is attained.

As in many previous works, e.g., [1], [2], we will consider throughput as the performance measure assuming that residual errors (occurring due to truncation of the HARQ) are taken care of by the upper layers [3]. We consider here the "canonical" problem defined in [1], where the CSI is available at the receiver but not at the transmitter, which knows only its distribution. The essential part of HARQ is the channel coding, which is done over many channel blocks, whose number varies according to the feedback (ACK/NACK).

It was shown in [1] that the throughput of HARQ may approach the ergodic capacity of the channel with sufficiently high "nominal" coding rate per round. However, a large number of HARQ rounds is then required; this approach has thus a limited value: long buffers are necessary which becomes a limiting factor for implementation of HARQ [4]. On the other hand, using finite nominal coding rate and truncated HARQ, the gap between the throughput achievable using HARQ and the theoretical limits i.e., the ergodic capacity, may be large. This is especially the case, when we target throughput which is close to the nominal rate [2], [5]. Thus, there is an operation range where the conventional HARQ is not useful to increase the throughput.[1]

B. Adaptive HARQ

The disappointing throughput-wise behaviour of the conventional HARQ is due to the fact that the number of symbols attributed to each HARQ round and the number of transmitted information bits are fixed, and the only adaptation to the channel realizations is done varying the (integer) number of rounds.

To address this problem, various adaptive versions of HARQ were proposed in the literature. For example, [7]–[13] suggested to vary the length of the codewords transmitted over the blocks so as to strike the balance between the number of symbols used and the probability of successful decoding. Their obvious drawback is that decreasing the number of symbols may leave an "empty" space within the block. To deal with this issue, it was proposed to share the block between various packets e.g., in [3], [14]–[17], to encode many

M. Jabi, A. Benyouss, and L. Szczecinski are with INRS-EMT, University of Quebec, Montréal, QC, H5A1K6, Canada. e-mail: {benyouss,jabi,leszek}@emt.inrs.ca,

E. Pierre-Doray is with Polytechnique de Montréal, QC, H3T1J4, Canada; he was also with INRS-EMT when this work was carried out. e-mail: {etipdoray@gmail.com}

M. Le Treust is with ETIS - UMR 8051 / ENSEA - Université de Cergy-Pontoise - CNRS, 95011 Cergy-Pontoise cedex, France. e-mail: {mael.letreust@ensea.fr.}. This work was supported by ENSEA under grant BRQ-HARQ-2014 and conducted as part of the project Labex MME-DII (ANR11-LBX-0023-01).

Part of this work was presented at the IEEE Wireless Commun. Network. Conf. (WCNC), Doha, Qatar, April 2016.

[1] It is interesting to note that this conclusion becomes even more radical if we assume that the CSI is available at the transmitter. Then, the conventional HARQ may even *decrease* the throughput [6].

packets into predefined size blocks as done in [18], [19], or to group variable-length codewords to fill the channel block [12], [20]. A simplified approach was also proposed in [21], [22] to transmit the redundancy using two-step encoding of two packets.

C. Contributions and organization

The main insight we obtain from the works cited above, is that a joint coding of many packets into a single channel block is necessary. However, such a *cross-packet* (XP) coding was implemented up to now only implicitly. Here, we want to address this issue up front via cross-packet HARQ (XP-HARQ); the idea is to get rid of the restricting assumptions proper to various heuristics developed before. We want to use a generic joint encoder which explicitly encodes many information packets into a codeword filling the channel block. The challenge will be then to optimize the rates to be used in various HARQ rounds so as the throughput is maximized.

The contributions of this work are thus the following:

- **We propose a general framework to analyze a joint encoding/decoding of multiple packets for HARQ.** In the intermediate steps we also derive i) the relationship between the coding rates and the throughput and, ii) the decoding conditions in the case of joint encoding/decoding.
  Our approach to cross-packet coding is similar to the one shown in [23], which, however, did not optimize the coding parameters. The optimization was proposed in [24], however, due to complex decoding rules, it was very tedious and thus limited to the case of a simple channel model. In our work we simplify the problem assuming the use of capacity achieving codes; this leads to a compact description of the decoding criteria. Practical codes which implement XP-HARQ were proposed in [25]–[27] focusing, however, on increasing the reliability (not the throughout) without a formal optimization framework. We quickly note that, in the case of persistent XP-HARQ, the rate optimization problem was solved in [28].
- **We propose to adapt the coding rates to the CSI experienced by the receiver in the past transmission rounds of HARQ**. Using such *outdated* CSI simplifies the optimization problem which can be then solved using the Markov decision process (MDP) formulation [29, Chap. 4]. Moreover, it yields results which may be treated as the ultimate throughput limits of any XP-HARQ adaptation schemes when the instantaneous CSI is not available at the transmitter.
  To transmit the outdated CSI-related information a multi-bit feedback signaling must be implemented. This idea was already exploited e.g., in [3], [7], [10], [12], [13], [30]–[35].
- **We derive closed form expressions for the attainable throughput in particular cases.** This is possible when i) HARQ is limited to two rounds, or ii) the simplified rate-adaptation is applied. The advantage is that the closed forms are directly comparable to the attainable limit.

snr_k	SNR in the kth round, (1)
$\overline{\mathsf{snr}}$	average SNR; (10)
N_{s}	number of symbols in each block
Φ_k	encoding function at the kth round
R	nominal transmission rate in IR-HARQ
K	maximum number of HARQ rounds
R_k^{Σ}	accumulated transmission rate in XP-HARQ; (24)
I_k	mutual information (MI) in the kth round; Sec. II-A
I_k^{Σ}	accumulated MI after k rounds; (4)
f_k	probability of k successive errors; Sec. II-A
η_K^{ir}	throughput of IR-HARQ; (6)
$\overline{C}$	ergodic capacity; (7)
η_K^{xp}	throughout of XP-HARQ; (25)
$\hat{\eta}_K^{\mathrm{xp}}$	throughout of rate-adaptive XP-HARQ; (28)
$\tilde{\eta}_K^{\mathrm{xp}}$	throughout with heuristic rate adaptation; (37)
η_K^{ts}	throughput of TS-HARQ; Sec. IV-C

TABLE I
SUMMARY OF THE MAIN NOTATIONS

The remainder of the paper is organized as follows. We define the transmission model as well as the basic performance metrics in Sec. II. The idea of cross-packet coding is explained in Sec. III. The optimization of the rates in the cross-packet coding strategy is presented in Sec. IV. Short examples are presented throughout the work to illustrate the main ideas. Also, the example of using practical encoders/decoders is shown in Sec. V and we conclude the work in Sec. VI. The optimization methods used to obtain the numerical results and the proof of decoding conditions are presented in appendices.

II. Channel model and HARQ

We consider a point-to-point incremental redundancy HARQ (IR-HARQ) transmission of a packet m over a block fading channel. After each transmission, using a feedback/acknowledgement channel, the receiver tells the transmitter whether the decoding of m succeeded (ACK) or failed (NACK). We thus assume that error detection is possible (e.g., via cyclic redundancy check (CRC) mechanisms) and that the feedback channel is error-free. For simplicity, we ignore any loss of resources due to the CRC and the acknowledgement feedback.

The transmission of a single packet may thus require many transmission rounds which continue till the Kth round is reached or till ACK is received. When K is finite, we say that HARQ is *truncated*, otherwise we say it is *persistent*. We define a HARQ *cycle* as the sequence of transmission rounds of the same packet m.

The received signal in the kth round is given by

$$\boldsymbol{y}_k = \sqrt{\mathsf{snr}_k}\boldsymbol{x}_k + \boldsymbol{z}_k, \quad k = 1, \ldots, K \tag{1}$$

where $\boldsymbol{z}_k$ and $\boldsymbol{x}_k$ modelling, respectively, the noise and the transmitted codeword are N_{s}-dimensional vectors, each containing independent, identically distributed (i.i.d.) zero mean, unit-variance random variables; snr_k is thus the signal-to-noise ratio (SNR) at the receiver. The elements of $\boldsymbol{z}_k$ are drawn from a complex Gaussian distribution, and elements of $\boldsymbol{x}_k$ –

from the uniform distribution over the set (constellation) $\mathcal{X}$. For convenience, the most relevant notation is summarized in Table I.

During the kth round, snr_k is assumed to be perfectly known/estimated at the receiver and unknown at the transmitter; it varies from one round to another and we model $\mathsf{snr}_k, k = 1, \ldots, K$ as the i.i.d. random variables SNR with distribution $p_{\mathsf{SNR}}(\mathsf{snr})$, which is known at both the transmitter and the receiver.

A. Conventional HARQ

In the conventional IR-HARQ, a packet $\mathsf{m} \in \{0,1\}^{RN_\mathrm{s}}$ is firstly encoded into a codeword $\boldsymbol{x} = \Phi[\mathsf{m}] \in \mathcal{X}^{KN_\mathrm{s}}$ composed of KN_s complex symbols taken from a constellation $\mathcal{X}$ where $\Phi[\cdot]$ is the coding function and R denotes the *nominal* coding rate per block.[2] Then, the codeword $\boldsymbol{x}$ is divided into K disjoint subcodewords $\boldsymbol{x}_k$ of length N_s composed of different symbols, i.e., $\boldsymbol{x} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_K]$. After each round k, the receiver tries to decode the packet m

$$\hat{\mathsf{m}}_k = \mathrm{DEC}[\boldsymbol{y}_{[k]}] \tag{2}$$

using all received channel outcomes till the kth block

$$\boldsymbol{y}_{[k]} = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{k-1}, \boldsymbol{y}_k]. \tag{3}$$

Following [1], [31], we assume N_s large enough to make the random coding limits valid. Then, knowing the MI $I_k = \mathsf{I}(X_k; Y_k|\mathsf{snr}_k)$ between the random variables X_k and Y_k modeling respectively, the channel input and output in the kth block, allows us to determine when the decoding is successful or not. The decoding error, i.e., $\mathsf{ERR}_k \triangleq \{\hat{\mathsf{m}}_k \neq \mathsf{m}\}$, occurs in the kth round if the accumulated MI at the receiver is smaller than the coding rate, that is

$$\mathsf{ERR}_k = \{I_k^\Sigma < R\}, \tag{4}$$

where $I_k^\Sigma \triangleq \sum_{l=1}^k I_l$ is the MI accumulated in k rounds. Of course, the MI depends on the SNR, i.e., $I_k \equiv I_k(\mathsf{snr}_k)$. For large N_s, this idealization may be approached using the so-called capacity-achieving codes.

On the other hand, the NACK is observed if the sequence of errors occurs

$$\begin{aligned}\mathsf{NACK}_k &\triangleq \{\mathsf{ERR}_1 \wedge \mathsf{ERR}_2 \wedge \ldots \wedge \mathsf{ERR}_k\} \\ &= \left\{I_k^\Sigma < R\right\}, \end{aligned} \tag{5}$$

where we used the obvious implication $\{I_k^\Sigma < R\} \implies \{I_{k-1}^\Sigma < R\}$.

IR-HARQ can be modelled as a Markov process where the transmission rounds correspond to the states, and the HARQ cycle corresponds to a renewal cycle in the process. Thus, the long-term average throughput, defined as the average number of correctly received bits per transmitted symbol, may be calculated from the renewal-reward theorem: it is a ratio between the average reward (number of bits successfully decoded per cycle, normalized by N_s) and the average renewal time (the expected number of transmissions needed to deliver the packet with up to K transmission rounds) [1].

Let $f_k \triangleq \Pr\{\mathsf{NACK}_k\}, k \geq 1$ be the probability of k successive errors so the probability of *successful* decoding in the kth round is given by $\Pr\left\{\mathsf{NACK}_{k-1} \wedge \overline{\mathsf{ERR}_k}\right\} = f_{k-1} - f_k$ [1], where the event $\overline{E}$ is the complement of E. The throughput is then calculated as follows [1]

$$\begin{aligned}\eta_K^{\mathrm{ir}} &= \frac{R(1-f_1) + R(f_1 - f_2) + \ldots + R(f_{K-1} - f_K)}{1\cdot(1-f_1) + 2\cdot(f_1 - f_2) + \ldots + K\cdot(f_{K-1})} \\ &= \frac{R(1-f_K)}{1 + \sum_{k=1}^{K-1} f_k}. \end{aligned} \tag{6}$$

Because the instantaneous CSI is not available at the transmitter, the highest achievable throughput is given by the ergodic capacity[3] of the channel [1], [36]

$$\overline{C} \triangleq \mathbb{E}_{\mathsf{SNR}}[I(\mathsf{SNR})]. \tag{7}$$

However, achieving $\overline{C}$ is not obvious: as shown in [1], it can be done by growing simultaneously R and K to infinity but this approach is impractical due to large memory requirements: Large K means that many channel blocks $\boldsymbol{x}_k$ are transmitted and the observations $\boldsymbol{y}_k$– stored at the receiver. We emphasize that this problem arises independently from our assumption of using large N_s, which is introduced to simplify the expression of the decoding condition and to make the problem tractable.

To clarify the ideas we will use two running examples. The first one considers a two-state channel where the throughput results can be derived by hand, and the idea of the cross-packet coding – directly appreciated. In the second one, we assume the transmission is carried out using the 16-points quadrature amplitude modulation (16-QAM) constellation over Rayleigh fading channel. This more "practical" case should allow the reader to relate to the often used wireless communications models.

Example 1 (Two-states channel)**.** *Consider a block-fading channel where the MI can only take two values, I_a and I_b, where $\Pr\{I = I_\mathrm{a}\} = 1-p$ and $\Pr\{I = I_\mathrm{b}\} = p$. The ergodic capacity is given by $\overline{C} = I_\mathrm{a}(1-p) + I_\mathrm{b}p$. We force the HARQ to deliver the packet by the last transmission, i.e., $f_K = 0$, which means that we impose the constraints on the coding rate $R \leq KI_\mathrm{a}$ if we assume that $I_\mathrm{a} < I_\mathrm{b}$.*

Let $I_\mathrm{a} = 1, I_\mathrm{b} = 1.5$, and $p = 0.75$. We obtain then $\overline{C} = 1.375$ and, for $K = 2, 3$ we calculate the throughput[4] as

$$\eta_2^{\mathrm{ir}} = \begin{cases} R, & \text{if} \quad R \leq 1 \\ 0.8R, & \text{if} \quad 1 < R \leq 1.5 \\ 0.5R, & \text{if} \quad 1.5 < R \leq 2 \end{cases}, \tag{8}$$

and

$$\eta_3^{\mathrm{ir}} = \begin{cases} \eta_2^{\mathrm{ir}}, & \text{if} \quad R \leq 2 \\ 0.48R, & \text{if} \quad 2 < R \leq 2.5 \\ 0.41R, & \text{if} \quad 2.5 < R \leq 3 \end{cases}. \tag{9}$$

[2]We define the nominal rate as the coding rate per channel block because HARQ is a variable-rate transmission: the number of used channel blocks is random so the final transmission rate is random as well.

[3]We use the term "capacity" to denote the achievable rate for a *given* distribution of X.

[4]For $R \leq 1$ we obtain $f_1 = 0$. For $1 < R \leq 1.5$ we have $f_1 = 1 - p$ and $f_2 = 0$. Setting $1.5 < R \leq 2$ yields $f_1 = 1$, $f_2 = 0$, etc.

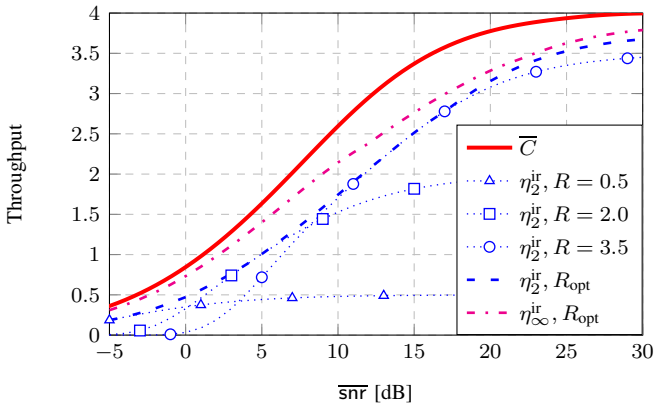


Fig. 1. Throughput of the conventional IR-HARQ, compared to the ergodic capacity, $\overline{C}$, in Rayleigh block-fading channel with 16-QAM modulation. The R_{opt} curve is an envelope of the throughputs η_K^{ir} obtained with different coding rates per block $R \in \{0.25, 0.5, \ldots, 7.75\}$.

The optimum throughput-rate pairs are then $(\eta_2^{\mathrm{ir}} = 1.2, R = 1.5)$ *and* $(\eta_3^{\mathrm{ir}} = 1.23, R = 3)$. *First, the benefit of using HARQ is clear: we are able to transmit without errors with a finite number of channel blocks and go beyond the obvious limit of* I_{a}. *Second, we note that for* $K = 2$, *after two transmissions, the accumulated MI always satisfies* $I_2^{\Sigma} \geq 2$, *while the condition* $I_2^{\Sigma} \geq 1.5$ *is sufficient to decode the packet. This may be seen as a "waste" which will be removed with the idea of cross-packet coding introduced in Sec. III.*

Example 2 (16-QAM over Rayleigh fading channel)**.** *Assume now that the transmission is done using symbols drawn uniformly from 16-QAM constellation* $\mathcal{X}$ *[37, Ch. 2.5] and that the channel gains follow Rayleigh distribution, i.e.,*

$$p_{\mathsf{SNR}}(\mathsf{snr}) = 1/\overline{\mathsf{snr}} \exp(-\mathsf{snr}/\overline{\mathsf{snr}}). \tag{10}$$

In order to compute the throughput of IR-HARQ, we need to calculate $I_k = I(\mathsf{SNR}_k)$ *and evaluate the probability* $f_k = \Pr\left\{I_k^{\Sigma} < R\right\}$*; the former is done using the numerical method shown in [37, Ch. 4.5], the latter – using the Monte-Carlo integration.*[5]

The average $\overline{C}$ *is compared with the throughput* η_K^{ir} *when* $K \in \{2, \infty\}$[6] *in Fig. 1. We show the throughput obtained with the fixed nominal rate* $R \in \{0.5, 2.0, 3.5\}$, *as well as the the achievable throughput when maximizing over the nominal rate* R, *these results are labeled by* R_{opt}. *In other words, we consider also the possibility of varying the nominal rate in function of the average SNR,* $\overline{\mathsf{snr}}$.

The results indicate that i) there is a significant loss with respect to the ergodic capacity when using truncated HARQ, and ii) increasing the number of transmission rounds ($K = \infty$) *helps recovering the loss for a small-medium range of throughput (e.g., for* $\eta^{\mathrm{ir}} = 1$ *we gain* ~ 3dB *and the gap to* $\overline{C}$ *is less than* 1dB*), but it is less useful in the region of high* η_K^{ir}*, i.e., in the vicinity of the maximum attainable throughput (e.g., for* $\eta^{\mathrm{ir}} = 3$*, we gain* 1dB *but the gap to* $\overline{C}$ *is still* ~ 5dB*). We highlight this well-known effect [2] to emphasize later the gains of the new coding strategy.*

[5]The alternative numerical method shown in [2, App. B] is difficult to implement as it requires evaluation of the distribution of $p_{I_k}(x)$ which has a singularity for $x \to 4$.

[6]For truncated HARQ, η_K^{ir} can also be computed without evaluating f_k using the proposed method in Appendix B by taking the policy $\pi(\mathsf{s}) = R$ if $\mathsf{s} \in \{(0,0,0,\mathsf{ACK}), (0,0,0,\mathsf{NACK})\}$ and $\pi(\mathsf{s}) = 0$ otherwise. $\eta_\infty^{\mathrm{ir}}$ can be calculated by taking K large enough in (6) as suggested in [2] or by evaluating the throughput using the method outlined in the Appendix B and considering the policy $\pi(\mathsf{s}) = R$ if $\mathsf{s} = (0,0)$ and $\pi(\mathsf{s}) = 0$ otherwise. We opt for the latter method.

III. Cross-packet HARQ

The examples shown previously indicate that the conventional coding cannot bring the throughput of HARQ close to the capacity unless the nominal coding rate R and the number of rounds K increase. We would like now to exploit a new coding possibility consisting in joint coding of packets during the HARQ cycle.

Let us start with the case of two transmission rounds. In the first round, we use the nominal rate R_1, i.e., the packet $\mathsf{m}_1 \in \{0,1\}^{R_1 N_{\mathrm{s}}}$ is encoded as

$$\boldsymbol{x}_1 = \Phi_1[\mathsf{m}_1] \in \mathcal{X}^{N_{\mathrm{s}}}, \tag{11}$$

and transmitted over the channel (1) yielding $\boldsymbol{y}_1 = \sqrt{\mathsf{snr}_1}\boldsymbol{x}_1 + \boldsymbol{z}_1$, where $\Phi_k[\cdot]$ denotes the encoding operation at the kth round. As in the conventional IR-HARQ, we may adjust R_1 according to the average SNR.

The decoding in the first round follows (2), and if $\hat{\mathsf{m}}_1 = \mathsf{m}_1$ (i.e., the packet m_1 is decoded correctly, which occurs if $I_1 \geq R_1$), a new cycle HARQ starts. However, if the decoding fails ($\mathsf{ERR}_1 = \{\hat{\mathsf{m}}_1 \neq \mathsf{m}_1\}$), we construct a longer packet concatenating m_1 with a packet m_2 which contains $R_2 N_{\mathrm{s}}$ new bits

$$\mathsf{m}_{[2]} \triangleq [\mathsf{m}_1, \mathsf{m}_2] \in \mathbb{B}^{(R_1+R_2)N_{\mathrm{s}}}. \tag{12}$$

It is next encoded

$$\boldsymbol{x}_2 = \Phi_2[\mathsf{m}_{[2]}] \in \mathcal{X}^{N_{\mathrm{s}}}, \tag{13}$$

and transmitted yielding $\boldsymbol{y}_2 = \sqrt{\mathsf{snr}_2}\boldsymbol{x}_2 + \boldsymbol{z}_2$ as shown in Fig. 2.

After this second round, the receiver tries to decode the packet $\mathsf{m}_{[2]}$ using the observations $\boldsymbol{y}_{[2]} = [\boldsymbol{y}_1, \boldsymbol{y}_2]$

$$\hat{\mathsf{m}}_{[2]} = \mathrm{DEC}[\boldsymbol{y}_{[2]}]. \tag{14}$$

We assume that the codebook used in Φ_2 is designed independently of the codebook used in Φ_1. This coding strategy is introduced without any claim of optimality but has the undeniable advantage of being simple to implement. We note that the idea of using Φ_2 independent of Φ_1 was also proposed in [23], [24].

The decoding of $\mathsf{m}_{[2]}$ based on the channel outcomes $\boldsymbol{y}_{[2]}$ succeeds if

$$I_2^{\Sigma} \geq R_2^{\Sigma} = R_1 + R_2, \tag{15}$$

$$I_2 \geq R_2, \tag{16}$$

where (15) is a constraint over the sum-rate that guarantees the joint decoding of the packet $\mathsf{m}_{[2]} = [\mathsf{m}_1, \mathsf{m}_2]$. We need the condition (16) to ensure the correct decoding of the packet

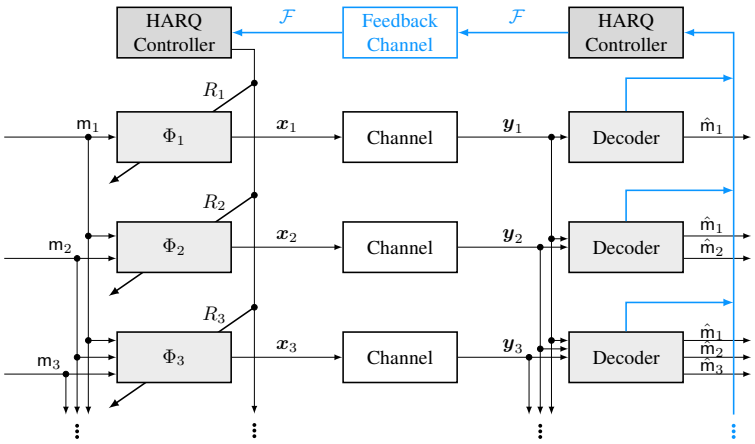


Fig. 2. Model of the adaptive XP-HARQ transmission: the HARQ controller uses the information $\mathcal{F}$ obtained over the feedback channel to choose the rate for the next round; $\mathcal{F}$ represent ACK/NACK acknowledgement in the case of one bit feedback, or, it carries the index of the coding rate in the case of rate-adaptive transmission (Sec. IV-A).

m_2. This means, the MI must be accumulated to decode each of the packets even though the decoding is done jointly. The formal proof of (15) and (16) is presented in the Appendix A. Similar decoding conditions were presented in the context of physical layer (PHY) security in [38].

We can now compare (15) with the decoding condition $I_2^{\Sigma} > R_1$ obtained in the conventional IR-HARQ, see (5). Intuitively, by introducing rate-R_2 message m_2 we want to prevent the "waste" of MI, which happens if, after the second round, I_2^{Σ} is much larger than R_1, cf. Example 1.

The operation described above can be then generalized as follows for all the rounds $k = 1, \dots, K$:

- At the transmitter
  1) If ACK was received in the previous round or the K rounds are terminated, start a new cycle ($k = 1$): transmit the packet m_1 encoded with the rate R_1; see (11).
  2) In the round $k \in (2, K)$, if NACK was received in the previous round, concatenate the packet $\mathsf{m}_k \in \mathbb{B}^{N_s R_k}$ with $\mathsf{m}_{[k-1]}$ to form $\mathsf{m}_{[k]}$ which is encoded into $\boldsymbol{x}_k$

$$\mathsf{m}_{[k]} = [\mathsf{m}_{[k-1]}, \mathsf{m}_k] \in \mathbb{B}^{N_s R_k^{\Sigma}}, \tag{17}$$

$$\boldsymbol{x}_k = \Phi_k[\mathsf{m}_{[k]}], \tag{18}$$

  where the coding rate is defined as $R_k^{\Sigma} = R_{k-1}^{\Sigma} + R_k$.
- At the receiver
  1) Decode the packet $\hat{\mathsf{m}}_k = \mathrm{DEC}[\boldsymbol{y}_{[k]}]$.
  2) If $\hat{\mathsf{m}}_k \neq \mathsf{m}_{[k]}$ (i.e., decoding failed): transmit NACK, and
     - If $k < K$: store $\boldsymbol{y}_{[k]}$ for the decoding in the next round;
     - If $k = K$: remove $\boldsymbol{y}_{[k]}$;
  3) If $\hat{\mathsf{m}}_k = \mathsf{m}_{[k]}$ (decoding succeeded): transmit ACK to the transmitter, and remove $\boldsymbol{y}_{[k]}$.

The rate R_k are computed off-line and are saved in a table at the receiver and transmitter; they depend only on the average SNR, $\overline{\mathsf{snr}}$, which is known at both communicating parties.

A. Throughput

To calculate the throughput of XP-HARQ, we adopt a similar approach as in (6) but first, we need to generalize the decoding conditions (15) and (16). This can be done by redefining the events NACK_k.

Starting from $k = 2$, we see that while the event NACK_1 remains unchanged with respect to the conventional coding, the event NACK_2 means that NACK_1 and ERR_2 were observed, the latter means that (15) and (16) are not satisfied, that is

$$\begin{aligned}
\mathsf{NACK}_2 &= \Big\{\mathsf{NACK}_1 \wedge \mathsf{ERR}_2\Big\} && (19)\\
&= \Big\{\big(I_1 < R_1\big) \wedge \overline{\big((I_2^{\Sigma} \geq R_2^{\Sigma}) \wedge (I_2 \geq R_2)\big)}\Big\}\\
&= \Big\{\big(I_1 < R_1\big) \wedge \big((I_2^{\Sigma} < R_2^{\Sigma}) \vee (I_2 < R_2)\big)\Big\} && (20)\\
&= \Big\{\big(I_1 < R_1\big) \wedge \big(I_2^{\Sigma} < R_2^{\Sigma}\big)\Big\}. && (21)
\end{aligned}$$

To pass from (20) to (21) we used the decoding failure implication

$$\{I_1 < R_1 \wedge I_2 < R_2\} \implies \{I_1 < R_1 \wedge I_2^{\Sigma} < R_2^{\Sigma}\}, \tag{22}$$

which means that NACK_1 combined with (15) implies (16).

For any $k > 1$ we generalize (21) as follows

$$\mathsf{NACK}_k = \{\mathsf{NACK}_{k-1} \wedge \big(I_k^{\Sigma} < R_k^{\Sigma}\big)\} \tag{23}$$

where

$$R_k^{\Sigma} \triangleq \sum_{l=1}^{k} R_l \tag{24}$$

is the accumulated transmission rate.

We can now appreciate the difference in the conditions defining NACK_k (23) for XP-HARQ and those defined in (5) for the conventional IR-HARQ. While in the latter, the probability of NACK depends only on I_k^{Σ}, in the former, we must take into account all values of $I_l^{\Sigma}, l = 1, \dots, k$.

To evaluation the throughput we also need to take into account the fact that the reward in the kth transmission round is given by R_k^{Σ}. So, the throughput is calculated as

$$\eta_K^{\text{xp}} = \frac{R_1^{\Sigma}(1-f_1) + R_2^{\Sigma}(f_1-f_2) + \ldots + R_K^{\Sigma}(f_{K-1}-f_K)}{(1-f_1) + 2\cdot(f_1-f_2) + \ldots + K\cdot(f_{K-1})} = \frac{\sum_{k=1}^{K} R_k\big(f_{k-1}-f_K\big)}{1+\sum_{k=1}^{K-1} f_k}, \tag{25}$$

where $R_1^{\Sigma} = R_1$ and, again, $f_k = \Pr\{\mathsf{NACK}_k\}, k \geq 1$ with NACK_k defined by (23).

As a sanity check we can set $R_k = 0, k = 2, \ldots, K$, and recover the conventional "single-packet" IR-HARQ, i.e., (25) will be equivalent to (6).

The fundamental difference of the proposed XP-HARQ with respect to the conventional IR-HARQ appears now clearly in the numerator of (25) which expresses the idea of variable rate transmission due to encoding of multiple packets. Of course, not only the numerator changed with respect to (6) but also the denominator is different due to the new definition of NACK_k in (23). To take advantage of the new degrees of freedom offered by XP-HARQ, we have to find the throughput-maximizing rates $R_1, \ldots, R_K$. This issue will be addressed in Sec. IV.

We emphasize here, that while we have to store the outcomes $\boldsymbol{y}_k$ at the receiver, this memory requirement is the same as in the case of the conventional IR-HARQ. Of course, in practice, complexity and memory requirement increase at the decoding stage because there is simply more information bits to be recovered. On the other hand, the encoding function may be implemented in many different ways but, in general, we do not need to store the codewords $\boldsymbol{x}_k$ at the transmitter.

Example 3 (Two-state channel and XP-HARQ)**.** *We consider now the proposed XP-HARQ in the scenario of Example 1. Let us start, as before, with $K = 2$ and $R_1 = 1.5$. After a decoding failure (which means that we obtained $I_1 = I_{\text{a}} = 1$), we are free to define any rate R_2. In the absence of any formal criterion (more on that in Sec. IV), we take the following auxiliary (and somewhat ad-hoc) condition: we want to guarantee a non-zero successful decoding probability, i.e., $f_2 < 1$. Here, since $I_2^{\Sigma} \in (2, 2.5)$, any $R_2 \leq 1$ can ensure that $f_2 < 1$. In particular, if the rate $R_2 \leq 0.5$ we guarantee a much stronger condition $f_2 = 0$.*

For the case when $K = 2$ and using $R_2 = 0.5$, we obtain $f_1 = 0.25$ and $f_2 = 0$. The throughput is then given by

$$\eta_2^{\text{xp}} = \frac{R_1 + 0.25R_2}{1+0.25} = 1.3. \tag{26}$$

Thus, we used exactly the same channel resources as in the conventional HARQ, obtained the same guarantee of successful decoding ($f_2 = 0$) after two transmission rounds, but the throughput is larger.

The difference is that, while we still have $I_2^{\Sigma} \in (2, 2.5)$, we now use $R_2^{\Sigma} = 2$ to eliminated the "waste" of MI in the conventional IR-HARQ, where $R_2^{\Sigma} = 1.5$. The improvement may be seen as the increase in the throughput (from $\eta_2^{\text{ir}} = 1.2$ to $\eta_2^{\text{xp}} = 1.3$) or as the reduction in the memory requirements (i.e., we obtain a better throughput with smaller K, see $\eta_3^{\text{ir}} = 1.23$ in Example 1). The price to pay for this advantage is the possible increase in the complexity of cross-packet encoding/decoding.

Similarly, for $K = 3$, we can use the larger value of R_2 (that guarantees our objective of decodability, $f_2 < 1$), i.e., $R_2 = 1$. In this case, $f_1 = 0.25$, and $f_2 = \Pr\{I_1 < 1.5 \wedge I_2^{\Sigma} < 2.5\} = 0.0625$. After the third transmission we observe $I_3^{\Sigma} \in (3, 3.5)$ so, using $R_3 = 0.5$, we obtain $f_3 = 0$ and thus the throughput is calculated as

$$\eta_3^{\text{xp}} = \frac{R_1 + 0.25R_2 + 0.0625R_3}{1+0.25+0.0625} \approx 1.36, \tag{27}$$

which is already quite close to $\overline{C} = 1.375$.

The improvement of the throughput in XP-HARQ is due to the way the codebook is constructed. While the conventional IR-HARQ, see Sec. II-A, is blind to the channel realizations, in XP-HARQ we match the information content of the codebook following the outcome of the transmissions.

IV. Optimization of the coding rates

Our goal now is to evaluate how well the XP-HARQ can perform. To this end, we will have to find the optimal coding rates $R_1, R_2, \ldots, R_K$ which maximize throughput (25).

Since the objective function is highly non linear, we will use the exhaustive search: for a truncated HARQ this can be done with a manageable complexity.

Example 4 (16QAM, Rayleigh fading – continued)**.** *In Fig. 3 we show the results of the exhaustive-search optimization of η_K^{xp} with η_K^{ir}; for implementability, we limited the search space: IR-HARQ uses $R_1 \in \{0, 0.25, \ldots, 3.75\}$ and XP-HARQ uses rates which satisfy $R_K^{\Sigma} \leq R_{\max}$, with $R_{\max} = 8$; $R_1 \in \{0.25, \ldots, 3.75\}$, $R_k \in \{0, 0.25, \ldots, 3.75\}$ $\forall k \in \{2, \ldots, K\}$.*

We used here an additional constraint which requires each transmission to have non zero probability of being decodable, that is $R_k < \log_2 M, \forall k = 1, \ldots, K$, where $M = 16$. In fact, these constraints were always satisfied in XP-HARQ so they only affect IR-HARQ; we will relax them in the next example.

In terms of SNR required to attain $\eta = 3$, the gain of XP-HARQ over IR-HARQ varies from 1.5dB *(for $K = 2$) to* 2.5dB *(for $K = 3$). For $\overline{\mathsf{snr}} \leq 5$ [dB], η_K^{xp} is practically equal to η_K^{ir} while the gap between them decreases significantly for $\overline{\mathsf{snr}} > 25$ [dB]. So, in order to highlight the differences among the results, we only show them in the interval of SNR* $(5, 25)$dB.

For low SNR, after optimization, we obtained $R_k = 0, k = 2, \ldots, K$, which means that the XP-HARQ is, de facto, transformed into IR-HARQ, see also the comment after (25)*. These effect will reappear throughout the paper in various examples: the results of IR-HARQ and XP-HARQ will be always similar in low SNR. The intuitive understanding is that, the observed mutual information I_1 is likely to be much smaller than R_1 and thus, to improve the decoding chances it is necessary to fill the channel blocks with incremental redundancy for the packet m_1; in other words, $R_k^{\Sigma} = R_1$, and XP-HARQ is equivalent to IR-HARQ.*

A. Rate adaptation

The possibility of varying the rates during the HARQ cycle opens a new optimization space. Here, we want to exploit it

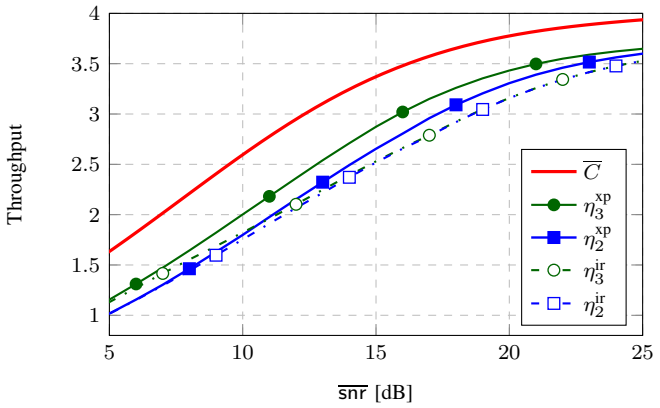


Fig. 3. Throughput of the conventional IR-HARQ (η_K^{ir}) compared to XP-HARQ (η_K^{xp}) in Rayleigh block-fading channel. The ergodic capacity ($\overline{C}$) is shown for reference.

following the idea of adapting the transmission parameters in HARQ on the basis of outdated CSI, as it was considered, e.g., in [7], [9]–[12], [33], [39].

The idea is to *adapt* the coding rates R_k using outdated CSIs, $I_1, I_2, \ldots, I_{k-1}$. This concept remains compatible with the assumption of transmitter operating without CSI knowledge because the outdated CSIs $I_1, I_2, \ldots, I_{k-1}$ cannot be used in the kth round to infer anything about I_k (due to i.i.d. model of the SNRs). On the other hand, we must assume that we can communicate over the feedback channel more than one-bit ACK/NACK messages. The assumption of such a multi-bit feedback not only simplifies the optimization but also yields the results which may be treated as the ultimate performance limits of any adaptation schemes when the instantaneous CSI is not available at the transmitter.

Using this approach, the rate R_k will depend on the MIs $I_1, \ldots, I_{k-1}$ but also – on the past rates $R_1^{\Sigma}, \ldots, R_{k-1}^{\Sigma}$ which determine the probability of the decoding success, see (23). This recursive dependence may be dealt with using the MDP framework, where the states of the Markov chain not only indicate the transmission number but also gather all information necessary to decide on the rate—which, in the language of the MDP, is called an *action*— taken from the predefined *action space* $\mathcal{R}$. The state has to be defined so that i) knowing the action (chosen rate), the state-transition probability can be determined after each transmission, and ii) the reward may be calculated knowing the state and the action. The state defined as a pair $\mathsf{s}_k = (R_k^{\Sigma}, I_k^{\Sigma})$ satisfies these two requirements, where we only need to consider the pairs which satisfy $R_k^{\Sigma} > I_k^{\Sigma}$, otherwise the decoding is successful and the HARQ cycle terminates.

Therefore, the rates R_k are now functions (called *policies*) of the states s_k, i.e., $R_k = R_k(\mathsf{s}_{k-1})$, and the rate adaptation problem consists in finding the throughput-maximizing policies. We calculate the throughput generalizing the expression (25)

$$\hat{\eta}_K^{\mathrm{xp}} = \frac{\mathbb{E}\big[\sum_{k=1}^{K} \xi_k R_k^{\Sigma}\big]}{1 + \sum_{k=1}^{K-1} f_k}, \tag{28}$$

where

$$\xi_k = \mathbb{I}\big[I_1 < R_1 \wedge \ldots \wedge I_{k-1}^{\Sigma} < R_{k-1}^{\Sigma} \wedge I_k^{\Sigma} \geq R_k^{\Sigma}\big], \tag{29}$$

indicates the successful decoding in the kth round, and

$$R_k^{\Sigma} = R_{k-1}^{\Sigma} + R_k(\mathsf{s}_{k-1}) \tag{30}$$

is the accumulated rate depending in a recursive fashion on the states of the Markov chain. The probability of k successive errors, f_k, may be expressed as (23) considering the dependence of the rates on the states given by (30). All the expectations are taken with respect to the states – or equivalently – with respect to $I_1, \ldots, I_K$.

The expression (28) will be useful in Sec. IV-B, however, its maximization with respect to the policies $R_k(\mathsf{s}_{k-1}), k = 1, \ldots, K$ will be done using efficient specialized algorithms as explained in Appendix B. In the particular case of two HARQ rounds ($K = 2$), the optimal rate adaptation policy can be derived in closed form as shown in Appendix C.

The operation of the adaptive XP-HARQ is the same as we described it between (17) and (24): the optimal rate adaptation policy is again computed off-line and is saved in tables at the receiver and the transmitter. The only difference with what we assumed before is that the tables depend not only on the average SNR (as it was the case for the non-adaptive XP-HARQ) but also on the accumulated rate, R_k^{Σ} and on the accumulated MI, I_k^{Σ}.

The MI is available/measured at the receiver so the adaptation is done by the receiver which i) finds the index of the table entry of R_k and ii) sends the index to the transmitter over the feedback channel. In this manner, the feedback overhead depends only on a *number* of available rates and not on their *values*.

Therefore, even if the outdated MI, I_k^{Σ} is discretized with a high resolution when optimizing the throughput (cf. Appendix B), the feedback load is affected by the cardinality of the action space, $\mathcal{R}$.

Example 5 (16QAM, Rayleigh fading channel – continued)**.** *To run the optimization algorithms outlined in Appendix B, we need to discretize the variables involved (states and actions). As for the rates (actions), we use a relatively course discretization step equal to* 0.25 *and define the action space as* $\mathcal{R} = \{0.25, 0.5, \ldots, R_{\max}\}$*, where we also enforce an additional constraint* $R_K^{\Sigma} \leq R_{\max}$*, which is compatible with the setup used in Example 4. As we will show later, the results are notably affected by* $R_{\max}$*; however, using a finer discretization step did not change them significantly. For instance, a minor enhancement appears only in high values of SNR (*$\overline{\mathsf{snr}} \geq$ 20dB*) when using* $\mathcal{R} = \{0.125, 0.25, \ldots, 8\}$ *instead of* $\mathcal{R} = \{0.25, 0.5, \ldots, 8\}$*.*

The throughput of adaptive XP-HARQ, $\hat{\eta}^{\mathrm{xp}}$*, is compared to the throughput of the conventional IR-HARQ in Fig. 4 for* $K = \infty$*, while Fig. 5 shows the comparison for truncated HARQ.*

Here, for IR-HARQ, we removed the constraints on the initial coding rate, $R_1 < \log_2 M$*, which were applied in Example 4. It allows us to increase the throughput* η_3^{ir} *at the cost of first transmission not being decodable. In our*

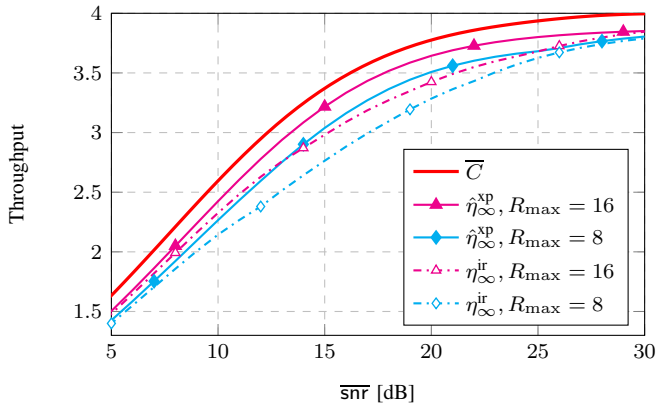


Fig. 4. Optimal throughput of the conventional IR-HARQ ($\eta^{\mathrm{ir}}_{\infty}$) compared to the proposed XP-HARQ ($\hat{\eta}^{\mathrm{xp}}_{\infty}$) in Rayleigh block-fading channel. The ergodic capacity ($\overline{C}$) is shown for reference.

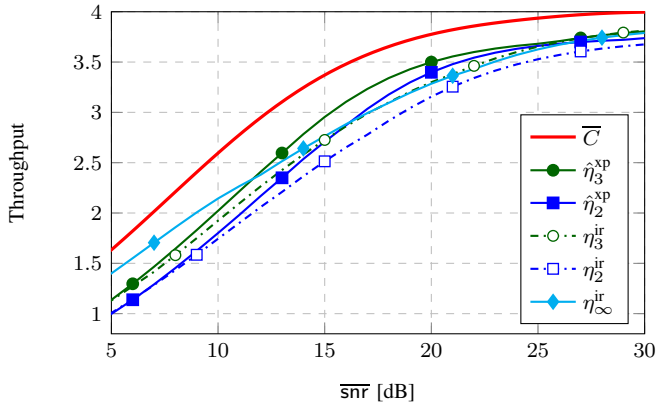


Fig. 5. Throughput of the conventional IR-HARQ (η^{ir}_{K}) compared to the proposed XP-HARQ ($\hat{\eta}^{\mathrm{xp}}_{K}$) for a truncated HARQ, $K \in \{2, 3\}$ in Rayleigh block-fading channel; $R_{\max} = 8$. The ergodic capacity ($\overline{C}$) and the optimal throughput of the persistent conventional IR-HARQ ($\eta^{\mathrm{ir}}_{\infty}$) are shown for reference.

view this is a potentially serious drawback but we show such results to complement those already shown in Fig. 3, where the decodability condition was imposed. Again, XP-HARQ was insensitive to the decodability constraints and always provided results with decodable transmissions.

The improvements due to adaptive XP-HARQ are most notable for high values of the throughput. In particular we observe that

- *The persistent XP-HARQ halves the gap between the ergodic capacity and the conventional IR-HARQ. For example, the SNR gap between $\hat{\eta}^{\mathrm{xp}}_{\infty} = 3$ and the ergodic capacity, $\overline{C} = 3$, is reduced by more than 50% when comparing to the gap between $\eta^{\mathrm{ir}}_{\infty} = 3$ and $\overline{C} = 3$, which is equal to 5dB when $R_{\max} = 8$. We note that the throughput of XP-HARQ increases when $R_{\max}$ increases: the SNR gap between $\overline{C}$ and $\hat{\eta}^{\mathrm{xp}}_{\infty}$ is reduced by half when $R_{\max} = 16$ is used instead of $R_{\max} = 8$.*
- *For any value of throughput $\eta > 3$, two rounds of XP-HARQ yield higher throughput than the conventional persistent IR-HARQ. Thus, in this operation range we may improve the performance and yet decrease the memory requirements at the receiver.*

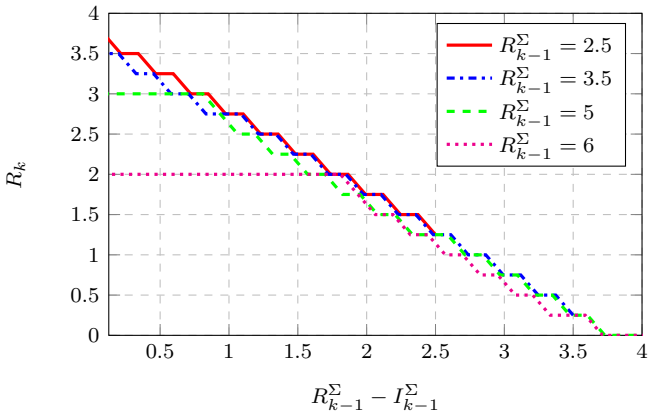


Fig. 6. Optimal rate R_k as a function of $R^{\Sigma}_{k-1} - I^{\Sigma}_{k-1}$ for different values of R^{Σ}_{k-1}; $K = \infty$, $\overline{\mathrm{snr}} = 20$dB, $R_{\max} = 8$.

B. Heuristic rate adaptation

Fig. 6 shows the optimal rate adaptation as a function of $R^{\Sigma}_{k-1} - I^{\Sigma}_{k-1}$ for different values of R^{Σ}_{k-1}, where we note a quasi-linear behaviour of the adaptation function with the saturation which occurs to guarantee $R^{\Sigma}_{k-1} + R_k \leq R_{\max}$.

To exploit this very regular form, which was also observed solving the related problems in [12], [33], we propose to use the following heuristic function inspired by Fig. 6

$$R_k = R_1 - (R^{\Sigma}_{k-1} - I^{\Sigma}_{k-1}), \tag{31}$$

where only the rate R_1 needs to be optimized (from Fig. 6 we find $R_1 \approx 3.5$). Furthermore, applying (31) recursively we obtain $R_2 = I_1, R_3 = I_2, \ldots, R_k = I_{k-1}$; the identical rate-adaptation strategy may be derived from [21, Sec. III].

The simplicity of the adaptation function allows us to evaluate analytically the throughput of XP-HARQ. To this end we need to calculate f_l in the denominator of (28) and the expectation in its numerator.

We first note that, from (31) we obtain

$$\big(I^{\Sigma}_k < R^{\Sigma}_k\big) \iff (I_k < R_1), \tag{32}$$

which means that the probability of decoding failure is independent in all transmission rounds. Thus

$$f_k = (f_1)^k, \tag{33}$$

and (29) may be formulated as

$$\xi_k = \Big(\prod_{l=1}^{k-1} \mathbb{I}\big[I_l < R_1\big]\Big)\mathbb{I}\big[I_k \geq R_1\big]. \tag{34}$$

From (31) we also obtain $R^{\Sigma}_k = R_1 + \sum_{l=1}^{k-1} I_l$, which allows us to calculate the expectation in the numerator of (28) as

$$\mathbb{E}[\xi_k R^{\Sigma}_k] = \mathbb{E}[\xi_k (R_1 + I_1 + \ldots, I_{k-1})] \tag{35}$$

$$= \big(R_1 f_1 + (k-1)\tilde{C}\big)(f_1)^{k-2}(1 - f_1), \tag{36}$$

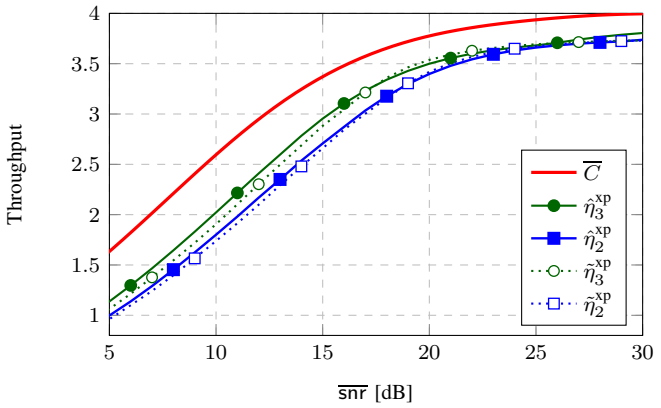


Fig. 7. Throughput of the optimal XP-HARQ ($\hat{\eta}_K^{\mathrm{xp}}$) is compared to the throughput of XP-HARQ with the heuristic policy ($\tilde{\eta}_K^{\mathrm{xp}}$) in Rayleigh block-fading channel. The ergodic capacity ($\overline{C}$) is shown for reference.

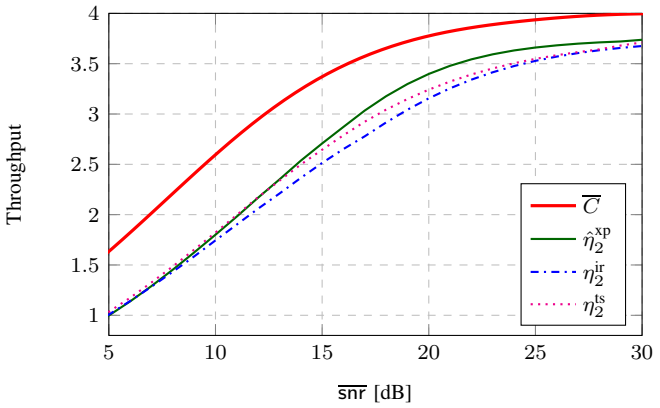


Fig. 8. Optimal throughput of the proposed XP-HARQ ($\hat{\eta}_K^{\mathrm{xp}}$) compared to TS-HARQ ($\eta_K^{\mathrm{ts}}$) when the outdated MI is available at the transmitter. We assume a finite number of rounds, i.e., $K = 2$ in Rayleigh block-fading channel. The set of available rate is $R_k^{\Sigma} \in \{0.25, 0.5, \dots, 8\}$ for XP-HARQ and, for TS-HARQ, the set of available sharing fractions is $p \in \{0, 0.125, 0.25, \dots, 1\}$. The ergodic capacity ($\overline{C}$) is also shown.

where $\tilde{C} = \mathbb{E}_{I_1}\big[I_1 \cdot \mathbb{I}\big[I_1 < R_1\big]\big]$ is a "truncated" expected MI.

Using (36) and (33) in (28), the throughput is calculated as

$$\tilde{\eta}_K^{\mathrm{xp}} = R_1(1-f_1) + \frac{\tilde{C}(1-f_1)}{1-f_1^K} \times \Big(-(K-1)f_1^{K-1} + \frac{1-f_1^{K-1}}{1-f_1} \Big). \tag{37}$$

Example 6 (16QAM, Rayleigh fading – continued)**.** *We compare in Fig. 7 the throughput of optimal XP-HARQ with the heuristic policy* (31)*, which is optimized over R_1. As expected, the optimal solution outperforms the heuristic policy but the gap is very small (less than* 0.5dB*). Moreover, since $\hat{\eta}_K^{\mathrm{xp}}$ was optimized over a finite set of rates $\mathcal{R} = \{0.25, 0.5, \dots, R_{\max}\}$, and the heuristic policy assumes that $\mathcal{R}$ is continuous and unbounded, $\tilde{\eta}_K^{\mathrm{xp}}$ slightly outperforms $\hat{\eta}_K^{\mathrm{xp}}$ above $\overline{\mathsf{snr}} = 20$*dB*. This gap can be reduced increasing the value of $R_{\max}$; decreasing the discretization step below* 0.25 *had much lesser influence on the results.*

We note that, in the limit, $K \to \infty$, (37) becomes

$$\tilde{\eta}_{\infty}^{\mathrm{xp}} = R_1(1-f_1) + \tilde{C}, \tag{38}$$

which is the same as [21, Eq. (12)].

This is quite an intriguing result which suggests that the strategy of [21] based on a double-layer encoding[7] and a transmission-by-transmission decoding (as opposed to the joint decoding required in XP-HARQ), asymptotically yield the same throughput as the heuristic cross-packet HARQ, whose throughput is also very close to the optimal XP-HARQ.

We cannot follow that path here but this relationship should be studied in more details; in particular, the effect of removing the idealized assumption of using a continuous set of rates $\mathcal{R}$, necessary to implement (31), should be analyzed.

[7] [21] proposes double-step encoding: to form $\mathsf{m}_{[k]}$ the bits m_k and the parity bits of $\mathsf{m}_{[k-1]}$ are first "mixed", and next, the channel encoder is used.

C. XP-HARQ vs. variable-length HARQ

The previous results indicate that XP-HARQ is a promising solution leading to a significant improvement over the conventional IR-HARQ. It is also instructive to compare XP-HARQ to other joint coding strategies. We use here the time sharing HARQ (TS-HARQ) proposed in [5] which can be applied in the same transmission model we used in this work. In general terms, TS-HARQ is the implementation of a variable-length coding [12] we already mentioned in Sec. I-B.

The main idea of TS-HARQ is to split the channel block, $\boldsymbol{x}$ into two parts assigned to the codewords of two different packets. In the simplest case, of $K = 2$, if a packet m_1 is not decoded correctly in the first round, the redundancy codeword $\boldsymbol{x}_2$ will use pN_{s} symbols in the next channel block, while the remaining $(1-p)N_{\mathrm{s}}$ symbols will be used by a new packet m_2. The "sharing fractions" defined by p must be optimized to maximize the throughput; more details can be found in [5].

Example 7 (16-QAM over Rayleigh fading channel (continued))**.** *Fig. 8 compares the results of XP-HARQ with those obtained using the optimized* TS-*HARQ. It is clear that XP-HARQ outperforms* TS-*HARQ in middle-to-high values of* $\overline{\mathsf{snr}}$*. This is expected since, in contrast with XP-HARQ,* TS-*HARQ does not use the joint encoding and decoding.*

However, we notice that for low values of $\overline{\mathsf{snr}}$*,* TS-*HARQ is slightly better than XP-HARQ by a fraction of dB. This can be explained by the fact that allowing different values for p means having additional rates. In fact, the nominal transmission rates in the second round of the packets m_1 and m_2 are R/p and $R/(1-p)$ respectively. So varying p for each allowed value of R generates a bigger set of rates. In this sense, the comparison is not entirely "fair" and advantages* TS-*HARQ. Nevertheless, the XP-HARQ still outperforms* TS-*HARQ in the regions of high SNR.*

V. Example of a practical implementation

Until now, we have adopted the perfect decoding assumption, i.e., the decoding error in the kth round is equivalent to the event $\{I_1 < R_1 \wedge \ldots \wedge I_k^{\Sigma} < R_k^{\Sigma}\}$. We will remove now this idealization to highlight also the practical aspect of XP-HARQ.

We thus implement the cross-packet encoders in Fig. 2 using turbo encoders. To this end, as shown in Fig. 9 we separate each encoder Φ_k into i) a bit-level multiplexer, $\mathcal{M}$, whose role is to interleave the input packets $\mathsf{m}_1, \ldots, \mathsf{m}_k$ and produce the packet, $\mathsf{m}_{[k]}$, ii) a conventional turbo-encoder (TC), iii) the rate-matching puncturer, $\mathcal{P}$, which ensures that all binary codewords $\boldsymbol{c}_k$ have the same length, N_{c}, and iv) a modulator, denoted by $\mathcal{X}$ as it maps the codewords $\boldsymbol{c}_k$ onto the codewords $\boldsymbol{x}_k$ from the constellation $\mathcal{X}$; since we use 16-QAM, $N_{\mathrm{c}} = N_{\mathrm{s}} \log_2(M)$, where $M = 16$.

The multiplexers $\mathcal{M}_k$ are implemented using pseudo-random interleaving. The encoders (TC) are constructed via parallel concatenation of two recursive convolutional encoders with polynomials $[13/15]_8$. Each TC produces a $N_{\mathrm{b},[k]} = N_{\mathrm{s}} R_1 + \ldots + N_{\mathrm{s}} R_k$ systematic (input) bits and $N_{\mathrm{p}} = 2N_{\mathrm{b},[k]}$ parity bits $\boldsymbol{p}_k$.[8] The bits $\boldsymbol{c}_k$ are obtained via concatenation of: i) deterministically punctured $\mathsf{m}_{[k]}$ (only "fresh" systematic bits m_k survive the puncturing, i.e., those which were not transmitted in the previous rounds), and ii) the parity bits selected from $\boldsymbol{p}_k$ via a periodic puncturing.

Such a construction of the encoders is of course not optimal; convolutional encoders well suited to the multidimensional concatenation may be sought; similarly, interleavers and puncturers may be further optimized. However, the optimal design represents a challenge of its own and must be considered out of scope of the example we present here.

The decoding can be implemented using conventional elements adapted to deal with outcomes of all transmissions, $\boldsymbol{y}_{[k]}$. From this perspective, we may see the binary codewords $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_k$ as an outcome of $2k$ concatenated convolutional encoders (two per TC in each channel encoder Φ_k). The decoding from the outcome of multiple encoding units was already addressed before [40] [41] and requires implementation of $2k$ Bahl–Cocke–Jelinek–Raviv (BCJR) decoders (two decoders, $\mathrm{BCJR}_{\mathrm{a},k}$ and $\mathrm{BCJR}_{\mathrm{b},k}$, for each of the encoders Φ_k).

We implement the serial scheduling, that is, we *activate*[9] $\mathrm{BCJR}_{\mathrm{a},1}, \mathrm{BCJR}_{\mathrm{b},1}, \mathrm{BCJR}_{\mathrm{a},2}, \mathrm{BCJR}_{\mathrm{b},2}, \ldots$, and finally $\mathrm{BCJR}_{\mathrm{a},k}, \mathrm{BCJR}_{\mathrm{b},k}$. This defines one decoding *iteration*.

The BCJR decoders are activated in the same order in the second iterations, and this continues till the maximum number of iterations is reached. Each iteration thus requires $2k$ activations; we kept the number of decoding iterations constant, so the number of activations increases with k. The implementation requires some care because the length of encoded binary sequences $\mathsf{m}_{[k]}$ grows with k and so is the length of the trellis defining the encoded sequences $\boldsymbol{p}_k$.

The exchange of the so-called extrinsic information is well known when two BCJR decoders are used, but slightly less obvious when we have to use $2k$ of them. For example, when the decoder $\mathrm{BCJR}_{\mathrm{a},t}$ is activated in the lth iteration, it uses extrinsic probabilities calculated for the information bits by the decoders $\mathrm{BCJR}_{\mathrm{a},\tau}, \mathrm{BCJR}_{\mathrm{b},\tau}, \tau = 1, \ldots, t-1$ in the lth iteration, and by the decoders $\mathrm{BCJR}_{\mathrm{b},t}, \mathrm{BCJR}_{\mathrm{a},\tau}, \mathrm{BCJR}_{\mathrm{b},\tau}, \tau = t+1, \ldots, k$ in the previous iteration $l-1$ (unless $l = 1$, then no information is used).

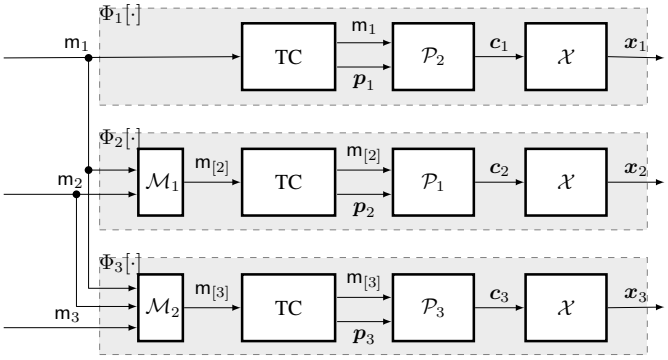


Fig. 9. Implementation of the encoders $\Phi_k[\cdot]$ using turbo codes (TC), bit multiplexing ($\mathcal{M}_k$), puncturing ($\mathcal{P}$), and modulation ($\mathcal{X}$).

Since we do not have the closed-form formula which describes the probability of error under particular channel conditions, especially when multiples transmissions are involved, the rate-adaptation approach seems to be out of reach and we focus on finding the fixed coding rates $R_k, k = 1, \ldots, K$. We use the brute search over the space of available coding rates which verifies the following conditions $\sum_{k=1}^{K} R_k \leq 8$, $R_1 \in \{1.5, 1.75, 2, \ldots, 3.75\}$, $R_k \in \{0, 0.25, \ldots, 3.75\}, \forall k > 1$.

The results we present, obtained for $N_{\mathrm{s}} = 1024$ and four decoding iterations using algorithm from the library [42], are shown in Fig. 10 where the SNR gap (for the throughput $\eta = 3$) between XP-HARQ and the conventional IR-HARQ is ~ 1.5dB for $K = 2$ and ~ 2dB for $K = 3$. We attribute a small improvement of the throughput η_3^{xp} over η_2^{xp} to the suboptimal encoding scheme we consider in this example.

We also note that the improvement of η_3^{ir} with respect to η_2^{ir} does not materialize. This is because IR-HARQ is optimized for R_1 but, due to limitation of the turbo encoder which generates only $3N_{\mathrm{b}}$ bits, a full redundancy cannot be always obtained and, in such a case, we are forced to repeat the systematic and parity bits. This explains why η_3^{ir} and η_2^{ir} are very similar for low throughput. On the other hand, they should be, indeed, similar for high throughput as we have seen in the numerical examples before.

We show in Fig. 10 the ergodic capacity where the gap to the throughput of the TC-based transmission is increased by additional 3dB which should be expected when using relatively-short codewords and practical decoders.

VI. Conclusions

In this work we proposed and analyzed a coding strategy tailored for HARQ protocol and aiming at the increase of the throughput for transmission over block fading channel. Unlike many heuristic coding schemes proposed previously, our goal was to address explicitly the issue of joint coding

[8] We neglects the effect of the trellis terminating bits.

[9] By activation we mean carrying out the calculation by the BCJR decoder.

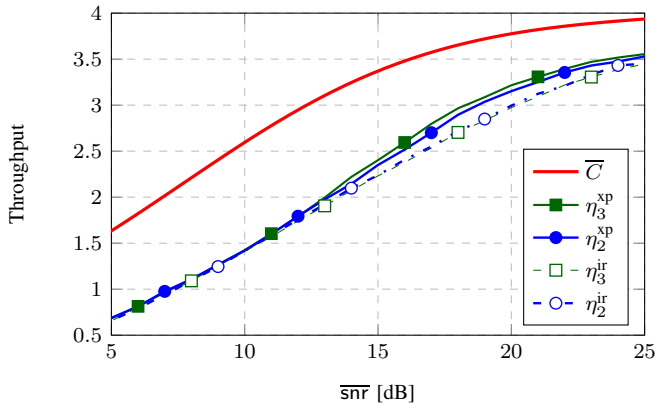


Fig. 10. Turbo-coded transmission: the conventional IR-HARQ (η_K) is compared to XP-HARQ (η_K^{xp}) in Rayleigh block-fading channel.

of many packets into the channel block of predefined length. With such a setup, the challenge is to optimize the coding rates for each packet which we do efficiently assuming existence of a multi-bits feedback channel which transmit the outdated CSI experienced by the receiver.

The throughput of the resulting XP-HARQ is compared to the conventional IR-HARQ indicating that significant gains can be obtained using the proposed coding strategy. The gains are particularly notable in the range of high throughput, where the conventional HARQ fails to offer any improvement with increasing number of transmission rounds. The proposed encoding scheme may be seen as a method to increase the throughput, or as a mean to diminish the memory requirements at the receiver; the price for the improvements is paid by a more complex joint encoding/decoding.

We also proposed an example of a practical implementation based on turbo codes. This example highlights the practical aspects of the proposed coding scheme, where the most important difficulties are i) the need of tailoring the encoder to provide the jointly coded symbols with the best decoding performance, and ii) the design of the simple decoder. Moreover, the real challenge is to leverage the possibility of adaptation to the outdated CSI. To do so, simple techniques for performance evaluation (e.g., the packet error rate (PER)) based on the expected CSI, must be used; such as, for example those studied in [43].

Appendix A
Decoding conditions of XP-HARQ

We outline the proof of the decoding conditions (15) and (16), stated in the following Lemma 1. The HARQ-code refers to the encoding functions stated in (11) and (13) and the joint decoding of the pair $[\mathsf{m}_1, \mathsf{m}_2]$.

Lemma 1 (Decoding conditions)**.** *For all $\varepsilon > 0$, there exists $\bar{n} \in \mathbb{N}$ such that for all $n \geq \bar{n}$, there exists an HARQ-code $c^\star$ such that for all SNR realizations $(\mathsf{snr}_1, \mathsf{snr}_2)$ that satisfy*

$$R_1 + R_2 \leq \mathsf{I}(X_1; Y_1|\mathsf{snr}_1) + \mathsf{I}(X_2; Y_2|\mathsf{snr}_2) - \varepsilon, \quad (39)$$
$$R_2 \leq \mathsf{I}(X_2; Y_2|\mathsf{snr}_2) - \varepsilon, \quad (40)$$

the error probability is bounded by

$$\Pr\left\{[\mathsf{m}_1, \mathsf{m}_2] \neq [\hat{\mathsf{m}}_1, \hat{\mathsf{m}}_2] \middle| c^\star, \mathsf{snr}_1, \mathsf{snr}_2\right\} \leq \varepsilon. \quad (41)$$

Proof of Lemma 1. We consider the random HARQ-code:

- *Random codebook:* we generate $2^{N_s \cdot R_1}$ codewords $\boldsymbol{x}_1$ and $2^{N_s \cdot (R_1+R_2)}$ codewords $\boldsymbol{x}_2$, drawn from the uniform distribution over the constellation $\mathcal{X}$.
- *Encoding function:* as explained in Sec. III, the encoder starts by sending $\boldsymbol{x}_1$ which corresponds to the packet (or *message* in the language of information theory) m_1. If the encoder receives a feedback NACK_1, it sends $\boldsymbol{x}_2$ corresponding to the pair of messages $[\mathsf{m}_1, \mathsf{m}_2]$. Otherwise a new transmission process starts.
- *Decoding function:* if the SNR realizations $(\mathsf{snr}_1, \mathsf{snr}_2)$ satisfy equations (40) and (39), then the decoder finds a pair of messages $[\mathsf{m}_1, \mathsf{m}_2]$ such that the following sequences of symbols are jointly typical:

$$\left(\Phi_1[\mathsf{m}_1], \boldsymbol{y}_1\right) \in A_\varepsilon^{\star N_s}, \ \left(\Phi_2[\mathsf{m}_1, \mathsf{m}_2], \boldsymbol{y}_2\right) \in A_\varepsilon^{\star N_s}. \quad (42)$$

- *Error* is declared when sequences are not jointly typical.

Error events. We define the following error events:

- $E_0 = \left\{\left(\Phi_1[\mathsf{m}_1], \boldsymbol{y}_1\right) \notin A_\varepsilon^{\star N_s}\right\} \cup \left\{\left(\Phi_2[\mathsf{m}_1, \mathsf{m}_2], \boldsymbol{y}_2\right) \notin A_\varepsilon^{\star N_s}\right\}$,
- $E_1 = \Big\{\exists [\mathsf{m}_1', \mathsf{m}_2'] \neq [\mathsf{m}_1, \mathsf{m}_2], \text{ s.t. } \left\{\left(\Phi_1[\mathsf{m}_1'], \boldsymbol{y}_1\right) \in A_\varepsilon^{\star N_s}\right\} \cap \left\{\left(\Phi_2[\mathsf{m}_1', \mathsf{m}_2'], \boldsymbol{y}_2\right) \in A_\varepsilon^{\star N_s}\right\}\Big\}$,
- $E_2 = \Big\{\exists \mathsf{m}_1' \neq \mathsf{m}_1, \text{ s.t. } \left\{\left(\Phi_1[\mathsf{m}_1'], \boldsymbol{y}_1\right) \in A_\varepsilon^{\star N_s}\right\} \cap \left\{\left(\Phi_2[\mathsf{m}_1', \mathsf{m}_2], \boldsymbol{y}_2\right) \in A_\varepsilon^{\star N_s}\right\}\Big\}$,
- $E_3 = \left\{\exists \mathsf{m}_2' \neq \mathsf{m}_2, \text{ s.t. } \left(\Phi_2[\mathsf{m}_1, \mathsf{m}_2'], \boldsymbol{y}_2\right) \in A_\varepsilon^{\star N_s}\right\}$.

The properties of the typical sequences imply that, for N_s large enough, $\Pr\{E_0\} \leq \varepsilon$, and the Packing Lemma [44, pp. 46] implies that the probabilities of the events E_1, E_2, E_3 are bounded by ε if the following conditions are satisfied

$$R_1 + R_2 \leq \mathsf{I}(X_1; Y_1|\mathsf{snr}_1) + \mathsf{I}(X_2; Y_2|\mathsf{snr}_2) - \varepsilon, \quad (43)$$
$$R_2 \leq \mathsf{I}(X_2; Y_2|\mathsf{snr}_2) - \varepsilon, \quad (44)$$
$$R_1 \leq \mathsf{I}(X_1; Y_1|\mathsf{snr}_1) + \mathsf{I}(X_2; Y_2|\mathsf{snr}_2) - \varepsilon. \quad (45)$$

Since (43)-(44) are the hypothesis (39)-(40) of Lemma 1, there exists HARQ-code $c^\star$ with small error probability. □

Appendix B
Optimization via MDP

To obtain the MDP formulation it is convenient to replace packet-wise notation of (1) with a time-wise model

$$\boldsymbol{y}[n] = \sqrt{\mathsf{snr}[n]}\boldsymbol{x}[n] + \boldsymbol{z}[n], \quad (46)$$

where n is the index of the channel block.

At each time n, the HARQ controller observes the *state* $\mathsf{s}[n]$, and takes an *action* $\mathsf{a}[n] = \pi(\mathsf{a}[n])$, according to the policy π. The transition probability matrix, $\boldsymbol{Q}(\mathsf{a})$, has the elements

$$Q_{\mathsf{s},\mathsf{s}'}(\mathsf{a}) \triangleq \Pr\{\mathsf{s}[n+1] = \mathsf{s}' | \mathsf{s}[n] = \mathsf{s}, \mathsf{a}[n] = \mathsf{a}\}, \quad (47)$$

defining the probabilities of the system moving to the state $\mathsf{s}' \in \mathcal{S}$ at time $n+1$ conditioned on the system being in the state $\mathsf{s} \in \mathcal{S}$ at time n and the controller taking the action $\mathsf{a} \in \mathcal{A}(\mathsf{s})$, where $\mathcal{A}(\mathsf{s})$ is the set of actions allowed in a state s and $\bigcup_{\mathsf{s}\in\mathcal{S}} \mathcal{A}(\mathsf{s}) = \mathcal{A}$. In our case, the actions are the coding rates, which we assume may take any positive value, and thus $\mathcal{A}(\mathsf{s}) = \mathbb{R}_+$.

A *policy* π is defined as a mapping $\pi : \mathcal{S} \mapsto \mathcal{A}$ between the state space, $\mathcal{S}$, and the action space, $\mathcal{A}$. We aim at finding a policy π which maximizes the long-term average throughput

$$\eta(\pi) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mathbb{E}\big[\mathsf{R}(\mathsf{s}[n], \pi(\mathsf{s}[n]))\big], \tag{48}$$

where $\mathsf{R}(\mathsf{s},\mathsf{a})$ is the average reward obtained when taking action a in the state s and the expectations are taken with respect to the random states $\mathsf{s}[n]$. In our case the reward is the number of decoded bits normalized by the duration of the channel block, N_s.

The optimal policy thus solves the following problem:

$$\hat{\pi} = \operatorname*{argmax}_{\pi(\cdot)} \eta(\pi) \tag{49}$$

and the optimal throughput of XP-HARQ

$$\hat{\eta}_K^{\mathrm{xp}} = \eta(\hat{\pi}), \tag{50}$$

may be found solving the Bellman equations [29, Prop. 4.2.1]

$$\hat{\eta}_K^{\mathrm{xp}} + h(\mathsf{s}) = \max_{\mathsf{a}\in\mathcal{A}(\mathsf{s})} \left[\mathsf{R}(\mathsf{s},\mathsf{a}) + \sum_{\mathsf{s}'\in\mathcal{S}} Q_{\mathsf{s},\mathsf{s}'}(\mathsf{a}) h(\mathsf{s}') \right], \quad \forall \mathsf{s} \in \mathcal{S}, \tag{51}$$

where $h(\mathsf{s})$ is a difference reward associated with the state. To calculate the optimal $\hat{\eta}_K^{\mathrm{xp}}$, we use here the policy iteration algorithm whose details may be found in [29, Sec. 4.4.1] and which guarantees to reach the solution after a finite number of iterations.

The unique optimal throughput $\hat{\eta}_K^{\mathrm{xp}}$ exists and is independent of the initial state, $\mathsf{s}[0]$ if, for any state $\mathsf{s}'[t] \in \mathcal{S}$, we can find a policy, which starting with arbitrary state $\mathsf{s}[0]$ reaches the state $\mathsf{s}'[t]$ in a finite time $t < \infty$, with non-zero probability [29, Prop. 4.2.6 and Prop. 4.2.4]. For our problems, finding such a policy is indeed possible, proof of which we skip for sake of brevity.

In order to define the state space and the average reward, we deal separately with the truncated and persistent XP-HARQ but in both cases we must track the accumulated rate, $R^\Sigma[n]$ (it defines the reward, $\mathsf{R}(\mathsf{s},\mathsf{a})$), and the accumulated MI, $I^\Sigma[n]$ (it defines the matrix $\boldsymbol{Q}$). Thus these two variables must enter the definition of the state, $\mathsf{s}[n]$.

A. Persistent HARQ

For the persistent XP-HARQ, the state can be defined as a pair

$$\mathsf{s}[n] \triangleq (I^\Sigma[n], R^\Sigma[n]), \tag{52}$$

and the transition to the state at time $n+1$ is defined as

$$\mathsf{s}[n+1] = \begin{cases} \big(I^\Sigma[n] + I[n], R^\Sigma[n] + R[n]\big), \\ \qquad \text{if} \quad R^\Sigma[n] + R[n] \geq I^\Sigma[n] + I[n] \\ (0,0), \quad \text{otherwise.} \end{cases}. \tag{53}$$

A non-zero reward is obtained only by terminating the HARQ cycle, i.e., moving to the state $\mathsf{s}[n+1] = (0,0)$,

$$\mathsf{R}(\mathsf{s}[n],\mathsf{a}) = \big(R^\Sigma[n] + \mathsf{a}\big) F_I^{\mathrm{c}}(R^\Sigma[n] - I^\Sigma[n] + \mathsf{a}), \tag{54}$$

where $F_I^{\mathrm{c}}(x) \triangleq 1 - F_I(x)$ and $F_I(x)$ is the cumulative density function (CDF) of I.

B. Truncated HARQ

In the truncated HARQ, a new HARQ cycle starts also if the maximum number of allowed rounds is attained (even if the message is not decoded correctly). Thus i) the index of the transmission round, k, must enter the defining of the state, ii) we need to make a distinction between the decoding success/failure of the last round. We thus define the state as

$$\mathsf{s}[n] \triangleq (I^\Sigma[n], R^\Sigma[n], \mathsf{k}[n], \mathsf{M}[n]), \tag{55}$$

where $\mathsf{k}[n]$ and $\mathsf{M}[n] \in \{\mathsf{ACK}, \mathsf{NACK}\}$ are respectively, the number of rounds and the decoding result after the transmission in block n. The system dynamic is described as follows:

$$\mathsf{s}[n+1] = \begin{cases} \big(0,0,0,\mathsf{ACK}\big), & \text{if} \quad \mathcal{E}_{\mathsf{ACK}}[n] \\ \big(0,0,0,\mathsf{NACK}\big), & \text{if} \quad \mathcal{E}_{\mathsf{NACK}}[n] \\ \big(I^\Sigma[n] + I[n], R^\Sigma[n] + R[n], \mathsf{k}[n]+1, \mathsf{NACK}\big), \\ \qquad \text{otherwise} \end{cases} \tag{56}$$

where

$$\mathcal{E}_{\mathsf{ACK}}[n] \triangleq \{R^\Sigma[n] + R[n] \leq I^\Sigma[n] + I[n]\}$$
$$\mathcal{E}_{\mathsf{NACK}}[n] \triangleq \{R^\Sigma[n] + R[n] > I^\Sigma[n] + I[n] \ \wedge \ \mathsf{k}[n] + 1 = K\}$$

are respectively, the conditions indicating a successful decoding and a decoding failure at the end of the HARQ cycle.

Thus, the state space is defined as: $\mathcal{S} = \mathbb{R}_+ \times \mathbb{R}_+ \times \{0,1,\ldots,K-1\} \times \{\mathsf{ACK},\mathsf{NACK}\}$ and the reward is defined by (54).

Appendix C
Optimal MDP for $K = 2$

Knowing the rate of the first transmission, R_1, the optimization problem (50) may be solved analytically for $K = 2$ using (28)

$$\hat{\eta}_2^{\mathrm{xp}} = \max_{R_2(I_1)} \frac{\mathbb{E}\Big[R_1 \mathbb{I}\big[I_1 \geq R_1\big]\Big]}{1+f_1} + \frac{\mathbb{E}\Big[(R_1 + R_2(I_1))\mathbb{I}\big[I_1 \leq R_1 \wedge I_2^\Sigma \geq R_1 + R_2(I_1)\big]\Big]}{1+f_1}. \tag{57}$$

Since f_1 is independent of $R_2(\cdot)$, solving (57) is equivalent to finding, for each value of $I_1 < R_1$, the optimal $R_2(\cdot)$ as follows

$$R_2(I_1) = \underset{R}{\operatorname{argmax}} \ (R_1 + R) \cdot F^{\mathrm{c}}_{I_2}(R_1 + R - I_1). \qquad (58)$$

which is a one-dimension optimization problem, that can be solved analytically, provided $F^{\mathrm{c}}_{I_2}(\cdot)$ is known.

In the case of Gaussian codebook, i.e., when the MI is given by $I_k = \log_2(1 + \mathsf{snr}_k)$, the optimal rate adaptation policy is given by the following closed-form

$$R_2(I_1) = \max\big(0, \frac{W(2^{I_1}\overline{\mathsf{snr}})}{\log(2)} - R_1\big), \qquad (59)$$

where $W(.)$ is Lambert W function defined as the solution of $x = W(x)\mathrm{e}^{W(x)}$.

REFERENCES

[1] G. Caire and D. Tuninetti, "The throughput of hybrid-ARQ protocols for the Gaussian collision channel," *IEEE Trans. Inf. Theory*, vol. 47, no. 5, pp. 1971–1988, Jul. 2001.

[2] P. Larsson, L. K. Rasmussen, and M. Skoglund, "Throughput analysis of ARQ schemes in Gaussian block fading channels," *IEEE Trans. Commun.*, vol. 62, no. 7, pp. 2569–2588, Jul. 2014.

[3] M. Jabi, M. Benjillali, L. Szczecinski, and F. Labeau, "Energy efficiency of adaptive HARQ," *IEEE Trans. Commun.*, vol. 64, no. 2, pp. 818–831, Feb. 2016.

[4] W. Lee, O. Simeone, J. Kang, S. Rangan, and P. Popovski, "HARQ buffer management: An information-theoretic view," *IEEE Trans. Commun.*, vol. 63, no. 11, pp. 4539–4550, Nov. 2015.

[5] M. Jabi, A. El Hamss, L. Szczecinski, and P. Piantanida, "Multi-packet hybrid ARQ: Closing gap to the ergodic capacity," *IEEE Trans. Commun.*, vol. 63, no. 12, pp. 5191–5205, Dec. 2015.

[6] R. Sassioui, M. Jabi, L. Szczecinski, L. B. Le, M. Benjillali, and B. Pelletier, "HARQ and AMC: Friends or foes?" *IEEE Trans. Commun. (to appear; available online: arxiv.org/abs/1606.05177)*, 2016.

[7] J.-F. Cheng, Y.-P. Wang, and S. Parkvall, "Adaptive incremental redundancy," in *IEEE Veh. Tech. Conf. (VTC Fall)*, Orlando, Florida, USA, Oct. 2003, pp. 737–741.

[8] E. Visotsky, V. Tripathi, and M. Honig, "Optimum ARQ design: a dynamic programming approach," in *IEEE Inter. Symp. Inf. Theory (ISIT)*, Jun. 2003, p. 451.

[9] R. Liu, P. Spasojevic, and E. Soljanin, "On the role of puncturing in hybrid ARQ schemes," in *IEEE Inter. Symp. Inf. Theory (ISIT)*, Jun. 2003, p. 449.

[10] E. Visotsky, Y. Sun, V. Tripathi, M. Honig, and R. Peterson, "Reliability-based incremental redundancy with convolutional codes," *IEEE Trans. Commun.*, vol. 53, no. 6, pp. 987–997, Jun. 2005.

[11] S. M. Kim, W. Choi, T. W. Ban, and D. K. Sung, "Optimal rate adaptation for hybrid ARQ in time-correlated Rayleigh fading channels," *IEEE Trans. Wireless Commun.*, vol. 10, no. 3, pp. 968–979, Mar. 2011.

[12] L. Szczecinski, S. R. Khosravirad, P. Duhamel, and M. Rahman, "Rate allocation and adaptation for incremental redundancy truncated HARQ," *IEEE Trans. Commun.*, vol. 61, no. 6, pp. 2580–2590, June 2013.

[13] S. Pfletschinger, D. Declercq, and M. Navarro, "Adaptive HARQ with non-binary repetition coding," *IEEE Trans. Wireless Commun.*, vol. 13, no. 8, pp. 4193–4204, Aug. 2014.

[14] R. Zhang and L. Hanzo, "Superposition-coding-aided multiplexed hybrid ARQ scheme for improved end-to-end transmission efficiency," *IEEE Trans. Veh. Technol.*, vol. 58, no. 8, pp. 4681 –4686, Oct. 2009.

[15] F. Takahashi and K. Higuchi, "HARQ for predetermined-rate multicast channel," in *IEEE 71st Vehicular Technology Conference (VTC 2010-Spring)*, May 2010, pp. 1–5.

[16] T. V. K. Chaitanya and E. G. Larsson, "Superposition modulation based symmetric relaying with hybrid ARQ: Analysis and optimization," *IEEE Trans. Veh. Technol.*, vol. 60, no. 8, pp. 3667–3683, Oct. 2011.

[17] A. Steiner and S. Shamai, "Multi-layer broadcasting hybrid-ARQ strategies for block fading channels," *IEEE Trans. Wireless Commun.*, vol. 7, no. 7, pp. 2640–2650, July 2008.

[18] M. El Aoun, R. Le Bidan, X. Lagrange, and R. Pyndiah, "Multiple-packet versus single-packet incremental redundancy strategies for type-ii hybrid ARQ," in *6th International Symposium on Turbo Codes and Iterative Information Processing (ISTC), 2010*, 226-230, Ed., Sep. 2010.

[19] M. El Aoun, "Optimisation des techniques de codage et de retransmission pour les systèmes radio avec voie de retour," *PhD thesis, Telecom Bretagne*, 2012.

[20] X. Wang, Q. Liu, and G. Giannakis, "Analyzing and optimizing adaptive modulation coding jointly with ARQ for QoS-guaranteed traffic," *IEEE Trans. Veh. Technol.*, vol. 56, no. 2, pp. 710–720, Mar. 2007.

[21] P. Popovski, "Delayed channel state information: Incremental redundancy with backtrack retransmission," in *IEEE Inter. Conf. Comm. (ICC)*, June 2014, pp. 2045–2051.

[22] M. Jabi, E. Pierre-Doray, L. Szczecinski, and M. Benjillali, "How to boost the throughput of HARQ with off-the-shelf codes," *IEEE Trans. Commun. (under review; available online: arxiv.org/abs/1607.06879)*, 2016.

[23] K. Trillingsgaard and P. Popovski, "Block-fading channels with delayed CSIT at finite blocklength," in *IEEE Inter. Symp. Inf. Theory (ISIT)*, June 2014, pp. 2062–2066.

[24] K. D. Nguyen, R. Timo, and L. K. Rasmussen, "Causal-CSIT rate adaptation for block-fading channels," in *IEEE Inter. Symp. Inf. Theory (ISIT)*, Jun. 2015, pp. 351–355.

[25] C. Hausl and A. Chindapol, "Hybrid ARQ with cross-packet channel coding," *IEEE Commun. Lett.*, vol. 11, no. 5, pp. 434–436, May 2007.

[26] J. Chui and A. Chindapol, "Design of cross-packet channel coding with low-density parity-check codes," in *IEEE Information Theory Workshop on Information Theory for Wireless Networks*, July 2007, pp. 1–5.

[27] D. Duyck, D. Capirone, C. Hausl, and M. Moeneclaey, "Design of diversity-achieving LDPC codes for H-ARQ with cross-packet channel coding," in *IEEE Inter. Symp. Pers. Indoor and Mob. Comm. (PIMRC)*, Sept. 2010, pp. 263–268.

[28] A. Benyouss, M. Jabi, M. Le Treust, and L. Szczecinski, "Joint coding/decoding for multi-message HARQ," in *IEEE Wireless Communications and Networking Conference (WCNC'16), 3-6 April, Doha, Qatar*, April 2016.

[29] D. Bertsekas, *Dynamic Programming and Optimal Control*, 3rd ed. Athena Scientific, 2007, vol. 2.

[30] D. Tuninetti, "On the benefits of partial channel state information for repetition protocols in block fading channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 5036–5053, Aug. 2011.

[31] K. Nguyen, L. K. Rasmussen, A. Guillén i Fàbregas, and N. Letzepis, "MIMO ARQ with multi-bit feedback: Outage analysis," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 765–779, Feb. 2012.

[32] A. Karmokar, D. Djonin, and V. Bhargava, "Delay constrained rate and power adaptation over correlated fading channels," in *IEEE Global Comm. Conf. (GLOBECOM)*, vol. 6, Nov. 2004, pp. 3448–3453.

[33] M. Jabi, L. Szczecinski, M. Benjillali, and F. Labeau, "Outage minimization via power adaptation and allocation in truncated hybrid ARQ," *IEEE Trans. Commun.*, vol. 63, no. 3, pp. 711–723, Mar. 2015.

[34] D. Djonin, A. Karmokar, and V. Bhargava, "Joint rate and power adaptation for type-I hybrid ARQ systems over correlated fading channels under different buffer-cost constraints," *IEEE Trans. Commun.*, vol. 57, no. 1, pp. 421–435, Jan. 2008.

[35] N. Gopalakrishnan and S. Gelfand, "Rate selection algorithms for IR hybrid ARQ," in *2008 IEEE Sarnoff Symposium*, Princeton, NJ, USA, Apr. 2008, pp. 1–6.

[36] P. Wu and N. Jindal, "Performance of hybrid-ARQ in block-fading channels: A fixed outage probability analysis," *IEEE Trans. Commun.*, vol. 58, no. 4, pp. 1129–1141, Apr. 2010.

[37] L. Szczecinski and A. Alvarado, *Bit-Interlaved Coded Modulation : Fundamentals, Analysis and Design*. Wiley, 2015.

[38] M. Le Treust, L. Szczecinski, and F. Labeau, "Secrecy & rate adaptation for secure HARQ protocols," in *IEEE Information Theory Workshop (ITW)*, Sep. 2013, pp. 1–5.

[39] S. Pfletschinger and M. Navarro, "Adaptive HARQ for imperfect channel knowledge," in *2010 International ITG Conference on Source and Channel Coding (SCC)*, Jan. 2010, pp. 1–6.

[40] S. Huettinger and J. Huber, "Design of multiple-turbo-codes with transfer characteristics of component codes," *Proc. Conf. Inform. Sciences and Syst. (CISS'02)*, pp. 10–5, 2002.

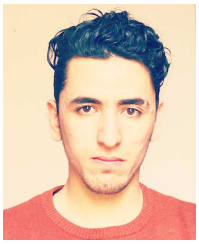
[41] D. Divsalar and F. Pollara, "Multiple turbo codes for deep-space communications," *TDA Progress Report*, vol. 42, p. 121, 1995.

[42] E. Pierre-Doray and L. Szczecinski. (2015) "FeCl channel coding library". [Online]. Available: https://github.com/eti-p-doray/FeCl/wiki

[43] I. Latif, F. Kaltenberger, R. Knopp, and J. Olmos, "Link abstraction for variable bandwidth with incremental redundancy HARQ in LTE," in *11th Inter. Symp. on Modeling Optimiz. in Mobile, Ad Hoc Wireless Networks (WiOpt)*, May 2013, pp. 23–28.

[44] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, 2011.

Mohammed Jabi (S'13), obtained the B.Sc. degree from INPT, Rabat, Morocco, in 2011, and the M.Sc. in telecommunications from INRS, Montreal, Canada, in 2013. Since 2013, he has been a Ph.D. candidate in telecommunications at INRS. His research interests include resource allocation in wireless networks and communication theory.



Abdellatif Benyouss obtained the B.Sc. degree from INPT, Rabat, Morocco, in 2015. Since 2015, he is a M.Sc. student at INRS, Montreal, Canada. His research interests include resource allocation in wireless networks, digital communications and development of advanced wireless technologies.

Maël Le Treust earned his Diplôme d'Etude Approfondies (M.Sc.) degree in Optimization, Game Theory & Economics (OJME) from the Université de Paris VI (UPMC), France in 2008 and his Ph.D. degree from the Université de Paris Sud XI in 2011, at the Laboratoire des signaux et systèmes (joint laboratory of CNRS, Supélec, Université de Paris Sud XI) in Gif-sur-Yvette, France. Since 2013, he is a CNRS researcher at ETIS laboratory (CNRS, ENSEA, Université de Cergy-Pontoise) in Cergy, France. In 2012, he was a post-doctoral researcher at the Institut d'électronique et d'informatique Gaspard Monge (Université Paris-Est) in Marne-la-Vallée, France. In 2012-2013, he was a post-doctoral researcher at the Centre Énergie, Matériaux et Télécommunication (INRS, Quebec University) in Montréal, Canada. From 2008 to 2012, he was a Math T.A. at the Université de Paris I (Panthéon-Sorbonne), Université de Paris VI (UPMC) and Université Paris Est Marne-la-Vallée, France. His research interests are empirical coordination, information theory, Shannon theory, game theory, physical layer security and wireless communications.

Etienne Pierre-Doray is currently pursuing the B.Eng. degree in computer engineering at Polytechnique de Montréal. He worked as a research intern at INRS from 2015 to 2016. His main interests are in the areas of information theory and signal processing.

Leszek Szczecinski (M'98-SM'07), obtained M.Eng. degree from the Technical University of Warsaw in 1992, and Ph.D. from INRS-Telecommunications, Montreal in 1997. From 1998 to 2001, he held the position of Assistant Professor at the Department of Electrical Engineering, University of Chile. He is now Professor at INRS-EMT, University of Quebec, Canada. In 2009-2010, as a Marie Curie Research Fellow, he was with CNRS, Laboratory of Signals and Systems, Gif-sur-Yvette, France.

His research interests are in the area of communication theory, modulation and coding, ARQ, wireless communications, and digital signal processing. He co-authored the book "Bit-Interleaved Coded Modulation: Fundamental, Analysis and Design" (Wiley, 2015).