Transactions Papers

Markov Analysis of Selective Repeat Type II Hybrid ARQ Using Block Codes

Leonardo Badia, *Member, IEEE,* Marco Levorato, *Student Member, IEEE,* and Michele Zorzi, *Fellow, IEEE*

***Abstract*—This paper presents an analytical model for the study of Hybrid ARQ techniques on Discrete Time Markov Channels by means of an appropriate Markov chain, which tracks the transmission outcome and can be used to evaluate several performance metrics, including throughput, loss probability, number of retransmissions, and delay. The analysis is carried out with the assumptions that the information frame is encoded by the source with a linear block code and hard decoding is used at the receiver side. We finally present numerical evaluations for the performance of a truncated Type II Hybrid ARQ technique based on Reed Solomon erasure codes.**

***Index Terms*—Queueing analysis, automatic repeat request, Markov processes, error analysis.**

I. Introduction

WIRELESS communications are prone to channel impairments, thus error control techniques play a key role in determining link performance and reliability. Investigation through adequate analytical frameworks is particularly suited to provide deep insight on the impact of these strategies on the main performance metrics, allowing for accurate system design that properly takes into account the underlying channel constraints.

The two basic error control techniques over time-varying channels are Forward Error Correction (FEC), which provides error protection in an open-loop fashion by means of error-correction codes, and pure ARQ, where erroneous packets are retransmitted in response to negative acknowledgements sent over a feedback channel by the receiver [1]. However, FEC often requires excessive redundancy to guarantee transmission reliability under harsh propagation effects, and usually leads to inefficiencies under time-varying channel conditions. On the other hand, ARQ techniques often fail to provide satisfactory delay performance, as they deliver data in the presence of errors using multiple transmissions for the same packet.

Hybrid ARQ (HARQ) schemes have been proposed as a solution to these shortcomings. HARQ techniques combine classic ARQ, since they involve retransmission of the erroneous data, and FEC, i.e., data are protected from channel impairments by error-correcting codes, in addition to the error-detecting codes which are typically used in pure ARQ for parity check and subsequent acknowledgement (ACK) / not acknowledgement (NACK) of transmissions.

There exists a wide literature on various HARQ schemes in different environments. However, most of the contributions focus on the physical layer, and the assessment of the protocol performance relies on simulation, e.g., [2]–[4]. Analytical approaches generally assume block fading with independent channel coefficients in different blocks. For example, in [5] the authors investigate the performance of a slotted Direct-Sequence Spread-Spectrum Multiple Access system with a Type II HARQ scheme. In [6] the packet discarding probability of Type II HARQ scheme with block codes and maximum-likelihood detection is derived. The authors of [7] analyze throughput and packet error rate of HARQ schemes modeling a fading channel with a Markov chain.

In all these papers, the HARQ scheme operates on one packet at a time, which is appropriate only for Stop-and-Wait (SW) ARQ. In most practical cases, between a transmission and the reception of its feedback message, there is still time to perform other transmissions. Thus, SW ARQ wastes several transmission opportunities. Instead, it is more appropriate to think of a Selective Repeat (SR) system, where the sender transmits continuously, and not acknowledged packets are selectively identified for retransmission. Differently from related papers, we consider a SR Type II [8] HARQ scheme to deliver information frames to the intended destination. In such a system, we derive several performance metrics, such as throughput, delivery delay, average number of retransmissions and frame discarding probability.

We try not to depend on a specific implementation but rather to keep a general approach where several specific kinds of SR Type II HARQ can be framed. Our framework assumes finite round trip time and fading envelope modeled with a Markov chain. Channel correlation is a fundamental issue in the evaluation of the performance of HARQ schemes, and must be taken into account in the system design. To this end, we assume the availability of a Markov chain, called in the following *channel chain*, modeling the channel, where different states correspond to different quality levels of the received packet in terms of incorrectly received bits.

Paper approved by L. K. Rasmussen, the Editor for Iterative Detection Decoding and ARQ of the IEEE Communications Society. Manuscript received June 23, 2006; revised February 15, 2007. Part of this work has been presented at IEEE Globecom 2006.

L. Badia (corresponding author) is with the IMT Lucca Institute for Advanced Studies, Piazza S. Ponziano 6, 55100 Lucca, Italy (e-mail: l.badia@imtlucca.it).

M. Levorato and M. Zorzi are with the Dept. of Information Engineering, University of Padova, via Gradenigo 6/B, 35131 Padova, Italy.

Digital Object Identifier 10.1109/TCOMM.2008.060374.

Through manipulations of the channel chain we derive the so-called *ARQ chain*, whose states represent the whole system. The representation of the system through the ARQ chain enables the evaluation of the system throughput and delay performance.

Note that in [9] a pure SR ARQ scheme is analyzed with a similar framework, i.e., by introducing an ARQ chain to jointly track the channel and the outcome of previous transmissions. However, the analytical model is not the same, since HARQ requires an entirely different Markov model than pure ARQ. Moreover, that contribution focuses on the delivery delay only, whereas here we derive several other performance metrics.

The analytical model used to derive the ARQ chain assumes that block codes with hard decision are used by HARQ. This choice is made for conceptual simplicity, but it is possible to consider extensions to other cases of interest, such as Convolutional or Turbo codes and/or soft decision [10], which on the other hand are to be seen as major undertakings and are left for further research.

Finally, we show how the model can quantitatively evaluate the performance of Type II HARQ systems. As a concrete example of this, we consider a multiple-level channel chain as based on an underlying Discrete Time Markov channel. For this situation, we derive an entirely analytical formulation and we assess the aforementioned metrics for different cases of channel correlation, showing that this parameter has a heavy impact on the overall performance of HARQ systems.

The remainder of this paper is organized as follows: in Section II we analytically describe the HARQ system by means of a Markov framework, deriving the ARQ chain from an N-State Discrete Time Markov Channel. The solution of the ARQ chain allows to evaluate different metrics, as shown in Section III. Finally, Section IV shows numerical evaluations and Section V concludes the paper.

II. Analytical Framework

We consider the transmission of an indefinitely long message whose fundamental unit is an *information frame*. HARQ dictates that each information frame is associated with a number of *HARQ packets*, which are sent over the channel; the connection between the information frame and the associated HARQ packets is different for Type I and Type II HARQ [8].

In Type I HARQ, an information frame is associated with only one HARQ packet, obtained applying a FEC code to the frame to increase its robustness against errors. If the number of errors exceeds the correction threshold, the same packet is sent again over the channel, similarly to pure ARQ. In Type II HARQ, which is the focus of this paper, a given information frame is associated with multiple HARQ packets. When a NACK is received for a specific HARQ packet, a physically different packet of the set is sent over the channel, since, according to the HARQ principle, a "retransmission" refers to sending additional redundancy for a given information frame, rather than repeating the corrupted packet.

Within this paper, we focus on the case where block codes are used to obtain the HARQ packets from an information frame. More specifically, we consider an $(L(F+1),k)$ block code with $L \geq k$, called C_F. It is not restrictive to consider this code as systematic, i.e., the first part of the codeword contains the information bits, whereas the latter part contains redundant bits. We assume that an information frame of length k is mapped into a codeword by means of C_F. For simplicity, let us assume that this codeword is subdivided in $F+1$ packets of size L. Indeed, considering HARQ packets of different sizes would be possible with a similar rationale, but the formulation would be much more cumbersome with no significant additional insight.

With this choice of parameters, it is possible to perform up to F retransmissions of the same information frame, as follows. The first HARQ packet associated with an information frame is sent at what we conventionally call retransmission 0, after which an acknowledgment is already possible if the receiver is able to decode the received message by seeing it as a codeword of an (L,k) block code C_0, which is a shortened version of C_F. More in general, let C_i denote an $((i+1)L,k)$ block code obtained as a shortened version of C_F. Hence, for every $i = 0,1,\ldots,F$, if retransmission i occurs for a given information frame, $i+1$ HARQ packets have been sent over the channel (including the current one) and their juxtaposition can be seen as a codeword of code C_i. In this way, the transmission of further HARQ packets associated to the same information frame improves the error-correction capability.

We consider in the following a discrete (slotted) time, where a slot equals the time required for transmitting one packet, and the round trip time equals a fixed number of slots m, in general greater than 1. The assumption of a fixed round trip time is well justified if the transmitting and receiving nodes are fixed, or their distance does not change significantly over time.

We assume that the case of undetected errors, i.e., misinterpretation of the codeword due to excessively high number of errors, can be neglected: in general, codes are properly designed exactly to make these situations very unlikely. Furthermore, we assume that the receiver's feedback is error-free. This is realistic if erasures are contrasted by using a time-out and, as happens in the forward channel, misinterpretations of acknowledgements can be ignored. Finally, our work assumptions include that the receiver's buffer is unlimited and the sender transmits continuously. These simplifications have been shown in the literature to significantly simplify the analysis without changing the qualitative behavior (see, e.g., [11]).

We focus on a hard decision process at the receiver, which is analyzed through a Markov approach. To define the system state, we quantize the number of errors contained in a HARQ packet into $K+1$ levels, i.e., a HARQ packet can be received with an *error level* equal to 0, 1, ..., K. Received HARQ packets associated with the same information frame and transmitted in different slots are juxtaposed in the order of transmission at the receiver's side to form a longer codeword. The *error level* of an information frame is also defined as the sum of the error levels of all associated packets which have been transmitted. Each time a retransmission occurs for a given information frame, its error level is increased by adding the error level of the newly transmitted HARQ packet. Thus, after J retransmissions, $0 \leq J \leq F$, of the same information frame, the possible values for its error level are $(J+1)K+1$.

For every $i = 0,1,2,\ldots,F$ we define θ_i (which satisfies $0 \leq \theta_i \leq K \cdot (i+1)$) as the error correction threshold of the

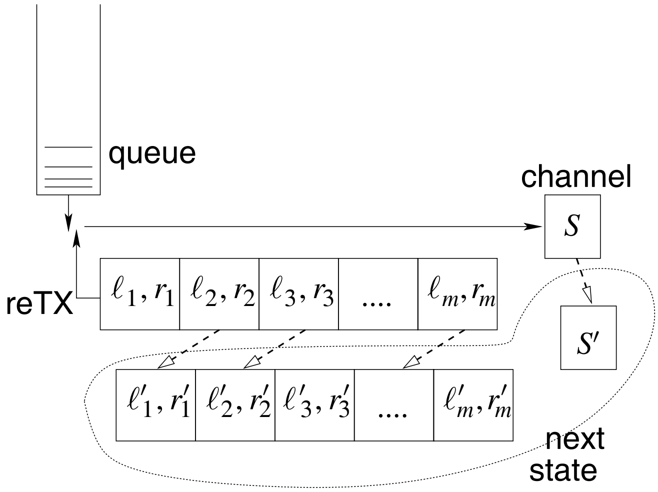


Fig. 1. The ARQ chain and its evolution.

code C_i. That is, an information frame is acknowledged and no longer retransmitted already after its first transmission if its error level is less than or equal to θ_0. If it is not, a retransmission will occur m slots later, when the NACK message will be received back at the transmitter's side. Analogously, at the $(i+1)$th transmission, the error level must be less than or equal to θ_i in order for the packet to be acknowledged. The definition of the θ_is depends on the actual way in which information frames are coded and HARQ packets are obtained. This point will be further discussed in Section IV.

We describe the error process of the HARQ packets with the already mentioned *channel chain*. This is assumed to be an N-state Markov Chain, whose $N \times N$ transition matrix is $\mathbf{T} = (t_{ij})$, where t_{ij}, $0 \le i, j \le N-1$, is the transition probability from state i to j. We also define an integer function ξ mapping a channel state $0, 1, \dots, N-1$ into an error level $0, 1, \dots, K$, i.e., the error level of a HARQ packet is equal to ξ_i when the channel state is i.

This Markov chain describes the channel transitions only, but, as already discussed, the selective repetition induces higher order memory. To derive the Markov chain describing the entire process, called ARQ chain, we follow an approach akin to the one used in [9], where we proved that one can use a Markov approach in which the status of the last m transmissions plus the channel state is tracked. However, the approach for the HARQ case is drastically different from [9], since we have to consider more possibilities depending on the error level quantization and the maximum number of transmissions.

The derivation of the ARQ chain works as follows. We consider a state vector $\mathbf{v} = ((\ell_1, r_1), \dots, (\ell_i, r_i), \dots, (\ell_m, r_m), S)$, where the last element S denotes the channel state; therefore, it takes a value in $0, 1, \dots, N-1$ and evolves according to matrix $\mathbf{T}$. Also, each state S is associated with its error level ξ_S, taking values in $0, 1, \dots, K$. The rest of $\mathbf{v}$ is an m-sized array describing the outcome of the last m packet transmissions. Every entry of the array contains an ordered pair of integers, (ℓ_i, r_i), with $i = 1, \dots, m$. In particular, the right-most pair, i.e., (ℓ_m, r_m), refers to the HARQ packet currently under transmission, whereas (ℓ_1, r_1) describes the outcome of the oldest considered HARQ packet, i.e., the one transmitted $m-1$ slots earlier. In general, (ℓ_i, r_i) refers to the HARQ packet transmitted at slot $t-m+i$, where t is the current time index. The element r_i, between 0 and F, corresponds to the number of retransmissions performed for the information frame the packet is associated to. Values higher than F are not permitted since if the frame is still not acknowledged after the Fth retransmission it is discarded. The first element of the pair, ℓ_i, is the error level of the related information frame, thus it is between 0 and $(r_i+1)K$ (at most, there are $(F+1)K+1$ possible values when $r_i = F$). To help understanding this process and the employed notation, Fig. 1 sketches the system state at a given time and its cyclical evolution. In the figure, it is shown that, as time goes by, the pairs (ℓ_i, r_i) cyclically shift, since, by definition, (ℓ_i, r_i) describes the HARQ packet transmitted $m-i-1$ slots earlier when the time index is $t+1$. The channel state S also evolves according to a Markov chain.

Note that a further simplification is possible, in order to describe the state with a lower number of possible values. In fact, since the impact of all values of ℓ_i lower than θ_{r_i} is the same (they describe a frame which is anyway correctly received), we might collapse all pairs (ℓ_i, r_i) for which $\ell_i \le \theta_{r_i}$ into $(0, r_i)$.

Proposition 1: On the aforementioned discrete time axis, the evolution of state $\mathbf{v}$ is fully described by a Discrete-Time Markov chain.

Proof: See Appendix A. ■

This justifies the following balance equations. If $\sigma(\mathbf{v})$ is the steady-state probability that the state vector is $\mathbf{v}$:

$$\text{if } (\xi_S \le \theta_0): \tag{1}$$
$$\sigma((\ell_1, r_1), \dots, (0,0), S) = \sum_{c=0}^{N-1} t_{cS} \Bigg(\sum_{x=0}^{F} \sigma((0,x), (\ell_1, r_1), \dots, (\ell_{m-1}, r_{m-1}), c) + \sum_{x=\theta_F+1}^{K(F+1)} \sigma((x,F), (\ell_1, r_1), \dots, (\ell_{m-1}, r_{m-1}), c) \Bigg)$$

$$\text{if } (\xi_S > \theta_0): \tag{2}$$
$$\sigma((\ell_1, r_1), \dots, (\xi_S, 0), S) = \sum_{c=0}^{N-1} t_{cS} \Bigg(\sum_{x=0}^{F} \sigma((0,x), (\ell_1, r_1), \dots, (\ell_{m-1}, r_{m-1}), c) + \sum_{x=\theta_F+1}^{K(F+1)} \sigma((x,F), (\ell_1, r_1), \dots, (\ell_{m-1}, r_{m-1}), c) \Bigg)$$

$$\text{if } 0 < x \le F \text{ and } \theta_x - \xi_S > \theta_{x-1}: \tag{3}$$
$$\sigma((\ell_1, r_1), \dots, (0,x), S) = \sum_{c=0}^{N-1} t_{cS} \cdot \sum_{y=\theta_{x-1}+1}^{\theta_x - \xi_S} \sigma((y, x-1), (\ell_1, r_1), \dots, (\ell_{m-1}, r_{m-1}), c)$$

$$\text{if } 0 < x \le F, \quad \theta_x < y \le xK + \xi_S: \tag{4}$$

$$\sigma((\ell_1, r_1), \ldots, (y, x), S) = \sum_{c=0}^{N-1} t_{cS} \cdot$$
$$\cdot \sigma((y - \xi_S, x-1), (\ell_1, r_1), \ldots, (\ell_{m-1}, r_{m-1}), c)$$

in all other cases : (5)

$$\sigma((\ell_1, r_1), \ldots, (\ell_m, r_m), S) = 0$$

Note that collecting all these equations, the transition matrix $\mathbf{G}$ of the ARQ chain can be obtained. This set of equations can be explained by following the proof of Proposition 1. In particular, (5) is justified by the observation that some vectors $\mathbf{v}$ do not actually represent a feasible state, which is the case for example when the number of errors in a given position ℓ_i is higher than $K(r_i + 1)$. The same also holds for situations where $1 \leq \ell_i \leq \theta_{r_i}$ since, as previously discussed, we represent all these cases with the aggregate state denoted by $\ell_i = 0$. The other equations are motivated as follows. A generic state vector $\mathbf{v}$ has a steady-state probability which comes from all possible channel transitions between a generic channel state c and the current state S (outer sum term in all equations). Equivalently, all equations have that every pair (ℓ_i, r_i) except the last one is deterministically derived from (ℓ_{i+1}, r_{i+1}) at the previous time instant. Additionally, the equations characterize different conditions for (ℓ_m, r_m). In (3) it is stated that an acknowledgement feedback ($\ell_m = 0$) is sent back at the xth transmission ($r_m = x$), for x greater than 0, if and only if the error level is less than or equal to the threshold θ_x but at the previous transmission the error level was above θ_{x-1}. This explains why the inner sum on the error level of the previous transmission y goes from $\theta_{x-1} + 1$ to $\theta_x - \xi_S$, so that the addition of the error level ξ_S due to the current channel state S is small enough to make correction possible (remember that the error level of a HARQ packet is added to the global error level of the information frame it corresponds to). The same considerations can be made in (4) but for the case where the packet is still not acknowledged ($\ell_m = y > \theta_{r_m}$). Remember the cases $0 < \ell_i \leq \theta_{r_i}$ are aggregated to the case $\ell_i = 0$. Eqs. (1) and (2) identically follow, but for the condition where $r_m = 0$, so that they correspond to a packet transmitted for the first time, which is acknowledged in (1) and not acknowledged in (2). Since there are two possibilities of transmitting a new frame, namely an acknowledgement is received (i.e., the previous value of ℓ_1 is 0) or a frame has reached the maximum number of transmissions (i.e., the previous value of r_1 is F), two terms are considered within the outer sum, which respectively describe these two possibilities. Note that the second term only considers the error level starting from $\theta_F + 1$, otherwise the case of a frame acknowledged after exactly F transmissions would be counted twice. In other words, either the frame is acknowledged or the frame is *still in error* after the Fth transmission.

III. Analytical Evaluation of Performance Metrics

We can now proceed to the evaluation of several metrics of interest for the SR ARQ analysis. The steady-state probabilities can be derived from the set of balance equations (1)–(5) and the normalization condition, i.e.

$$\sum_{\text{all } \mathbf{v}} \sigma(\mathbf{v}) = 1 \tag{6}$$

since the balance equations are homogeneous. The following performance metrics can then be evaluated: average throughput T, average number of frame retransmissions N_{fr}, probability of frame discarding P_{fd}.

The average throughput, defined as the average fraction of the slots in which a frame is acknowledged, can be evaluated as the sum of the steady-state probabilities of the states in which $\ell_1 = 0$. If we define the set $\mathcal{A}$ as $\mathcal{A} = \{\mathbf{v} \mid \mathbf{v} = ((0, r_1), (\ell_2, r_2), \ldots, (\ell_m, r_m), S)\}$, we have:

$$T = \sum_{\mathbf{v} \in \mathcal{A}} \sigma(\mathbf{v}) \tag{7}$$

An equivalent description of this metric may be obtained by considering a different position than the first, i.e., ℓ_i with $1 < i \leq m$ instead of ℓ_1, due to the fact that the shift from m through 1 is deterministic. However, we indicate ℓ_1 in the previous expression since it corresponds to the instant when the correct reception is known at the transmitter, so that a new packet is sent. Equivalently, the condition $\ell_m = 0$ would have the physical meaning of describing when the destination node correctly receives the packet.

Analogously, as r_1 is the number of retransmissions undergone by a given frame, the average total number of retransmissions per correctly received information frame can be computed as

$$N_{fr} = \frac{\sum_{\mathbf{v} \in \mathcal{A}} \left(r_1 \, \sigma(\mathbf{v}) \right)}{\sum_{\mathbf{v} \in \mathcal{A}} \sigma(\mathbf{v})} = T^{-1} \sum_{\mathbf{v} \in \mathcal{A}} \left(r_1 \, \sigma(\mathbf{v}) \right) \tag{8}$$

Similar to the calculation of the average number of packet retransmissions, the condition of frame discarding is instead described by having $(\ell_1, r_1) = (x, F)$ as the first pair, where $\theta_F < x \leq (F+1)K$. Defining $\mathcal{B}$ as the set of states where the information frame associated with the received HARQ packet is going to be discarded, i.e., $\mathcal{B} = \{\mathbf{v} | \mathbf{v} = ((x, F), (\ell_2, r_2), \ldots, (\ell_m, r_m), S)$ with $\theta_F < x \leq (F+1)K\}$, the probability of frame discarding P_{fd} is

$$P_{fd} = \sum_{\mathbf{v} \in \mathcal{B}} \sigma(\mathbf{v}). \tag{9}$$

Finally, we can evaluate the delivery delay τ_D, defined as the time elapsed between the first transmission of an information frame and its final release to the upper layers from the re-sequencing buffer of the receiver [11], which happens when all packets with lower identifier have been correctly received or discarded.[1] To derive the statistics of this term, we refer to the analysis presented in [9], where the delivery delay of a pure SR ARQ packet was determined from the preliminary evaluation of the steady state probabilities of the ARQ chain. First of all, note that it is simpler to evaluate

[1] Since we take the point of view of frame delivery, and we focus on a finite maximum number of retransmissions, we need to consider a frame as released even in the event that another one with lower id is discarded (rather than correctly delivered), or at some point no more packets would be delivered.

the delivery delay at the transmitter's side. This differs from the evaluation at the receiver's side only by a constant term t_c which is the propagation delay (approximately half the round-trip time). After the computation of the steady state probabilities of the ARQ chain, we can observe that the first transmission of a packet can only occur if in the current time slot either (i) an acknowledgement is received or (ii) a packet is discarded due to too many retransmissions, i.e., the state before the transmission belonged to either $\mathcal{A}$ or $\mathcal{B}$, for case (i) or (ii), respectively. Conditioned on being in one of these cases, one can consider a fictitious Markov chain which is identical to the previous one except for the fact that the arrivals of new packets are "turned off." This is in order to reflect the fact that, after the first transmission of the packet of interest, all subsequent transmissions of new packets are irrelevant for τ_D, which is only affected by packets with *lower id*. For the case considered in the present paper, this simply means to merge the right-hand part of (1) in (2), since erroneous transmissions of a newly arrived packet are now neglected.

The evolution of this chain can be used to determine the time instant of the delivery of the information frame, which corresponds to the first passage to one of the states where all previous frames are acknowledged or discarded; in other words, $\ell_i \leq \theta_{r_i}$ and/or $r_i = F$, for all i. Notice that these states form an absorbing set, i.e., once the resolution condition is reached, it is kept indefinitely. Formally, take $\mathbf{s_0}$ as a column vector with as many entries as the states of the ARQ chain. Let all entries be equal to zero except for the states where $\ell_i \leq \theta_{r_i}$ or $r_i = F$ for all i, which equal 1. Moreover, construct a vector $\boldsymbol{\alpha} = (\alpha)_{\mathbf{v}}$ indexed by the ARQ chain states $\mathbf{v}$ as follows:

$$\alpha_{\mathbf{v}} = \begin{cases} \frac{\sigma(\mathbf{v})}{\sum_{\mathbf{w} \in \mathcal{A} \cup \mathcal{B}} \sigma(\mathbf{w})} & \text{if } \mathbf{v} \in \mathcal{A} \cup \mathcal{B} \\ 0 & \text{if } \mathbf{v} \notin \mathcal{A} \cup \mathcal{B} \end{cases} \tag{10}$$

According to what stated above, we can find the probability distribution of the delivery delay as:

$$\mathcal{P}_c[t] = P[\tau_D \leq t] = \boldsymbol{\alpha} \mathbf{G}^{t+1} \mathbf{s_0}, \quad t \geq 0 . \tag{11}$$

IV. Numerical Results

The evaluation presented in the previous sections relies on the availability of an N-state Markov chain which describes the channel so that every state i is characterized by an error level ξ_i. Moreover, for any number of retransmissions j a threshold θ_j is needed to be compared with the error level of the frame. The derivation of a multiple level Markov chain from a physical channel is a deeply studied subject, e.g., [12], [13], thus it will not be investigated directly here. Instead, we adopt a simple and practical approach, which derives an N-state channel chain from a simple two-state chain. This is just an example to directly validate the protocol model presented before and to show how this can be used to evaluate and compare different Type II Hybrid SR ARQ strategies with hard decision.

We consider an (n, k) Reed Solomon (RS) erasure block code, with symbols from the Galois Field $\mathbb{Z}_{2^M}$, where k and n are the number of symbols of the uncoded and coded message, respectively. In order to encode a binary message $\mathcal{F}$ of kM bits, we first split $\mathcal{F}$ in k groups of M bits, corresponding to k symbols of $\mathbb{Z}_{2^M}$, and then apply the code to the k symbols, obtaining a coded message C of nM bits. The described code is equivalent to a binary linear code (nM, kM). The minimum distance of an (n, k) RS code is $d_{min} = n - k + 1$; thus, assuming that in a coded message of n symbols there are c erasures and e unknown errors at the symbol level, the message is successfully decoded if [10]

$$2e + c \leq d_{min} - 1 = n - k. \tag{12}$$

Our numerical evaluation is referred to a case similar to [14], where a Cyclic Redundancy Check (CRC) code is tailed to each symbol. Thus, the decoder knows the location of the symbol errors by detecting the bit errors contained in each symbol thanks to the CRC code, that is, there are only erasures, and $e = 0$. Hence, the correct reception of at least k symbols is sufficient for the message reconstruction. For sufficiently large values of M, the increased correction capability outweighs the throughput loss due to the CRC code overhead.

Assuming that an information packet contains k symbols, for the HARQ system under analysis, we take $n = (F + 1)k$, recalling that F is the maximum number of retransmissions before frame discarding. In this way we may divide the overall codeword into $F + 1$ packets, containing k symbols each, to be transmitted one at a time without repetition. With the notation used in Section II, L, K, k are all the same value, for simplicity. Thus, code C_0 corresponds to a (k, k) code, i.e., to information symbols only, C_1 is a $(2k, k)$ RS code and so on, and proper thresholds θ_i are defined as $\theta_i = ik$.

To model the channel with a Markov approach, we consider the errors at the symbol level to be described by a two-state Markov process with transition matrix $\mathbf{P} = \{p_{ij}\}$, $i, j \in \{0, 1\}$, where state 0 means error-free channel and 1 on the other hand describes always erroneous condition. For this model, the steady-state error probability is $\varepsilon = p_{01}/(p_{10}+p_{01})$ and the average error burst length is $B = 1/p_{10}$. In this way, the $K+1$ error levels on a packet are obtained by considering its K symbols, each of which could be correct or not according to the outcome of the two-state Markov chain. This indeed determines $K + 1$ possible outcomes for what concerns the number of errors which are present in a single packet, that is level 0 is obtained when all symbols are correct (i.e., the two-state chain stays in state 0 for K subsequent instances), level 1 corresponds to all K possibilities where only one symbol is erroneous, and so on. However, to keep the Markov property of the model, we also need to memorize the outcome of the last symbol separately. In fact, all transitions to the next error level only depend on the outcome of the *last* symbol, since the two-state chain is Markov.

Thus, this approach determines a suitable $N \times N$ transition matrix $\mathbf{T}$ with $N = 2K$, since the N states describe all possibilities of error level ($K + 1$ values) and last symbol outcome (2 possibilities), but two cases never happen, since all correct (erroneous) symbols always imply that the last one is also correct (erroneous). We assume that the states are numbered so that 0 means that the error level of the packet is 0 (which implies last symbol is correct), $2K - 1$ means that the error level is K (which implies last symbol is erroneous). For every other intermediate case $0 < j < 2K - 1$, state

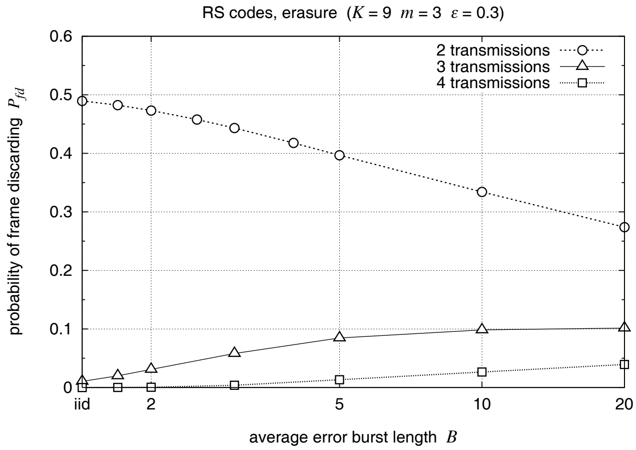


Fig. 2. SR Type II HARQ performance for RS erasure codes: impact of channel burstiness on the probability of frame discarding.

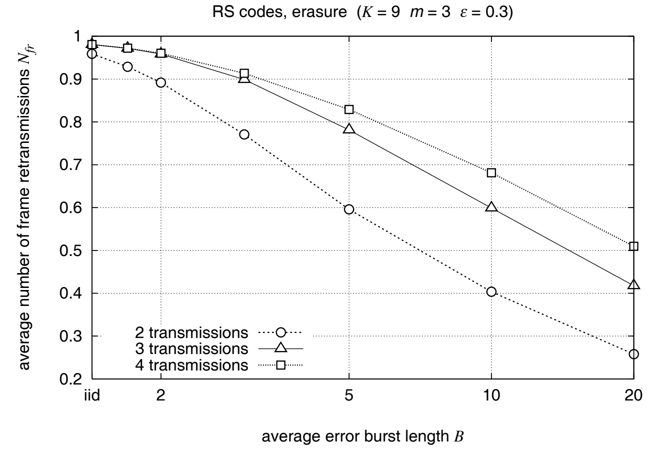


Fig. 3. SR Type II HARQ performance for RS erasure codes: impact of channel burstiness on the average number of frame retransmissions.

j means that the error level is $\lceil j/2 \rceil$ and the last symbol is correct or erroneous according to $((j))_2$ being 0 or 1, where $((\cdot))_2$ denotes the modulo 2 operation.

Considering a transmission of K consecutive symbols, with time indices $1, 2, \ldots, K$, one can denote with $\varphi_{xy}(s, K)$, $x, y \in \{0, 1\}$ the probability that s symbols out of K are successful and the channel state is y for the Kth, given that the channel state was x at time 0 (i.e., for the last symbol transmitted before the sequence of K symbols starts), which is a well-known function that can be derived as shown in [15]. This allows to promptly compute the matrix $\mathbf{T}$ since t_{ij} is set equal to

$$t_{ij} = \varphi_{xy}(j, K), \qquad \text{where } x = ((i))_2,\ y = ((j))_2. \quad (13)$$

About the numerical choice of the parameters, note the following. It is known [16] that channel correlation heavily affects the performance of SR ARQ. It is therefore interesting to see the extent of this phenomenon in the case of Type II HARQ. To this end, under the aforementioned model we have to choose the value of B (average length of symbol error bursts) as the independent variable. Note that the case where $B = 1/\varepsilon$, thus corresponding to i.i.d. (independent and identically distributed) symbol errors, will be denoted as "iid" on the x-axis of the graphs. The investigations performed by changing the value of B give the most relevant insight that can be obtained through the simple model described above. Our numerical computations have shown that the impact of the average error rate ε and the round-trip time m is qualitatively less significant, as it results in a shift of the curve without any deviation from the expected behavior. For this reason, we only present here results for a representative case, i.e., $m = 3$ and $\varepsilon = 0.3$.

Fig. 2 highlights that in Type II HARQ a sufficiently large number of allowed retransmissions reduces the probability of discarding a frame, P_{fd}, almost to 0. For the case of uncorrelated channel, when 4 transmissions are allowed the value of P_{fd} can be pushed down to less than 10^{-5}. However, the channel correlation has a severe impact on the performance. In fact, in bursty channels ($B > 10$) the probability that the receiver fails to recover the information frame even with 4

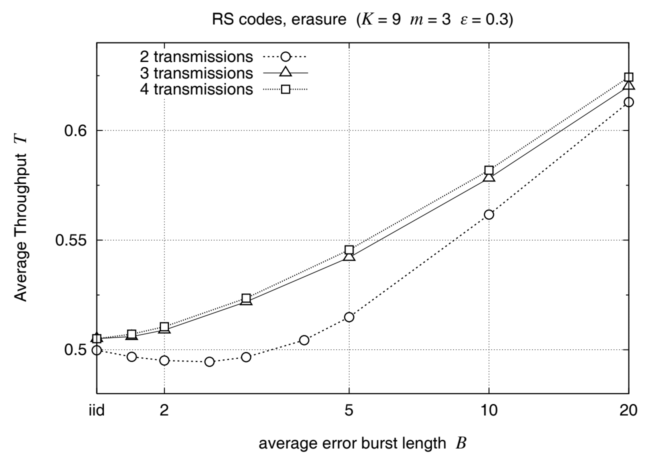


Fig. 4. SR Type II HARQ performance for RS erasure codes: impact of channel burstiness on the average throughput.

transmissions is still significant.

Allowing a higher number of retransmissions has a very different effect for correlated and uncorrelated channels also for what concerns the average number of frame retransmissions N_{fr}, as emphasized in Fig. 3. If low correlation is present, the number of retransmissions does not increase, or increases very mildly, by increasing F. On the other hand, for highly correlated channels, N_{fd} is significantly increased by allowing a higher number of maximum retransmissions. Also, in Fig. 3 it is visible that the average number of frame retransmissions is a decreasing function of the channel burstiness. A possible explanation of this behavior is that when the channel is correlated it is also likely to stay in the good state for a longer time; thus, it is more likely that the frame is delivered by transmitting fewer packets.

Finally, Fig. 4 shows that strong error correlation implies better throughput performance, even though this also corresponds to a larger number of discarded frames. We observe that increasing F from 2 to 3 obtains a higher throughput, but the further improvement obtained when $F = 4$ is not significant. Especially, the situation is critical for *moderately correlated* channels, which can have performance problems

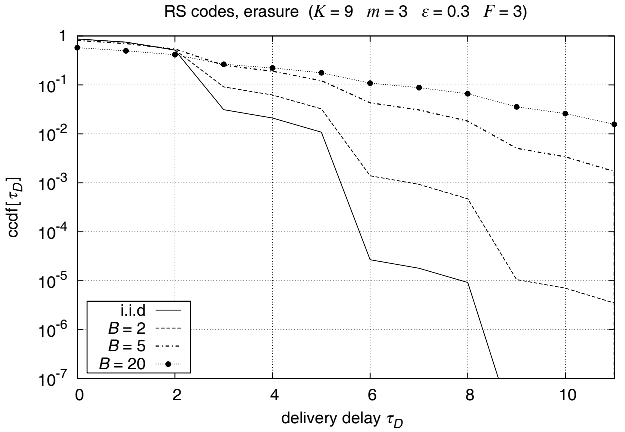


Fig. 5. SR Type II HARQ performance for RS erasure codes: ccdf of the delivery delay for the 4 transmission case.

due to both not so low frame discarding (see Fig. 2) and high number of retransmissions per frame (Fig. 3); Fig. 4 shows that the throughput has even a minimum around $B = 3$, due to the joint effect of these phenomena.

Thus, we conclude that channel correlation has a significant impact on HARQ performance, which our model is able to describe. On the other hand, a characterization of the HARQ process only through the average error probability is not appropriate, since it neglects the impact of error correlation.

Moreover, in Fig. 5 we report the complementary cumulative distribution function (ccdf) of the delivery delay, in the case of 4 transmissions (i.e., $F = 3$). Note that the distribution is meaningful only between 0 and 11 since, if the frame is delivered, this must happen within $F + 1$ round-trip times. Thus, $\tau_D \leq 11$ with probability 1. These figures show a strongly variable behavior of this metric. For low channel correlation the HARQ mechanism is able to provide low delivery delay with a sufficiently high probability: for example, 99.9% of the frames are delivered in 5.5 and 7 slots in the i.i.d and $B = 2$ cases, respectively. However, as the channel correlation increases, the tail of the distribution becomes heavier, so that for the case $B = 20$ there is a probability higher than 1% that the frame is delivered on the last slot before discarding. This might be a problem for applications with strict delay requirements, which might tolerate some loss as long as the delivery is timely for most frames.

Not surprisingly, when the channel is correlated it is also more likely that the frame is delivered immediately (the value of ccdf$[0]$ for $B = 20$ is lower than for the other cases). This is because of the same reason discussed for Fig. 4, i.e., a correlated channel also stays in a good state for a longer time.

As a general conclusion, these results show that the implementation of HARQ may strongly affect the resulting performance. To this end, our analytical framework might be useful in quantifying the numerical behavior. In addition, we also showed that some system parameters, especially channel burstiness, have a critical effect on the performance, which leads to very different behaviors, even more than for pure ARQ strategies.

V. Conclusions

We presented a Markov analysis for Selective Repeat Type II Hybrid ARQ techniques, which allows to study from a general perspective the behavior in terms of throughput, number of retransmissions and delay. The presented analytical framework is entirely tunable and adaptable to different channel models; moreover, it can be promptly extended to consider also different assumptions for what concerns the transmission process or the employed coding. Exact results have been presented in order to evaluate Selective Repeat truncated Type II HARQ for the case of Reed Solomon linear erasure block codes. These results can be useful to gain detailed understanding about the behavior of HARQ mechanisms.

Appendix A: Proof of Proposition 1

Proof: The statement can be constructively proven by showing that the probability of every transition depends only on the current state. We consider the transition from $\mathbf{v} = ((\ell_1, r_1), \ldots, (\ell_m, r_m), S)$ to $\mathbf{v}' = ((\ell'_1, r'_1), \ldots, (\ell'_i, r'_i), S')$. S evolves into S' according to a Markov process with transition matrix $\mathbf{T}$. Also the $m - 1$ left-most pairs (ℓ'_i, r'_i) can be seen as inherited from the previous $m - 1$ right-most pairs (ℓ_i, r_i) contained in the previous time sample. In other words, at every time sample the m-sized window is simply shifted to the left by one slot, so that $(\ell'_i, r'_i) = (\ell_{i+1}, r_{i+1}) \quad \forall i = 1, 2, \ldots, m - 1$. Thus, since all (ℓ'_i, r'_i) for $1 \leq i < m$ are deterministically equal to a component of $\mathbf{v}$, they have only 1-step memory. The right-most pair (ℓ'_m, r'_m) describes instead, at every time instant, the outcome of a new transmission, which however depends in part on the previous left-most pair, (ℓ_1, r_1). In more detail, this last transition happens as follows. If the error level ℓ_1 describes an acknowledged packet, i.e., $\ell_1 \leq \theta_{r_1}$, a new frame is transmitted in the next slot, so that $r'_m = 0$ and ℓ'_m is its error level, obtained through the transition matrix $\mathbf{T}$ and the function ξ, and is exactly equal to $\xi_{S'}$. Similarly, a new frame is also transmitted, regardless of ℓ_1, if r_1 is equal to the maximum number of allowed retransmissions F (since in this case the frame is discarded). Otherwise, the number of retransmissions r'_m is simply equal to r_1+1, and the error level ℓ'_m derives from ℓ_1 via the addition of the error level of the new packet to the previous error level of the frame, which causes the value of $\xi_{S'}$ to be *summed* to ℓ_1 in order to obtain ℓ'_m. In every case, (ℓ'_m, r'_m) depends only on components of $\mathbf{v}$, which proves the Markov property. ■

References

[1] H. O. Burton and D. Sullivan, “Errors and error control,” *Proc. IEEE*, vol. 60, no. 11, pp. 1293–1301, Nov. 1972.

[2] J.-F. Cheng, “Coding performance of hybrid ARQ schemes,” *IEEE Trans. Commun.*, vol. 54, no. 6, pp. 1017–1029, June 2006.

[3] E. Visotsky, Y. Sun, V. Tripathi, M. L. Honig, and R. Peterson, “Reliability–based incremental redundancy with convolutional codes,” *IEEE Trans. Commun.*, vol. 53, no. 6, pp. 987–997, June 2005.

[4] E. Soljanin, N. Varnica, and P. Whiting, “LDPC code ensembles for incremental redundancy hybrid ARQ,” in *Proc. International Symposium on Information Theory (ISIT)*, Adelaide, Australia, 2005, pp. 995–999.

[5] Q. Zhang, T. F. Wong, and S. Lehnert, “Performance of a type–II hybrid ARQ protocol in slotted DS–SSMA packet radio systems,” *IEEE Trans. Commun.*, vol. 47, no. 2, pp. 281–290, Feb. 1999.

[6] E. Malkamaki and H. Leib, “Performance of truncated type–II hybrid ARQ schemes with noisy feedback over block fading channels,” *IEEE Trans. Commun.*, vol. 48, no. 9, pp. 1477–1487, Sept. 2000.

[7] Q. Zhang and S. A. Kassam, "Hybrid ARQ with selective combining for fading channel," *IEEE J. Select. Areas Commun.*, vol. 17, no. 5, pp. 867–880, May 1999.

[8] S. Lin and D. J. Costello, *Error Control Coding: Fundamentals and Applications*. Englewood Cliffs, NJ: Prentice-Hall, 1983.

[9] M. Rossi, L. Badia, and M. Zorzi, "Exact statistics of ARQ packet delivery delay over Markov channels with finite round-trip delay," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1858–1868, July 2005.

[10] S. B. Wicker, *Error Control Systems for Digital Communication and Storage*. Englewood Cliffs, NJ: Prentice-Hall, 1994.

[11] L. Badia, M. Rossi, and M. Zorzi, "SR ARQ packet delay statistics on Markov channels in the presence of variable arrival rate," *IEEE Trans. Wireless Commun.*, vol. 5, no. 7, pp. 1639–1644, July 2006.

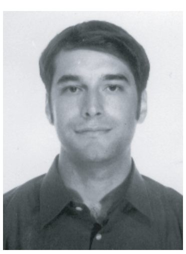
[12] C. C. Tan and N. C. Beaulieu, "On first-order Markov modeling for the Rayleigh fading channel," *IEEE Trans. Commun.*, vol. 48, no. 12, pp. 2032–2040, Dec. 2000.

[13] Q. Zhang and S. A. Kassam, "Finite-State Markov Model for Rayleigh Fading Channels," *IEEE Trans. Commun.*, vol. 47, no. 11, pp. 1688–1692, Nov. 1999.

[14] L. Rizzo, "Effective erasure codes for reliable computer communication protocols," *ACM SIGCOMM Computer Communication Review*, vol. 27, no. 2, pp. 24–36, Apr. 1997.

[15] R. A. Howard, *Dynamic Probabilistic Systems*. New York: John Wiley & Sons, INC., 1971.

[16] W. Luo, K. Balachandran, S. Nanda, and K. Chang, "Delay analysis of selective-repeat ARQ with applications to link adaptation in wireless packet data systems," *IEEE Trans. Wireless Commun.*, vol. 4, no. 3, pp. 1017–1029, May 2005.



Leonardo Badia (S'02, M'04) was born in Ferrara, Italy, in 1977. He received the Laurea Degree (with honors) in Electrical Engineering and the Ph.D. in Information Engineering from the University of Ferrara, Italy, in 2000 and 2004, respectively. During 2002 and 2003 he was on leave at the Radio System Technology Labs (now Wireless@KTH), Royal Institute of Technology of Stockholm, Sweden. After having been with the Engineering Department of the Università di Ferrara, Italy, he joined in 2006 the "Institutions Markets Technologies" (IMT) Lucca Institute for Advanced Studies, in Lucca, Italy, where he is currently a Research Fellow. He also collaborates with DEI, University of Padova, Italy. His research interests include energy efficient Ad Hoc Networks, transmission protocol modeling, Admission Control and economic modeling of Radio Resource Management for Wireless Networks. Dr. Badia serves also as reviewer for several periodicals in the communication area.

Marco Levorato (S'06) was born in Venice on March 18th, 1980. He obtained both the BE (Electronics and Telecommunications Engineering) and the ME (Telecommunications Engineering) *summa cum laude* from the University of Ferrara (Italy) in 2002 and 2005, respectively. During 2005 he held a fellowship at the University of Padova (Italy), and from January 2006 he has been a Ph.D. student in Information Engineering at the University of Padova under the supervision of Prof. Michele Zorzi. His research interests include cooperative communications, design of ad hoc networks with multiuser detection and analysis of Hybrid ARQ techniques.

Michele Zorzi (S'89, M'95, SM'98, F'07) was born in Venice, Italy, in 1966. He received the Laurea degree and the Ph.D. degree in Electrical Engineering from the University of Padova, Italy, in 1990 and 1994, respectively. During the Academic Year 1992/93, he was on leave at the University of California, San Diego (UCSD), attending graduate courses and doing research on multiple access in mobile radio networks. In 1993, he joined the faculty of the Dipartimento di Elettronica e Informazione, Politecnico di Milano, Italy. After spending three years with the Center for Wireless Communications at UCSD, in 1998 he joined the School of Engineering of the University of Ferrara, Italy, and in 2003 joined the Department of Information Engineering of the University of Padova, Italy, where he is currently a Professor. His present research interests include performance evaluation in mobile communications systems, random access in mobile radio networks, ad hoc and sensor networks, and energy constrained communications protocols.

Dr. Zorzi was the Editor-In-Chief of the IEEE WIRELESS COMMUNICATIONS MAGAZINE from 2003 to 2005, is currently the Editor-In-Chief of the IEEE TRANSACTIONS ON COMMUNICATIONS, and serves on the Steering Committee of the IEEE TRANSACTIONS ON MOBILE COMPUTING, and on the Editorial Boards of the IEEE TRANSACTIONS ON COMMUNICATIONS, the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, the WILEY JOURNAL OF WIRELESS COMMUNICATIONS AND MOBILE COMPUTING and the ACM/URSI/KLUWER JOURNAL OF WIRELESS NETWORKS. He was also guest editor for special issues in the IEEE PERSONAL COMMUNICATIONS MAGAZINE (Energy Management in Personal Communications Systems) and the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS (Multi-media Network Radios).