

# SILP: A Stochastic Imitative Learning Protocol for Multi-Carrier Spectrum Access

Stefano Iellamo, Marceau Coupechoux, and Zaheer Khan

**Abstract**—Decentralized wireless networks require efficient channel access protocols to enable wireless nodes (WNs) to access dedicated frequency channels without any coordination. In this paper, we develop a distributed spectrum access protocol for the case where the WNs are equipped with multiple radio transceivers. We consider the case where the channels are identical and duly separated so that each of the users' antenna can access only one of the available channels. To model the competition amongst WNs, we formulate a particular multi-agent multi-carrier spectrum access game, where each WN has to decide at each iteration how many antennas and which frequency channels it has to access. To study the resulting equilibrium, we solve a multi-objective optimization problem and design a bi-level learning algorithm which is proven to converge towards a socially efficient and max-min fair equilibrium state.

## I. INTRODUCTION

In decentralized wireless networks, a fundamental while challenging task is the design of distributed spectrum (channel) access mechanisms enabling wireless nodes (WN) to access the available channels in an efficient way without any coordination. In this paper, we develop and analyze a spectrum access control mechanism, where each user is able to access multiple channels at the same time (e.g., by means of multiple antennas) and aims at maximizing its marginal contribution over the accessed channels. To this end, it follows a two-dimensional revision protocol based on imitation and learning that can orient the network towards an efficient and fair equilibrium state.

When multiple channels can be accessed by a network of WNs via random access MAC protocols (such as CSMA/CA or ALOHA) it is common practice for the WNs to access all of them. Such operation can however bring congestion so that it is important to find a way to tell the WNs how many and which channels to access in order to keep the system working at the optimal operation point. Moreover, the resources should be allocated in a fair manner so that all WNs get an equal split of the system total delivered throughput.

In this paper we propose a spectrum access protocol which allows the WNs to take decisions distributedly and to drive the system towards a fair and effective equilibrium point. Our solution is of generic applicability and requires only the WNs

to be able to access and aggregate multiple channels (common feature in today's WN and user terminals). Hence, there are many possible practical use cases one could mention: for example, it could be combined with standards relying on random access MAC protocols (like e.g., Wi-Fi and Zigbee) to prevent congestion and maximize offered throughput and fairness.

The multi-resource spectrum access problem we consider can be viewed as a power set congestion game where the players' strategy can be any subset of the available resources. Such games were firstly studied in the seminal works of Rosenthal [1] and Monderer [2], where it was shown that if the costs (or payoffs) for using a given resource is the same for all players and are only dependent on the number of players including that resource in their strategy (i.e., on the congestion level), then the resulting congestion game possesses at least a Pure Nash Equilibrium (PNE). In [3] it is shown that any power set *graphical* congestion game, where each player calculates the level of congestion for using a resource by taking into account only neighboring players using that same resource (an interference graph is often used for this purpose), possess at least one PNE, and that such equilibrium can be reached by means of lazy best response sequential updates. In [4] the authors define a new class of games, power set congestion games with load-dependent failures: they redesign the players' utility function as the sum of the payoffs associated to the used resources minus the cost for using them. While the latter is solely dependent on the congestion level like for the classical case, the payoff for using a certain resource is constant but can go to zero with a certain probability if the resource fails. For this model the authors of [4] propose an algorithm that converges to a PNE and which is guaranteed to exist if the failure probability does not depend on the congestion level and, in general, if costs are non-decreasing.

Another class of games, where players have to select multiple resources are network congestion games, where the resources are the road segments to be chosen in order to reach a given destination node departing from a given source node. In this case the strategy is limited to the sets of resources forming a source-destination path, meaning that the selection of one resource within a strategy profile may be dependent on the selection of other resources. It has been proved that weighted network congestion games and (unweighted) network congestion games with player specific costs can represent every finite game, and that the intersection of these two classes can represent all exact potential games [5].

Stefano Iellamo is with Amadeus, Airlines R&D Unit, stefano.iellamo@amadeus.com

Marceau Coupechoux is with Telecom ParisTech, INFRES Laboratory, marceau.coupechoux@telecom-paristech.fr

Zaheer Khan is with the University of Oulu, Centre for Wireless Communications (CWC), zaheer.khan@oulu.fi

Fairness and efficiency of multi-resource allocation has been studied in the context of cloud computing where resource providing entities have to allocate resources (e.g., bandwidth, compute servers, sensors) to their users in a fair manner and schedule tasks efficiently. Differently to our case, the resource providers receive resource requests from the users and they (not the users) have to decide the allocation of resources and tasks in a fair and efficient manner; with global information settings and by limiting the information exchange between them. In [6] and [7] for instance, the authors build on the concepts of  $\alpha$ -proportional fairness and dominant resource fairness in order to strike a balance between efficiency and fairness of the allocated resources.

In the wake of this literature, we introduce in this paper spectrum access policies based on learning by imitation, a behavior rule widely applied in human societies consisting of imitating successful behaviors. Firstly studied in economics (see, e.g., [8], [9] and references therein), imitation rules have been recently adopted to design distributed evolutionary algorithms for wireless and cognitive radio networks [10], [11]. The problem of distributed spectrum access has been widely addressed in the literature of cognitive radio networks (CRNs). A first set of papers assumes that the number of secondary users (SUs) is smaller than the number of channels. In this case, the problem is closely related to the classical Multi-Armed Bandit (MAB) problem [12]. Some work has investigated the issue of adapting traditional MAB approaches to the CRN context, among which Anandkumar *et al.* proposed two algorithms with logarithmic regret, where the number of SUs is known or estimated by each SU [13]. Complementary, other works assume large population of SUs and study the system dynamics under asymptotic assumptions. In [14] and [15], the authors propose respectively a distributed learning procedure for spatial spectrum access which is proven to converge to a Nash Equilibrium (NE) in the asymptotic case, and a channel switch strategy aimed at maximizing expected throughput. Spatial spectrum access games can be seen as graphical congestion games and are also studied in [16] where players are allowed to select a subset (of fixed cardinality) of the available channels (their strategy is the transmission attempt probability for each selected channel). The authors show that simple algorithms based on best response updates and log-linear learning can be used for the non-cooperative and cooperative case respectively (i.e., the users aim to maximize their own throughput or the resource allocation fairness respectively).

Different from this literature, we study a model where a set of wireless nodes have to decide how many and which channels to access in order to maximize the system offered throughput and the allocation fairness. To this end, we propose a spectrum access protocol allowing to distributedly converge to a social optimum and fair equilibrium state.

With this paper, our contribution is quad-fold:

- We formulate the multi-carrier spectrum access problem as a multi-objective optimization problem, where the sum data rate as well as the system fairness have to be maximized.

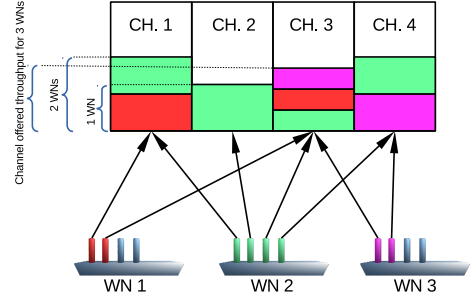


Fig. 1. System model: Wireless Nodes equipped with multiple antennas are accessing frequency channels. Active antennas equally share the bandwidth on a channel. The total throughput on a channel is a concave function of the number of active antennas on it.

- Faced to the challenge of solving the problem in a distributed way, we re-formulate the problem using a bi-level game theoretical framework.
- We propose a spectrum access protocol which we prove to converge to a stochastically stable equilibrium which is both max-min fair and socially-efficient.
- We validate our theoretical results by simulation and compare our solution with two state-of-the-art algorithms from the literature [3], [17].

The paper is organized as follows: In Section II we present the system model and formulate the multi-objective optimization problem in Section III. In Section IV we re-formulate the problem using a game theoretical framework. We describe and analyze our learning algorithms in Sections V and VI, respectively. In Section VII we validate our theoretical results by simulation. Section VIII concludes the paper. In Appendix I, we provide a mathematical background on models of evolution with noise. All proofs are given in Appendix II.

## II. SYSTEM MODEL AND ACCESS PROTOCOL

### A. Network Model

In this paper, we consider a set  $\mathcal{N}$  of  $N$  of Wireless Nodes (WN) pairs made of a transmitter and a receiver trying to access the radio spectrum (Fig. 1). Such spectrum consists of a set  $\mathcal{C}$  of  $C$  identical frequency channels, each with bandwidth  $B$ . A WN transmitter  $j \in \mathcal{N}$  is equipped with  $A \leq C$  reconfigurable antennas. Every antenna can be in any of the following modes: *active*, i.e., it is accessing one of the channels in  $\mathcal{C}$  or *sensing*, i.e., it is performing measurements on a channel. We assume that at least one antenna is always active for all WNs. We assume a *full buffer* traffic model, in which WNs have always data to transmit and a complete interference graph, in which all WNs can hear each other. In practice, this means that the considered network is of small or medium size. When accessing a channel, WNs share the medium using a random access protocol, e.g., CSMA/CA and thus get a share of the available bandwidth for their transmissions, see Figure 1.

Notations of the paper are summarized in Table I.

TABLE I  
NOTATIONS

General	
Set of Wireless Nodes (WN)	$\mathcal{N} = \{1, \dots, j, \dots, N\}$
Set of channels	$\mathcal{C} = \{1, \dots, i, \dots, C\}$
Set of antennas	$\mathcal{A} = \{1, \dots, a, \dots, A\}$
Internal Game $\mathcal{G}^{in}$	
WN $j$ strategy set	$\mathcal{S}_j$
WN $j$ strategy set at a Pareto PNE	$\mathcal{S}_j^*$
Strategy profile	$\mathbf{s} = \{\mathcal{S}_1, \dots, \mathcal{S}_j, \dots, \mathcal{S}_N\}$
Strategy profile at a Pareto PNE	$\mathbf{s}^*$
Set of strategies	$\mathcal{S}$
Number of iterations	$T$
WN $j$ marginal contribution (MC) on channel $i$	$\pi_{ji}^{in}$
WN $j$ MC on channel $i$ at a Pareto PNE	$\pi_{ji}^{in*}$
WN $j$ utility	$U_j^{in}$
System state at $t$	$\mathbf{z}^{in}(t)$
Union of PNE	$\Psi^*$
External Game $\mathcal{G}^{out}$	
Strategy profile	$\mathbf{r}$
WN $j$ strategy	$r_j$
WN $j$ throughput on channel $i$	$\pi_{ji}^{out}$
WN $j$ utility	$U_j^{out}$
System state at $\kappa$	$\mathbf{z}^{out}(\kappa)$
Limit Set (LS)	$\omega$
Union of max-throughput LSs	$\Upsilon^*$
Enlarged union of Pareto LSs	$\Omega^*$

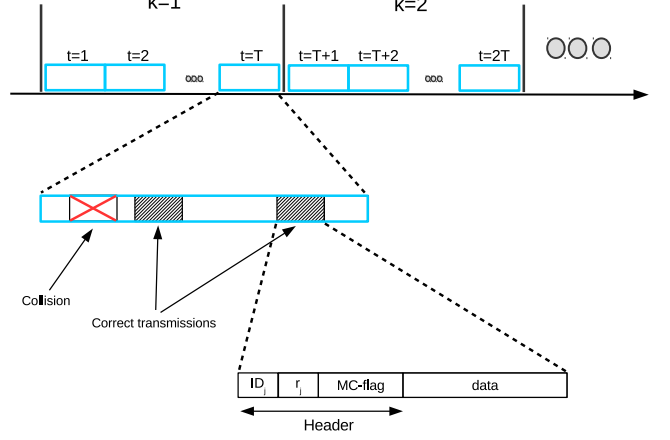


Fig. 2. Frame structure utilized by the WNs on a channel.

### B. Frame Structure

We assume that on every channel time is divided into blocks of equal duration and every block is divided in  $T$  slots (Figure 2). During a block, we impose that the mode of an antenna (active or sensing) is fixed. During a slot, the channel accessed or sensed by an antenna is fixed. Accessed or sensed channels can thus change only at slot boundaries and antenna modes can change only at block boundaries. As shown in Figure 2, transmit packets are augmented with a header containing a field for the MC-flag (1 bit), for the number  $r_j$  of active antennas ( $\lceil \log_2(A) \rceil$  bits), and for the WN ID.

### C. Throughput

At a given time instant, let  $\mathcal{S}_j \subset \mathcal{C}$  be the subset of channels accessed by WN  $j$  and  $\mathbf{s} = \{\mathcal{S}_1, \dots, \mathcal{S}_N\}$  the channel allocation profile of all WNs. The profile  $\mathbf{s}$  is an element of  $[\mathcal{C}]^{\leq A}$ , i.e., all the subsets of  $\mathcal{C}$  with cardinality less or equal to  $A$ .

The throughput achieved by WN  $j$  is the sum of the throughputs achieved by every active antenna<sup>1</sup>. On a given channel  $i \in \mathcal{C}$ , we assume that the random access protocol is fair and thus depends only on the number of contending antennas. Let  $p(n_i)$  the throughput achieved by every antenna on channel  $i$  when  $n_i$  active antennas are contending. We assume that the function  $p(n_i)$  is decreasing and that  $n_i p(n_i)$  is concave as a function of  $n_i$ . The total throughput on a channel has thus a unique maximum denoted  $n_{OPT}$ . For a WN desiring to selfishly maximize its throughput, there is an incentive to activate all its antennas. From a system point of view however, if all WNs activate all their antennas, this may result in a high congestion, so that the overall throughput is not maximized.

<sup>1</sup>If WNs are secondary users opportunistically accessing channels where a primary user has strict priority, throughput should be understood as the throughput achieved with the *remaining* resources left by the primary user.

## III. MULTI-OBJECTIVE FORMULATION

In this section, we first formulate our problem as a multi-objective maximization problem and we characterize its solution.

### A. Multi-Objective Maximization Problem

For a given channel allocation profile  $\mathbf{s}$ , the throughput of WN  $j$  is given by  $u_j(\mathbf{s}) = \sum_{i \in \mathcal{S}_j} p(n_i(\mathbf{s}))$ , where  $n_i(\mathbf{s}) = \sum_{j \in \mathcal{N}} \mathbf{1}\{i \in \mathcal{S}_j\}$ , and  $\mathbf{1}\{\cdot\}$  is the indicator function. The system throughput is given by  $U(\mathbf{s}) = \sum_{j \in \mathcal{N}} u_j(\mathbf{s})$ . We consider the max-min fairness criterion and we define the fairness of a profile  $\mathbf{s}$  as follows:  $\eta(\mathbf{s}) = \min_{j \in \mathcal{N}} u_j(\mathbf{s})$ . Our multi-objective maximization problem is now:

$$\max_{\mathbf{s} \in [\mathcal{C}]^{\leq A}} (U(\mathbf{s}), \eta(\mathbf{s})) \quad (1)$$

**Definition 1.** A feasible solution  $\mathbf{s}^1 \in [\mathcal{C}]^{\leq A}$  Pareto dominates another solution  $\mathbf{s}^2 \in [\mathcal{C}]^{\leq A}$  if  $U(\mathbf{s}^1) \geq U(\mathbf{s}^2)$ ,  $\eta(\mathbf{s}^1) \geq \eta(\mathbf{s}^2)$  and one of the inequalities is strict. A solution  $\mathbf{s}^o$  is called Pareto optimal if there does not exist another solution that dominates it.

**Remark:** As all channels are identical, all WNs have the same number of antennas and all antennas active on a given channel get the same throughput, we approximate the fairness criterion as follows  $\eta(\mathbf{s}) \approx \min_{j \in \mathcal{N}} |\mathcal{S}_j|$ . Because the problem is discrete, a WN with less active antennas may indeed have a higher throughput than another with more antennas at a Pareto optimal allocation profile.

For a given channel allocation profile  $\mathbf{s}$ , let  $y_s(k)$  be the number of channels with  $k \leq N$  active antennas on them. Note that  $\sum_{k=0}^N y_s(k) = C$  and recall that we have assumed that all WNs have at least one antenna active and that the function  $n \mapsto np(n)$  is concave with a unique maximum, denoted  $n_{OPT}$ .

**Lemma 1.** For the non-trivial case  $n_{OPT} > 0$ , a Pareto optimal allocation profile  $\mathbf{s}$  has the following properties:

- If  $AN \geq Cn_{OPT}$  and  $N \leq Cn_{OPT}$ , then  $y_s(n_{OPT}) = C$  and  $y_s(k) = 0$  for  $k \neq n_{OPT}$ . The number of active antennas per user is  $\lfloor \frac{Cn_{OPT}}{N} \rfloor$  or  $\lfloor \frac{Cn_{OPT}}{N} \rfloor + 1$  and the sum of active antennas is  $Cn_{OPT}$ . In this case, there are enough antennas to fill all channels at  $n_{OPT}$ .
- If  $N > Cn_{OPT}$ , then all WNs will activate exactly one antenna, and there will be  $C - \text{mod}(\frac{N}{C})$  channels with  $\lfloor \frac{N}{C} \rfloor$  WNs and the remainder  $\text{mod}(\frac{N}{C})$  channels with  $\lceil \frac{N}{C} \rceil$  WNs. In this case, there is at least one channel with more than  $n_{OPT}$  antennas.
- Otherwise, if  $AN < Cn_{OPT}$ , all WNs will activate all of their antennas, and there will be  $C - \text{mod}(\frac{AN}{C})$  channels with  $\lfloor \frac{AN}{C} \rfloor$  WNs and the remainder  $\text{mod}(\frac{AN}{C})$  channels with  $\lceil \frac{AN}{C} \rceil$  WNs. In this case, there is at least one channel with fewer than  $n_{OPT}$  antennas.

*Proof.* See Appendix II-A.  $\square$

As a consequence, a Pareto optimal solution can be computed by a central entity provided that  $N$ ,  $A$  and  $n_{OPT}$  are known. Moreover, the central entity would be also expected to provide each WN with the number of antennas it has to activate. In the rest of this paper, we look for a distributed approach in which the WNs don't know  $N$  and we rely on game theory to reformulate the problem. The cases where  $n_{OPT}$  is known and unknown are both studied.

#### IV. BI-LEVEL GAME THEORETICAL FORMULATION

In this section, we reformulate the problem as a spectrum access game in order to reach this solution in a distributed way.

##### A. Bi-Level Approach

The problem of maximizing and load balancing the throughput in multi-resource system in a distributed way is a difficult one. Unlike [3], which proposes distributed algorithms allowing convergence to a PNE, we aim to design a distributed algorithm allowing convergence to an equilibrium point which is Pareto optimal, i.e., efficient and fair.

Our approach is to decouple the problem into an internal game  $\mathcal{G}^{in}$  and an external game  $\mathcal{G}^{out}$ . In the internal game, the number of active antennas is fixed per WN and the strategies consist in the set of accessed channels. The internal game is thus a channel allocation game. In the external game, the strategies consist of the number of active antennas. Every slot within a block is an iteration of the internal game, whereas every block is an iteration of the external game. The utilities of the external game at the end of a block depend on the utilities obtained after the  $T$  slots of this block in the internal game.

##### B. Internal Game $\mathcal{G}^{in}$

The internal game is played at every slot  $t$  over a block duration<sup>2</sup> (see Figure 2).

**Definition 2** (Internal game  $\mathcal{G}^{in}$ ). For any vector of number of active antennas  $\mathbf{r} = [r_1, \dots, r_N]$ ,  $r_j \in \{1, \dots, A\}$ , the internal

game  $\mathcal{G}^{in}(\mathbf{r})$  is a 3-tuple  $(\mathcal{N}, \{C^{r_j}\}_{j \in \mathcal{N}}, \{U_j^{in}\}_{j \in \mathcal{N}})$ , where  $\mathcal{N}$  is the set of players,  $C^{r_j}$  and  $U_j^{in}$  are the strategy set and the utility of player  $j$ , respectively. The strategy  $\mathcal{S}_j \subset C^{r_j}$  of player  $j$  is the set of channels accessed by  $j$  and is such that  $|\mathcal{S}_j| = r_j$ . For a strategy profile  $\mathbf{s} = \{\mathcal{S}_1, \dots, \mathcal{S}_N\}$ , the utility of player  $j$  is defined as:

$$U_j^{in}(\mathbf{s}) = \sum_{i \in \mathcal{S}_j} \pi_{ij}^{in}(\mathbf{s}), \quad (2)$$

where

$$\pi_{ij}^{in}(\mathbf{s}) = n_i(\mathbf{s})p(n_i(\mathbf{s})) - (n_i(\mathbf{s}) - 1)p(n_i(\mathbf{s}) - 1) \quad (3)$$

and  $n_i(\mathbf{s}) = \sum_{j \in \mathcal{N}} \mathbf{1}\{i \in \mathcal{S}_j\}$ .

Note that  $\pi_{ij}^{in}$  is the marginal contribution of player  $j$  on channel  $i$  in terms of throughput and thus its utility is the sum of the marginal contributions on the accessed resources. The sum throughput of the system  $U(\mathbf{s})$  is a welfare function as defined in [18]. When this function is separable and utilities are marginal contributions, there exists at least one PNE that maximizes the welfare function [18]. As we will show in Section VI, such PNEs also share equally the throughput among active antennas.

**Lemma 2** ([18]). The internal game  $\mathcal{G}^{in}(\mathbf{r})$  is a distributed welfare game with potential function  $\sum_{i \in \mathcal{C}} n_i(\mathbf{s})p(n_i(\mathbf{s}))$ . There exists at least one PNE, which maximizes the sum throughput of the system and which is a global maximizer of the potential function.

##### C. External game $\mathcal{G}^{out}$

The external game  $\mathcal{G}^{out}$  is played over iteration blocks  $\kappa^3$  (see Figure 2). For this game a player  $j$  strategy at  $\kappa$  is the number  $r_j(\kappa)$  of active antennas. At the end of each iteration block  $\kappa$ , the utility of a player in  $\mathcal{G}^{out}$  is the throughput achieved at the end of the internal game with profile  $\mathbf{s}^{in}(\kappa T)$ . See Figure 3 for an overview of the bilevel approach.

**Definition 3** (External game  $\mathcal{G}^{out}$ ). The external game  $\mathcal{G}^{out}$  is a 3-tuple  $(\mathcal{N}, \{1, \dots, A\}^N, \{U_j^{out}\}_{j \in \mathcal{N}})$ , where  $\mathcal{N}$  is the set of players and  $\{1, \dots, A\}$  is the strategy set for a player. A strategy  $r_j \in \{1, \dots, A\}$  is the number of antennas activated by player  $j$ . For a strategy profile  $\mathbf{r} = [r_1, \dots, r_N]$ , the utility of player  $j$  is defined as:

$$U_j^{out}(\mathbf{r}) = u_j(\mathbf{r}, \mathbf{s}^{in}) \quad (4)$$

where  $u_j$  and  $\mathbf{s}^{in}$  are the throughput and the profile achieved by player  $j$  in the internal game at the end of the current block.

**Definition 4** (Power set congestion game [2]). A power set congestion game is a congestion game where 1) players' pure strategy is any subset of the resources, 2) the payoff associated to the use of each resource is the same for all players and it depends solely on the number of players using that resource, 3) players' utility function is the sum of the payoffs for using each resource included in the pure strategy.

<sup>2</sup>The dependence on  $t$  of the utilities and strategies are omitted when it is clear from the context.

<sup>3</sup>The dependence of utilities and strategies on  $\kappa$  is omitted when it is clear from the context

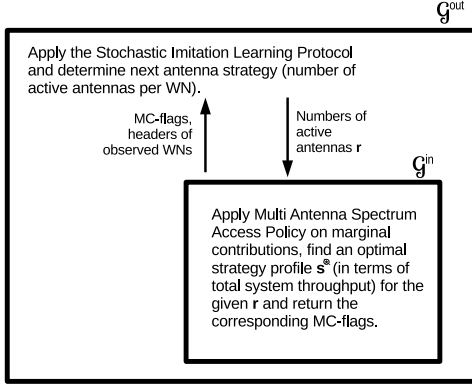


Fig. 3. Bi-level Approach with internal and external games and algorithms.

**Lemma 3** (Theorem 3.1 of [2]). *The external game  $\mathcal{G}^{out}$  can be viewed as a power set congestion game. As a consequence, it is a potential game possessing at least one PNE.*

## V. LEARNING ALGORITHM

In this section, we propose a learning algorithm to reach a PNE of the external game and achieve an efficient solution for the system, which is also fair for the players. We split the algorithm into a MAIN part and two sub-routines: MASAP (Multi-Antenna Spectrum Access Protocol) for the internal game and SILP (Stochastic Imitative Learning Protocol) for the external game.

### A. MAIN

The MAIN algorithm is played at every player. It proceeds by iterations. At each block boundary, the routine MASAP is called to play the internal game at every slot until the next block. Then the routine SILP is called to play the external game. The output of MASAP consists in a flag and a set.

The binary flag MC-flag (for marginal contribution flag) is red if the player has obtained at least one negative marginal contribution during the internal game and white otherwise. The set  $\mathcal{H}_j$  is a set of WNs headers observed by player  $j$  during the internal game.

**Algorithm 1** MAIN: Multi-carrier spectrum access (executed at each WN  $j$ )

- 
- 1: **Initialization:** Set  $T$  and define functions  $\epsilon(t)$  and  $\epsilon(\kappa)$ . Randomly choose  $r_j$ .
  - 2: **while** true (at each iteration block  $\kappa \geq 1$ ) **do**
  - 3:    $\{\text{MC-flag}(j), \mathcal{H}_j\} \leftarrow \text{MASAP}(r_j)$  (Algorithm 2)
  - 4:    $r_j \leftarrow \text{SILP}(r_j, \text{MC-flag}(j), \mathcal{H}_j)$  (Algorithm 3)
  - 5: **end while**
- 

### B. MASAP for $\mathcal{G}^{in}$

We now detail MASAP (Multi-Antenna Spectrum Access Protocol, Algorithm 2).

- The algorithm is executed at every WN  $j$ , proceeds by iterations and take decisions at the end of every slot.
- During a slot  $t$ , no channel switch is allowed and WN  $j$  accesses the chosen set of channels  $\mathcal{S}_j$ .
- At the end of each iteration  $t$ , each WN  $j$  calculates its marginal contribution  $\pi_{ij}^{in}$  on every channel  $i \in \mathcal{S}_j$  (line 8). It may deduce it from the number of WNs on the channel, see Section VII for implementation details.
- The strategy for the next iteration  $t + 1$  is chosen according to a learning rule based on exploitation and exploration (mutation). In the literature of evolutionary models with noise, a *mutation* or *tremble* is defined as a random strategy selection occurring with small probability eventually tending to zero. In our case, a mutation is a switch from a worst performing channel towards a best known channel which is not already occupied by any of  $j$ 's active antennas at the current iteration. In other words, upon mutation a WN revises its strategy by dropping one of the most congested channel and accessing one of the least known congested one.
- If the WN performed a mutation at the last iteration, it exploits its knowledge of the previous iteration and come back to channel  $i'$  if it is worth (lines 10-14).
- Otherwise, it performs exploration, i.e., a mutation, with a probability  $\epsilon(t)$  tending to zero (lines 15-18). This phase could be completely random in the choice of the target channel. We however introduce the heuristic that allows to speed up the convergence in practice (line 17).
- At the end of the block, when  $t = T - 1$  and  $t = T$  (lines 20-28), the WN collects the information about the channels on which the marginal contribution is negative. On these channels, there are too many active antennas with respect to the objective of maximizing the system throughput. This information will be communicated to the external game through the MC-flag. At  $t = T - 1$ , the WN observes its own channels (lines 20-21), while at  $t = T$ , it collects the same information from a subset of WNs chosen at random (lines 23-28) by observing the packets headers.

### C. SILP for $\mathcal{G}^{out}$

In this section, we provide detail of the Stochastic Imitation Learning Protocol (SILP) used in  $\mathcal{G}^{out}$  that achieves a fair and socially-efficient PNE. The algorithm is based on the concept of *imitation*, which is a mixed of the concepts of *mutation* and *imitation*. Like a mutation, it occurs with a decreasing probability  $\epsilon(\kappa)$  tending to zero. Like in an imitation, a decision is taken by observing the strategies of a subset of other players. In our case, imitations can be seen as mutations where the next strategy is not chosen uniformly at random but rather differs from the previous one by at most one active antenna based on the observed number of active antennas of other WNs.

---

**Algorithm 2** MULTI ANTENNA SPECTRUM ACCESS PROTOCOL (MASAP, executed at each WN  $j$ )

---

```

1: Input: Number of active antennas  $r_j$  for player  $j$ .
2: Output: MC-flag( $j$ ),  $\mathcal{H}_j$ .
3: Initialization:
4:  $\pi_{ij}^{in}(0) \leftarrow 0$ , for all  $i \in \mathcal{C}$ 
5: MC-flag( $j$ )  $\leftarrow$  'white'
6: Set  $m_j \in \mathbf{N}$ . %  $m_j \in \{1, 2, 3\}$  in simulations.
7: Randomly choose  $\mathcal{S}_j \subseteq \mathcal{C}$  with  $|\mathcal{S}_j| = r_j$ .
8:  $mutation \leftarrow 0$  % is 1 if a mutation has been performed
9: for  $t = 1$  to  $T$  do
10:   for  $i \in \mathcal{S}_j$  do
11:     Compute  $\pi_{ij}^{in}(t)$ .
12:   end for
13:   if  $mutation = 1$  %  $j$  did perform mutation at  $t - 1$ 
       switching from channel  $i'$  to  $i$  then
14:     if  $\pi_{ij}^{in}(t - 1) > \pi_{ij}^{in}(t)$  then
15:       %  $j$  migrates its antenna from  $i$  to  $i'$ 
16:        $\mathcal{S}_j \leftarrow \mathcal{S}_j \cup \{i'\} \setminus \{i\}$ 
17:     end if
18:      $mutation \leftarrow 0$ 
19:   else
20:     %  $j$  migrates a worst performing antenna towards
       a best known channel.
21:     Let  $i_1 \in \operatorname{argmax}_{i \in \mathcal{S}_j} n_i(t)$  and  $i_2 \in$ 
        $\operatorname{argmin}_{i \notin \mathcal{S}_j} n_i(t)$ 
22:     With probability  $\epsilon(t)$  do
23:        $\mathcal{S}_j \leftarrow \mathcal{S}_j \cup \{i_2\} \setminus \{i_1\}$ 
24:     end if
25:   end if
26:   if  $t = T - 1$  &  $\exists i \in \mathcal{S}_j$ , s.t.  $\pi_{ij}(t) < 0$  then
27:     MC-flag( $j$ )  $\leftarrow$  'red'
28:   end if
29:   if  $t = T$  then
30:     Sense and observe the header of a set  $\mathcal{I}_j$  of  $m_j$ 
       WNs chosen at random.
31:     Store the headers in the set  $\mathcal{H}_j$  (contains MC-flag
       and number of active antennas).
32:     if  $\exists j' \in \mathcal{I}_j$  s.t. MC-flag( $j'$ ) = 'red' then
33:       MC-flag( $j$ )  $\leftarrow$  'red'
34:     end if
35:   end if
36: end for

```

---

SILP algorithm is shown in Algorithm 3 and can be summarized as follows:

- The algorithm is executed at every WN  $j$  and proceeds by iterations at the end of every block.
- During a block the number of active antennas is fixed to  $r_j(\kappa)$ .
- There is an imutation when the number of active antennas has changed in the last iteration (lines 2-3).
- If the WN has not performed imutation (line 5), it takes a decision depending on the MC-flag and the numbers of antennas of other players. If the MC-flag is 'white', this means that the system throughput is likely to increase by activating an antenna. This will however be done only if fairness is improved (lines 7-9). If the MC-flag is 'red'

---

**Algorithm 3** STOCHASTIC IMITATION LEARNING PROTOCOL (SILP, EXECUTED AT EACH WN  $j$ )

---

```

1: Input: MC-flag( $j$ ),  $\mathcal{H}_j$  (from MASAP),  $r_j$ 
2: if  $r_j(\kappa - 1) \neq r_j(\kappa)$  then  $imutation \leftarrow 1$ 
3: else  $imutation \leftarrow 0$ 
4: end if
5: if  $imutation = 0$  then
6:   With probability  $\epsilon(\kappa)$  do
7:     if  $r_j(\kappa) \leq \min_{r_{j'} \in \mathcal{H}_j(\kappa)} \{r_{j'}(\kappa)\}$  && MC-
       flag( $j$ ) = 'white' &&  $r_j(\kappa) < C$  then
8:       % If system throughput can be increased and  $j$  has
       less active antennas than observed WNs
9:        $r_j(\kappa + 1) \leftarrow r_j(\kappa) + 1$ 
10:    else
11:      if  $r_j(\kappa) \geq \max_{r_{j'} \in \mathcal{H}_j(\kappa)} \{r_{j'}(\kappa)\}$  && MC-
        flag( $j$ ) = 'red' &&  $r_j(\kappa) > 1$  then
12:        % If system throughput cannot be increased
        and  $j$  has more active antennas than others
13:         $r_j(\kappa + 1) \leftarrow r_j(\kappa) - 1$ 
14:      end if
15:    end if
16:  else
17:    if  $r_j(\kappa) > \max_{r_{j'} \in \mathcal{H}_j(\kappa - 1)} \{r_{j'}(\kappa - 1)\}$  && MC-
      flag( $j$ ) = 'red' &&  $r_j(\kappa) > r_j(\kappa - 1)$  then
18:      % By increasing the number of active antennas,
      system throughput has decreased and  $j$  has more antennas
      than others had
19:       $r_j(\kappa + 1) \leftarrow r_j(\kappa) - 1$ 
20:    end if
21:  end if

```

---

on the contrary, the number of active antennas should be decreased (lines 11-13).

- If the WN has performed imutation at the last iteration by increasing its number of active antennas and the MC-flag is 'red' (line 17), this means that WN  $j$  is potentially responsible for decreasing the system throughput. In this case, it should decrease its number of active channels (line 19).

## VI. CONVERGENCE ANALYSIS

### A. $\mathcal{G}^{in}$ : MASAP Equilibria and Convergence

We model the dynamics generated by MASAP as a model of evolution with noise (or mistakes model). A background on the latter is provided in Appendix I. Let us define the state of the system as follows:

**Definition 5** (System State). *We define the state  $z^{in}(t)$  of the system at iteration  $t$  as:*

$$z^{in}(t) \triangleq \{\mathcal{S}_j(t), \mathcal{S}_j(t - 1)\}_{j \in \mathcal{N}},$$

or equivalently:

$$z^{in}(t) \triangleq (\mathbf{s}(t), \mathbf{s}(t - 1)).$$

**Definition 6** (Single-Player Single-Antenna Improvement). *A strategy profile  $\mathbf{s}' = (\mathcal{S}'_1, \dots, \mathcal{S}'_j, \dots, \mathcal{S}'_N)$  is a single-player single-antenna improvement over the strategy profile  $\mathbf{s} =$*

$(S_1, \dots, S_j, \dots, S_N)$  if it coincides with  $s$  in every coordinate except one, say coordinate  $S_j$ , the strategy set  $S'_j$  in turn coincides with  $S_j$  in all elements (i.e., channels) except one, and the utility of player  $j$  is higher under  $s'$  than under  $s$ .

**Proposition 1.** In  $\mathcal{G}^{in}$  and under MASAP all limit sets (LS) are made of a single state, which is an absorbing state. All absorbing states are limit sets.

*Proof.* See Appendix II-B. The definition of LS is given in Definition 13 in Appendix I-A.  $\square$

**Proposition 2.** In  $\mathcal{G}^{in}$  and under MASAP there always exists an improvement path, made of single-player single-antenna improvements, which is finite and ends in a PNE.

*Proof.* See Appendix II-C.  $\square$

Let denote  $Z^*$  the union of all the states corresponding to the PNEs. As PNEs are absorbing states, every state in  $Z^*$  is a limit set. Let  $R(\Omega)$  be the radius of a set  $\Omega$  as defined in Definition 18 in Appendix I. Let  $CR^*(\Omega)$  be the modified coradius of  $\Omega$  as defined in Definition 21 in Appendix I-B. The following Lemma follows directly from Proposition 2.

**Lemma 4.** For all  $z^{in} \notin Z^*$  that are LS, it holds that  $R(z^{in}) = 1$ . Moreover,  $Z^*$  can be reached from any LS  $z^{in}$  by stepwise mutations.

**Lemma 5.** It holds that  $CR^*(Z^*) = 1$

*Proof.* It follows from Lemma 4 and the definition of modified cost. From any state, we reach a LS at zero cost and then, there is a path of LSs towards  $Z^*$ , each with a radius of 1.  $\square$

**Lemma 6.** It holds that  $R(Z^*) > 1$ .

*Proof.* We show that a single mutation is not sufficient to leave the basin of attraction of  $Z^*$ . As in a LS at a PNE no user has incentive to deviate unilaterally with one of its antennas, any single-player single-antenna mutation leads to a decrease of payoff for the concerned player, say  $j$ . Thus, at the next iteration  $t+1$  of MASAP, WN  $j$  will switch its antenna, which performed a mutation at  $t-1$ , back to the previous channel accessed at  $t-1$ . Hence, one mutation is not enough to quit the basin of attraction of  $Z^*$ .  $\square$

We can now state the main result of this section:

**Theorem 1.** For  $\mathcal{G}^{in}$  and under MASAP:

- 1) The system dynamics converge a.s. to the PNE union  $Z^*$  as the temperature  $\epsilon(t)$  vanishes. The maximum wait until such union of limit sets is reached from any state outside its domain of attraction is  $O(\epsilon^{-1})$  as  $\epsilon \rightarrow 0$ .
- 2) At all PNEs the sum data rate of the system is maximized. Moreover, the throughput allocation to every active antenna is max-min fair; i.e., MASAP load balances the system throughput amongst active antennas.
- 3) The computational complexity at the WNs is  $O(TC + N)$  at every call of the algorithm.

*Proof.* See Appendix II-D.  $\square$

## B. $\mathcal{G}^{out}$ : SILP Convergence

Similarly to what has been done for the analysis of  $\mathcal{G}^{in}$ , we firstly provide a few definitions and propositions which will pave the way towards the main theorem. Let us define the state of the system as follows:

**Definition 7** (System State). We define the state  $z^{out}(\kappa)$  of the system at iteration block  $\kappa$  as:

$$z^{out}(\kappa) \triangleq \{r_j(\kappa), \{r_{j'}(\kappa), r_{j'}(\kappa-1)\}_{j' \in \mathcal{I}_j(\kappa)}\}_{j \in \mathcal{N}}$$

In other words, the system state includes for every WN  $j$  its number of active antennas as well as the number of active antennas of all observed WNs in iterations  $\kappa$  and  $\kappa-1$ . We say that we are in a *complete information* setting if the set  $\mathcal{I}_j$  contains all WNs except  $j$ , i.e.,  $\mathcal{I}_j(\kappa) = \mathcal{N} \setminus \{j\}$  for all  $j$  and all  $\kappa$ . Otherwise we are in an *incomplete information* setting. In this case, we still assume that if a WN has a red MC-flag, this information is known by every WN<sup>4</sup>.

In SILP, the unperturbed dynamics is such that the transition probability from any state to any other state is zero, i.e., all states are absorbing states. Although this is a very specific case, the conditions of Lemma 9 showing the existence of a limit distribution as  $\epsilon \rightarrow 0$  are verified. As a consequence all states are limit sets.

**Definition 8** (Union of max-throughput Limit Sets  $\Upsilon$ ). We define  $\Upsilon$  as the union of limit sets including the LSs which are characterized by maximum system offered throughput.

**Definition 9** (Enlarged Union of max-throughput Limit Sets  $\Upsilon^*$ ). We define  $\Upsilon^*$  as the union of limit sets including the LSs in  $\Upsilon$  plus those which can be reached from  $\Upsilon$  by a single player imutation.

**Definition 10** (Enlarged Union of Pareto Limit Sets  $\Omega^*$ ). We define  $\Omega^*$  as the union of limit sets including:

- the LSs from  $\Upsilon$  which are max-min fair.
- the LSs that can be reached from the LSs above by a single-player imutation

**Proposition 3.** Regardless of the level of information (complete or incomplete), there always exists a path, i.e., a finite sequence of stepwise imutations, that leads to  $\Omega^*$  from any LS. Also, there always exists a path that leads to  $\Upsilon^*$  from any LS.

*Proof.* See Appendix II-E.  $\square$

**Proposition 4.** For any LS  $\omega$ , its radius is  $R(\omega) = 1$ .

*Proof.* Consider an imutation from  $\omega$  to a state  $\omega'$ . As all states are absorbing for the unperturbed dynamics, there is no possible transition from  $\omega'$  to  $\omega$ .  $\square$

**Lemma 7.** The modified coradii of  $\Omega^*$  and  $\Upsilon^*$  are equal to one. That is,  $CR^*(\Omega^*) = CR^*(\Upsilon^*) = 1$ .

*Proof.* It follows from Proposition 3, Proposition 4 and the definition of modified coradius (see Definition 21 in Appendix I-B).  $\square$

<sup>4</sup>In practice, as we will see in simulations, this happens with a high probability since the information is propagated by every antenna of all WNs.



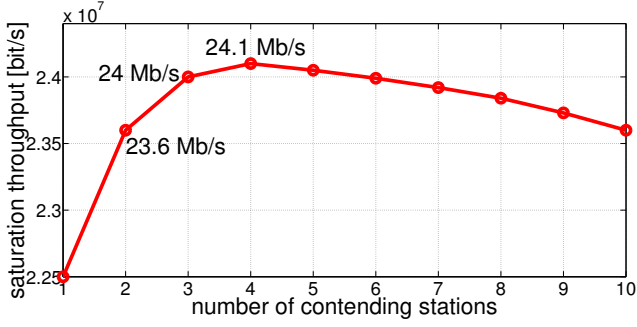


Fig. 4. Saturation throughput for CSMA/CA with RTS/CTS as a function of the number of the contending stations operating on the same channel [20]. All channels deliver the largest amount of throughput when there are 4 contending WNs on them; the throughput starts to decrease afterwards.

**Lemma 8.** *In the incomplete information case  $R(\Omega^*) = 1$  and  $R(\Upsilon^*) > 1$ . In the complete information case  $R(\Omega^*) > 1$ .*

*Proof.* See Appendix II-E.  $\square$

**Theorem 2.** *SILP with complete and incomplete information converges a.s. to  $\Omega^*$  and  $\Upsilon^*$  respectively as the noise term  $\epsilon(\kappa)$  vanishes. The maximum wait until such unions of limit sets are reached from any state outside their domain of attraction is  $O(\epsilon^{-1})$  as  $\epsilon \rightarrow 0$ . The computational complexity of SILP at every WN is  $O(N)$  at every call.*

*Proof.* As  $R(\Omega^*) > CR^*(\Omega^*)$  and  $R(\Upsilon^*) > CR^*(\Upsilon^*)$  in the complete and incomplete information case respectively (Lemma 8 and Lemma 7) the result follows from [19]. The algorithm always executes steps 1-4, and then executes one of the following: steps 6-9, steps 10-14, steps 16-20. Steps 1-4 involves a comparison and an assignment, its complexity is thus  $O(1)$ . The most demanding operation in each of the other possible steps is finding a minimum or a maximum in a set of cardinality at most  $N$ . As a consequence, the overall complexity is in  $O(N)$  at every call.  $\square$

## VII. NUMERICAL ANALYSIS

### A. Payoff estimation

It has been so far implicitly assumed that marginal contributions are observable and deterministic at each iterations and only depend on the number of contending WNs. However, in real implementations the WNs derive their current payoff relying on environmental factors which may bring uncertainty and errors.

We here discuss two possible methods for payoff (i.e., marginal contribution) estimation which can be used by the WNs in combination with random access MAC protocols.

**Lookup table** The WNs are in possession of a lookup table matching the number of contending stations with the corresponding saturation throughput. Thus, with this method it is enough for the WNs to count the number of contending WNs on each accessed channel (the header

ID field is public - see Fig. 2) to derive the corresponding marginal contribution. Uncertainty and errors can be brought in this case by the estimation of the number of contending stations. In a small or medium size network and a reasonable throughput accuracy, only few tens of bits are required to store a look-up table. As a consequence, it can be easily pre-launched by the device manufacturer or downloaded before operation.

**Collision count** If the WNs do not know neither  $N$  nor  $n_{OPT}$ , they can estimate the payoff (i.e., the marginal contribution)  $\pi_{ji}^{in}$  of each of their active antenna at iteration  $t$  with the following approximation:

$$\pi_{ji}^{in} \simeq \tilde{\pi}_{ji}^{in} = \frac{\xi_{ji} - c_{ji}}{n_t}$$

where  $n_t$  is the number of time slots in one iteration  $t$ ,  $\xi_{ji}$  and  $c_{ji}$  are the total number of  $j$ 's correct transmissions and collisions on channel  $i$  respectively. Note that the value obtained by means of the formula above is a lower bound of the real value (i.e.,  $\pi_{ji}^{in} \leq \tilde{\pi}_{ji}^{in}$ ), as collisions with more than 2 antennas involved should not be subtracted (intuitively, collision would have occurred anyway if one of the 3 or more antennas involved had stepped back from transmitting). However, the presented formula suits a fully distributed scenario (can be obtained through own transmission statistics and sensing operations while not transmitting); Moreover, the approximation is accurate if the probability of multiple collisions is sufficiently low. This holds true for loads not exceeding too much the random access protocol saturation point.

Note that the system designer can cope with uncertainty and errors (e.g., error-prone channels) by extending the duration of the blocks  $\kappa$  at the cost of a slower convergence. Increasing  $\kappa$  indeed implies that more packets are sent per block, so that a better estimation of the correct payoff can be achieved in both look-up table and collision count implementation approaches.

### B. Simulation settings

We consider a wireless network made of 8 channels and 10 identical WNs, each of which equipped with 8 antennas. The WNs apply CSMA/CA as random access protocol on each accessed channel and the parameters are taken according to [20] (we equip the WNs with a lookup table based on Fig. 4). Each WN is characterized by a generic strictly decreasing noise function  $\epsilon(\cdot)$  which tends to 0 in MASAP and to a very small constant value in SILP.

We compare our algorithm to the lazy best response (lazy BR) approach proposed in [3] and to the payoff-based version of the binary log-linear learning (BLLL) algorithm in [17], [21]. We compare also the performance to the Pareto NE, obtained using Lemma 1 and by performing an exhaustive search over the returned solutions.

With lazy BR, the WNs revise their strategy sequentially: at each iteration block a pre-determined WN updates its strategy via a best response update that minimizes the number of new channels occupied minus the number of old channels released (old and new are with reference to the previous iteration



block). With this technique, we obtain an asynchronous full information game where 1) a scheduler is necessary for the WNs to revise their strategies and apply the lazy best response sequentially and 2) no equilibrium selection is performed in terms of fairness.

With BLLL, the WN revise their strategy simultaneously at the end of each iteration. Every WN calculates its own utility, which in our case is the sum of the marginal contributions obtained from the accessed channels. A WN strategy consists of a pair  $(a_j, \mathbf{v}_j)$ , where  $\mathbf{v}_j$  is a list of channels to access, of length  $a_j$ . The strategy is chosen according to the payoff-based binary log-linear rule with probability  $1 - \epsilon(t)$  and uniformly randomly with probability  $\epsilon(t)$  eventually tending to zero or to a very small value. For each  $j$ , the payoff based binary log-linear rule consists of observing the utility obtained at the current and previous iteration and going back to the previous strategy with a probability proportional to the utility difference, see [17] for the details of the algorithm.

For comparison, we assume that SILP iteration blocks, lazy BR iterations and BLLL iterations have equal duration. Unless specified, the SILP curves assume a complete information setting.

### C. Simulation results

Fig. 5 displays a realization of the MASAP algorithm in terms of aggregate system throughput as a function of the iteration  $t$ . As one can see, the PNE is reached fairly quickly (by the 60th iteration).

In Fig. 6 and Fig. 7 the SILP algorithm is compared to lazy BR and BLLL in terms of total system throughput. In the SILP case, we see that as soon as the noise vanishes a stochastically stable limit set contained in the Enlarged Union of Pareto LSs is attained. With best response update, convergence is much quicker and a PNE is attained after just 6 iteration blocks. Differently, BLLL does not converge completely by the set horizon. This is due to the fact that the WNs strategy set is very large and the probability that a unilateral single player strategy revision is improving (in other words, the probability that a unilateral strategy deviation causes a system throughput increase) is very low. Notice that the SILP curve depicted in Fig. 7 displays a slightly lower average throughput at the equilibrium wrt lazy BR. This can be explained by recalling that the Enlarged Union of Pareto limit sets (Definition 10) contains not only the Pareto-efficient LS, but also some LSs which are in the immediate proximity (they can be reached if, starting from a Pareto LS, a WN activates an additional antenna).

Since a Pareto allocation is achieved when the total system throughput is maximum and equally distributed among WN, we choose the Jain's fairness index [22] to study the fairness of our algorithm. The Jain's fairness index  $\mathcal{J}(\cdot)$  ranges between 0 and 1 and is maximum if the available resources are perfectly aligned. The index is calculated at each iteration  $\kappa$  as  $\mathcal{J}(\kappa, U_1^{out}, \dots, U_N^{out}) = \frac{(\sum_{j \in \mathcal{N}} U_j^{out})^2}{N \cdot \sum_{j \in \mathcal{N}} (U_j^{out})^2}$ . Fig. 8 and Fig. 9 show that the equilibrium selection is very effective. In fact, SILP converges to a very high fairness index within 10 iteration

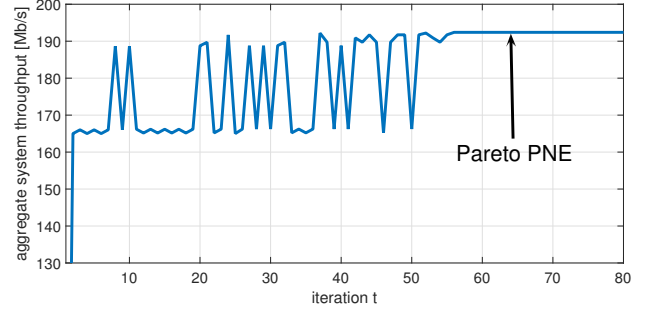


Fig. 5. MASAP: convergence trend in terms of the total system throughput (one realization). Input antenna vector  $\mathbf{r} = [5 \ 2 \ 6 \ 6 \ 3 \ 3 \ 2 \ 4 \ 3 \ 6]$ .

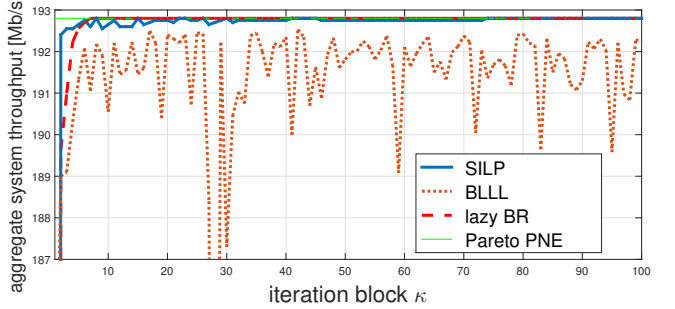


Fig. 6. Aggregate system throughput vs. number of iterations for SILP, BLLL, lazy BR and at Pareto NE for one realization.

blocks, while lazy BR and BLLL fairness has a decreasing trend from the initial state.

Fig. 10 shows the fairness degradation in the incomplete information case, i.e., when the WNs have the ability to imitate only a limited number of other WNs. In the incomplete information setting, we assume that every WN is able to observe 1, 3 or 5 other WNs (chosen uniformly at random at each iteration block in our simulation). We notice here that there is indeed a degradation in terms of Jain's fairness index, but this degradation is quite limited, especially if compared with the fairness achieved by lazy best response (Fig. 9). Moreover, the convergence time is not affected. Note that we don't report the system throughput trend under incomplete information regime. Indeed, even by imitating only one WN it is possible to achieve a very good information on all the channels, so that the curves result practically superimposed. This occurs because a WN flag is representative for all its accessed channels.

Fig. 11 depicts the number of active antennas per user as a function of the iteration block  $\kappa$ . We see that at the Pareto equilibrium the total number of active antennas oscillated between 32 and 33 when  $\epsilon(t)$  becomes very small. This is in line with our definition of Enlarged Union of Pareto limit sets (Definition 10).

## VIII. CONCLUSION

We have considered the problem faced by a set of wireless nodes having to decide how many and which channels to access in order to maximize the system offered throughput and the allocation fairness. The sets of wireless nodes and frequency channels are assumed homogeneous and we have

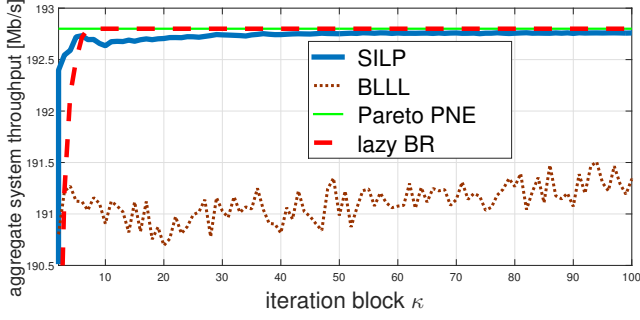


Fig. 7. Aggregate system throughput vs. number of iterations for SILP, BLLL, lazy BR and at Pareto NE. Each point is an average over 100 realizations.

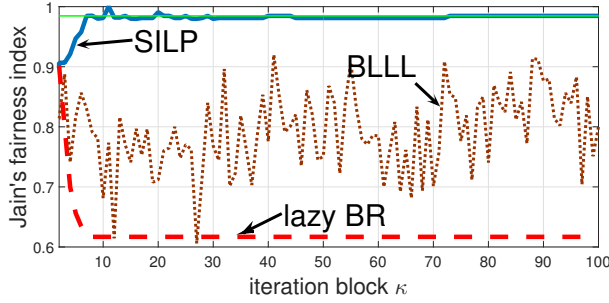


Fig. 8. Jain's fairness index vs. number of iterations for SILP, BLLL, lazy BR and at Pareto NE for one realization.

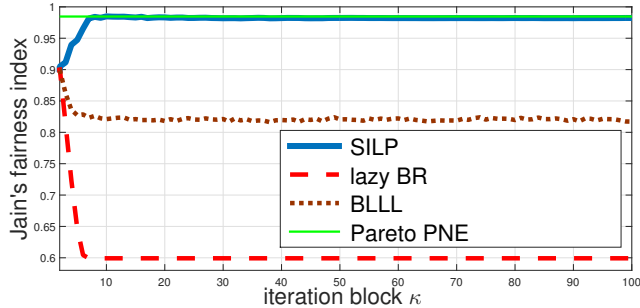


Fig. 9. Jain's fairness index vs. number of iterations for SILP, BLLL, lazy BR and at Pareto NE. Each point is an average over 100 realizations.

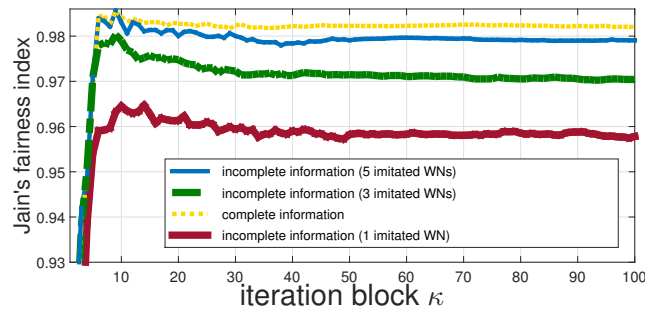


Fig. 10. Jain's fairness index vs. number of iterations for SILP in the complete and incomplete information cases. Each point is an average over 100 realizations.

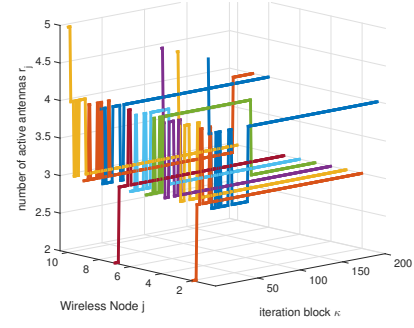


Fig. 11. Number of active antennas for every WN vs. number of iterations with SILP.

formulated the multi-carrier spectrum access problem as a multi-objective optimization problem, and have derived the centralized solution. We have then reformulated the problem as a distributed spectrum access game and have designed a bi-level learning algorithm which we have proven to converge towards an equilibrium state where both fairness and system throughput are maximized. A possible future work is the implementation of the proposed approach in a test-bed to confront theory and practice.

## APPENDIX I BACKGROUND

### A. Model of Evolution with noise

**Definition 11** (Model of evolution with noise [19]). A model of evolution with noise or mistakes model is a triple  $(Z, P, P(\epsilon))$  where:

- 1)  $Z$  is the state space of a stochastic process  $X$  and is supposed to be finite;
- 2)  $P = (p_{zz'})_{(z,z') \in Z^2}$  is a Markov transition matrix defined on  $Z$ ;
- 3)  $P(\epsilon) = (p_{zz'}(\epsilon))_{(z,z') \in Z^2}$  is a family of Markov transition matrices on  $Z$  indexed by  $\epsilon \in [0, \bar{\epsilon})$  s.t.:
  - a)  $P(\epsilon)$  is ergodic for  $\epsilon > 0$ ;
  - b)  $P(\epsilon)$  is continuous in  $\epsilon$  and  $P(0) = P$ ;
  - c) there is a cost function  $c : Z^2 \rightarrow \mathcal{R}^+ \cup \{\infty\}$  s.t. for any pair of states  $(z, z')$ ,  $\lim_{\epsilon \rightarrow 0} \frac{p_{zz'}(\epsilon)}{\epsilon^{c_{zz'}}}$  exists and is strictly positive for  $c_{zz'} < \infty$  and  $p_{zz'}(\epsilon) = 0$  for small  $\epsilon$  if  $c_{zz'} = \infty$ .

**Definition 12** (Unperturbed and perturbed Markov chain). In a model of evolution with noise  $(Z, P, P(\epsilon))$ ,  $(Z, P)$  is called the unperturbed Markov chain and, for any  $\epsilon$ ,  $(Z, P(\epsilon))$  is a perturbed Markov chain. The family of perturbed Markov chains indexed by  $\epsilon$  is called a regular perturbation.

**Remark.** The fact that  $P(\epsilon)$  is ergodic ensures that from any state  $z \in Z$ , we can reach any state  $z' \in Z$  in a finite number of steps with positive probability. The unperturbed Markov chain is however not necessarily ergodic. If not, the Markov chain  $(Z, P)$  has one or more limit sets.

Note that the particular design of the perturbation does not compromise the existence of the limit distribution, since the specification of the perturbation provided in [19] is sufficiently flexible so as to allow any stationary specification of noise in

which one may be interested to be represented, provided that the conditions in Definition 11 still hold. A separate study on state-dependent mutation rates is also presented in [23].

**Definition 13** (Limit set). *A limit set (or recurrent class)  $L$  of a Markov chain  $X = (Z, P)$  is a set of states of  $X$  such that  $\forall z \in L, P[X_{t+1} \in L | X_t = z] = 1$  and  $\forall z, z' \in L$ , there exists  $\tau > 0$  s.t.  $P[X_{t+\tau} = z' | X_t = z] > 0$ .*

The unperturbed Markov chain can be interpreted as the evolution of the system when players follow a predefined rule of evolution like *Best Response*. Noise  $\epsilon$  can be interpreted as a probability that players do not follow the rule of the dynamics. For example, if the rule is *Best Response*, players choose the best response strategy at the next iteration step with probability  $1 - \epsilon$  and choose any other strategy at random with probability  $\epsilon$ . When a player does not follow the predefined rule, we say that there is a *mutation* by analogy with what happens in species evolution.

**Definition 14** (State transition cost). *The cost or resistance  $c_{zz'}$  of the transition  $z \rightarrow z'$  is the rate at which the transition probability  $p_{zz'}(\epsilon)$  tends to zero as  $\epsilon$  vanishes:*

$$c_{zz'} = \begin{cases} 0 & \text{if } P_{zz'}(0) > 0 \\ k & \text{if } P_{zz'}(\epsilon) = (a_{zz'} + o(1))\epsilon^k \\ \infty & \text{if } P_{zz'}(\epsilon) = 0 \quad \forall \epsilon \in [0, \bar{\epsilon}] \end{cases}$$

for some  $\bar{\epsilon} > 0$  and constants  $a_{zz'}$ .

Let  $\mu(\epsilon)$  be the stationary probability distribution of the perturbed Markov chain  $(Z, P(\epsilon))$ .

**Lemma 9** (Existence of limit distribution [24]). *There exists a limit distribution  $\mu^* = \lim_{\epsilon \rightarrow 0} \mu(\epsilon)$ .*

Thus,  $\mu^*$  is a stationary probability of the unperturbed Markov chain  $(Z, P)$ :

**Lemma 10** ([19]). *The set of stochastically stable states is included in the limit sets of the unperturbed Markov chain  $(Z, P)$ .*

**Definition 15** (Long-run stochastically stable set). *A state  $z \in Z$  is said to be long-run stochastically stable if and only if  $\mu_z^* > 0$ .*

Let  $\Omega$  be a union of one or more limit sets of  $(Z, P)$ . We now want to study the conditions for  $\Omega$  to be stochastically stable. We also want to know the speed at which  $\Omega$  is reached. For this purpose, [19] defines  $W(x, \Omega, \epsilon)$  to be the expected wait until set  $\Omega$  is reached knowing that we start in state  $x$  and that the system follows the perturbed Markov chain  $(Z, P(\epsilon))$ . The goal is to characterize  $\max_{x \in Z} W(x, \Omega, \epsilon)$ .

## B. Radius and Coradius Theorem

We start with some definitions of concepts illustrated in Fig. 12 before giving the main theorem. Define a path  $(z_1, z_2, \dots, z_\tau)$  as a sequence of states.

**Definition 16** (Basin of attraction). *Let  $\Omega$  be a union of one or more limit sets of  $(Z, P)$  and let  $(z_1, z_2, \dots, z_\tau)$  be a sequence of states. The basin of attraction  $D(\Omega)$  of  $\Omega$  is the set of initial*

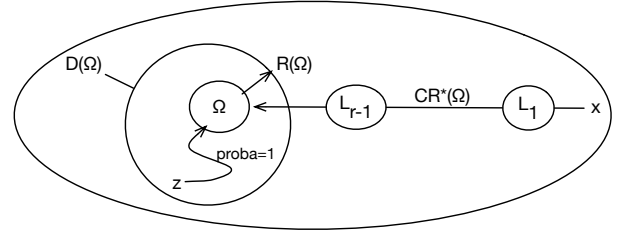


Fig. 12. Illustration of the main concepts: basin of attraction  $D(\Omega)$ , radius  $R(\Omega)$ , modified coradius  $CR^*(\Omega)$ ,  $L_1, \dots, L_{r-1}$  are limit sets,  $x$  is a state that maximizes the modified cost to  $\Omega$ ,  $z$  is a state in  $D(\Omega)$ .

states from which the unperturbed Markov chain converges to  $\Omega$  with probability 1, i.e.:

$$D(\Omega) = \{z \in Z | \Pr[\exists \beta \text{ s.t. } \forall \tau > \beta, z_\tau \in \Omega | z_0 = z] = 1\}$$

**Definition 17** (Path cost). *For two sets  $X$  and  $Y$ , a path in  $Z$  is a sequence of states  $(z_1, z_2, \dots, z_\tau)$  with  $z_1, z_2, \dots \in X$  and  $z_\tau \in Y$ . The cost of the path is the sum*

$$c(z_1, z_2, \dots, z_\tau) = \sum_{i=1}^{\tau-1} c_{z_i, z_{i+1}}.$$

Let  $S(X, Y)$  be the set of all paths from  $X$  to  $Y$  and

$$C(X, Y) = \min_{(z_1, \dots, z_\tau) \in S(X, Y)} c(z_1, \dots, z_\tau)$$

be the set-to-set cost between  $X$  and  $Y$ . The radius of the basin of attraction of  $\Omega$  is defined as the minimum number of mutations needed to leave  $D(\Omega)$  given that we start in  $\Omega$ .

**Definition 18** (Radius). *The radius  $R(\Omega)$  of  $\Omega$  is the minimum cost of any path from  $\Omega$  out of  $D(\Omega)$ , i.e.:*

$$R(\Omega) = C(\Omega, Z - D(\Omega)).$$

**Definition 19** (Coradius). *The coradius  $CR(\Omega)$  of  $\Omega$  is defined by:*

$$CR(\Omega) = \max_{x \notin \Omega} \min_{(z_1, \dots, z_\tau) \in S(x, \Omega)} c(z_1, \dots, z_\tau).$$

In other words, the coradius is the minimum number of mutations needed to reach  $\Omega$  from the most unfavorable state.

**Definition 20** (Modified path cost). *Let  $(z_1, \dots, z_\tau)$  be a path from  $x$  to  $\Omega$ . Let  $L_1, \dots, L_r$  be a set of consecutive limit sets with  $L_r \subset \Omega$  and  $L_i \not\subset \Omega$  for all  $i < r$ , through which the path passes. The modified cost of the path is obtained by subtracting from the initial cost function the intermediate radii of the limit sets  $L_i$ :*

$$c^*(z_1, \dots, z_\tau) = c(z_1, \dots, z_\tau) - \sum_{i=2}^r R(L_i). \quad (5)$$

**Definition 21** (modified coradius). *The modified coradius of the basin of attraction of  $\Omega$  is defined as:*

$$CR^*(\Omega) = \max_{x \notin \Omega} \min_{(z_1, \dots, z_\tau) \in S(x, \Omega)} c^*(z_1, \dots, z_\tau). \quad (6)$$

where  $\min_{(z_1, \dots, z_\tau) \in S(x, \Omega)} c^*(z_1, \dots, z_\tau)$  is the modified cost between a state  $x$  and  $\Omega$ .

The theorem proposed by Ellison in [19] is a sufficient condition to identify a long-run stochastically stable set of the system. It also gives an lower bound on convergence rate.

**Theorem 3** (Convergence to long-run stochastically stable set with modified cost [19]). *Let  $(Z, P, P(\epsilon))$  be a model of evolution with noise, and suppose that for some set  $\Omega$  which is a union of limit sets  $R(\Omega) > CR^*(\Omega)$ . Then:*

- 1) *The long-run stochastically stable set of the model is contained in  $\Omega$ .*
- 2) *For any  $y \notin \Omega$ ,  $W(y, \Omega, \epsilon) = O(\epsilon^{-CR^*(\Omega)})$  as  $\epsilon \rightarrow 0$ .*

In other words, if it is more difficult to leave  $\Omega$  and its basin of attraction than to come back to it, the long-run stochastically stable set is contained in  $\Omega$ .

## APPENDIX II PROOFS

### A. Proof of Lemma 1

We first consider the case  $Cn_{OPT} \leq AN \leq ACn_{OPT}$ . For a given channel allocation profile  $s$ , let  $y_s(k)$  be the number of channels with  $k \leq N$  active antennas on them. Note that  $\sum_{k=0}^N y_s(k) = C$ . A solution  $s$  maximizes  $U(s)$  if and only if  $y_s(k_0) = C$  where  $k_0 = n_{OPT}$  and  $y(k) = 0$  for  $k \neq k_0$ . Indeed, in this cases the throughput is maximized in every channel. Among all the  $s$  satisfying this condition, we can pick the ones maximizing fairness as the set of feasible solutions is finite. Let  $s^o$  be one of them. Since  $U(s^o)$  is a global maximum characterized by the highest fairness  $\eta(s^o)$ , it is not possible to increase one of the two objectives without decreasing the other, which is the definition of Pareto optimality for multi-objective maximization problems.

We now consider the case  $N \geq Cn_{OPT}$ . In this case, there exists at least a channel with more  $n_{OPT}$  WNs on it, regardless of how the antennas are distributed. Since a WN's marginal contribution is negative if the accessed channel has more than  $n_{OPT}$  WNs, a WN with more than one active antenna causes a decrease in the total system throughput wrt the case with only one antenna, regardless of how the antennas are distributed. Hence, a solution  $s$  maximizes  $U(s)$  only if all WNs have exactly one active antenna. Among all the  $s$  satisfying this condition, we can pick the solution  $s^o$  maximizing fairness as the set of feasible solutions is finite. Let  $s^o$  be one of them. Since  $U(s^o)$  is a global maximum characterized by the highest fairness  $\eta(s^o)$ , it is not possible to increase one of the two objectives without decreasing the other, which is the definition of Pareto optimality for multi-objective maximization problems.

Let us now consider the case  $AN < Cn_{OPT}$ . That means, there exist solutions where all WNs activate all of their  $A$  antennas active and all antennas bring a positive marginal contribution. Hence, a solution  $s$  maximizes  $U(s)$  only if all WNs have all of their  $A$  antenna antennas active. Among all the  $s$  satisfying this condition, we can pick the solution  $s^o$  maximizing fairness as the set of feasible solutions is finite. Let  $s^o$  be one of them. Since  $U(s^o)$  is a global maximum characterized by the highest fairness  $\eta(s^o)$ , it is not possible to increase one of the two objectives without decreasing the other, which is the definition of Pareto optimality for multi-objective maximization problems.

### B. Proof of Proposition 1

First note that every absorbing state is obviously a LS. We proceed by contradiction by assuming that a LS has two or more distinct states over which the system cycles in the unperturbed dynamics. MASAP allows users at  $t$  to only switch back a single antenna to a channel accessed at  $t-1$ . As a consequence, WNs antennas may only switch between two channels in the unperturbed dynamics (there is no exploration). Any antenna that stays on a channel during two iterations won't move anymore because the algorithm enters the "else" condition in steps 16-18 and in the unperturbed dynamics  $\epsilon = 0$ . As the LS cycles, there is thus a non empty set of WNs antennas that switch at every iteration from one channel to another for ever in a deterministic way. This means that the system evolves between only two states. As WNs antennas can compare their utility between these two states, they are able to choose one of them and stay on the corresponding channel, this contradicts the fact that these antennas change endlessly.

### C. Proof of Proposition 2

We want to show that a single-player single-antenna improvement is always possible via unilateral better response if we are not at a PNE, and that such unilateral improvements always prompt a system throughput increase. Suppose we are not at a PNE. Consider two channels  $i$  and  $i'$  with  $n_i$  and  $n_{i'}$  WNs, respectively. Take a WN  $j$  on channel  $i$  and moves its antenna on channel  $i'$ . Then its utility changes as follows:

$$\begin{aligned} \Delta_{ii'} &= (n_{i'} + 1)p(n_{i'} + 1) - n_{i'}p(n_{i'}) - n_i p(n_i) \\ &\quad + (n_i - 1)p(n_i - 1) \end{aligned}$$

If  $n_i - n_{i'} = 1$ , then  $\Delta_{ii'} = 0$ . If  $n_i - n_{i'} = 0$ ,  $\Delta_{ii'} = [(n_i + 1)p(n_i + 1) - n_i p(n_i)] - [n_i p(n_i) - (n_i - 1)p(n_i - 1)]$ , which is negative because  $np(n)$  is concave. As a consequence, as we are not at a PNE, necessarily there exist two channels  $i$  and  $i'$  such that  $n_i - n_{i'} \geq 2$ . Choose any WN on channel  $i$  not having another antenna on  $i'$  and move its antenna to channel  $i'$ . Such a WN exists because  $n_i - n_{i'} \geq 2$ . The utility of this WN changes by  $\Delta_{ii'} = [(n_{i'} + 1)p(n_{i'} + 1) - n_{i'} p(n_{i'})] - [n_i p(n_i) - (n_i - 1)p(n_i - 1)]$ , which is positive because  $np(n)$  is concave and  $n_{i'} < n_i - 1$ . We thus found a single-player single-antenna improvement. After this improvement, the system throughput as changed by:  $(n_i - 1)p(n_i - 1) + (n_{i'} + 1)p(n_{i'} + 1) - n_i p(n_i) + n_{i'} p(n_{i'}) > 0$  because of the concavity of  $np(n)$  and  $n_i - 1 > n_{i'}$ . This means that along the improvement path, the system throughput is strictly increasing. As the strategy sets are finite, the improvement path is finite. At the end of the path, as no further improvement is possible, we are at a PNE.

### D. Proof of Theorem 1

- 1) It follows from Lemma 5 and Lemma 6 that  $R(\Psi^*) > CR^*(\Psi^*)$  and the application of Theorem 3 in Appendix I.
- 2) Let us first of all show that the internal game can be seen a distributed welfare game. To this end, let us express the total throughput of the system in the form of separable welfare functions [18]:

$$W(s) = \sum_{i \in C} W_i(\{s\}_i)$$

where  $\{s\}_i = \{j \in \mathcal{N} | i \in \mathcal{S}_j\}$  is the set of users using channel  $i$  and  $W_i(\{s\}_i) = n_i p(n_i)$ , i.e., the total (cumulative) throughput available at  $i$  upon current strategy profile  $s$ . As  $n_i p(n_i)$  is concave,  $W_i$  is submodular. In a distributed welfare game, each agent's utility is defined as some fraction of the welfare generated at each resource the agent is using. In our case, the utility of WN  $j$  playing strategy  $\mathcal{S}_j \in \mathcal{S}$  for a given allocation  $s$  can be written:

$$U_j^{in}(s) = \sum_{i \in \mathcal{S}_j} d_j(n_i(s))$$

where  $d_j(\cdot)$  is the distribution rule at resource  $i$ . For the internal game  $\mathcal{G}^{in}$ , we have adopted the *marginal contribution distribution rule*:

$$d_j(n_i(s)) = W_i(\{s\}_i) - W_i(\{s_{-j}\}_i)$$

Thus,  $d_j(\cdot)$  can be interpreted as a user's contribution to the total experienced welfare.

We now show that the Price of Stability (PoS) and the Price of Anarchy (PoA) are both equal to 1, from which we will conclude that all PNEs are Pareto efficient. Recall from [18] that PoS and PoA for distributed welfare games are defined as:

$$\begin{aligned} PoS(G) &:= \max_{s^* \in \mathcal{E}} \frac{W(s^*)}{W(s^o)} \\ PoA(G) &:= \min_{s^* \in \mathcal{E}} \frac{W(s^*)}{W(s^o)} \end{aligned} \quad (7)$$

where  $\mathcal{E}$  is the set of PNEs and  $s^o$  is a Pareto efficient strategy profile. We know from [25] and [18] that for distributed welfare games characterized by players' utility function in the form of marginal contributions the PoS is equal to one and that the PoA is at least  $\frac{1}{2}$ . We now want to show that all PNEs have the same welfare value so that  $PoA = 1$ . Indeed, as at a PNE, every two channels  $i$  and  $i'$  are such that  $|n_i - n_{i'}| \leq 1$  (see Proposition 2) and as all channels are identical, all PNEs have the same total throughput.

Let now show by contradiction that the antenna throughput allocation is max-min fair at a PNE. Suppose that at a PNE, we can find another allocation such that the minimum throughput of all antennas is increased. If this allocation is such that  $|n_i - n_{i'}| \leq 1$  for all channels  $i$  and  $i'$ , then this is another PNE. In this case, the minimum antenna throughput is unchanged and we didn't increase it. If this allocation is such that there exist  $i$  and  $i'$  such that  $n_i - n_{i'} > 1$ , then we can move one antenna from channel  $i'$  to channel  $i$  and improve the minimum antenna throughput. The allocation is thus not max-min fair.

3) Let us evaluate the complexity of MASAP at each call of the algorithm and as executed by a single WN. MASAP executes  $T$  iterations. At every iteration, the payoff for each accessed channel is computed. Since at most  $A$  channels can be accessed, this operation takes  $O(A)$ . Steps 13-19 has  $O(A)$  complexity since the most demanding operation is to search and remove an element from a set of cardinality  $A$ . Steps 20-25 have a complexity  $O(C)$  because it requires finding a minimum and a maximum from a set of  $C$  elements at most. Steps 26-28 is an assignment and thus have complexity  $O(1)$ . Steps 29-34 are run only when  $t = T$ . They consist in reading

the headers of at most  $T$  packets (one per slot) and finding an element in a set of at most  $N$  elements. With  $C \geq A \geq 1$ , the overall complexity is thus  $O(TC + N)$ .

### E. Proof of Proposition 3

Let us categorize the system limit sets at iteration block  $\kappa$  into unions with similar characteristics:

- **LS Union 1** The WNs have experienced at least a negative marginal contribution (red flag) so that if one active antenna is turned off the total system offered throughput increases. This occurs because the channels are characterized by decreasing marginal contributions.
- **LS Union 2** All the marginal contributions are positive (white flag) for all WNs AND the system offered throughput can be increased by one additional active antenna (activated by any of the WNs).
- **LS Union 3:** All the marginal contribution of all the WNs are positive (white flag) AND the system offered throughput can NOT be increased by one additional active antenna (activated by any of the WNs). Thus, Union 3 contains max-throughput LSs (see Definition 8).
- **LS Sub-Union 3.1:** WNs' utilities are not allocated in a fair manner. That is, Sub-Union 3.1 does not contain Pareto LSs.
- **LS Sub-Union 3.2:** This is the Enlarged Union of Pareto LSs  $\Omega^*$ . The states in sub-unions 3.1 and 3.2 are the states of  $\Upsilon^*$ .

We first consider the complete information case and we will show that there is a single-improvement path to LS sub-union 3.2, which corresponds to  $\Omega^*$ . First not that we have paths of the following kinds:  $1 \rightarrow 2 \rightarrow 3$ ,  $1 \rightarrow 3$ ,  $2 \rightarrow 3$ . Indeed, in 1 and 2 the throughput can be increased and is upper bounded, so that we finish in a state in the union 3. At this stage, either we reached 3.2 and we found a single-improvement path leading to  $\Omega^*$ , or we reached 3.1. From 3.1, there are paths of the kind:  $3.1 \rightarrow 1 \rightarrow 3.1 \rightarrow \dots \rightarrow 1 \rightarrow 3.2$ . Indeed, following SILP, the WN with the minimum number of active antennas activate one antenna. As we were in a state maximizing the sum throughput, we come back to a state belonging to union 1. In this state, the WN with the maximum number of antennas deactivates one antenna. As a consequence, in a finite number of steps of the type  $3.1 \rightarrow 1$ , all WNs have a number of antennas that differ by one at most. As we have reached a state maximizing the sum throughput and we verify the conditions of Lemma 1, we have reached 3.2.

In the incomplete information setting, the paths  $1 \rightarrow 2 \rightarrow 3$ ,  $1 \rightarrow 3$ ,  $2 \rightarrow 3$  are still feasible as the MC-flag information is supposed to be known by every WN. However, as WNs do not know if they have the minimum or maximum number of active antennas among WNs, we have no guarantee to reach 3.2. In any case,  $\Upsilon^*$  is reached as it is the union of 3.1 and 3.2.

### F. Proof of Lemma 8

We start by the complete information case. If we start from a state  $\omega \in \Omega^*$  that is also max-min fair (i.e. Pareto),



with one imutation, we reach a state in  $\Omega^*$  by definition of this set. Otherwise,  $\omega$  is not max-min fair and is at one imutation from a Pareto state  $\omega'$ . In  $\omega'$ , the MC-flag is thus white. One imutation from  $\omega'$  necessary increases by one the number of active antennas. The MC-flag in  $\omega$  is thus red. Following SILP, any imutation from  $\omega$  desactivates one active antenna. This brings back the system to a Pareto state. In the incomplete information case, the same arguments hold for  $\Upsilon^*$  by distinguishing states that maximizes the sum throughput and those that are at one imutation away from them.

## REFERENCES

- [1] Robert W. Rosenthal. A class of games possessing pure-strategy Nash equilibria. *Int. J. Game Theory*, 2(1):65–67, Dec 1973.
- [2] Dov Monderer and Lloyd S. Shapley. Potential Games. *Games and Economic Behavior*, 14(1):124–143, May 1996.
- [3] Richard Southwell, Yanjiao Chen, Jianwei Huang, and Qian Zhang. Convergence Dynamics of Graphical Congestion Games. In Vikram Krishnamurthy, Qing Zhao, Minyi Huang, and Yonggang Wen, editors, *Game Theory for Networks*, number 105 in Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, pages 31–46. Springer Berlin Heidelberg, January 2012.
- [4] Michal Penn, Maria Polukarov, and Moshe Tennenholtz. Congestion games with load-dependent failures: Identical resources. *Games and Economic Behavior*, 67(1):156–173, Sep 2009.
- [5] Igal Milchtaich. Representation of finite games as network congestion games. *Int. J. Game Theory*, 42(4):1085–1096, Nov 2013.
- [6] Jalal Khamse-Ashari, Ioannis Lambadaris, George Kesidis, Bhuvan Urganekar, and Yiqiang Zhao. An Efficient and Fair Multi-Resource Allocation Mechanism for Heterogeneous Servers. *IEEE Trans. Parallel Distrib. Syst.*, 29(12):2686–2699, Dec 2018.
- [7] Qinyun Zhu and Jae C. Oh. Equality or Efficiency: A Game of Distributed Multi-type Fair Resource Allocation on Computational Agents. *2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, 2:139–142, Dec 2015.
- [8] Carlos Alós-Ferrer and Fei Shi. Imitation with asymmetric memory. *Economic Theory*, 49(1):193–215, January 2012.
- [9] Carlos Alós-Ferrer and Karl H. Schlag. Imitation and Learning. pages 271–297, 2009.
- [10] Stefano Iellamo, Lin Chen, and Marceau Coupechoux. Proportional and double imitation rules for spectrum access in cognitive radio networks. *Computer Networks*, 2013.
- [11] X. Chen and J. Huang. Imitative spectrum access. In *2012 10th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt)*, pages 145–152, May 2012.
- [12] A. Mahajan and D. Teneketzis. *Foundations and Applications of Sensor Management*, chapter Multi-armed Bandit Problems, pages 121–151. Springer-Verlag, 2007.
- [13] A. Anandkumar, N. Michael, and A. Tang. Opportunistic Spectrum Access with Multiple Users: Learning under Competition. In *Proc. IEEE International Conference on Computer Communication (INFOCOM)*, San Diego, CA, Apr. 2010.
- [14] Xu Chen and Jianwei Huang. Spatial spectrum access game: nash equilibria and distributed learning. In *Proceedings of the thirteenth ACM international symposium on Mobile Ad Hoc Networking and Computing, MobiHoc '12*, pages 205–214, New York, NY, USA, 2012. ACM.
- [15] Richard Southwell, Jianwei Huang, and Xin Liu. Spectrum mobility games. *2012 Proceedings IEEE INFOCOM*, pages 37–45, Mar 2012.
- [16] Kobi Cohen, Angelia Nedić, and R. Srikant. Distributed Learning Algorithms for Spectrum Sharing in Spatial Random Access Wireless Networks. *IEEE Trans. Autom. Control*, 62(6):2854–2869, Jun 2017.
- [17] Mohammadhosein Hasanbeig and Lacra Pavel. On Synchronous Binary Log-Linear Learning and Second Order Q-learning\*\*This work was supported by an NSERC grant. *IFAC-PapersOnLine*, 50(1):8987–8992, Jul 2017.
- [18] Jason R. Marden and Adam Wierman. Distributed Welfare Games. *Operations Research*, 61(1):155–168, February 2013.
- [19] G. Ellison. Basins of Attraction, Long-Run Stochastic Stability, and the Speed of Step-by-Step Evolution. *Review of Economic Studies*, 67, 2000.
- [20] B. Mawlawi, J.-B. Dore, N. Lebedev, and J.-M. Gorce. Multiband CSMA/CA with RTS-CTS strategy. In *2014 IEEE 10th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, pages 628–633, October 2014.
- [21] Jason R. Marden and Jeff S. Shamma. Revisiting log-linear learning: Asynchrony, completeness and payoff-based implementation. *Games and Economic Behavior*, 75(2):788–808, Jul 2012.
- [22] R. Jain, D. Chiu, and W. Hawe. A Quantitative Measure of Fairness and Discrimination for Resource Allocation in Shared Computer Systems. Research Report TR-301, DEC, 1984.
- [23] James Bergin and Barton L. Lipman. Evolution with State-Dependent Mutations. *Econometrica*, 64(4):943–956, 1996.
- [24] H. P. Young. The Evolution of Conventions. *Econometrica*, 61(1):57–84, Jan. 1993.
- [25] A. Vetta. Nash equilibria in competitive societies, with applications to facility location, traffic routing and auctions. In *The 43rd Annual IEEE Symposium on Foundations of Computer Science, 2002. Proceedings.*, pages 416–425, 2002.



and product management.



He has been General Co-Chair of WiOpt 2017 and Gamenets 2019. In the Computer and Network Science department of Telecom ParisTech, he is working on cellular networks, wireless networks, ad hoc networks, cognitive networks, internet of things, focusing mainly on performance evaluation, optimization and resource management.



2007-2008. His research interests include implementation of advanced signal processing and wireless communications algorithms on Xilinx FPGAs and Zynq System-on-Chip (SoC) boards, application of game theory to model distributed wireless networks, prototyping access protocols for wireless networks, IoT location tracking systems, cognitive and cooperative communications, and wireless signal design.

**Stefano Iellamo** is Product Definition Analyst at Amadeus IT Services UK Limited. He obtained his Master from Politecnico di Milano (2009) and his Ph.D. from Telecom ParisTech (2014). Since 2015 he has been working for Vodafone Group (2015), FORTH-ICS (2015-2017) and Sorbonne University (2017-2019); and has recently joined Amadeus IT Services UK Limited where he serves as a Product Definition Analyst in the London R&D division. His interests revolve around spectrum access, critical communications, internet connectivity, game theory

**Prof. Marceau Coupechoux** is Professor at Telecom ParisTech and Professeur Chargé de Cours at Ecole Polytechnique. He obtained his Master from Telecom ParisTech (1999) and from University of Stuttgart (2000), his Ph.D. from Institut Eurecom (2004), his Habilitation from University Pierre et Marie Curie (2015). From 2000 to 2005, he was with Alcatel-Lucent (in Bell Labs former Research & Innovation and then in the Network Design department). He was Visiting Scientist at the Indian Institute of Science, Bangalore, India, in 2011-2012.

**Zaheer Khan** received his Dr.Sc in electrical engineering from the University of Oulu, Finland, and his M.Sc degree in electrical engineering from University College Borås, Sweden, in 2011 and 2007, respectively. Currently, he is an Adjunct Professor at the University of Oulu, Finland. He has also worked for a Tenure Track Lecturer position at the University of Liverpool, United Kingdom (2016-2017), and as a research fellow/principal investigator at the University of Oulu between 2011-2016. He was the recipient of the Marie Curie fellowship for