Modulation and Coding Schemes Selection for Type-II HARQ in Time-Correlated Fading Channels

Nassar Ksairi* and Philippe Ciblat†

*Mathematical and Algorithmic Sciences Lab, France Research Center, Huawei Technologies Co. Ltd.
†Telecom ParisTech, Paris, France

***Abstract*—Assuming the use of Hybrid ARQ, we propose to find out the best modulation and coding scheme at each retransmission round when partially outdated channel state information is available at the transmitter. Our analysis and proposed algorithm are based on Markov decision processes.**

I. Introduction

It is now well-spread to combine packet retransmission and forward error coding in order to obtain reliable wireless links. This leads to the so-called Hybrid ARQ (HARQ) technique which operates with a one-bit feedback. In order to choose more relevantly the design parameters of the retransmissions, more information on the channel than the 1-bit feedback can be reported to the transmitter side at each retransmission round. Based on this framework, many works have been devoted to the optimization of the HARQ mechanism.

Most of these works capture the quality of transmission through information-theoretic tool such as the outage probability, the instantaneous capacity or the resource consumption (the degradation of the transmission is then only due to the randomness of the channel) [1], [2], [3], [4]. In this case, the parameters to be optimized may typically be the scheduling (number of subcarriers, which subcarriers), the power (which power at each retransmission step, which average power, etc), the coding rate (which size of retransmitted packet), etc. Finally, the parameters are modified dynamically as soon as the quality of the previous retransmission is available at the transmitter side with the goal to optimize a long-term performance, see, [2], [3], [4], [5] and references therein. For instance, a well-adapted tool for handling this optimization is the so-called Markovian Decision Process (MDP) [6], [7]. For instance, in [3], the power and coding rate of a Type-II HARQ based secondary user is optimized via MDP tools given some information from the primary user behavior.

Only few works consider a real Modulation and Coding Scheme (MCS) to be optimized. In that case, it is more difficult to capture the link performance since an accurate approximation of the physical layer is needed. To the best of our knowledge, the most important works in this topic are done by [8], [9]. The best MCS is chosen according to the instantaneous Bit Error Rate (BER) in [8], and to the instantaneous expected throughput in [9]. There works do not ensure the long-term average throughput optimization.

Therefore the contribution of the paper is to optimize the MCS retransmission round by retransmission round for maximizing the long-term average throughput for any HARQ mechanism and for any time-correlated fading channel. The appropriate tool to fix our problem is the MDP. Nevertheless some assumptions explained later prevent us from using exactly the same MDP based tools as in [3], [2] which also contributes to the originality of the paper.

The paper is organized as follows: in Section II, the system model is described. In Section III, the underlying Markov chain model for the HARQ mechanism is exhibited. In Section IV, we give the optimization problem and we prove that it has a solution. In Section V, an algorithm solving the optimization issue is introduced. Section VI is devoted to numerical results. Concluding remarks are drawn in Section VII.

II. System model

Consider a cellular network with K active mobile nodes into a cell. We consider a OFDMA scheme with N subcarriers covering a total bandwidth of WHz.

A. Channel model

The link between the base station and the k-th user ($k \in \{1, 2, \ldots, K\}$) is assumed to be block fading and frequency selective with a M-long channel impulse response $\mathbf{h}_{k,t} = [h_{k,t}(0), \ldots, h_{k,t}(M-1)]^{\mathrm{T}}$ during any block t that is independent-tap Rayleigh distributed. The superscript $(.)^{\mathrm{T}}$ stands for the transposition operator. The stochastic process associated with this time variation of each channel impulse response is assumed to be correlated and to follow a first-order Gauss-Markov model

$$\mathbf{h}_{k,t} = \alpha \mathbf{h}_{k,t-1} + \sqrt{1-\alpha^2}\mathbf{w}_{k,t}(m),\ t \geq 0\,. \qquad (1)$$

where α is a constant satisfying $0 < \alpha < 1$ and $\mathbf{w}_{k,t} = [w_{k,t}(0), \ldots, w_{k,t}(M-1)]^{\mathrm{T}}$ is an i.i.d zero-mean circularly-symmetric Gaussian vector with variance $\varsigma_k^2(m)$ for the m-th component.

Let $\mathbf{H}_{k,t} = [H_{k,t}(0), \ldots, H_{k,t}(N-1)]^{\mathrm{T}}$ be the discrete Fourier transform of $\mathbf{h}_{k,t}$. Conditioned on the channel impulse response during the previous block, the subcarriers of the k-th link during the t-th block are thus distributed as follows: $H_{k,t}(n)|\mathbf{h}_{k,t-1}$ is a Gaussian circularly-symmetric

This work was supported by LEXNET European grant and by ACE ANR grant partly. Moreover N. Ksairi is on leave from Telecom ParisTech and is now with France Research Center, Huawei Technologies.

random variable with mean $\alpha H_{k,t-1}(n)$ and with variance $(M/N)(1-\alpha^2)\varsigma_k^2$ where $\varsigma_k^2 = \sum_{m=0}^{M-1} \varsigma_k^2(m)$.

If the n-th subcarrier is assigned to the k-th link during the transmission of the j-th OFDM symbol ($j \in \{1, \ldots, J\}$) of the t-th block, the received signal at this subcarrier is

$$Y_{k,t}(j,n) = H_{k,t}(n) X_{k,t}(j,n) + Z_{k,t}(j,n)\,, \qquad (2)$$

where $X_{k,t}(j,n)$ is the symbol transmitted at subcarrier n of the j-th OFDM symbol of the t-th block belonging to the k-th link, and where $Z_{k,t}(j,n)$ is a zero-mean Gaussian additive noise with variance $\sigma_k^2 = N_0 W$ where N_0 is the noise power spectral density.

B. Block structure

We assume that the transmitter (base station if downlink; mobile if uplink) during the t-block has only the knowledge of the channel coefficients corresponding to the $(t-1)$-th block via a feedback link *i.e.,* $\{\mathbf{h}_{k,t-1}\}_{k=1\cdots K}$. A block of J consecutive OFDM symbols is split into B resource blocks (RB) of identical length. Each RB offers Q resource elements, typically, subcarriers. For the t-th block, the network manager assigns a set $\mathcal{B}_{k,t}$ of RBs to the k-th user. Due to the orthogonality property of OFDMA, we have $\mathcal{B}_{k_1,t} \cap \mathcal{B}_{k_2,t} = \emptyset, \forall k_1, k_2 \in \{1, \ldots, K\}, \forall t$.

In order to facilitate the optimization problem (as actually done in LTE [10] for the so-called semi-persistent scheduling mode), we assume

- the assigned power by the network manager to the k-link during the t-th block is independent of the subcarrier index, i.e., we have $P_{k,t} \stackrel{\text{def}}{=} \mathbb{E}\left[|X_{k,t}(j,n)|^2\right]$ for its $Q|\mathcal{B}_{k,t}|$ subcarriers. We denote $E_{k,t} \stackrel{\text{def}}{=} NP_{k,t}/W$ the energy consumed to transmit one symbol on one subcarrier on the link of the k-th user.
- the RB and power assignments are not done at each block, i.e., these assignments are constant within a large time window and we can set $\mathcal{B}_{k,t} = \mathcal{B}_k$ and $E_{k,t} = E_k$, $\forall t$. From now on, $B_k \stackrel{\text{def}}{=} |\mathcal{B}_k|$.

C. HARQ description

Let us move on to the description of the HARQ mechanism. Having K users, we consider a *K parallel stop-and-wait* [11] with at most L transmissions. During one block, each active user sends one packet corresponding to one HARQ round. We assume that each active link k receives an infinite stream of information bits coming from the upper layer while arranged in data packets of variable length $D_{k,t}$ transmitted by the current HARQ process. Any HARQ mechanism (ARQ, Type-I HARQ, Type-II HARQ) will be supported by our work. Hereafter, due to page limitation, we will only focus on Incremental Redundancy HARQ (IR-HARQ) but extension to other above-mentioned mechanisms are straightforward. The data packet is firstly encoded by a FEC code of rate R_0. The resulting codeword of size $D_{k,t}/R_0$ bits, assuming to be a multiple of QB_k (i.e., all the subcarriers included in the available RB are used) is interleaved and then split into L variable-size segments that are referred to as Redundancy Versions (RVs) in the LTE standard. These RV are sequentially sent to the receiver according to the positive or negative acknowledgment (ACK or NACK, respectively). Let $\ell_{k,t}$ be the random variable representing the number of attempts that have been made so far (before the current t-th block) to transmit the *latest* data packet destined to user k. For any t, $\ell_{k,t}$ is as follows: $\ell_{k,t} = 0$ if the transmitter has received an ACK; $\ell \in \{1, \cdots, L-1\}$ if the transmitter has received ℓ NACKs after ℓ attempts, $\ell_{k,t} = L$ if the latest data transmission has failed after L attempts. Denote by $R_{k,t}$ the coding rate as it will be seen by the receiver after getting the t-th block. One can easily check that $R_{k,t} \geq R_0, \forall t$ and that $R_{k,t-(\ell_{k,t} \bmod L)}, \ldots, R_{k,t-1}, R_{k,t}$ is a decreasing sequence. The term $t-(\ell_{k,t} \bmod L)$ corresponds to the index of block coinciding with the first attempt to transmit the *current* data packet.

Focus first on the first round of a new data packet transmission, i.e., when $\ell_{k,t} = 0, L$. We select the constellation and the coding rate. We will assume that this selection is the same for each subcarrier and RB index of the user k. Consequently, we choose the symbols from a $2^{m_{k,t}}$-QAM constellation and the coding rate $R_{k,t}$, via the puncturing of the mother code. The modulation and coding scheme (MCS) associated with the link k during the t-th block is so represented by the couple $\mathrm{MCS}_{k,t} \stackrel{\text{def}}{=} (m_{k,t}, R_{k,t})$. The number of available MCS couples is finite and denoted by C'. In the following, we use $d_{k,t} \in \{0, \ldots, C'-1\}$ to designate the index of the MCS chosen during the first round of the current HARQ process. During retransmissions of the same process, $d_{k,t}$ evidently stays the same. Moreover, as $D_{k,t}$ is the number of information bits of the k-th user during the t-th block, we have $D_{k,t} = m_{k,t} R_{k,t} B_k$ when $\ell_{k,t} \in \{0, L\}$. So $D_{k,t}$ is now determined and can not be modified by the retransmission process but only by the first round of the next data packet.

Let us now focus on the round associated with retransmission of current data packet, i.e., $\ell_{k,t} \in \{1, \ldots, L-1\}$. Before going further, notice that the coding rate, the modulation index and the number of assigned RBs are all related as follows:

$$R_{k,t} = \frac{D_{k,t}}{\sum_{j=t-(\ell_{k,t} \bmod L)}^{t} m_{k,j} B_k}\,. \qquad (3)$$

Consequently, the modulation index $m_{k,t}$ is the only MCS parameter to be modified since the coding rate directly comes from Eq. (3). The number of available modulation indices is $(C - C' + 1)$ (with $C \geq C'$). For convenience, we use the values $\{C', \cdots, C\}$ to designate the index of modulation schemes used in retransmission. This index should be chosen such that the number of transmitted bits in block t does not exceed the number of remaining bits of the mother codeword that has not been sent yet. Here, $m(\mathrm{MCS})$ designates the modulation index associated with $\mathrm{MCS} \in \mathcal{M} \stackrel{\text{def}}{=} \{0, \ldots, C\}$. When $\ell \in \{1, \cdots, L-1\}$, this leads to

$$Q \sum_{j=t-\ell_{k,t}}^{t} m(\mathrm{MCS}_{k,j}) B_k \leq D_{k,t}/R_0. \qquad (4)$$

D. Performance metrics

Let $\mathcal{E}_{k,t}$ be the event that decoding the current information packet upon receiving the t-th block leads to an error and define $\pi_{k,t} \overset{\text{def}}{=} \mathbb{P}\{\mathcal{E}_{k,t}\}$ as the Block Error Rate (BLER) based on the codeword associated with the concatenation/combining of the first received $(l_{k,t} \bmod L)+1$ rounds. In [12], $\pi_{k,t}$ can be approximated by

$$\pi_{k,t} = 1 - de^{-c_F \overline{\epsilon}_{k,t}^F - \cdots - c_1 \overline{\epsilon}_{k,t}} , \tag{5}$$

where $F \in \mathbb{N}^*$ is the approximation order and where $\overline{\epsilon}_{k,t}$ is an approximation of the average physical-layer Bit Error Rate (BER) associated with the hard-decision made on the coded bits upon the reception of the $(\ell_{k,t} \bmod L)+1$ rounds at the end of the t-th block. The parameters $d, F, c_1, \ldots, c_F$ must be adjusted such that the approximation fit the BLER simulated curves. If $m_{k,j}$-QAM modulation (where $m_{k,j} = m(\mathrm{MCS}_{k,j})$) is used for the transmission of the j-th block of user k ($j \in \{t-(l_{k,t} \bmod L), \ldots, t\}$), we have

$$\overline{\epsilon}_{k,t} = \frac{\sum_{j=t-\tilde{\ell}_{k,t}}^{t} m_{k,j} \sum_{c\in\mathcal{B}_k, n\in\mathcal{N}_c} e^{-1.6\frac{E_k |H_{k,j}(n)|^2}{(2^{m_{k,j}}-1)N_0}}}{5\tilde{N}\sum_{j=t-\tilde{\ell}_{k,t}}^{t} m_{k,j} B_k} . \tag{6}$$

with $\tilde{\ell}_{k,t} = (\ell_{k,t} \bmod L)$, $\mathcal{N}_c$ the set of subcarriers associated with the c-th RB. Notice that we assume constant-length RB, so $\tilde{N} = |\mathcal{N}_c|$ is independent of c and $\tilde{N} = Q/J$.

E. Application to LTE

The previous system model fits perfectly well with LTE [10]. In terms of terminology, the block can be replaced with Transmission Time Interval (TTI), data packet with transport block ($D_{k,t}$ is called the Transport Block Size -TBS$_{k,t}$-), $\pi_{k,t}$ with post-HARQ BLER. Each block is composed of $J = 14$ OFDM symbols. Each RB has $Q = 168$ resource elements. If $W = 20$MHz, then $B = 110$. Moreover, the block is 1ms long. Concerning HARQ, we have $L = 4$ and $R_0 = 1/3$. Finally, $C' = 29$ and $C = 31$ and the modulation for retransmission is only QPSK, 16-QAM and 64-QAM.

III. Proposed Markov chain model

As the HARQ can be modeled with an underlying Markov chain and as our goal is to maximize long-term average throughput, the MDP will be the relevant framework : the decision-making entity (here, the transmitter of the k-th link at the t-th block) has to choose an **action** $a_{k,t}$ from a set of available actions $\mathcal{A}$ called the **action space**. As seen in Section II, the only possible action is the selection of $\mathrm{MCS}_{k,t}$ i.e., $a_{k,t} \overset{\text{def}}{=} \mathrm{MCS}_{k,t}$ which means that $\mathcal{A} = \mathcal{M}$. The action $a_{k,t}$ will affect the evolution of a random variable called the **state** from its current value $s_{k,t}$ to $s_{k,t+1}$ belonging to the set of values $\mathcal{S}$ called the **space space**. The definition of the state should include all the link-related variables that are relevant for the computation of the throughput and for the optimization of the action selection while ensuring that the ensuing random process $(s_{k,t})_{t\in\mathbb{N}}$ is a Markov process. Therefore we propose

$$s_{k,t} \overset{\text{def}}{=} (\ell_{k,t}, \mathbf{h}_{k,t-1}, \epsilon_{k,t}, \delta_{k,t}, \mu_{k,t}) \tag{7}$$

with

$$\epsilon_{k,t} \overset{\text{def}}{=} \mathbb{1}_{\{1\cdots L-1\}}(\ell_{k,t}) \times \overline{\epsilon}_{k,t-1} , \tag{8a}$$

$$\delta_{k,t} \overset{\text{def}}{=} \mathbb{1}_{\{0,1\cdots L-1\}}(\ell_{k,t}) \times d_{k,t-1} , \tag{8b}$$

$$\mu_{k,t} \overset{\text{def}}{=} \mathbb{1}_{\{0,1\cdots L-1\}}(\ell_{k,t}) \times \sum_{j=t-(\ell_{k,t} \bmod L)}^{t-1} m_{k,j} , \tag{8c}$$

where $\mathbb{1}_{\mathcal{X}}(.)$ is the indicator function for any set $\mathcal{X}$. Note that $\delta_{k,t}$ is only reset (to zero, conventionally) in the case of a failed transmission ($\ell_{k,t} = L$), while $\epsilon_{k,t}$ and $\mu_{k,t}$ are reset in both cases of a failed and a successful transmission ($\ell_{k,t} = 0, L$). Let us move on the action space. Actually, we define the set $\mathcal{A}(s) \subset \mathcal{A}$ of really possible actions for $a_{k,t}$ when $s_{k,t} = s$. Let $\ell(s)$ be the first component of the state s. In case of new transmission (i.e, $\ell(s) \in \{0, L\}$), we have

$$\mathcal{A}(s) = \{0, \ldots, C'-1\} . \tag{9}$$

Let $\mu(s)$ be the fifth component of the state s and $D(\delta)$ be the number of information bits when the MCS $\delta \in \{0, \ldots, C'-1\}$ has been chosen during the first round. In case of retransmission (i.e., $\ell(s) \in \{1, \ldots, L-1\}$), we have

$$\mathcal{A}(s) = \left\{ \{C', \ldots, C\} \middle| m(\mathrm{MCS}) \leq \frac{D(\delta(s))}{QB_kR_0} - \mu(s) \right\} . \tag{10}$$

Before going further, we introduce the following definition and technical assumption.

Definition 1. *A* deterministic stationary Markov policy *is a measurable function $f : \mathcal{S} \to \mathcal{A}$ with $f(s) \in \mathcal{A}(s), \forall s \in \mathcal{S}$. We denoted by $\mathcal{F}$ the set of all such policies. The function f is also called the* decision function*.*

Assumption 1. *The policy f is such that $\forall \mathrm{MCS}, m \in \{0, \ldots, C'-1\}$, the sets $\left\{\mathbf{h} \in \mathbb{C}^M | f(0, \mathbf{h}, 0, 0, 0) = \mathrm{MCS}\right\}$ and $\left\{\mathbf{h} \in \mathbb{C}^M | f(0, \mathbf{h}, 0, m, 0) = \mathrm{MCS}\right\}$ are non-negligible with respect to the Lebesgue measure on $\mathbb{C}^M$.*

Assumption 1 is not restrictive since none of the possible C' first-round MCSs is impossible, or equivalently, any one of them could be selected for a new HARQ data transmission.

To complete the MDP characterization, we need to describe the evolution rules from the state $s_{k,t}$ to the random state $s_{k,t+1}$ given an action $a_{k,t}$. In addition to Eq. (1), we have

$$\ell_{k,t+1} = \begin{cases} 0 & \text{if ACK at block } t \\ (\ell_{k,t} \bmod L) + 1 & \text{if NACK at block } t, \end{cases} \tag{11a}$$

$$\epsilon_{k,t+1} = \mathbb{1}_{\{1\cdots L-1\}}(\ell_{k,t+1}) \times \left(\frac{\mu_{k,t}}{m(a_{k,t}) + \mu_{k,t}} \epsilon_{k,t} + \frac{m(a_{k,t})}{5\tilde{N}B_k(m(a_{k,t}) + \mu_{k,t})} \sum_{c\in\mathcal{B}_k, n\in\mathcal{N}_c} e^{\frac{-1.6E_k|H_{k,t}(n)|^2}{(2^{m(a_{k,t})}-1)N_0}} \right) , \tag{11b}$$

$$\delta_{k,t+1} = \mathbb{1}_{\{0,1\cdots L-1\}}(\ell_{k,t+1}) d_{k,t} , \tag{11c}$$

$$\mu_{k,t+1} = \mathbb{1}_{\{1\cdots L-1\}}(\ell_{k,t+1}) \times (\mu_{k,t} + m(a_{k,t})) . \tag{11d}$$

Moreover, Eq. (11b) was derived from Eq. (6) by using $\mu_{k,t}$ given in Eq. (8c). The randomness in Eq. (1) and Eqs. (11a)-(11d) is accounted for by the *probability transition kernel* $Q_f(.|s)$ defined for any state $s \in \mathcal{S}$ as

$$Q_{f_k}(\mathcal{X}|s) \stackrel{\text{def}}{=} \mathbb{P}_{f_k,s_{k,0}}\left[s_{k,t+1} \in \mathcal{X} \middle| s_{k,t} = s, a_{k,t} = f_k(s_{k,t})\right], \tag{12}$$

where $\mathbb{P}_{f_k,s_{k,0}}$ stands for the probability measure on $\mathcal{S}$ that underlies the evolution of the Markov process when policy f_k is applied and the initial state is $s_{k,0} = (0, \mathbf{h}_{k,-1}, 0, 0, 0)$. For lack of space, the expression of Q_{f_k} is omitted.

IV. Proposed optimization issue

Using MDP terminology, the number of information bits the receiver gets when it succeeds in decoding the current codeword is called the **instantaneous reward**. In our case, the reward is nonzero only when the current state coincides with the first round of a new HARQ process following the reception of an ACK. More precisely, the instantaneous reward (in bits/s) associated with state $s = (\ell, \mathbf{h}, \epsilon, \delta, \mu)$ is defined as

$$r(s) \stackrel{\text{def}}{=} \begin{cases} \frac{D(\delta)}{T}, & \text{if } \ell = 0, \\ 0, & \text{otherwise,} \end{cases} \tag{13}$$

where T is the duration in seconds of the transmission block. Note that the instantaneous reward in our model can take different values depending on the current state unlike [3]. reward is constant. This apparently slight different has a significant effect of making the proofs in our case much more involved. Now that the instantaneous reward is defined, the average throughput (in bits/s) of the link can be expressed as the policy-dependent long-term **average reward**:

$$\eta_k^{f_k}(s_{k,0}) = \liminf_{t\to\infty} \frac{1}{t}\mathbb{E}_{f_k,s_{k,0}}\left[\sum_{j=0}^{t-1} r(s_{k,j})\right]. \tag{14}$$

Maximizing the average throughput of the link is thus equivalent to solving the following problem.

Problem 1. *Compute the optimal average reward* $\eta_k^*(s_{k,0}) \stackrel{def}{=} \sup_{f_k\in\mathcal{F}} \eta_k^{f_k}(s_{k,0})$.

V. Proposed algorithm

A. Problem solvability

Problem 1 is said to be *solvable* if and only if there exists a policy $f_k^* \in \mathcal{F}$ satisfying $\eta_k^{f_k^*}(s) = \eta_k^*(s)$ for all $s \in \mathcal{S}$. A *sufficient* condition for this solvability is if there exists a positive constant η_k^* and a real-valued bounded measurable function g on $\mathcal{S}$ such that [6], $\forall s \in \mathcal{S}$,

$$\eta_k^* + g(s) = \max_{a\in\mathcal{A}(s)}\left[r(s) + \int_{\mathcal{S}} g(y)Q(dy|s,a)\right], \tag{15}$$

where $Q(.|s,a)$ is the transition kernel given a state and an action, and then the optimal average reward is independent of the initial state

$$\eta_k^*(s) = \eta_k^*, \ \forall s \in \mathcal{S}, \tag{16}$$

and the optimal decision function f_k^* is obtained as follows.

$$f_k^*(s) \stackrel{\text{def}}{=} \arg\max_{a\in\mathcal{A}(s)}\left[r(s) + \int_{\mathcal{S}} g(y)Q(dy|s,a)\right], \forall s \in \mathcal{S}. \tag{17}$$

To prove the validity of the condition in Eq. (15), an approach has been proposed in the literature when the following condition [13, p.56] is satisfied:

$$\exists \beta \in (0,1) | \sup_{(s,a),(s',a')\in\mathcal{K}} \|Q(.|s,a) - Q(.|s',a')\|_{\text{TV}} \leq 2\beta, \tag{18}$$

where $\|.\|_{\text{TV}}$ is the *total variation norm* of signed measures and $\mathcal{K}$ is the set of admissible state-actions pairs *i.e.,* $\mathcal{K} \stackrel{\text{def}}{=} \left\{(s,a) \in \mathcal{S}\times\mathcal{A} \middle| a \in \mathcal{A}(s)\right\}$. Unfortunately, mainly as our instantaneous reward function $r(s)$ is *not* constant with respect to the state s, the above condition is not satisfied by our problem. We therefore propose to resort to the so-called *vanishing-discount* approach [6] to prove the solvability. First, we need the following assumption.

Assumption 2. *The function* $(\mathbf{h}_{k,t}, \ldots, \mathbf{h}_{k,t+L-1}) \mapsto \mathbb{P}\left[\mathcal{E}_{k,t}, \ldots, \mathcal{E}_{k,t+L-1} | \mathbf{h}_{k,t}, \ldots, \mathbf{h}_{k,t+L-1}\right]$ *is strictly positive.*

Assumption 2 is mild as it means that the probability of failure of the transmission of a data packet is non-zero whatever the channel realization. We then obtain the main Theorem. Due to lack of space, proof is omitted.

Theorem 1. *Let Assumptions 1 and 2 hold. There exists a bounded function* h *on* $\mathcal{S}$ *and a decision function* f_k^* *such that Eqs.* (15) *and* (17) *hold. Thus Problem 1 is solvable.*

B. Proposed algorithm based on the Value Iteration approach

We now want to design an algorithm for exhibiting decision functions leading to the optimal long-term average reward η_k^*. Let $v_{k,0}(s) = 0$ for all $s \in \mathcal{S}$ and define for any $t \geq 1$:

$$v_{k,t}(s) \stackrel{\text{def}}{=} \max_{a\in\mathcal{A}(s)}\left[r(s) + \int_{\mathcal{S}} v_{k,t-1}(y)Q(dy|s,a)\right]. \tag{19}$$

Let $f_{k,0} : \mathcal{S} \to \mathcal{A}$ be arbitrary and define $f_{k,t} : \mathcal{S} \to \mathcal{A}$ for any $t \geq 1$ as the decision function obtained as the argument of the maximum in the RHS of Eq. (19), i.e., we get

$$v_{k,t}(s) = r(s) + \int_{\mathcal{S}} v_{k,t-1}(y)Q(dy|s, f_{k,t}(s)). \tag{20}$$

The existence of such a decision function is ensured since the instantaneous reward function r is bounded and the set of actions $\mathcal{A}$ is finite. The Markov policy $\{f_{k,t}\}_{t\in\mathbb{N}}$ is called a *Value Iteration* (VI) policy. Theorem 2 states that the VI algorithm converges to the solution of Problem 1.

Theorem 2. *Let Assumptions 1 and 2 hold, then:*

$$u_{k,t}(s) \to_{t\to\infty} g(s), \qquad w_{k,t}(s) \to_{t\to\infty} \eta_k^* \tag{21}$$

where $u_{k,t}(s) \stackrel{def}{=} v_{k,t}(s) - v_{k,t}(z)$ *and* $w_{k,t}(s) \stackrel{def}{=} v_{k,t}(s) - v_{k,t-1}(s)$ *with* z *a fixed state chosen arbitrarily, and where* η_k^* *and* g *are the solutions of Eq.* (15)*. Thus* η_k^* *is the optimal long-term average reward.*

Proof: We define the error function associated with the t-th step of the VI algorithm as:

$$e_{k,t}(s) \stackrel{\text{def}}{=} v_{k,t}(s) - t\eta_k^* - g(s). \tag{22}$$

One can easily show that

$$\begin{aligned} w_{k,t}(s) &= \eta_k^* + e_{k,t}(s) - e_{k,t-1}(s), \\ u_{k,t}(s) &= g(s) + e_{k,t}(s) - e_{k,t}(z). \end{aligned} \tag{23}$$

Moreover, from Lemma 4.6 in [13, p.63], we have

$$\|e_{k,t}\| \stackrel{\text{def}}{=} \sup_{s\in\mathcal{S}} |e_{k,t}(s)| \leq \|e_{k,0}\| \stackrel{\text{def}}{=} \sup_{s\in\mathcal{S}} |e_{k,0}(s)|. \tag{24}$$

The rest of the proof relies on the fact the sequence of functions $\{e_{k,t}\}_{t\in\mathbb{N}}$ is *equicontinuous* [6] thanks to the properties of $Q(.|.)$ in our model. This equicontinuity combined with Eq. (24) helps us to prove that the sequence $e_{k,t}(s)$ converges uniformly in s to a constant identical for any $s \in \mathcal{S}$. This convergence, plugged into Eq. (23), concludes the proof. ■

Remark 1. *In practice, it is advantageous from a computational-complexity point of view to first execute the VI algorithm 'off-line' (using a sufficiently large number of different values of α and of the average SNR). The resulting steady-state decision functions can then be stored to be used later as lookup tables during real-time operation.*

VI. Numerical results

Simulations results were obtained using the system parameters given in Subsection II-E, except for C', C and L set to 9, 12 and 3 respectively. We put $M = 1$ (flat fading channel), the initial coding rates are $3/8, 1/2, 3/4$ and $N_0 = -170$ dBm/Hz. We consider SNR=30 dB and the value of α is 0.969 corresponding to a mobile terminal moving at 25 km/h.

Practical implementation of the VI algorithm requires that the continuous-valued variables of the state vector i.e., $\mathbf{h}_{k,t-1}$ and $\epsilon_{k,t}$, should be discretized so that the size of the state space becomes finite. The validity of this discretization procedure can be established using [14]. Discretization has been done using 8 uniform quantization levels for ϵ and 4 uniform levels for each real dimension of each component of $\mathbf{h}_{k,t-1}$.

Remark 2. *In principle, if the probability distribution of $\overline{\epsilon}_{k,t}$is known, it is possible to drop the variable $\mathbf{h}_{k,t-1}$ from the definition of the MDP, thus reducing both the size of the state space and the feedback overhead. This issue will be addressed in future work.*

We now compare the performance of the proposed algorithm to three other MCS adaptation schemes. The first scheme consists in choosing at each time the MCS randomly from the set of available MCSs. The second scheme consists in choosing at each time the MCS only based on the (outdated) CSI $\mathbf{h}_{k,t-1}$ i.e., without taking into account neither the fact that HARQ is used nor the statistics of the associated Markov chain [11]. Actually the highest possible MCS that has an uncoded error probability smaller or equal to 10% is chosen. The third scheme is less trivial as it takes into account the use of HARQ, but without accounting for the future consequences of the current action i.e., without taking into account the statistics of the associated Markov chain. Actually the MCS leading to the smallest BER at each time is chosen.

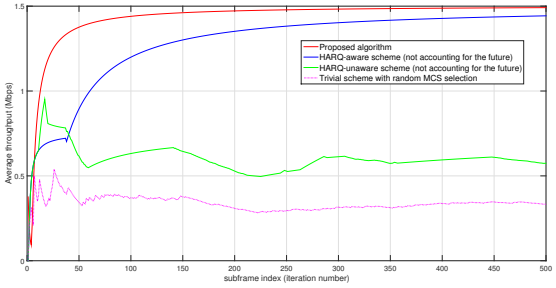

Figure 1. Average throughput of the proposed and existing algorithms

VII. Conclusion

Markov Decision Process tool enables us to propose an algorithm for selecting the best modulation and coding scheme within a HARQ mechanism. In further works, we would like to encompass power and subcarrier allocation as well and to make the proposed algorithm work with minimal loss in performance when feedback is reduced to the scalar value $\overline{\epsilon}_{k,t-1}$ rather than the vector $\mathbf{h}_{k,t-1}$.

References

[1] T. V. K. Chaitanya and E. G. Larsson, "Adaptive power allocation for HARQ with chase combining in correlated rayleigh fading channels," *IEEE Commun. Lett.*, vol. 3, no. 4, pp. 169–172, April 2014.

[2] L. Szczecinski, S. 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.

[3] R. Tajan, C. Poulliat, and I. Fijalkow, "Interference management for cognitive radio systems exploiting primary IR-HARQ: A constrained markov decision process approach," in *Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, November 2012.

[4] J. Perret and D. Tuninetti, "Repetition protocols for block fading channels that combine transmission requests and state information," in *IEEE International Conference on Communications (ICC)*, May 2008.

[5] T. Villa, R. Merz, and R. Knopp, "Dynamic resource allocation in heterogeneous networks," in *IEEE Global Communications Conference (GLOBECOM)*, December 2013.

[6] O. Hernández-Lerma and J. B. Lasserre, *Discrete-Time Markov Control Processes: Basic Optimality Criteria*. Springer, 1996.

[7] E. Altman, *Constrained Markov Decision Processes*. Chapman and Hall/CRC, 1999.

[8] J. Harsini, F. Lahouti, M. Levorato, and M. Zorzi, "A type II hybrid ARQ protocol with adaptive modulation and coding for time-correlated fading channels: Analysis and design," in *IEEE International Conference on Communications (ICC)*, May 2010.

[9] I. Stupia, V. Lottici, F. Giannetti, and L. Vandendorpe, "Link resource adaptation for multiantenna bit-interleaved coded multicarrier systems," *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3644–3656, June 2012.

[10] E. Dahlman, S. Parkvall, and J. Skold, *4G: LTE/LTE-Advanced for Mobile Broadband*. Academic Press, 2011.

[11] F. Khan, *LTE for 4G Mobile Broadband: Air Interface Technologies and Performance*. Cambridge University Press, 2009.

[12] J. Escudero-Garzás, B. Devillers, L. Vandendorpe, and A. García-Armada, "Subchannel, bit and power allocation in multiuser OFDM systems for goodput optimization with fairness," in *Symposium on Wireless Personal Multimedia Communications (WPMC)*, April 2009.

[13] O. Hernández-Lerma, *Adaptive Markov Control Processes*. Springer-Verlag, 1989.

[14] C. Chow, "Multigrid algorithms and complexity results for discrete-time stochastic control and related fixed-point problems," Ph.D. dissertation, MIT, 1989.