Information-theoretic multi-user power adaptation in retransmission schemes

Romain Tajan*, Philippe Ciblat†,
*IMS UMR 5218, Bordeaux INP, Univ. of Bordeaux, CNRS, F-33405, Talence, France
†LTCI, CNRS, Télécom ParisTech, Université Paris Saclay, F-75013, Paris, France

***Abstract*—In this article, we address the problem of adapting power among users and their packet retransmissions in a downlink OFDMA context for optimizing a weighted sum throughput. Retransmission protocols are modeled using information-theoretic considerations. We propose a Constrained Markov Decision Process approach to optimize the weighted sum throughput subject to an average power constraint. Finally, we introduce structured policies to reduce the complexity of solving the CMDP.**

I. Introduction

In modern telecommunication standards such as High Speed Downlink Packet Access (HSDPA) or 4G LTE (Long Term Evolution), reliability of data transmission is ensured using Hybrid Automatic Repeat reQuest (HARQ) protocols. HARQ protocols are a combination of Forward Error Correction (FEC) and Automatic Repeat reQuest (ARQ) protocol. In recent telecommunication standards, feedback signals are used to optimize HARQ parameters such as the rate or the power of transmission.

Resource allocation for HARQ protocols has been well investigated in a single user context. In [1], [2], allocating power is shown to improve the throughput in presence of partial Channel State Information (CSI) at the transmitter. Other works consider outdated CSI, the transmitter has information on past fading only [3]–[8]. These contributions exploit a Markov chain structure of the HARQ protocol to provide either a linear or a dynamic programming approach for solving optimization problems. In particular, [3] and [9] exploit time correlation of the fading channel to propose respectively a Partially Observable Markov Decision Process POMDP) and a Markov Decision Process (MDP) for solving resource allocation problems considering HARQ protocols.

Resource allocation has been considered in multiuser OFDMA context in [10] [11]. However, in these cases, resource allocation does not vary across the different transmission rounds of HARQ protocols. Both [10] and [11] solved a resource allocation problem using the Lagrangian approach. This approach was possible since when HARQ properties (power, bandwidth) do not vary across HARQ rounds, analytic expressions for throughput or outage and their derives exist. When HARQ characteristics are allowed to vary, expression of throughput or outage rarely exist. Furthermore, when they exist, these expressions are generally difficult to exploit without approximations [5] [4].

In this article, we optimize the spectral efficiency of a downlink Orthogonal Frequency Division Multiple Access (OFDMA) network by allocating power across users and across the different HARQ rounds. We propose to model the power allocation problem as a Constrained Markov Decision Process (CMDP) as it is often done in the single user context. This model could be thought as the multiuser extension of the work proposed in [6]. However, the Linear Programming approach proposed in [6] cannot directly be applied in our context since the dimension of the CMDP increases because of the multiuser context. Hence we propose an alternative approach that replaces solving a large LP by solving small LP and a separable concave program.

The paper is organized as follows: in Section II, the system model is presented. In Section III, we build a Markov chain to model the state of the different HARQ protocols. Section IV, is dedicated to the optimization of the network spectral efficiency under average power constraint. Simulation results are presented in Section V. Conclusions are drawn in Section VI.

II. Network and system model

A. Channel Model

Consider a slotted downlink OFDMA network where K users share N orthogonal subcarriers. At the BaseStation (BS), a scheduler has assigned N_k subcarriers to user k. We will assume that $N = \sum_k N_k$. The set of subcarriers allocated to user k is denoted by $\mathcal{N}_k$ and the fraction of bandwidth occupied by user k is denoted $\beta_k = \frac{N_k}{N}$.

The BS has only outdated Channel State Information (CSI) and statistical CSI. We consider that fading coefficients associated with the different subcarriers of user k are independent and identically distributed, hence the the BS allocates its power equally among subcarriers of user k:

$$P_{k,n,t} = P_{k,t} \quad \forall n \in \mathcal{N}_k. \tag{1}$$

The input-output relationship between the BS and user k on subcarrier n, for all time u in slot t is

$$Y_{k,n,t}(u) = \sqrt{P_{k,t}} H_{k,n,t} X_{k,n,t}(u) + Z_{k,n,t}(u), \tag{2}$$

where $Y_{k,n,t}$ is the received complex signal, $X_{k,n,t}$ is the transmitted signal distributed as $\mathcal{CN}(0,1)$, $P_{k,t}$ is the power allocated to user k by the BS, $H_{k,n,t}$ is a complex channel gain, and the noise $Z_{k,n,t}$ is complex valued Gaussian random variable $\mathcal{CN}(0,1)$. The noise on subcarrier n is independent from the noise on other tones. $H_{k,n,t}$ is independent from channel gains on other subcarriers or other slots and $H_{k,n,t}$ remains constant over the slot. Furthermore, for every user k we consider a Rayleigh fading channel. The channel coefficients under coherent detection $\gamma_{k,n,t} = |H_{k,n,t}|^2$ are independent and identically distributed with distribution

$$f_{\gamma_{k,n,t}}(\gamma) = \frac{e^{-\gamma/\bar{\gamma}_k}}{\bar{\gamma}_k} \quad n \in \mathcal{N}_k, \quad t \in \mathbb{N} \tag{3}$$

where $\bar{\gamma}_k$ is the average Signal to Noise Ratio (SNR) of user k.

B. IR-HARQ protocol

The Incremental Redundancy HARQ uses jointly ARQ and FEC to ensure reliability of data transmission. The BS handles K IR-HARQ protocols in parallel (one for each user). The IR-HARQ protocol of user k has at most L_k transmission attempts to limit delay.

We present now IR-HARQ protocol. Before sending a data packet to user k, the BS encodes this packet using a code of rate $\frac{R_k}{L_k}$ and divide this codeword into L_k frames of constant length. Each time user k fails decoding the information packet, it sends a Negative ACKnowledgement to the BS and the BS transmits a new frame. Let $\ell_{k,t}$ be a random variable accounting for the number of frames that has been unsuccessfully sent up to slot t for the transmission of the current code word. If $\ell_{k,t} \in \{0, \ldots, L_k - 1\}$ and user k still fails to decode, then $\ell_{k,t+1} = \ell_{k,t} + 1$ and a new frame is transmitted by BS. On the contrary, if user k successfully decodes the information packet, then $\ell_{k,t+1} = 0$ and BS starts the transmission of a new information packet at slot $t + 1$. If $\ell_{k,t} = L_k$, then the BS has no more frames to transmit, an *outage* is declared and the BS starts the transmission of a new information packet.

The random variable $\ell_{k,t}$ can be used in a stochastic modeling of IR-HARQ protocols, as it is proposed in the next section.

III. Markov chain model of the network

If outdated CSIT is available (channel at time $t - 1$ is known or related extra mutual information and we are sending the frame t), the IR-HARQ of every user can be represented as a Markov chain with two variables: i) $\ell_{k,t}$, and ii) the accumulated mutual information. We will then use this model to propose a Constrained Markov Decision Processes (CMDP).

If no outdated CSIT is available, then only the first variable is required for describing the IR-HARQ protocols. The considered problem would be a Partially Observable CMDP, but this is out of the scope of this paper.

A. State space

It is now admitted that for large codewords, IR-HARQ protocols behave as mutual information accumulators (see [12]). The extra mutual information provided by the transmission at slot t is

$$\Delta(\boldsymbol{\gamma}_{k,t}, P_{k,t}) = \frac{1}{N_k} \sum_{n\in\mathcal{N}_k} \log_2\left(1 + P_{k,t}\gamma_{k,n,t}\right)$$

where $\boldsymbol{\gamma}_{k,t} = \{\gamma_{k,n,t}\}_{n\in\mathcal{N}_k}$.

The IR-HARQ protocol for user k at slot t is represented by the couple $s_{k,t} = (\ell_{k,t}, I_{k,t})$, where $I_{k,t}$ is the ACcumulated Mutual Information (ACMI). The transition rule for $I_{k,t}$ depends on $\ell_{k,t}$. Indeed, given $\ell_{k,t}$, a packet is always sent at time t. Its nature depends on the value of $\ell_{k,t}$. If $\ell_{k,t} \neq 0, L_k$, the packet corresponds to redundancy. If $\ell_{k,t} = 0$ or $\ell_{k,t} = L_k$, this is a new data packet. Therefore, we have

$$I_{k,t+1} = \mathbb{1}_{1,\ldots,L-1}\left(\ell_{k,t}\right) I_{k,t} + \Delta(\boldsymbol{\gamma}_{k,t}, P_{k,t}). \tag{4}$$

Equivalently, the transition rule for $\ell_{k,t+1}$ depends on $\ell_{k,t}$ and $I_{k,t}$. Since, successful decoding at time $t+1$ happens if and only if $I_{k,t+1} > R_k$ the transition function for $\ell_{k,t+1}$ is

$$\ell_{k,t+1} = \mathbb{1}_{I_{k,t+1}>R_k}\left(I_{k,t+1}\right)\left(\mathbb{1}_{\ell_{k,t}<L_k}\left(\ell_{k,t}\right)\ell_{k,t} + 1\right). \tag{5}$$

The vector $\mathbf{s}_t = (s_{1,t}, ..., s_{K,t})$ represents the state of the global network at the BS at time t. The set of all possible states is called the state space.

B. Action space and policies

Let $\mathcal{P}_k$ be the set of power admissible for user k. The BS allocates at time t the power $P_{k,t} \in \mathcal{P}_k$ to every subcarrier in $\mathcal{N}_k$. The complete power allocation at time t is then given by $\mathbf{P}_t = (P_{1,t}, P_{2,t}, \ldots, P_{K,t})$. The set of admissible power vector is $\mathcal{P} = \mathcal{P}_1 \times \mathcal{P}_2 \times \cdots \times \mathcal{P}_K$.

At time t, the BS knows s_t but does not know $\gamma_{k,t}$ (outdated CSI). We consider that the power allocation π is defined by the conditional probability $\pi(d\mathbf{P}_t|\mathbf{s}_t)$. Such allocation policies are called *randomized stationary policies* (the conditional probability does not change with t). A special case using the Dirac measure, are power allocations of the form $\mathbf{P}_t = \pi_D(\mathbf{s}_t)$ which can also be written as $\pi(d\mathbf{P}_t|\mathbf{s}_t) = \delta_{\pi_D(\mathbf{s}_t)}(d\mathbf{P}_t)$. These policies are named *deterministic policies*.

C. Transition kernel for $\mathbf{s}_t$

Combining equations (4) and (5) for all k, we can express $\mathbf{s}_{t+1}$ as

$$\mathbf{s}_{t+1} = f(\mathbf{s}_t, \mathbf{P}_t, \boldsymbol{\gamma}_t) \tag{6}$$

where $\boldsymbol{\gamma}_t = \left(\boldsymbol{\gamma}_{1,t}, \ldots, \boldsymbol{\gamma}_{K,t}\right)$. From (6), it can be proved that for any measurable subset S

$$\mathbb{P}(\mathbf{s}_{t+1} \in S|\mathbf{s}_t, \mathbf{P}_t, \ldots, \mathbf{s}_0, \mathbf{P}_0) = \mathbb{P}(\mathbf{s}_{t+1} \in S|\mathbf{s}_t, \mathbf{P}_t). \tag{7}$$

For every randomized stationary policy π, $\{\mathbf{s}_t\}$ forms a Markov chain where the probability that $\mathbf{s}_{t+1} \in S$ from s_t is

$$\mathbb{P}_\pi(\mathbf{s}_{t+1} \in S|\mathbf{s}_t) = \int_{\mathcal{P}} \mathbb{P}(\mathbf{s}_{t+1} \in S|\mathbf{s}_t, \mathbf{P}_t)\pi(d\mathbf{P}_t|\mathbf{s}_t) \tag{8}$$

D. Optimization problem

Defining R_k as the equivalent code rate for the transmission on one slot for user k, we introduce the number bits correctly decoded by user k per seconds and per Hertz as

$$r_k(s_{k,t}) = \begin{cases} R_k = \frac{b_k}{B_k T_{slot}} \text{if } \ell_{k,t} = 0 \\ 0 \text{ otherwise} \end{cases} \tag{9}$$

where b_k is the number of information bits in a codeword, T_{slot} is the duration of a slot, and B_k is the bandwidth occupied by user k. The throughput of user k is then

$$\eta_k(\pi) = \liminf_{T\to\infty} \mathbb{E}_\pi \left[\frac{1}{T} \sum_{t=0}^{T-1} r_k(s_{k,t}) \right]. \tag{10}$$

where $\mathbb{E}_\pi[\cdot]$ is the expectation considering that the policy π is used.

The average power allocated to user k is defined in a similar way

$$\bar{P}_k(\pi) = \limsup_{T\to\infty} \mathbb{E}_\pi \left[\frac{1}{T} \sum_{t=0}^{T-1} P_{k,t} \right]. \tag{11}$$

Let $\{\omega_k\}$ be a set of priorities satisfying $\sum_{k=1}^{K} \omega_k = 1$, and $\eta(\pi) = \sum_{k=1}^{K} \omega_k \eta_k(\pi)$ be a weighted throughput network utility, we seek a randomized stationary policy $\pi^\star$ that solves

$$\begin{aligned} \pi^* &= \arg\max_{\pi} \eta(\pi) \quad (12)\\ s.t. \quad & \sum_{k=1}^{K} \beta_k \bar{P}_k(\pi) \leq P_{\max} \end{aligned}$$

where $\beta_k = \frac{N_k}{N}$ is the fraction of bandwidth occupied by user k.

Problem (12) belongs to the class of Constrained Markov Decision Process (CMDP) (see [13]). Usually, CMDPs are proposed over a set of policies that includes non-stationary policies (policies that depend on time). However, based on results of [13] and [6] we can prove that for problem (12) randomized stationary policies define a dominating class of policies. This means that for every policy ϕ (stationary or not, randomized or not), if ϕ is admissible for problem (12), there exists a randomized policy π admissible for (12) such that $\eta(\pi) \geq \eta(\phi)$. The proof of this point is omitted due to lack of space.

Usually in telecommunications, problem (12) would have been solved within the set of deterministic policies (as in [9] for unconstrained MDP). However proving the existence of an optimal deterministic policy for a general CMDP is difficult. Therefore, we consider this optimization problem within the set of randomized stationary policies.

IV. Solving problem (12)

Solving (12) is equivalent to an infinite dimensional Linear Programming (LP) [14]. A solution of this infinite LP can be approched numerically by discretizing the state and action spaces as done in [6]. In our case, the complexity (size of the discrete action and state spaces) is an issue. Indeed, consider that every HARQ protocol has at most L transmission rounds and consider a grid on I_k of size N_I, the discrete state space has a size of $N_I^K(L+1)^K$. Similarly for the action space, suppose that every set $\mathcal{P}_k$ contains N_P power levels, the complete set $\mathcal{P}$ has N_P^K elements. Following [6], we should solve a finite LP with $N_I^K(L+1)^K N_P^K$ variables and $N_I^K(L+1)^K + 2$ constraints. This size of LP becomes rapidly prohibitive. In the next section we propose structured policies called factorized policies that are optimal and that can be used to approximate the solution of the initial infinite LP with a lower complexity.

A. Optimality of factorized policies

A *factorized policy* is a randomized stationary policy that has the following structure:

$$\tilde{\pi}(d\mathbf{P}|\mathbf{s}) = \prod_{k=1}^{K} \tilde{\pi}(dP_k|s_k). \tag{13}$$

A factorized policy is such that the BS allocates power to user k considering the state of user k only. Let $\tilde{\pi}^\star$ be an optimal policy for problem (12) restricted to factorized policies, we now show that

$$\eta(\pi^\star) = \eta(\tilde{\pi}^\star). \tag{14}$$

Since factorized policies are randomized stationary policies we have $\eta(\pi^\star) \geq \eta(\tilde{\pi}^\star)$. To show that $\eta(\pi^\star) \leq \eta(\tilde{\pi}^\star)$, we propose to build $\tilde{\pi}^\star$ from $\pi^\star$.

First, note that problem (12) can be rewritten as

$$\begin{aligned} \max_{\Psi_1,\dots\Psi_K,\pi\in\Pi} \quad & \sum_{k=1}^{K} \omega_k \eta_k(\pi) \quad (15)\\ s.t. \quad & \sum_{k=1}^{K} \beta_k \Psi_k \leq P_{\max} \\ & \bar{P}_k(\pi) \leq \Psi_k,\ \forall k \in [1,K] \\ & \Psi_k \geq 0,\ \forall k \in [1,K] \end{aligned}$$

Indeed, if $\pi^\star$ is optimal for problem (12), $(\bar{P}_1(\pi^\star), \dots, \bar{P}_k(\pi^\star), \pi^\star)$ is optimal for problem (15) and the converse is also true.

Let π_{Ψ_k} be a single user policy for user k that solves

$$\begin{aligned} \pi_{\Psi_k} &= \arg\max_{\pi} \eta_k(\pi) \quad (16)\\ s.t. \quad & \bar{P}_k(\pi) \leq \Psi_k \end{aligned}$$

$\tilde{\pi}^\star$ is built as follows

$$\tilde{\pi}^\star(d\mathbf{P}|\mathbf{s}) = \prod_{k=1}^{K} \pi_{\bar{P}_k(\pi^\star)}(dP_k|s_k) \tag{17}$$

where $\pi_{\bar{P}_k(\pi^\star)}$ is a single user policy found by solving (16) with power constraint $\Psi_k = \bar{P}_k(\pi^\star)$.

By construction $\tilde{\pi}^\star$ is admissible for problem (12) and verifies that $\sum_{k=1}^{K} \omega_k \eta_k(\tilde{\pi}^\star) \geq \sum_{k=1}^{K} \omega_k \eta_k(\pi^\star)$. This proves that the policy $\tilde{\pi}^\star$ is optimal for problem (12).

B. Solving (12) on factorized policies

When restricted to factorized policies similar to those of equation (17), problem (15) becomes

$$\begin{aligned} \max_{\Psi_1,\dots\Psi_K} \quad & \sum_{k=1}^{K} \omega_k \eta_k(\Psi_k) \quad (18)\\ s.t. \quad & \sum_{k=1}^{K} \beta_k \Psi_k \leq P_{\max} \\ & \Psi_k \geq 0,\ \forall k \in [1,K] \end{aligned}$$

where $\eta_k(\Psi_k)$ is a shorthand notation for $\eta_k(\pi_{\Psi_k})$.

Analytic expression for $\eta_k(\Psi_k)$ rarely exists, hence finding a solution with a Lagrangian approach is not possible. However, we prove in Appendix A that for every k, $\eta_k(\Psi_k)$ is increasing and concave. Furthermore, problem (18) is separable since its objective and constraint functions are sums of K functions, each function depending only on an individual variable Ψ_k (see [15]). Concave and separable problems can be approximated by linear programs [15].

To obtain an approximation of the optimal solution for problem (18), every function $\eta_k(\Psi_k)$ is approximated by the piece-wise linear function defined by the points of the form $(\Psi_{k,j}, \eta(\Psi_{k,j}))$ where $0 = \Psi_{k,1} < \Psi_{k,2} < \cdots < \Psi_{k,J_k} = P_{max}$ are given grid points. When replacing every function $\eta_k(\cdot)$ with its piecewise linear approximation, (18) is equivalent to the following linear programming:

$$\begin{aligned} \max_{\lambda_{i,j}} \quad & \sum_{k=1}^{K} \omega_k \sum_{j=1}^{J_k} \lambda_{k,j} \eta_k(\Psi_{k,j}) \\ s.t. \quad & \sum_{k=1}^{K} \beta_k \sum_{j=1}^{J_k} \lambda_{k,j} \Psi_{k,j} \leq P_{\max} \\ & \sum_{j=1}^{J_k} \lambda_{k,j} = 1, \ \forall k \in [1, K] \\ & \lambda_{k,j} \geq 0, \ \forall k \in [1, K], \ \forall j \in [1, J_k]. \end{aligned} \tag{19}$$

Let $(\boldsymbol{\lambda}_1^\star, \boldsymbol{\lambda}_2^\star, \ldots, \boldsymbol{\lambda}_K^\star)$ with $\boldsymbol{\lambda}_k^\star = (\lambda_{k,1}^\star, \ldots, \lambda_{k,J_k}^\star)$ be the solution of problem (19), an approximate solution for (12) given by

$$\hat{\pi}(d\mathbf{P}|\mathbf{s}) = \prod_{k=1}^{K} \pi_{\hat{\Psi}_k}(dP_k|s_k) \tag{20}$$

where $\pi_{\hat{\Psi}_k}$ is a policy that solves problem (16) with power constraint $\hat{\Psi}_k = \sum_{j=1}^{J_k} \lambda_{k,j}^\star \Psi_{k,j}$. Because of the first constraint in problem (19), we have that $\sum_k \beta_k \hat{\Psi}_k \leq P_{max}$ hence $\hat{\pi}$ is an admissible policy for problem (15). Furthermore, using concavity of functions $\eta_k(\cdot)$ and the Jensen inequality, we have

$$\eta(\hat{\pi}) = \sum_{k=1}^{K} \omega_k \eta_k(\hat{\Psi}_k) \geq \sum_{k=1}^{K} \omega_k \sum_{j=1}^{J_k} \lambda_{k,j}^\star \eta_k(\Psi_{k,j}).$$

Note that solving (19) requires solving $\sum_k J_k$ single user LPs with $N_I(L+1)N_P$ variables and $N_I(L+1)+2$ constraints for computing all $\eta_k(\Psi_{k,j})$ and solving one LP with $\sum_k J_k$ variables and $K+1$ constraints. This complexity is much lower than solving a unique LP with $N_I^K(L+1)^K N_P^K$ variables and $N_I^K(L+1)^K + 2$ constraints.

V. Simulation Results

The network is composed of $K = 8$ users sharing $N = 64$ subcarriers equally ($N_k = 8$ and $\beta_k = \frac{1}{8}$). Each user experiences a Rayleigh channel with average channel gains $\bar{\gamma} \in \{0.9, 0.9, 0.8, 0.6, 0.4, 0.4, 0.3, 0.1\}$. Every user uses the same IR-HARQ protocol with $L_k = L = 7$ transmission attempts $R = 9$ bits/s/Hz. We consider $N_I = 16$ quantification levels for the ACMI corresponding to a 4-bits feedback channel. For every user, $\mathcal{P}_k$ is $N_P = 16$ levels in $[-10dBW, 20dBW]$. The average throughput is considered in problem (15) by taking $\omega_k = \frac{1}{8}$.

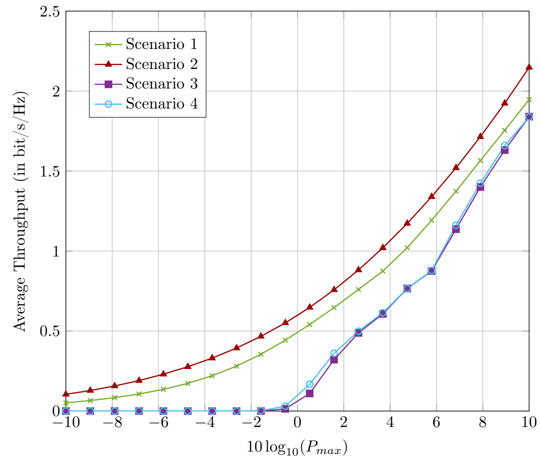

Figure 1. Comparison of average throughputs $K = 8$

In Figure 1, we consider 4 different scenarios. In Scenario 1, we compute the optimal power allocation for problem as proposed in Section IV with $J_k = 64$. In Scenario 2, problem (18) is solved by replacing $\eta_k(\Psi_k)$ with the ergodic capacity:

$$\eta_k(\Psi_k) = \mathbb{E}\left[\log_2\left(1 + \Psi_k |H_{k,n,t}|^2\right)\right].$$

In Scenario 3, power is allocated according to the solution found in Scenario 2 whereas the HARQ protocol described at the beginning of this section is used. In Scenario 4, a statistical CSI based water-filling is considered. This water-filling solution is based on average SNRs $\bar{\gamma}_k$ and is computed by replacing $\eta_k(\Psi_k)$ with

$$\eta_k(\Psi_k) = \log_2(1 + \Psi_k \bar{\gamma}_k).$$

This statistical CSI waterfilling is used to allocate power when the HARQ protocol presented at the beginning of this section is used by all users.

In Figure 1, we observe that the power allocation of Scenario 1 leads to a substantially better throughput than those of Scenarios 3 and 4. Furthermore, the average throughput of the power allocation of Scenario 1 is closed to the one of Scenario 2, which is a consequence of a result shown in [12]: asymptotically in L, the throughput of a single user tends to the ergodic capacity. Scenario 3 and 4 seem to give similar results, which is a consistent result with [16].

In Figure 2 we study the influence of the parameter J_k on the average throughput. We observe that values of J_k as close as 64 are sufficient for solving problem (19).

VI. Conclusion

In this article we consider a power allocation in a multi-user downlink OFDMA network. The BS exploits outdated CSI to allocate power for every user while taking into account their HARQ protocol. The proposed approach to solve the power

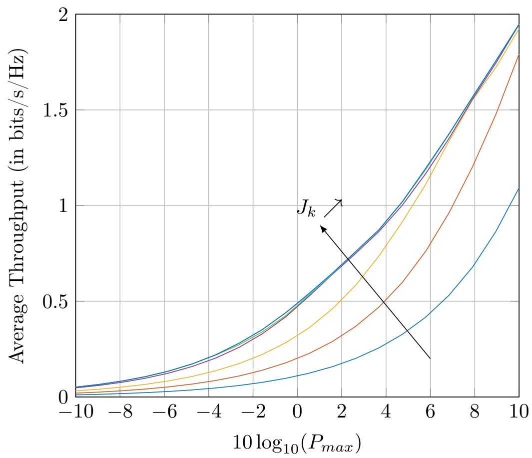

Figure 2. Sum spectral efficiency versus SNR for $J_k \in \{4, 8, 16, 32, 64, 128, 256\}$

allocation problem is firstly to model the whole problem as a Constrained Markov Decision Process and secondly to propose a policy structure that leads to a tractable way of solving the original problem while keeping the complexity reasonable (even for moderate number of users in the network).

Appendix A
Properties of η_k

In this appendix, we show that the function $\eta_k(\Psi)$ is increasing and concave. We recall that for every Ψ, $\eta_k(\Psi) = \eta_k(\pi_\Psi)$ where π_Ψ is a randomized stationary policy solving (16).

A. Proof that η_k is increasing

The proof that $\eta_k(\Psi)$ is increasing is trivial. Indeed, let $\Psi_1 \leq \Psi_2$. Since π_{Ψ_1} is an admissible policy for problem (16) with constraint Ψ_2, we have that $\eta_k(\Psi_1) \leq \eta_k(\Psi_2)$.

B. Proof that η_k is concave

To prove that $\eta_k(\Psi)$ is concave, we show that for every $\Psi_1 < \Psi_2$ and $\alpha \in [0, 1]$

$$\eta_k(\alpha\Psi_1 + (1-\alpha)\Psi_2) \geq \alpha\eta_k(\Psi_1) + (1-\alpha)\eta_k(\Psi_2). \quad (21)$$

To show (21) it is sufficient to provide a randomized stationary policy π verifying

$$\eta_k(\pi) = \alpha\eta_k(\Psi_1) + (1-\alpha)\eta_k(\Psi_2) \quad (22)$$
$$\bar{P}_k(\pi) = \alpha\Psi_1 + (1-\alpha)\Psi_2. \quad (23)$$

Indeed, since π verifies equation (23), it is admissible for problem (16) with constraint $\alpha\Psi_1 + (1-\alpha)\Psi_2$ and

$$\eta_k(\alpha\Psi_1 + (1-\alpha)\Psi_2) \geq \eta_k(\pi) \geq \alpha\eta_k(\Psi_1) + (1-\alpha)\eta_k(\Psi_2)$$

A policy (that is not stationary) ϕ verifying equations (22) and (23) can be built as follows: at each slot T, the BS computes

$$P_k^{(T)} = \frac{1}{T}\sum_{t=0}^{T-1} P_{k,t}.$$

If $P_k^{(T)} \geq \alpha\Psi_1 + (1-\alpha)\Psi_2$ BS uses π_{Ψ_1} else BS uses π_{Ψ_2}. In [17], such policy is called a "steering policy" and it is shown that equations (22) and (23) are verified when $T \to \infty$. The policy ϕ is not a randomized stationary policy. However, adapting results of [14] to problem (16) we can show that for every policy φ (stationary or not), there exists a randomized stationary policy π such that $\eta_k(\pi) \geq \eta_k(\varphi)$ and $\bar{P}_k(\pi) \leq \bar{P}_k(\varphi)$. Applying this result to ϕ leads to the existence of a randomized stationary policy π such that

$$\eta_k(\pi) \geq \eta_k(\phi) \text{ and } \bar{P}_k(\pi) \leq \bar{P}_k(\phi).$$

Since $\bar{P}_k(\phi)$ and $\eta_k(\phi)$ are given by equations (22) and (23), we have that π is also admissible for problem (16), and that

$$\eta_k(\alpha\Psi_1 + (1-\alpha)\Psi_2) \geq \eta_k(\pi) \geq \alpha\eta_k(\Psi_1) + (1-\alpha)\eta_k(\Psi_2)$$

References

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

[2] D. Barbieri and D. Tuninetti, "On repetition protocols and power control for multiple access block-fading channels," in *Communications (ICC), 2011 IEEE International Conference on*, June 2011, pp. 1–5.

[3] A. Karmokar, D. Djonin, and V. Bhargava, "POMDP-based coding rate adaptation for Type-I Hybrid ARQ systems over fading channels with memory," *Wireless Communications, IEEE Transactions on*, vol. 5, no. 12, pp. 3512 –3523, december 2006.

[4] L. Szczecinski, P. Duhamel, and M. Rahman, "Adaptive incremental redundancy for HARQ transmission with outdated CSI," in *IEEE 2011 Global Telecom. Conf. (GLOBECOM 2011)*, Dec 2011, pp. 1–6.

[5] T. Chaitanya and E. Larsson, "Optimal power allocation for Hybrid ARQ with chase combining in i.i.d. Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 61, no. 5, pp. 1835–1846, May 2013.

[6] R. Tajan, "Mécanismes de retransmission Hybrid-ARQ en radio-cognitive." Ph.D. dissertation, Univ. of Cergy-Pontoise, 2013.

[7] B. Makki, A. Graell i Amat, and T. Eriksson, "Green Communication via Power-Optimized HARQ Protocols," *IEEE Trans. Veh. Technol.*, vol. 63, no. 1, pp. 161–177, Jan 2014.

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

[9] N. Ksairi and P. Ciblat, "Modulation and coding schemes selection for type-II HARQ in time-correlated fading channels," in *Signal Processing Advances in Wireless Communications (SPAWC), 2015 IEEE 16th International Workshop on*, June 2015, pp. 91–95.

[10] S. Marcille, P. Ciblat, and C. Le Martret, "Optimal resource allocation in HARQ-based OFDMA wireless networks," in *Military Communication Conference, 2012 - MILCOM 2012*, Oct 2012, pp. 1–6.

[11] N. Ksairi, P. Ciblat, and C. Le Martret, "Optimal resource allocation for Type-II-HARQ-based OFDMA ad hoc networks," in *Global Conference on Signal and Information Processing (GlobalSIP), 2013 IEEE*, Dec 2013, pp. 379–382.

[12] G. Caire and D. Tuninetti, "The throughput of Hybrid-ARQ protocols for the gaussian collision channel," vol. 47, no. 5, pp. 1971–1988, 2001.

[13] E. Altman, *Constrained Markov decision processes*, ser. Stochastic modeling. Chapman & Hall/CRC, 1999.

[14] O. Hernández-Lerma, J. González-Hernández, and R. R. López-Martínez, "Constrained Average Cost Markov Control Processes in Borel Spaces," *SIAM J. Control Optim.*, vol. 42, no. 2, pp. 442–468, Feb. 2003.

[15] S. Stefanov, *Separable Programming: Theory and Methods*. Springer Science & Business Media, 2001, vol. 53.

[16] A. Lozano, A. Tulino, and S. Verdu, "Optimum power allocation for multiuser OFDM with arbitrary signal constellations," *IEEE Trans. Commun.*, vol. 56, no. 5, pp. 828–837, May 2008.

[17] D.-J. Ma and A. M. Makowski, "A class of steering policies under a recurrence condition," in *Proc. 27th IEEE Conf. Decision and Control*. IEEE, 1988, pp. 1192–1197.