Cross-layer optimization of adaptive multi-rate wireless networks using truncated Chase combining HARQ

Jaume Ramis, Guillem Femenias, Felip Riera-Palou and Loren Carrasco
Mobile Communications Group – University of the Balearic Islands (Spain)
Email: {jaume.ramis,guillem.femenias,felip.riera,loren.carrasco}@uib.es

Abstract—A cross-layer performance analysis of a wireless network using adaptive modulation and coding at the physical layer and a truncated Chase combining hybrid automatic repeat request (HARQ-CC) scheme for error control at the data link layer is developed. Based on a Markov chain queueing model, analytical expressions for performance metrics such as throughput, average packet delay and packet loss rate are derived and then used to formulate a constrained optimization problem to maximize the system throughput under the prescribed Quality-of-Service constraints. Numerical results reveal that HARQ-CC consistently outperforms the classical Type-I Hybrid forward error correction/automatic repeat request schemes.

I. INTRODUCTION

In the last decade, the explosive development of wireless services and applications has produced an unprecedented revolution in wireless communications networks. In order to combat the impact of wireless fading, a lot of transmission and reception strategies have been developed to improve the spectral and/or power efficiency of the physical layer, such as the adaptive modulation and coding (AMC) schemes.

The unique nature of AMC to improve the performance of upper-layer protocols has spurred the development of cross-layer designs between the physical layer (PHY) and data link control layer (DLC), which combine AMC with an automatic repeat request (ARQ) protocol (see, for instance, [1]–[9]). One of the main shortcomings of these works is that they rely on first-order amplitude-based finite-state Markov chains (AFSMC) to model the wireless fading channel. However, as was shown by Tan and Beaulieu in [10], first-order AFSMCs having an exponentially decaying auto-correlation function (ACF) can not fit the hypergeometric ACF of the statistical Rayleigh fading process used to model wireless flat-fading channels [11], thus compromising the design of higher layer protocols. In [12]–[14], based on a first-order two-dimensional finite-state Markov chain (FSMC) model, which is able to improve the ACF fitting of the first-order AFSMC, we proposed a novel cross-layer analytical framework for AMC-based wireless systems using either infinitely persistent or truncated Type-I hybrid forward error correction (FEC)/ARQ. Otherwise the received coded data block is discarded and a retransmission is requested by the receiver, similar to standard ARQ. A more sophisticated form of hybrid FEC/ARQ schemes is known as hybrid ARQ (HARQ) [15]. In Type-I HARQ, if the receiver fails to decode a packet, any previously received signal is stored in a buffer and a retransmission request in the form of a negative acknowledgment (NACK) is fed back to the transmitter. Upon reception of this NACK the transmitter sends the same coded packet again. At the receiver side, the optimal solution is to combine these multiple signals according to the maximal ratio combining (MRC) principle [16], [17]. The type-I HARQ with MRC is often referred to as the Chase combining (CC) scheme.

In this paper, based on the physical layer first-order two-dimensional FSMC model introduced in [13], we propose an analytical link level queueing model of a point-to-point adaptive multi-rate wireless system with a truncated Chase combining HARQ (HARQ-CC) scheme. This contribution generalizes and extends the analytical tools proposed in [8], [9], [13], [14], where either infinitely persistent or truncated Type-I hybrid FEC/ARQ were assumed. Using this approach, analytical expressions for performance metrics are derived, which allow the formulation of a cross-layer design, conceived as a constrained optimization problem, that can be used to exploit the joint impact on quality-of-service (QoS) performance measures of both AMC at the PHY layer and truncated HARQ-CC based error control at the DLC layer.

The rest of this paper is organized as follows. In Section II the system model and assumptions are introduced. In Section III our proposed Markov chain based model is presented and analytical expressions for the performance parameters are derived. A cross-layer optimization strategy to support QoS-guaranteed traffic is proposed in Section IV. Numerical results to assess the validity of our model and to illustrate the system performance are presented in section V. Finally, VI provides some concluding remarks.

II. SYSTEM MODEL AND ASSUMPTIONS

As in [1]–[3], a point-to-point wireless packet communication system is considered. The processing unit at the data link layer is a packet of fixed size equal to \( N_b \) bits, and the processing unit at the physical layer is a frame composed by a variable number of packets that depends on the transmission...
mode (TM) selected by the AMC scheme. The link is assumed to support QoS-guaranteed traffic characterized by a maximum average packet delay $D_{\text{max}}$ and a target link layer packet loss rate $P_{\text{loss}}$. At the transmitter side, the HARQ controller manages a buffer (queue) that operates in a first-in-first-out (FIFO) mode and can store up to $Q$ packets. We set the maximum number of ARQ retransmissions to $N_r$. Packets will be removed from the buffer either after being successfully received by the mobile host or after $N_r + 1$ failed attempts. The AMC scheme is assumed to have a set $\mathcal{M} = \{0, \ldots, M - 1\}$ of $M$ transmission modes, each of which corresponding to a particular combination of modulation and coding strategies, including the case in which the transmitter does not transmit.

Without loss of generality, convolutionally coded M-QAM schemes adopted from IEEE 802.16 standard [18] will be used in the AMC pool. All possible TMs (except the non-transmission mode) are listed in Table I. In this case, when using TM $n \in \mathcal{M}$, a rate-1/2 convolutional encoder generates a sequence $\mathbf{b} = \{b_{i}\}_{i=1}^{2N_{c}}$ of encoded bits and, after puncturing, the system transmits $N_{c} = N_{b}/R_{c}(n)$ coded bits per packet, where $R_{c}(n)$ denotes the code rate obtained after puncturing when using TM $n$. In HARQ schemes based on the CC strategy the same puncturing pattern is applied for each (re)transmission. For a generic packet, the sequence of punctured coded bits corresponding to (re)transmission $i \in \{0, \ldots, N_r\}$ can be denoted as $\mathbf{b}^{(i)} = \{b_{i}^{(i)}\}_{i=1}^{N_{c}(n)}$. These punctured coded bits are mapped onto a sequence of symbols $s^{(i)} = \{s_{k}^{(i)}\}_{k=1}^{N_{c}(n)}/\log_{2}M$ which are selected from the M-QAM constellation corresponding to TM $n$. Moreover, the number of transmitted packets per frame depends on the TM $n$ and it is given by $p_{n} = bR_{n}$, where $R_{n}$ denotes the number of information bits per symbol used by TM $n$ and $b$ is a parameter which is up to the designer’s choice. For convenience, we will consider that $p_{0} = \cdots = p_{M-1}$, with $p_{0} = 0$ (i.e., TM 0 corresponds to the absence of transmission) and $p_{M-1} \triangleq C$.

A Rayleigh block-fading channel model has been adopted [19], in which the channel gain $h_{n}$ corresponding to the $\nu$th frame transmission is characterized as a zero-mean circularly symmetric complex Gaussian random variable with unit power. That is, it is assumed that the channel duration $T_{f}$ is much smaller than the coherence time of the channel. Furthermore, although the channel is assumed to remain invariant over at least one time frame interval while it is allowed to vary across successive frame intervals, it is also highly probable that it will remain practically invariant over a large number of successive frame intervals.

The received signal corresponding to transmitted symbol $s_{k}^{(i)} = s_{k}^{(i)}h_{n} + n_{i,k}$, where, by convenient abuse of notation, $h_{i}$ denotes the channel gain of the frame period corresponding to the $i$th (re)transmission and $n_{i,k}$ is a zero-mean circularly symmetric complex Gaussian noise with variance $N_{0}/2$ per dimension. The instantaneous received signal-to-noise ratio (SNR) during the $\nu$th frame transmission is defined as $\gamma_{\nu} = E_{s}[h_{\nu}]^{2}/N_{0}$, which is distributed according to the probability density function $p_{\gamma_{\nu}}(\gamma) = (1/\sqrt{\gamma}) \exp(-\gamma/\bar{\gamma})$, $\gamma \geq 0$, where $\bar{\gamma} = E_{s}[h_{\nu}]^{2}/N_{0}$ is the average received SNR and $E_{s}$ is the average power of the received signal. According to [1], when implementing the AMC strategy, the entire SNR range is partitioned into a set of nonoverlapping intervals defined by the partition $\Sigma_{\nu} = \{\sum_{k=1}^{\nu}, \sum_{k=1}^{\nu+1}\}$ and mode $n$ is selected when $\gamma_{\nu} \in \Sigma_{\nu}$.

Assuming perfect channel state information to be available at the receiver side, the HARQ-CC scheme combines multiple received signals according to the MRC principle. Thus, assuming without loss of generality that the $\nu$th puncturing pattern has been used in all (re)transmissions, the combined signal after $i$ (re)transmissions can be expressed as

$$
\hat{r}_{k}^{(i)} = \sum_{j=0}^{i} r_{k}^{(j)}h_{j}^{r} = s_{k}^{(0)}\rho_{i} + v_{k}^{(i)},
$$

where $\rho_{i} = \sum_{j=0}^{i} |h_{j}^{r}|^{2}$ and $v_{k}^{(i)}$ is a zero-mean circularly symmetric complex Gaussian noise with variance $\sigma_{v_{k}^{(i)}}^{2} = \rho_{i} N_{0}$. The logarithmic likelihood ratio (LLR) corresponding to bit $b_{i}^{(0)}$ mapped onto symbol $s_{k}^{(0)}$ on the $i$th (re)transmission can be expressed as

$$
\lambda_{l}^{(i)} = \log \frac{\sum_{\tilde{s}_{k} \in S_{l}^{(i)}} \exp \left( -|\hat{r}_{k}^{(i)} - \tilde{s}_{k}\rho_{i}|^{2}/\sigma_{v_{k}^{(i)}}^{2} \right)}{\sum_{\tilde{s}_{k} \in \tilde{S}_{l}^{(i)}} \exp \left( -|\hat{r}_{k}^{(i)} - \tilde{s}_{k}\rho_{i}|^{2}/\sigma_{v_{k}^{(i)}}^{2} \right)},
$$

where $S_{l}^{(i)}$ and $\tilde{S}_{l}^{(i)}$ are, respectively, the sets of symbols $\tilde{s}_{k}$ with the bit indexed by $l$, corresponding to the $i$th transmission, equal to zero or one. These LLRs are subsequently depunctured to obtain the sequence $\lambda_{l}^{(i)} = \{\lambda_{l,i}^{(i)}\}_{i=1}^{2N_{c}}$ that is then passed to the soft Viterbi decoder.

When using TM $n$ on the $i$th (re)transmission, and with the assumption of a slow block-fading channel model, the instantaneous packet error rate (PER) at the output of the soft Viterbi decoder can be approximated as

$$
\text{PER}_{n,i}(\gamma) \approx \begin{cases} 
1 & 0 \leq \gamma < \frac{\gamma_{n,i}}{\gamma_{n,i-1}} \\
\frac{a_{n,i} e^{-a_{n,i}(i+1)}}{a_{n,i}} & \gamma \geq \frac{\gamma_{n,i-1}}{\gamma_{n,i-1}}
\end{cases},
$$

where $a_{n,i}$, $g_{n,i}$, and $\gamma_{n,i}$ listed in Table I, are the fitting parameters for TM $n$ and transmission number $0$ with a packet length of $N_{0} = 1080$ bits. These parameters have been obtained by least-squares fitting the above approximate expression of the PER to the curves obtained through simulation. Numerical results have been obtained as the ratio between the erroneously transmitted packets after transmission

<table>
<thead>
<tr>
<th>Mode</th>
<th>Code Rate</th>
<th>$R_{n}(\text{bits/symbol})$</th>
<th>$g_{n,0}$</th>
<th>$\gamma_{p,n,0}$(dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/2</td>
<td>1/2</td>
<td>11.104</td>
<td>3.315</td>
<td>-1.212</td>
</tr>
<tr>
<td>2/3</td>
<td>3/4</td>
<td>5.144</td>
<td>0.691</td>
<td>5.92</td>
</tr>
<tr>
<td>5/6</td>
<td>3/6</td>
<td>3.266</td>
<td>0.339</td>
<td>0.26</td>
</tr>
</tbody>
</table>

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE Globecom 2010 proceedings.
number 0 using TM \( n \) and the overall number of transmitted packets.

After soft Viterbi decoding, an error detection process (using, for instance, a cyclic redundancy check code) is performed and the corresponding ACK/NACK message is fed back to the ARQ controller. Given the short length of the ACK/NACK messages and the use of a high degree of FEC protection, an error-free and instantaneous ARQ feedback channel will be assumed in this paper.

III. DISCRETE TIME MARKOV CHAIN MODEL AND ANALYSIS

A. Arrival process

As in [13], it is assumed in this paper that the packet generation model adheres to a special case of the discrete batch Markovian arrival process (D-BMAP) [20], which is characterized by a transition probability matrix

\[
U = \sum_{a=0}^{\infty} U_a = \begin{bmatrix}
  u(0, 0) & \cdots & u(A-1, 0) \\
  \vdots & \ddots & \vdots \\
  u(0, A-1) & \cdots & u(A-1, A-1)
\end{bmatrix},
\]

where \( u(a_\mu, a_\mu') \) denotes the probability of a transition from phase \( a_\mu \) to phase \( a_\mu' \) with a batch arrival of size \( a_\mu \) new packets. The sub-stochastic matrices \( U_a \), for all \( a \in \{0, \ldots, A-1\} \) are constructed by keeping only the \((a+1)\)th row of \( U \) and setting all other rows to zero, and by definition of the D-BMAP, \( U_a = 0 \) for all \( a > A \).

Owing to the Markovian property of the arrival process we have that \( \omega = \omega U \) and \( \omega 1_A = 1 \), where \( \omega \) denotes the D-BMAP steady-phase probability vector and \( 1_A \) is an all-ones column vector of length \( A \). Then the average arrival rate \( \lambda \) can be calculated as

\[
\lambda = \omega \sum_{a=0}^{A-1} a U_a 1_A.
\]

B. Two-dimensional Markov channel modeling

Let us consider the Rayleigh block-fading channel quantities \( \gamma_\nu \) and \( \delta_\nu \). Let us also partition the ranges of \( \gamma_\nu \) and \( \delta_\nu \) into sets of nonoverlapping two-dimensional cells defined by the partitions \( \Gamma^c = \{[0, \gamma_0^c), [\gamma_0^c, \gamma_1^c), \ldots, [\gamma^{c}_{K-1}, \gamma^{c}_K)\} \) with \( \gamma_0^c = 0 \) and \( \gamma^{c}_K = \infty \), and \( \Delta = \{(-\infty, 0), [0, \infty)\}, \) respectively. The partition \( \Gamma^c \) is designed using the methodology introduced in [13]. Thus, a first-order two-dimensional Markov channel model can be defined where each state of the channel corresponds to one of such cells. That is, the Markov chain state of the channel at time instant \( t = \nu T_f \) can be denoted as \( \zeta_\nu = (\chi_\nu, \Delta_\nu) \), \( \nu = 0, 1, \ldots, \infty \), where \( \chi_\nu = k \) if and only if \( \gamma_0^c \leq \gamma_\nu < \gamma^{c}_{k+1} \) and \( \Delta_\nu = 0 \) (or \( \Delta_\nu = 1 \)) if and only if \( \delta_\nu < 0 \) (or \( \delta_\nu \geq 0 \)).

C. Physical layer two-dimensional Markov model

Keeping in mind both the TM selection process used by the AMC scheme and the first-order two-dimensional Markov channel model, let us now partition the range of \( \gamma_\nu \) into the set of non-overlapping intervals defined by the partition \( \Gamma^{m,c} = \{[0, \gamma_0^{m,c}), [\gamma_0^{m,c}, \gamma_1^{m,c}), \ldots, [\gamma^{m,c}_{K-1}, \gamma^{m,c}_K)\} \) with \( \gamma_0^{m,c} = 0 \) and \( \gamma^{m,c}_K = \infty \), where each partition interval \( [\gamma^{m,c}_k, \gamma^{m,c}_{k+1}) \) is characterized by a particular combination of TM and channel state. As in Subsection III-B, let us also consider the partition of \( \delta_\nu \) into the set of non-overlapping intervals \( \Delta = \{(-\infty, 0), [0, \infty)\} \). Using this two-dimensional partitioning, a first-order two-dimensional Markov model for the physical layer can be defined where each state corresponds to one of such two-dimensional rectangular-shaped cells. Furthermore, the physical layer Markov chain state at time instant \( t = \nu T_f \) can be denoted as \( \zeta_\nu = (\phi_\nu, \Delta_\nu) \), \( \nu = 0, 1, \ldots, \infty \), where \( \varphi_\nu \in \{0, \ldots, N_{PHY} - 1\} \) denotes the combination of TM and channel state in this frame interval and \( \Delta_\nu \in \{0, 1\} \) is used to denote the up or down\(^1 \) characteristic of the instantaneous SNR in time frame interval \( t = (\nu - 1)T_f \).

At any time instant \( t = \nu T_f \) the physical-layer state can be univocally characterized by an integer number \( n_\nu = 2\varphi_\nu + \Delta_\nu \) and obviously, \( n_\nu \in \{0, \ldots, 2N_{PHY} - 1\} \). The physical layer will be in a state \( n \in \{0, \ldots, 2N_{PHY} - 1\} \) with a steady-state probability \( p_{PHY}'(n) \), that can be calculated using [13, eqs. (8)-(9)], and each of these states will be characterized by a conditional average PER for the \( i \)th (re)transmission given by

\[
PER_{n,i} = \left\{ \begin{array}{ll}
p_{PHY}'(n) \int_{n/2}^{n/2+1} \int_0^x PER_{n,i-1}((x,y)) dy dx & \text{if } n \text{ even} \\
p_{PHY}'(n) \int_{(n+1)/2}^{(n+1)/2+1} \int_0^x PER_{n,i-1}((x,y)) dy dx & \text{if } n \text{ odd}
\end{array} \right.
\]

where \( \beta_n \) denotes the TM corresponding to the \( i \)th physical layer state, \( P_{\nu, \gamma_{\nu-1}}(x, y) \) is the joint pdf of the random variables \( \gamma_\nu \) and \( \gamma_{\nu-1} \), and \( PER_{n,i-1}(\cdot) \) represents the probability that the \( i \)th transmission attempt fails conditioned on that the previous transmissions have been erroneous.

\[
PER_{n,i-1}(\gamma) = \frac{PER_{n,i}(\gamma)}{PER_{n,i}(\gamma)}.
\]

Furthermore, the physical-layer FSMC will be characterized by a transition probability matrix

\[
P_S = \left[ P_S(n_\nu, n_\nu') \right]_{n_\nu, n_\nu'=0}^{2N_{PHY}-1},
\]

whose elements can be calculated using [13, eqs. (11)-(15)]. In this paper, the steady-state probabilities, the conditional average packet error rates and the state-transition probabilities have all been computed either numerically or by simulation. Clarke’s statistical Rayleigh fading process, characterized by a maximum normalized Doppler frequency \( f_d T_f \), has been used to model the wireless flat-fading channel [11].

D. Embedded Markov chain

The queueing process induced by both the truncated HARQ protocol and the AMC scheme can be formulated in discrete time with one time unit equal to one frame interval. The

\(^1\)If \( \gamma_\nu < \gamma_{\nu-1} \) then the instantaneous SNR is descending and it can be tagged as down; on the contrary, if \( \gamma_\nu \geq \gamma_{\nu-1} \) then the instantaneous SNR is ascending and it can be tagged as up.
system states are observed at the beginning of each time unit. Let \( \sigma_t = (q_{t}, a_{t}, \varphi_{t}, \Delta_{t}) \) denote the system state at time instant \( t = \nu T_f \), where \( q_{t} = (q_{t,0}, \ldots, q_{t,N_r}) \) transmits the queue state at this time instant, with \( q_{t,i} \) denoting the number of packets in the queue that have been already transmitted \( i \) times and \( Q_{\nu} \triangleq \sum_{i=0}^{N_r} q_{t,i} \in \{0, \ldots, Q\} \) denoting the total number of packets in the queue, \( a_{t} \in \{0, \ldots, A-1\} \) represents the phase of the D-BMAP, \( \varphi_{t} \in \{0, \ldots, N_{PHY} - 1\} \) represents the combination of TM and channel state in this frame interval and \( \Delta_{t} \in \{0, 1\} \) is used to denote the up or down characteristic of the instantaneous SNR in time frame interval \( t = (\nu-1)T_f \). If we just look at the set of time instants \( t = \nu T_f, \nu = 0, 1, \ldots, \infty \), the transitions between states are Markovian. Therefore, an embedded Markov chain can be used to describe the underlying queueing process. The state space of this embedded finite state Markov chain is \( S = \{S_n\}_{n=1}^{N_r} \) with size \( N_r = 2N_{PHY}A\sum_{k=0}^{Q}(k+1) \).

The transition probability from state \( S_{\mu} = (q_{\mu}, a_{\mu}, n_{\mu}) \in S \) to state \( S_{\mu'} = (q_{\mu'}, a_{\mu'}, n_{\mu'}) \in S \), where \( n_{\mu} = 2\varphi_{\mu} + \Delta_{\mu} \) and \( n_{\mu'} = 2\varphi_{\mu'} + \Delta_{\mu'} \) can be written as:

\[
P_{S_{\mu}, S_{\mu'}} = u(a_{\mu}, a_{\mu'})P_{s}(n_{\mu}, n_{\mu'})P_{q_{\mu}, q_{\mu'}}|a_{\mu}, n_{\mu},
\]

where \( P_{s}(n_{\mu}, n_{\mu'}) \) denotes the transition probability between D-BMAP phases \( a_{\mu} \) and \( a_{\mu'} \), which can be obtained from matrix \( U \), \( P_{s}, n_{\mu}, n_{\mu'} \) is the physical layer transition probability between states \( n_{\mu} \) and \( n_{\mu'} \), which can be derived from matrix \( P_{s} \), and \( P_{q_{\mu}, q_{\mu'}}|a_{\mu}, n_{\mu} \) is the transition probability from state \( q_{\mu} \) to state \( q_{\mu'} \) when the D-BMAP is in phase \( a_{\mu} \) and the physical layer state is \( n_{\mu} \).

Consider that the system state is \( S_{\mu} \), and that \( Q_{\mu,i} = \sum_{t=0}^{N_r} q_{t,i} \) represents the number of packets in the queue that have already been transmitted \( i \) or more times (obviously, \( Q_{\mu,0} = Q_{\mu} \)). Let \( \tau_{\mu} = \min \{Q_{\mu,c_{\mu}}\} \) denote the number of transmitted packets, \( \tau_{\mu,i} = \min \{q_{\mu,i}, \tau_{\mu} - Q_{\mu,i+1}\} \) the number of transmitted packets among those in the queue that have been already transmitted \( i \) times and \( \epsilon_{\mu,i} \) the number of packets erroneously transmitted among those in the queue that have been already transmitted \( i \) times. Using this notation, the feasible queue transitions can be expressed as:

\[
q_{\mu',N_r} = \begin{cases} 
q_{\mu,N_r} - \tau_{\mu}, & \tau_{\mu} \leq q_{\mu,N_r}, \\
\epsilon_{\mu,N_r-1}, & \tau_{\mu} > q_{\mu,N_r}.
\end{cases}
\]

\[
q_{\mu',i} = \begin{cases} 
q_{\mu,i} - \tau_{\mu}, & \tau_{\mu} \leq Q_{\mu,i+1}, \\
Q_{\mu,i} - \tau_{\mu}, & Q_{\mu,i+1} < \tau_{\mu} \leq Q_{\mu,i}, \\
\epsilon_{\mu,i-1}, & \tau_{\mu} > Q_{\mu,i}.
\end{cases}
\]

for \( i \in \{1, \ldots, N_r - 1\} \), and

\[
q_{\mu',0} = \begin{cases} 
\min \{Q - Q_{\mu,1} + q_{\mu,0} + a_{\mu} \}, & \tau_{\mu} \leq Q_{\mu,1}, \\
\min \{Q - Q_{\mu,1}, Q_{\mu} + a_{\mu} - \tau_{\mu} \}, & Q_{\mu,1} < \tau_{\mu} \leq Q_{\mu}.
\end{cases}
\]

Consequently, by defining \( P_{y}(z) \triangleq \frac{(z^{\mu} - x^{\mu-1} - y)}{y^{\mu} - x^{\mu-1} - y} \), and assuming a slow block-fading channel, the state transition probabilities can be safely approximated as

\[
P_{q_{\mu,N_r}, q_{\mu'}, N_r | q_{\mu}, n_{\mu}, n_{\mu'}} = \begin{cases} 
1, & q_{\mu,N_r} = q_{\mu',N_r} - \tau_{\mu}, \tau_{\mu} \leq q_{\mu,N_r} , \\
P_{\nu,T_{\mu}, N_r-1}^{PHY}(F_{\text{PER}}^{\text{PHY}}(q_{\mu,N_r} - \tau_{\mu})), & q_{\mu,N_r} > q_{\mu',N_r} , \\
0, & q_{\mu,N_r} > q_{\mu',N_r} - \tau_{\mu}, \tau_{\mu} \geq q_{\mu',N_r} , \\
1, & q_{\mu,i} = q_{\mu',i}, \tau_{\mu} \leq Q_{\mu,i+1} , \\
P_{\nu,T_{\mu}, i}^{PHY}(F_{\text{PER}}^{\text{PHY}}(q_{\mu,i} - \tau_{\mu})), & Q_{\mu,i+1} < \tau_{\mu} \leq Q_{\mu,i} , \\
0, & \tau_{\mu} \geq Q_{\mu,i} , \\
q_{\mu,i} = Q_{\mu,i} - \tau_{\mu}, & \tau_{\mu} \geq Q_{\mu,i} , \\
0, & \tau_{\mu} \geq Q_{\mu,i} + Q_{\mu,i} + a_{\mu} - \tau_{\mu}.
\end{cases}
\]

for \( i \in \{1, \ldots, N_r - 1\} \), and

\[
P_{\nu,T_{\mu}, 0}^{PHY}(F_{\text{PER}}^{\text{PHY}}(q_{\mu}, n_{\mu})), \tau_{\mu} \leq Q_{\mu,1} , \\
0, \tau_{\mu} > Q_{\mu,1}.
\]

C. Packet loss rate and throughput

The number of lost packets due to buffer overflow when the system changes from state \( S_{\mu} \) to state \( S_{\mu'} \) is given by:

\[
N_{\nu,T_{\mu}, N_r} = \sum_{i=1}^{N_r} \pi_{S_{\mu}, S_{\mu'}} N_{\nu,T_{\mu}, N_r} | S_{\mu}, S_{\mu'},
\]

and the packet loss rate \( P_{\nu,T_{\mu}, N_r} \) (measured in packets per frame) can then be obtained as \( P_{\nu,T_{\mu}, N_r} = N_{\nu,T_{\mu}, N_r} / \lambda \).

The number of lost packets due to exceeding the maximum number of allowed retransmissions when the system is \( S_{\mu} \) can be calculated as

\[
N_{\nu,T_{\mu}, N_r} = \sum_{l=0}^{\tau_{\mu}, N_r} l \pi_{S_{\mu}, S_{\mu'}}^{\nu,T_{\mu}, N_r} (F_{\text{PER}}^{\text{PHY}}(q_{\mu}, n_{\mu})),
\]

with \( \tau_{\mu,N_r} = \min \{q_{\mu,N_r}, c_{\mu} \} \), and the corresponding average number of lost packets can be obtained as:

\[
N_{\nu,T_{\mu}, N_r} = \sum_{\mu=1}^{N_r} \pi_{S_{\mu}, N_{\nu,T_{\mu}, N_r}} | S_{\mu}.
\]
Accordingly, the probability of packet loss due to exceeding $N_r$ retransmissions can be expressed as $P_{l_{ARQ}} = \bar{N}_{l_{ARQ}}/\lambda$.

In our finite buffering truncated ARQ-based error control system, the packet loss rate $P_l$ (measured in packets per frame) can be expressed as $P_l = P_{l_{BAO}} + P_{l_{ARQ}}$, and given the packet loss rate $P_l$, the average throughput can be calculated as $\eta = \lambda(1 - P_l)$.

**F. Average queue length and average packet delay**

Using the well-known Little’s formula [21], the average delay for our embedded Markov chain can be calculated as $D_l = L_q/\lambda(1 - P_{l_{BAO}})$, where $L_q$ denotes the average number of packets in the queue that can be obtained as

$$L_q = \sum_{\mu=1}^{N_q} \pi_{\mu} S_{\mu} Q_{\mu}.$$

**IV. CROSS-LAYER OPTIMIZATION**

As shown in previous sections, given a buffer size $Q$, an average SNR $\bar{\gamma}$ and a normalized maximum Doppler frequency $f_d T_f$, interesting performance measures include, for instance, throughput, average packet delay or packet loss rate. These measures are a function of the AMC TM switching levels $\Gamma^m \in \mathbb{R}_+^{M+1}$, where $\mathbb{R}_+$ denotes the set of non-negative real numbers, and the measured or estimated arrival packet rate $\lambda \in \Theta$, where $\Theta$ is the range of feasible arrival rate values.

In order to simplify the optimization approach, let us define the AMC switching thresholds as the instantaneous SNR values for which the value of $PER_{n,N_r} = P_0$, that is,

$$\gamma^m_n = \frac{1}{(N_r + 1) g_{n,0} \ln \left( \frac{a_{n,0}}{P_0} \right)}$$

for $n = 1, \ldots, M - 1$, with $\gamma^m_0 = 0$, and $\gamma^m_M = \infty$. In this case, the simplified constrained optimization problem can be formulated as

$$(P_{0_{opt}}, \lambda_{opt}) = \arg \max_{P_0 \in \mathbb{R}_+, \lambda \in \Theta} \eta (P_0, \lambda)$$

subject to the constraints

$$P_l (P_{0_{opt}}, \lambda_{opt}) \leq P_{l_{max}}, \quad D_l (P_{0_{opt}}, \lambda_{opt}) \leq D_{l_{max}}.$$

Obviously, because $P_0$ and $\lambda$ lie in a bounded space $\mathbb{R}_+ \times \Theta$, we can resort to a 2-D exhaustive search to numerically solve the proposed cross-layer optimization problem.

**V. NUMERICAL RESULTS**

In order to verify the validity of the proposed cross-layer framework, analytical results obtained with our proposed 2D-FSMC model will be confronted with computer simulation results obtained using Clarke’s statistical Rayleigh model. Unless otherwise specified, numerical results correspond to the following default parameters: normalized maximum Doppler frequency $f_d T_f = 0.02$, average received SNR $\bar{\gamma} = 8$ dB, buffer size $Q = 8$, number of channel states $K = 5$, parameter $b = 6$ and a D-BMAP that has been either characterized with the transition probability matrix $U$ defined in [13, Section VI] or parameterized to obtain a truncated Poisson process with variable arrival rate $\lambda$.

The dependence of the two components of the average packet loss rate, $P_{l_{BAO}}$ and $P_{l_{ARQ}}$, on the target average PER $P_0$ is depicted in Fig. 1. The analysis of these graphs reveals that, as expected, higher $P_0$ values imply an increase in the packet loss rate due to exceeding the maximum number of allowed retransmissions $P_{l_{ARQ}}$. Moreover, it also implies the utilization of higher order TMs, which leads to an increment of the queuing service rate and, consequently, to a decrease in the buffer overflow probability $P_{l_{BO}}$. Nevertheless, it can be appreciated that when the number of allowed retransmissions $N_r$ grows, the increase of the service rate for high $P_0$ values cannot cope with the huge number of required retransmissions resulting in more packets remaining in the queue and $P_{l_{BO}}$ increasing accordingly. A decrease in $P_{l_{BAO}}$ and $P_{l_{ARQ}}$ is obtained when higher $N_r$ values are allowed. Figure 2 shows an increment in the maximum throughput $\eta$ and a reduction in the minimum delay $D_l$ as the number of allowed retransmissions is increased; however the most significant gain is obtained in going from $N_r = 0$ to $N_r = 1$, and the advantage of higher $N_r$ values becomes marginal. These results are in agreement with those obtained in [14]. As it can be observed, in all cases the behaviour of the simulation of a FIFO queuing system under a truncated HARQ-CC protocol and with a physical layer based on Clarke’s model, is faithfully reproduced by the presented analytical physical-link layer 2D-FSMC model. The shape of the curves and the location of the optimum values obtained by

---

**Fig. 1.** Average packet loss rate vs. target PER.

**Fig. 2.** Average throughput and average packet delay vs. target PER.
simulation (Clarke’s) coincide with those obtained using the proposed analytical model (2D-FSMC), which is particularly important to ensure an optimal cross-layer design. QoS-guaranteed traffic characterized by a maximum average packet loss rate \( P_{\text{loss}} \) = 0.1 packets/frame and a maximum average packet delay \( D_{\text{max}} \) = 4 frames has been considered with the objective of analyzing the dependence of the proposed cross-layer design on the maximum normalized Doppler frequency \( f_d T_f \), with an average received SNR \( \gamma = 6 \) dB. Traffic has been generated using a truncated-Poisson process with a truncation length of 3 packets (e.g. \( \Theta = (0,3) \)). Figure 3 illustrates the dependence of the obtained QoS parameters \( \eta \), \( \hat{P} \) and \( D_1 \) on the maximum normalized Doppler frequency, as well as the optimum values of the target PER \( P_{\text{opt}}^{\text{PER}} \) and arrival rate \( \lambda_{\text{opt}} \). It shows that for larger \( f_d T_f \) values, which correspond to better channel conditions, the optimum sustainable arrival rate \( \lambda_{\text{opt}} \) increases while ensuring the fulfillment of both QoS requirements. Obviously, higher \( N_t \) results in an increase of the optimum throughput. As expected, HARQ-CC performs better than Type-I Hybrid FEC/ARQ scheme.

VI. CONCLUSION

We have proposed a novel link level queuing model that generalizes the analytical tools presented in our previous contribution [14] to AMC/HARQ-CC wireless systems. Using our proposed first-order two-dimensional FSMC model-based approach, analytical expressions for fundamental performance metrics have been derived. The analytical link level queuing model has then been used to formulate a cross-layer design conceived as a constrained optimization problem to exploit the joint impact on QoS performance of both AMC at the PHY-layer and HARQ-CC-based error control at the DLC-layer. Numerical examples have shown that the derived performance metrics of our analytical model faithfully reproduce simulation results based on Clarke’s statistical Rayleigh fading model.

ACKNOWLEDGMENTS

This work has been supported in part by the MEC and FEDER under project COSMOS (TEC2008-02422).

REFERENCES