LEARNING DISTRIBUTED POWER ALLOCATION POLICIES IN MIMO CHANNELS

Elena Veronica Belmega[†], *Samson Lasaulce*[†], *Mérouane Debbah*[*] *and Are Hjørungnes*[‡]

[†] LSS (joint lab of CNRS,
SUPELEC, Univ. Paris-Sud 11)
Gif-sur-Yvette Cedex, France
email: belmega@lss.supelec.fr,
lasaulce@lss.supelec.fr

[*]Alcatel-Lucent Chair on Flexible
Radio, SUPELEC
Gif-sur-Yvette Cedex, France
email: merouane.debbah@supelec.fr

[‡]UNIK - University Graduate Center
University of Oslo
Kjeller, Norway
email: arehj@unik.no

ABSTRACT

In this paper[1], we study the discrete power allocation game for the fast fading multiple-input multiple-output multiple access channel. Each player or transmitter chooses its own transmit power policy from a certain finite set to optimize its individual transmission rate. First, we prove the existence of at least one pure strategy Nash equilibrium. Then, we investigate two learning algorithms that allow the players to converge to either one of the NE states or to the set of correlated equilibria. At last, we compare the performance of the considered discrete game with the continuous game in [7].

1. INTRODUCTION

Game theory appears to be a suitable framework to analyze self-optimizing wireless networks. The transmitters, based on their knowledge on the environment and cognitive capabilities, allocate their own resources to optimize their individual performance with very little or no intervention from a central authority.

Game theoretical tools have recently been used to study the power allocation problem in networks with multiple antenna terminals. In [1],[2],[3],[4],[5], the authors studies the MIMO slow fading interference channel, in [6] the MIMO cognitive radio channel, and in [7] the multiple access channel. The main drawback of these approaches is the fact that the action sets (or possible choices) of the transmitters are the convex cones of positive semi-definite matrices. In practice, this is an unrealistic assumption and discrete finite action sets should be considered. Another raising issue is related to the iterative water-filling type algorithms that converge to the games' Nash equilibria (NE) states. In order to apply these algorithms, the transmitters are assumed to be strictly rational players that perfectly know the structure of the game (at least their own payoff functions) and the strategies played by the others in the past.

An alternative way of explaining how the players may converge to an NE is the theory of learning [14]. Learning algorithms are long-run processes in which players, with very little knowledge and rationality constraints, try to optimize their benefits. In [8], the authors propose two stochastic learning algorithms that converge to the pure strategy NE and to mixed strategy NE of the energy efficiency game in a single-input single-output (SISO) interference channel. In [10], the multiple access point wireless network is investigated where a large number of users can learn the correlated equilibrium of the game. A similar scenario is studied in [12]. In [9], learning algorithms are proposed in a wireless network where users compete dynamically for the available spectrum. In [11], the authors study learning algorithms in cellular networks where the links are modeled as collision channels. An adaptive algorithm was proposed in [1] for the MIMO interference channel. The proposed algorithm allows the users to converge to a Stackelberg equilibrium by learning the ranks of their own covariance matrices that maximize the system sum-rate.

In this paper, we study the power allocation game in fast fading multiple-input multiple-output (MIMO) multiple access channels (MAC), similarly to [7]. We assume that the action sets of the transmitters are discrete finite sets and consist in uniformly spreading their powers over a subset of antennas. Assuming the single user decoding scheme at the receiver, we show that the proposed game is a potential one and the existence of a pure strategy Nash equilibrium (NE) follows directly. However, the uniqueness of the NE cannot be ensured in general and, thus, several iterative algorithms that converge to one of the NE states are studied. A best-response type algorithm is compared with a reinforcement learning algorithm in terms of system performance, required information, and cognitive capabilities of players. To improve the system performance, we consider a second learning algorithm based on regret matching that converges to the set of correlated equilibria (CE).

We begin our analysis by describing the system model in Sec. 2 and introducing some basic game theoretical concepts. Then, in Sec. 3, we analyze the Nash equilibria of the power allocation game. First, we review the setting of [7] in Subsec. 3.1 and then, study the discrete game in Subsec. 3.2. In Sec. 4, we study two learning algorithms: One that allows the users to converge to one of the NE (see Subsec. 4.1) and another that allows the users to converge to the set of CE (see Subsec. 4.2). We analyze the performance of the different scenarios via numerical simulations in Sec. 5 and conclude with several remarks in Sec. 6.

2. SYSTEM MODEL

We consider a multiple access channel (MAC) composed of an arbitrary number of mobile stations (MS) $K \geq 2$ and a single base station (BS). We further assume that each mobile station is equipped with n_t antennas whereas the base station has n_r antennas. We assume the fast fading model where the receiver has perfect knowledge of the channel matrices. The knowledge required at the transmitters depends on the different scenarios and will be defined accordingly. The equivalent

[1] This work was supported by the Research Council of Norway and the French Ministry of Foreign Affairs through the AURORA project entitled "Communications under Uncertain Topologies".

baseband signal received at the base station is:

$$\underline{Y} = \sum_{k=1}^{K} \mathbf{H}_k \underline{X}_k + \underline{Z}, \qquad (1)$$

where the time index has been ignored and $\underline{X}_k$ is the n_t-dimensional column vector of symbols transmitted by user k, $\mathbf{H}_k \in \mathbb{C}^{n_r \times n_t}$ is the channel matrix (stationary and ergodic process) of user k and $\underline{Z}$ is a n_r-dimensional complex white Gaussian noise distributed as $\mathcal{N}(\underline{0}, \sigma^2 \mathbf{I}_{n_r})$.

In order to take into account the antenna correlation effects at the transmitters and receiver, we will assume the different channel matrices to be structured according to the unitary-independent-unitary model introduced in [23], $\forall k \in \mathscr{K}$, $\mathbf{H}_k = \mathbf{V}_k \tilde{\mathbf{H}}_k \mathbf{W}_k$, where $\mathscr{K} = \{1, ..., K\}$, $\mathbf{V}_k$ and $\mathbf{W}_k$ are deterministic unitary matrices. Also $\tilde{\mathbf{H}}_k$ is an $n_r \times n_t$ matrix whose entries are zero-mean independent complex Gaussian random variables with an arbitrary profile of variances, such that $\mathbb{E}|\tilde{H}_k(i,j)|^2 = \frac{\sigma_k(i,j)}{n_t}$. Note that the Kronecker propagation model (where the channel matrices are of the form $\mathbf{H}_k = \mathbf{R}_k^{1/2} \tilde{\Theta}_k \mathbf{T}_k^{1/2}$) is a special case of the UIU model. The BS is assumed to use a simple single user decoding (SUD) technique. The achievable ergodic rate of user $k \in \mathscr{K}$ is given by:

$$u_k(\mathbf{Q}_k, \mathbf{Q}_{-k}) = \mathbb{E}[i_k(\mathbf{Q}_k, \mathbf{Q}_{-k})], \qquad (2)$$

where $i_k(\mathbf{Q}_k, \mathbf{Q}_{-k})$ denotes the instantaneous mutual information

$$\begin{aligned} i_k(\mathbf{Q}_k, \mathbf{Q}_{-k}) \quad = \quad & \log_2 \left| \mathbf{I}_{n_r} + \rho \mathbf{H}_k \mathbf{Q}_k \mathbf{H}_k^H + \rho \sum_{\ell \neq k} \mathbf{H}_\ell \mathbf{Q}_\ell \mathbf{H}_\ell^H \right| - \\ & \log_2 \left| \mathbf{I}_{n_r} + \rho \sum_{\ell \neq k} \mathbf{H}_\ell \mathbf{Q}_\ell \mathbf{H}_\ell^H \right|. \end{aligned} \qquad (3)$$

In this paper, we study the power allocation game where the players are autonomous non-cooperative devices that choose their power allocation policies, $\mathbf{Q}_k$, to maximize their own transmission rates, $u_k(\mathbf{Q}_k, \mathbf{Q}_{-k})$.

2.1 Non-Cooperative Game Framework

In what follows, we briefly define some basic game theoretical concepts (see e.g. [13] for details) and standard notations that will be used throughout the paper. A normal-form game is defined as the triplet $\mathscr{G} = (\mathscr{K}, \{\mathscr{A}_k\}_{k \in \mathscr{K}}, \{u_k\}_{k \in \mathscr{K}})$ where $\mathscr{K}$ is the set of players (the K transmitters), $\mathscr{A}_k$ represents the set of actions (discrete or continuous) that player k can take (different power allocation policies), and $u_k : \mathscr{A} \rightarrow \mathbb{R}_+$ is the payoff function of user k that depends on his own choice but also the choices of the others (the ergodic achievable rate in (2)) where $\mathscr{A} = \times_{k \in \mathscr{K}} \mathscr{A}_k$ represents the overall action space. We denote by $\underline{a} \in \mathscr{A}$ a strategy profile and by a_{-k} the strategies of all the players except k.

The Nash equilibrium has been introduced in [15] and appears to be the natural solution in non-cooperative games. The mathematical definition of a pure-strategy NE is given by:

Definition 1 *A strategy profile $\underline{a}^* \in \mathscr{A}$ is a Nash equilibrium for the game $\mathscr{G} = (\mathscr{K}, \{\mathscr{A}_k\}_{k \in \mathscr{K}}, \{u_k\}_{k \in \mathscr{K}})$ if for all $k \in \mathscr{K}$ and all $a_k \in \mathscr{A}_k$: $u_k(a_k^*, a_{-k}^*) \geq u_k(a_k, a_{-k}^*)$.*

This definition translates the fact that the NE is a stable state from which no user has any incentive to deviate unilaterally. A mixed strategy for user k is a probability distribution over its own action set $\mathscr{A}_k$. Let $\Delta(\mathscr{A}_k)$ denote the set of probability distributions over the set $\mathscr{A}_k$. The mixed NE is defined similarly to pure-strategy NE by replacing the pure strategies with the mixed strategies. The existence of NE has been proven in [15] for all discrete games. If the action spaces are discrete finite sets, then $\underline{p}_k \in \Delta(\mathscr{A}_k)$ denotes the probability vector such that $p_{k,j}$ represents the probability that user k chooses a certain action $a_k^{(j)} \in \mathscr{A}_k$ and $\sum_{a_k^{(j)} \in \mathscr{A}_k} p_{k,j} = 1$.

We also define the concept of correlated equilibrium [16] which can be viewed as the NE of a game where the players receive some private signaling or playing recommendation from a common referee or mediator. The mathematical definition is as follows:

Definition 2 *A joint probability distribution $\underline{q} \in \Delta(\mathscr{A})$ is a correlated equilibrium if for all $k \in \mathscr{K}$ and all $a_k^{(j)}, a_k^{(i)} \in \mathscr{A}_k$*

$$\sum_{\underline{a} \in \mathscr{A} : a_k = a_k^{(j)}} q_{\underline{a}} \left[u_k(a_k^{(j)}, a_{-k}) - u_k(a_k^{(i)}, a_{-k}) \right] \geq 0, \qquad (4)$$

where $q_{\underline{a}}$ denotes the probability associated to the action profile $\underline{a} \in \mathscr{A}$.

At the CE, User k has no incentive in deviating from the mediator's recommandation to play $a_k^{(j)} \in \mathscr{A}_k$ knowing that all the other players follow as well the mediator's recommendation (a_{-k}). Notice that the set of mixed NE is included in the set of CE by considering independent p.d.f's. Similarly, the set of pure strategy NE is included in the set of mixed strategy NE by considering degenerate p.d.f.'s (i.e. $p_{k,j} \in \{0,1\}$) over the action sets of users.

3. NON-COOPERATIVE POWER ALLOCATION GAME

In this section, we analyse the NE of the power allocation game in fast fading MIMO MAC. First, we briefly review the case where the action sets of the users are continuous [7]. Then, we focus our attention on the practical case where the action sets of the users are discrete and finite. In this section, the players are assumed to be strictly rational transmit devices. Based on the available information, the transmitters choose the power allocation policy maximizing their own transmission rates. Furthermore, rationality is assumed to be common knowledge.

3.1 Compact and Convex Action Sets

We consider the same scenario as [7]. The transmitters are assumed to know only the statistics of the channels. The non-cooperative normal-form game is denoted by $\mathscr{G}_C = (\mathscr{K}, \{\mathscr{C}_k\}_{k \in \mathscr{K}}, \{u_k\}_{k \in \mathscr{K}})$. Each mobile station $k \in \mathscr{K}$ chooses its own input transmit covariance matrix $\mathbf{Q}_k \in \mathscr{C}_k$ to maximize its own achievable ergodic rate defined in (2). The action set of player $k \in \mathscr{K}$ is the convex cone of positive semi-definite matrices:
$\mathscr{C}_k = \left\{ \mathbf{Q}_k \in \mathbb{C}^{n_t \times n_t} | \mathbf{Q}_k \succeq 0, \mathrm{Tr}(\mathbf{Q}_k) \leq \overline{P}_k \right\}$. In [7], the authors proved the existence and uniqueness of NE using Theorems 1 and 2 in [17]. We provide here an alternative proof based on the notion of potential games [18].

Definition 3 *A normal form game* $\mathscr{G} = (\mathscr{K}, \{\mathscr{A}_k\}_{k\in\mathscr{K}}, \{u_k\}_{k\in\mathscr{K}})$ *is a* potential game *if there exists a potential function* $P : \mathscr{A} \to \mathbb{R}_+$ *such that, for all* $k \in \mathscr{K}$ *and every* $\underline{a}, \underline{b} \in \mathscr{A}$

$$u_k(a_k, \underline{a}_{-k}) - u_k(b_k, a_{-k}) = P(a_k, a_{-k}) - P(b_k, \underline{a}_{-k}). \quad (5)$$

Following [18], the local maxima of the potential function are the NE of the game. Thus, every potential game has at least one NE. For the game $\mathscr{G}_C$, the system achievable sum-rate:

$$R(\mathbf{Q}_1, \dots, \mathbf{Q}_K) = \mathbb{E}\log_2 \left| \mathbf{I} + \rho \sum_{k=1}^{K} \mathbf{H}_k \mathbf{Q}_k \mathbf{H}_k^H \right|, \quad (6)$$

is a potential function. It can be checked that $R(\mathbf{Q})$ is strictly concave w.r.t. $(\mathbf{Q}_1, \dots, \mathbf{Q}_K)$. Thus, it has a unique global maximizer which corresponds to the unique NE of the game. Furthermore, based on the finite improvement path (FIP) property [18], the iterative water-filling type algorithm in [7] converges to the unique NE. In [19], the author proves that for strict concave potential games, the CE is unique and consists in playing with one probability the unique pure NE. So the CE reduces to the unique NE of the game.

There are several drawbacks of this distributed power allocation framework: i) The action sets of users are assumed to be compact and convex sets (unrealistic in practical scenarios); ii) In order to implement the iterative water-filling algorithm, the transmitters need to know the global channel distribution information and to observe, at every iteration, the strategies chosen by the other players (very demanding in terms of information assumptions and signaling cost).

3.2 Finite Action Sets

Let us now consider the scenario where the action sets of users are discrete finite sets. The discrete game is very similar to $\mathscr{G}_C$ and is denoted by $\mathscr{G}_D = (\mathscr{K}, \{\mathscr{D}_k\}_{k\in\mathscr{K}}, \{u_k\}_{k\in\mathscr{K}})$. The action set of user k is a simple quantized version of $\mathscr{C}_k$:

$$\mathscr{D}_k = \left\{ \frac{\overline{P}_k}{\ell} \mathrm{Diag}(\underline{e}_\ell) \,\middle|\, \ell \in \{1, \dots, n_t\}, \underline{e}_\ell \in \{0,1\}^{n_t}, \sum_{i=1}^{n_t} e_\ell(i) = \ell \right\}. \quad (7)$$

$\mathscr{D}_k$ represents the set of diagonal matrices that consists in allocating uniform power over only a subset of ℓ eigenmodes. Note that the discrete game $\mathscr{G}_D$ remains a potential game with the same potential function in (6). Thus, the existence of at least one pure NE is guaranteed. However, the uniqueness property of the NE is lost in general.

We consider hereunder two particular scenarios that illustrate the extreme cases where either all strategy profiles in $\mathscr{D} = \times_k \mathscr{D}_k$ are NE or where the NE is unique.

3.2.1 Completely Correlated Antennas

Let us assume the Kronecker model where the transmit antennas and receive antennas are completely correlated, i.e., for all k, $\mathbf{R}_k = \mathbf{J}_{n_r}$ and $\mathbf{T}_k = \mathbf{J}_{n_t}$. The matrix $\mathbf{J}_n$ is a $n \times n$ matrix with all entries equal to one. In this case, the potential function is constant and independent of the users' covariance matrices:

$$R(\mathbf{Q}_1, \dots, \mathbf{Q}_K) = \mathbb{E}\log_2 \left| \mathbf{I}_{n_r} + \rho \overline{P} \sum_{k=1}^{K} \sum_{i=1}^{n_r} \sum_{j=1}^{n_t} |h_k(i,j)|^2 \mathbf{J}_{n_r} \right|. \quad (8)$$

This means that all the possible action profiles in $(\mathbf{Q}_1, \dots, \mathbf{Q}_K) \in \mathscr{D}$ are potential maximizers and thus NE of $\mathscr{G}_D$.

3.2.2 Independent Antennas

Now, we consider the other extreme case where the antennas at the terminals are completely uncorrelated, i.e., for all k, $\mathbf{R}_k = \mathbf{I}_{n_r}$ and $\mathbf{T}_k = \mathbf{I}_{n_t}$. In other words, $\mathbf{H}_k$ is a random matrix with i.i.d. complex Gaussian entries. Let us recall that in the continuous setting derived in Subsec. 3.1, if $\mathbf{H}_k$ are i.i.d. matrices, then the NE policy for all users is spread their powers uniformly over all the antennas: $\forall k, \mathbf{Q}_k^{(\mathrm{UPA})} = \frac{\overline{P}_k}{n_t} \mathbf{I}_{n_t}$. In the continuous case, the potential function is strictly concave. Thus, for that any user k the strategy $\mathbf{Q}_k^{(\mathrm{UPA})}$ strictly dominates all the other strategies in $\mathscr{C}_k$. From the fact that $\mathscr{D}_k \subset \mathscr{C}_k$, the strategy $\mathbf{Q}_k^{(\mathrm{UPA})}$ strictly dominates all the other strategies in $\mathscr{D}_k$ also. In conclusion, the NE is unique and corresponds to the same solution as in the continuous game. Note that this is a very particular case and occurs only because the NE profile in the continuous case, $(\mathbf{Q}_1^{(\mathrm{UPA})}, \dots, \mathbf{Q}_K^{(\mathrm{UPA})}) \in \mathscr{C} = \times_k \mathscr{C}_k$ happens to be also in the discrete set $\mathscr{D}$.

We see that, when quantizing the action sets of players, the uniqueness of the NE is no longer guaranteed. This raises an important issue when playing the one-shot game. There is a priori no explanation for users to expect the same equilibrium point. Because of this, their actions may not even correspond to an NE at all. A possible way to cope with this problem is to consider distributed iterative algorithms that converge to one of the NE points. Let us consider the iterative algorithm based on the best-response functions (similarly to [7]). Knowing that $\mathscr{G}_D$ is a potential game, by the FIP property, the users converge to one of the possible NE depending on the starting point. At each iteration, only one of the players updates his action by choosing its best action w.r.t. its own payoff. For exemple, at iteration t user k chooses $Q_k^{[t]} = \arg \max_{\mathbf{Q}_k \in \mathscr{D}_k} u_k \left(\mathbf{Q}_k, \mathbf{Q}_{-k}^{[t-1]} \right)$, while the other users don't do anything and $\mathbf{Q}_{-k}^{[t]} = \mathbf{Q}_{-k}^{[t-1]}$. Notice that user k is supposed to know the previous actions of the other players $\mathbf{Q}_k^{[t-1]}$. This involves a high amount of signaling between players. At the end of each iteration, the user that updated its choice needs to send it to all the other users. Furthermore, the users are assumed to be strictly rational and need to know the structure of the game and their own payoff in order to compute the best-response functions.

4. LEARNING ALGORITHMS

In this section, we discuss a different class of iterative algorithms that converge to the equilibrium points of the discrete game $\mathscr{G}_D$ described in Subsec. 3.2. As opposed to the best-response algorithm, the users are no longer rational devices but simple automata that know only their own action sets. They start at a completely naive state choosing randomly their action (following the uniform distribution over their own action sets for exemple). After the play, each users obtains a certain feedback from the nature (e.g., the realization of a random variable, the value of its own payoff). Based only on this value, each user applies a simple updating rule of its mixed strategy. It turns out that, in the long-run, the updating rules converge to some desirable system states (NE, CE). Note that the rationality assumption is no longer needed. The transmitters don't even need to know the structure of the game or even that a game is played at all. The price to pay will be reflected in slower convergence time.

4.1 A Reinforcement Learning Algorithm

We consider a stochastic learning algorithm similar to [20]. Let us index the elements of $\mathscr{D}_K = \{\mathbf{D}_k^{(1)}, \ldots, \mathbf{D}_k^{(m_k)}\}$ with $m_k = \text{Card}(\mathscr{D}_k)$ (i.e., the cardinal of $\mathscr{D}_k$). At step $t > 0$ of the iterative process, User k randomly chooses a certain action $\mathbf{Q}_k^{[t]} \in \mathscr{D}_k$ based on the probability distribution $\underline{p}_k^{[t-1]}$ from the previous iteration. As a consequence, it obtains the realization of a random variable, which is, in our case, the normalized instantaneous mutual information $i_k^{[t]} = \frac{\tilde{i}_k\left(\mathbf{Q}_k^{[t]}, \mathbf{Q}_{-k}^{[t]}\right)}{I_{\max}} \in [0,1]$. Where $\tilde{i}_k(\cdot,\cdot)$ is a finite approximation of the mutual information $i_k(\cdot,\cdot)$ such that:

$$\tilde{i}_k(\cdot,\cdot) = \begin{cases} i_k(\cdot,\cdot) \text{ , if } i_k(\cdot,\cdot) \leq I_{\max} \\ I_{\max} \text{ , otherwise} \end{cases} , \tag{9}$$

where $I_{\max}$ is chosen such that the expectation of $\tilde{i}_k(\cdot,\cdot)$ approximates the expected mutual information and thus depends on the system's parameters (n_r, n_t, ρ). Based on this value, User k updates its own probability distribution as follows:

$$p_{k,j}^{[t]} = \begin{cases} p_{k,j}^{[t-1]} - b i_k^{[t]} p_{k,j}^{[t-1]}, & \text{if} \quad \mathbf{Q}_k^{[t]} \neq \mathbf{D}_k^{(j)}, \\ p_{k,j}^{[t-1]} + b i_k^{[t]} (1 - p_{k,j}^{[t-1]}), & \text{if} \quad \mathbf{Q}_k^{[t]} = \mathbf{D}_k^{(j)}, \end{cases} \tag{10}$$

where $0 < b < 1$ is a step size and $p_{k,j}^{[t]}$ represents the probability that user k choses $\mathbf{D}_k^{(j)}$ at iteration t. Using well known results in weak convergence of random processes [20], the sequence will converge, when $b \to 0$ to the solution of a deterministic ordinary differential equation (ODE). Similarly to [21], it can be checked that the potential function in (6) is a Lyapunov function for this ODE. This means that the stationary stable points of the ODE correspond to the maxima of the potential and, thus, to the pure strategy NE of $\mathscr{G}_D$. In conclusion, when $t \to +\infty$, the updating rule (10) converge to one of the pure strategy NE. This means that the users learn their own NE strategies knowing only the realization of their mutual information and using a simple updating rule.

4.2 Learning Correlated Equilibria

In general, the performance at the NE for discrete games depends on the quantized choice of the action sets of users. In order to improve the users' performance, we study a different learning algorithm which allows them to converge towards a correlated equilibrium.

We consider the modified regret matching algorithm introduced in [22] which allows the players to converge to the set of correlated equilibria. Each user needs only the knowledge of its own payoff values received over the time.

At iteration t, User k choses randomly an action $\mathbf{Q}_k^{[t]}$ following the distribution $\underline{p}_k^{[t-1]}$ and obtains the value of its payoff $u_k^{[t]} = u_k(\mathbf{Q}_k^{[t]}, \mathbf{Q}_{-k}^{[t]})$. Without loss of generality, assume $\mathbf{Q}_k^{[t-1]} = \mathbf{D}_k^{(j)}$. The play probabilities are updated as follows:

$$\begin{cases} p_{k,i}^{[t]} &= \left(1 - \frac{\delta}{t^\gamma}\right) \min\left\{\frac{1}{\mu} M_k^{[t-1]}(j,i), \frac{1}{m_k - 1}\right\} + \frac{\delta}{t^\gamma}\frac{1}{m_k}, \quad \text{for} \quad i \neq j, \\ p_{k,j}^{[t]} &= 1 - \sum_{i \neq j} p_{k,i}^{[t]}, \end{cases} \tag{11}$$

where $0 < \delta < 1$, $0 < \gamma < 1/4$, $\mu > 0$ a sufficiently large parameter that ensures the probabilities are well defined. We observe that User k needs to know not only $u_k^{[t]}$ but also all the past values of its payoff $\left\{u_k^{[\tau]}\right\}_{\tau < t}$. The basic idea is that if at time t a player plays action $\mathbf{D}_k^{(j)}$ then the probability that at time $t+1$ the player chooses a different action $\mathbf{D}_k^{(i)}$ is proportional to the regret for not having chosen action $\mathbf{D}_k^{(i)}$ instead of $\mathbf{D}_k^{(j)}$. The regret is measured as an approximation of the increase in average payoff (if any) resulting if User k had chosen action $\mathbf{D}_k^{(i)}$ in all the past when $\mathbf{D}_k^{(j)}$ was chosen and is denoted by $M_k^{[t]}(j,i)$:

$$M_k^{[t]}(j,i) = \left[\frac{1}{t} \sum_{\tau \leq t, \mathbf{Q}_k^{[\tau]} = \mathbf{D}_k^{(i)}} \frac{p_{k,j}^{[\tau]}}{p_{k,i}^{[\tau]}} u_k^{[\tau]} - \frac{1}{t} \sum_{\tau \leq t, \mathbf{Q}_k^{[\tau]} = \mathbf{D}_k^{(j)}} u_k^{[\tau]}\right]^+ . \tag{12}$$

It turns out (see [22]) that the empirical distribution of play up to t denoted by $\underline{z}_t \in \Delta(\mathscr{D})$

$$z_t(\mathbf{Q}_1, \ldots, \mathbf{Q}_K) = \frac{1}{t}\text{Card}\{\tau \leq t : (\mathbf{Q}_1^{[\tau]}, \ldots, \mathbf{Q}_K^{[\tau]}) = (\mathbf{Q}_1, \ldots, \mathbf{Q}_K)\}, \tag{13}$$

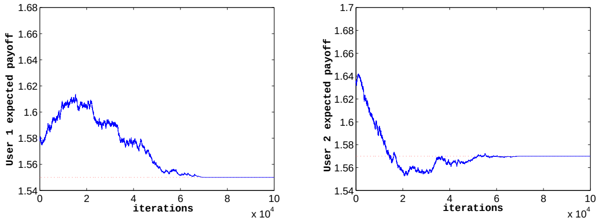
for all $(\mathbf{Q}_1, \ldots, \mathbf{Q}_K) \in \mathscr{D}$ converges almost surely as $t \to +\infty$ to the set of correlated equilibria.

There are several differences with the learning algorithm we discussed in Subsec. 4.1. Here, the learning process is no longer stochastic and the feedback each user gets at iteration t is the value of the deterministic payoff $u_k^{[t]} = u_k(\cdot,\cdot)$ instead of $i_k(\cdot,\cdot)$. The consequence is that the convergence is faster but the nature has to feedback not only the instantaneous mutual information but the ergodic achievable rate. Also, the updating rule for User k at iteration t depends on the whole history of received payoff values $\left\{u_k^{[\tau]}\right\}_{\tau \leq t}$ and not only on the current iteration $u_k^{[t]}$.

5. SIMULATION RESULTS

In what follows, we evaluate the gap between the results obtained at the equilibrium point of $\mathscr{G}_C$ in Subsec. 3.1 and $\mathscr{G}_D$ in Subsec. 3.2. We also analyze the performance of the two learning algorithms. We consider the following scenario: Two users ($K = 2$), $n_r = n_t = 2$, the Kronecker channel model where the transmit and receive correlation follow the exponential profile (i.e. $\mathbf{R}_k(i,j) = r_k^{|i-j|}$ and $\mathbf{T}_k = t_k^{|i-j|}$) characterized by the coefficients $r_1 = 0.7$, $r_2 = 0.5$, $t_1 = 0.2$, $t_2 = 0.4$, and $\sigma^2 = 1$ W.

First, we consider the discrete game in Subsec. 3.2. In Fig. 1, we plot the expected payoff depending on the probability distribution over the action sets at every iteration for User 1 in Fig. 1(a) and for User 2 in Fig. 1(b) assuming $\overline{P}_1 = \overline{P}_2 = 5$ W. We assume here that the stochastic reinforcement algorithm in Subsec. 4.1 is applied by both users in



(a) User 1.

(b) User 2.

Figure 1: Expected payoff vs. iteration number for $K = 2$ users.

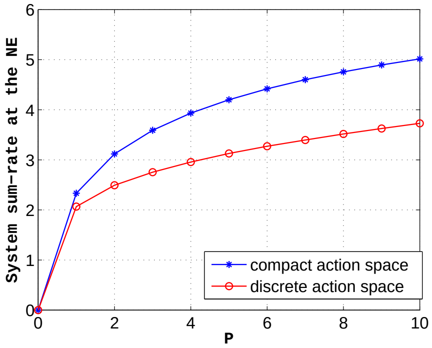


Figure 2: The achievable sum-rate at the NE. Compact action sets game vs. discrete action sets game. There is an optimality loss due to the quantization of the users' action sets.

order to learn their NE strategies. We observe that the users converge after approximatively $8 \cdot 10^4$ iterations. By using a based response algorithm the convergence is almost instantaneous (only 2 or 3 iterations). However, the rationality assumption and perfect knowledge of the game structure for each player are required.

At last, we compare the performance of the overall system in terms of achievable sum-rate of the two games discussed in Sec. 3 as function of $P \in \{0, \ldots, 10\}$ W, assuming $\overline{P}_1 = \overline{P}_2 = P$. In Fig. 2, we plot the achievable sum-rate obtained at the NE with the iterative water-filling type algorithm proposed in [7] for $\mathscr{G}_C$. Also, we plot the achievable sum-rate obtained at the NE point of $\mathscr{G}_D$ to which the users applying the learning algorithm in Subsec. 4.1 converge. We observe that there is a performance loss due to the quantization of the action sets of users. The discrete action sets $\mathscr{D}_k$ can be further refined and the results of the algorithms improved. However this will result in a higher complexity and computational costs.

6. CONCLUSIONS

We study the discrete non-cooperative power allocation game in MIMO MAC systems. In the long-run, the transmitters can learn their optimal subset of active antennas. The players are not assumed to be rational but automata which apply simple updating rules on the p.d.f.'s over their possible power allocation policies. We evaluate the performance gap between the convergence NE state of the learning procedure and the NE of the analogous game with rational players and assuming compact and convex action sets.

REFERENCES

[1] G. Arslan, M. F. Demirkol, and Y. Song, "Equilibrium efficiency improvement in MIMO interference systems: A decentralized stream control approach", *IEEE Trans. on Wireless Communications*, vol. 6, no. 8, pp. 2984–2993, Aug. 2007.

[2] G. Scutari, D. P. Palomar, and S. Barbarossa, "Optimal linear precoding strategies for wideband non-cooperative systems based on game theory-part I: Nash equilibria", *IEEE Trans. on Signal Processing*, vol. 56, no 3, pp. 1230–1249, Mar. 2008.

[3] G. Scutari, D. P. Palomar, and S. Barbarossa, "Optimal linear precoding strategies for wideband non-cooperative systems based on game theory-part II: Algorithms", *IEEE Trans. on Signal Processing*, vol. 56, no. 3, pp. 1250–1267, Mar. 2008.

[4] G. Scutari, D. P. Palomar, and S. Barbarossa, "Competitive design of multiuser MIMO systems based on game theory: A unified view", *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 7, pp. 1089–1103, Sep. 2008.

[5] E. G. Larsson, and E. A. Jorswieck, "Competition versus cooperation on the MISO interference channel", *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 7, pp. 1059–1069, Sep. 2008.

[6] G. Scutari, and D. P. Palomar, "MIMO cognitive radio: a game theoretical approach", *IEEE Transactions on Signal Processing*, vol. 58, no. 2, pp. 761–780, Feb. 2010.

[7] E. V. Belmega, S. Lasaulce, M. Debbah, M. Jungers, and J. Dumont, "Power allocation games in wireless networks of multi-antenna terminals", *Springer Telecommunications Systems Journal*, in press 2010.

[8] Y. Xing, and R. Chandramouli, "Stochastic learning solution for distributed discrete power constrol game in wireless data networks", *IEEE/ACM Trans. on Networking*, vol. 16, no. 4., pp. 932–944, Aug. 2008.

[9] F. Fu, and M. van der Schaar, "Learning to compete for resources in wireless stochastic games", *IEEE Trans. on Vehicular Technology*, vol. 58, no. 4, pp. 1904–1919, May 2009.

[10] P. Mertikopoulos, and A. L. Moustakas, "Correlated anarchy in overlapping wireless networks", *IEEE Journal on Sel. Areas in Communications*, vol. 26, no. 7, pp. 1160–1169, Sep. 2008.

[11] E. Sabir, R. El-Azouzi, V. Kavitha, Y. Hayel, and E.-H. Bouyakhf, "Stochastic learning solution for consttrained Nash equilibrium throughput in non saturated wireless collision channels", *Int. Conf. On Perf. Eval. Method. And Tools (Valuetools), Pisa, Italy*, Oct. 2009.

[12] P. Coucheney, C. Touati and B. Gaujal, "Fair and efficient user-network association algorithm for multi-technology wireless networks", *Conf. on Computer Communications (INFOCOM), Rio de Janeiro, Brazil*, Apr. 2009.

[13] D. Fudenberg and J. Tirole, "Game Theory", *The MIT Press*, 1991.

[14] D. Fudenberg, and D. K. Levine, "The theory of learning in games", *the MIT Press*, 1998.

[15] J. Nash, "Equilibrium points in n-person games",*Proc. of the Nat. Academy of Sciences*, vol. 36, pp. 48–49, 1950.

[16] R. J. Aumann, "Subjectivity and correlation in randomized strategies", *Journal of Mathematical Economocs*, vol. 1, pp. 67–96, 1974.

[17] J. Rosen, "Existence and uniqueness of equilibrium points for concave n-person games", *Econometrica*, vol. 33, pp. 520–534, 1965.

[18] D. Mondered and L. S. Shapley, "Potential Games", *Games and Economic Behavior*, vol. 14, pp. 124–143, 1996.

[19] A. Neyman, "Correlated equilibrium and potential games", *Int. Journal of Game Theory*, vol. 26, pp. 223–227, 1997.

[20] P. S. Sastry, V. V. Phansalkar and M. A. L. Thatchar, "Decentralized learning of Nash equilibria in multi-person stochastic games with incomplete information", *IEEE Trans. on Systems, Man, and Cybernetics*, vol. 24, no. 5, pp. 769–777, May 1994.

[21] W. H. Sandholm, "Potential games with continuous player sets", *Journal of Economic Theory*, vol. 97, pp. 81–108, 2001.

[22] S. Hart and A. Mas-Collel, "A reinforcement procedure leading to correlated equilibrium", *Economic Essays, Springer*, pp. 181–200, 2001.

[23] A. Tulino and S. Verdu, "Impact of antenna correlation on the capacity of multi-antenna channels", *IEEE Trans. on Inform. Theory*, vol. 51, no. 7, pp. 2491–2509, July 2005.