# Femtocell Systems with Self Organization Capabilities

Ana Galindo-Serrano, Lorenza Giupponi

# Femtocell Systems with Self Organization Capabilities

(Invited Paper)

Ana Galindo-Serrano and Lorenza Giupponi
Centre Tecnològic de Telecomunicacions de Catalunya (CTTC)
Parc Mediterrani de la Tecnologia, Av. Carl Friedrich Gauss 7, Barcelona, Spain 08860
e-mail: {ana.maria.galindo,lorenza.giupponi}@cttc.es

***Abstract*—In this paper we present a self-organization technique to manage interference in femtocell networks. We model femtocells as a multiagent system implementing a form of Reinforcement Learning (RL) known as Q-Learning (QL) to solve the aggregated interference problem originated due to the macro-femtocell systems coexistence. We discuss this approach and propose a modification in order to solve some of the potential implementation limits of the QL algorithm. In order to achieve a more accurate environment representation and therefore a more appropriate femtocell system behavior, we combine Fuzzy logic with QL, in the form of Fuzzy Q-Learning (FQL), which allows agents to represent continuously the state and action spaces. Results are presented for the proposed QL and FQL algorithms as well as for two other FQL approaches with less complex perceptions of the surrounding environment and targets and a smart power control proposed by the 3GPP standardization body.**

***Index Terms*—Femtocell network, interference management, multiagent system, FQL.**

## I. Introduction

In the past few years the use of data services in mobile networks has notably increased, which requires higher quality of services and data throughput capacity from operators. These requirements become much more demanding in indoor environments, where, due to the wall penetration losses, communications suffer a higher detriment. As a solution, short-range Base Stations (BSs), known as femtocells [1], are proposed. Femtocells are installed by the end consumer and communicate with the macrocell system through internet by means of a Digital Subscriber Line (DSL), fiber or cable connection. Due to this deployment model, number and location of femtocells are unknown for the operators and therefore, there is no possibility for a centralized network planning. In case of co-channel operation, which is the more rewarding option for operators in terms of spectral efficiency, an aggregated interference problem may arise due to the multiple, simultaneous and uncoordinated femtocell transmissions. Therefore, in this paper we propose to model the multiple femto BSs as a multiagent system [2], where each femtocell is able to learn a transmission power policy in such a way that the interference it is generating, added to the whole femtocell network interference, do not jeopardize the macrocell system performance.

To do this we propose a Reinforcement Learning (RL) approach [3], known as Q-Learning (QL), based on discrete state and action representation. In each learning iteration the agent evaluates the current state of the environment, based on this perception it selects an action from the available set of actions and the execution of the selected action is evaluated through a cost function. The acquired knowledge is stored and the process is repeated. In literature, several works can be found where QL applied to multiagent systems properly work [4], [5], [6], although, from our point of view it has an important disadvantage, the discrete state and action representation. On the one hand, it may not give accurate enough solutions or may preclude the use of this method when detailed states and/or actions are required, on the other hand, the definition of the discrete states introduces high subjectivity in the algorithm definition.

In this paper we propose a solution to solve the discrete state and action representation through a Fuzzy Q-Learning (FQL) scheme, which combines the advantages of Fuzzy Inference System (FIS) and multiagent RL. In particular, the FIS allows to generalize the state space and to generate continuous actions. Furthermore, it reduces the learning period since previous expert knowledge can be naturally embedded into the fuzzy rules. Additionally, in this paper we analyze the femtocell scenario performance in terms of system capacity. In particular, we propose an approach which, besides controlling the femtocells total transmission power and maintaining the macrocell capacity above a given threshold, it also focuses on the maximization of femtocells capacity, taking into account the femto-to-femto interference. In order to evaluate the performance, we compare five algorithms, the QL and three FQL-based schemes with different perceptions of the surrounding environment and different targets, so as to show that different objectives that can be achieved for both macro and femto systems. The benchmark algorithm is a Smart Power Control (SPC) based on interference measurements, proposed by 3rd Generation Partnership Project (3GPP) in [7].

This paper is structured as follows. The system model is presented in Section II. In Section III, we introduce the multiagent systems. In Section IV, we briefly present the QL and FQL algorithms. In Section V, we present the simulation scenario. In Section VI, we present the reference algorithms and describe relevant simulation results. Finally, in Section VII, we summarize the main conclusions.

## II. System Model

We consider a heterogeneous wireless network composed of a set of $\mathcal{M}$ macrocells that coexist with $\mathcal{F}$ femtocells. The $M{=}|\mathcal{M}|$ macrocells form a regular hexagonal network layout with inter-site distance $D$, and provide coverage over the entire network, comprising both indoor and outdoor users. The $F{=}|\mathcal{F}|$ femtocells are placed indoors within the macrocellular coverage area following the 3GPP dual-strip deployment model. Both, macrocells and femtocells, operate in the same frequency band, which allows to increase the spectral efficiency per area through spatial frequency reuse.

An Orthogonal Frequency Division Multiple Access (OFDMA) downlink is considered, where the system bandwidth $B$ is divided into $R$ Resource Blocks (RBs), with $B = R \cdot B_{\mathrm{RB}}$. A RB represents one basic time-frequency unit that occupies the bandwidth $B_{\mathrm{RB}}$ over time $T$. Associated with each macrocell and femtocell are $U^{\mathrm{M}}$ macro and $U^{\mathrm{F}}$ femto users, respectively. The multiuser resource assignment that distributes the $R$ RBs among the $U^{\mathrm{M}}$ macro and $U^{\mathrm{F}}$ femto users, is carried out by a proportional fair scheduler.

We denote by $\mathbf{p}_t^n = (p_{1,t}^n, \ldots, p_{R,t}^n)$ the transmission power vector of BS $n$ at time $t$, with $p_{r,t}^n$ denoting the downlink transmission power of RB $r$. The maximum transmission power for femtocells and macrocells are $P_{\mathrm{max}}^{\mathrm{F}}$ and $P_{\mathrm{max}}^{\mathrm{M}}$, with $P_{\mathrm{max}}^{\mathrm{F}} \ll P_{\mathrm{max}}^{\mathrm{M}}$, such that $\sum_{r=0}^{R} p_{r,t}^m \leq P_{\mathrm{max}}^{\mathrm{M}}$, $m{\in}\mathcal{M}$ and $\sum_{r=0}^{R} p_{r,t}^f \leq P_{\mathrm{max}}^{\mathrm{F}}$, $f{\in}\mathcal{F}$.

We analyze the system performance in terms of Signal to Interference Noise Ratio (SINR) and achieved data rate given in (bit/s). Assuming perfect synchronization in time and frequency, the SINR of macrouser $u$ who is allocated in RB $r$ of macrocell $m{\in}\mathcal{M}$ amounts to:

$$\gamma_{r,t}^m = \frac{p_{r,t}^m h_{r,t}^{mu}}{\sum\limits_{n\in\mathcal{M}, n\neq m} p_{r,t}^n h_{r,t}^{nu} + \sum\limits_{f\in\mathcal{F}} p_{r,t}^f h_{r,t}^{fu} + \sigma^2}$$

where $h_{r,t}^{mu}$ accounts for the link gain between the transmitting macrocell $m$ and its macrouser $u$; while $h_{r,t}^{nu}$ and $h_{r,t}^{fu}$ represent the link gain of the interference that BSs $n$ and $f$ imposes on macrouser $u$, respectively. Finally, $\sigma^2$ denotes the thermal noise power.

Likewise, the SINR of femtouser $v$ who is allocated in RB $r$ by femtocell $f{\in}\mathcal{F}$ is in the form:

$$\gamma_{r,t}^f = \frac{p_{r,t}^f h_{r,t}^{fv}}{\sum\limits_{m\in\mathcal{M}} p_{r,t}^m h_{r,t}^{mv} + \sum\limits_{n\in\mathcal{F}, n\neq f} p_{r,t}^n h_{r,t}^{nv} + \sigma^2}$$

where $h_{r,t}^{fv}$ represents the link gain between the transmitting femtocell $f$ and its femtouser $v$, $h_{r,t}^{nv}$ and $h_{r,t}^{mv}$ indicate the link gain between BS $n$ and $m$ and femtouser $v$, respectively.

The data rate that cell $n$ achieves at slot $t$ is the sum over the individual rates of all RBs, which is upper bounded by the Shannon's capacity:

$$C_t^n = \sum_{r=0}^{R} C_{r,t}^n = \sum_{r=0}^{R} \frac{B}{R} \log_2(1 + \gamma_{r,t}^n)$$

where $C_{r,t}^n$ is the capacity of RB $r$ at slot $t$, and $n{\in}\mathcal{M}$ determines the rate of a macrocell and $n{\in}\mathcal{F}$ that of a femtocell. Then, the total sum-rate of the network amounts to:

$$C_t^{\mathrm{sys}} = \sum_{m\in\mathcal{M}} C_t^m + \sum_{f\in\mathcal{F}} C_t^f$$

## III. Introduction to Multiagent systems

In Machine Learning literature, the ability of learning new behaviors online and automatically adapting to the temporal dynamics of the system is commonly associated with RL [8]. At each learning iteration, the agent perceives the state of the environment and takes an action to transit to a new state. A scalar cost is received, which evaluates the quality of this transition. In a wireless setting, the single agent approach can be applied in scenarios characterized by only one decision maker, as it is the case for the Radio Network Controller (RNC) or the serving gateway in cellular networks. On the other hand, the multiagent approach is applied in situations where the intelligence has to be distributed across multiple nodes, as it is the case of femtocell networks. In this paper we refer to the multiagent system. The theoretical framework can be found in stochastic games, which are non-cooperative games defined by the following quintuple $\{\mathcal{F}, \mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{C}\}$, where:

- $\mathcal{F}{=}\{1, 2, \ldots, F\}$ is the set of agents (femto BSs).
- $\mathcal{S}{=}\{s_1, \ldots, s_k\}$ is the set of possible states and $k$ is the agent's number of available states.
- $\mathcal{A}{=}\{\mathsf{a}_1, \ldots, \mathsf{a}_l\}$ is the set of actions and $l$ is the agent's number of available actions.
- $\mathcal{C} : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ is the cost function, which is fed back to each agent after the execution of a certain action.
- $\mathcal{P}$ is a probabilistic transition function, defining the probability of migrating from one state to another, provided the execution of a certain joint action.

For each independent agent, the state transition function probabilistically specifies the next state of the environment as a function of its current state and the joint action. The cost function specifies expected instantaneous cost as a function of current state and action. The model is a *Markov* model if the state transitions are independent of any previous environment state or agent actions. The objective is to find a policy that minimizes the cost of each state $x$. As a result, the aim is to find an optimal policy for the infinite-horizon discounted model, relying on the result that, in this case, there exists an optimal deterministic stationary policy [8].

We define the optimal value of state $x$ as the expected infinite discounted sum of costs that the agent gains if it starts in state $x$ and then executes the optimal policy. We indicate with $\pi$ the complete decision policy, so that the optimal value of state $x$ can be written as:

$$V^*(x) = \min_{\pi} \mathbb{E}\left(\sum_{t=0}^{\infty} \delta^t c(t)\right) \tag{1}$$

where $0 \leq \delta < 1$ is a discount factor and $c(t)$ is the value of function $\mathcal{C}$ in time $t$. According to the principle of Bellman's

optimality [8], this optimal value function is unique and can be defined as the solution to the equation:

$$V^*(x) = \min_a \left( \tilde{c} + \delta \sum_{y \in \mathcal{S}} P_{x,y}(a) V^*(y) \right) \quad (2)$$

which asserts that the value of state $x$ is the expected cost $\tilde{c}{=}\mathbb{E}\{c\}$, plus the expected discounted value of the next state, $y$, using the best available action. Given the optimal value function, we can specify the optimal policy as:

$$\pi^*(x) = \arg\min_a \left( \tilde{c} + \delta \sum_{y \in \mathcal{S}} P_{x,y}(a) V^*(y) \right) \quad (3)$$

In the literature, two ways have been identified to solve this problem. The first one consists of the knowledge of the state transition probability function $\mathcal{P}_{x,y}(a)$. The second one, conversely, does not rely on this previous knowledge and is based on RL. As a result, RL is primarily concerned with how an agent ought to take actions in an environment so as to minimize the notion of long-term cost, that is, so as to obtain the optimal policy, when the state transition probabilities are not known in advance. In multiagent settings where each agent has incomplete information about state transition probabilities and payoff functions of other agents, we can say that it learns independently from the other agents, therefore we approximate the other agents as part of the environment, and we still can apply the Bellman's criterion introduced above. In this case, the convergence to optimality proof does not hold strictly, but this independent learning approach has been shown to correctly converge in multiple applications [9]. As a result, we rely on RL, and in particular on QL which has the ability of learning optimal decision policies through real time interactions with the surrounding environment.

## IV. Learning Methods

In femtocells with co-channel operation the interference can not be handled by the operator by means of centralized network planning, because the number and location of femtocells are unknown and the absence of coordination with the macro network generates a distributed interference management problem. As a result, we study how femtocells distributively make decisions about resource allocation. These decisions strongly depend on those made by neighbor femtocells, since the system performances are limited by the aggregated interference generated by all femtocells simultaneously transmitting in the considered band. We map the problem onto a multiagent system, whose characteristics are: (1) the intelligent decisions are made by multiple intelligent and uncoordinated nodes; (2) the nodes partially observes the overall scenario; and (3) their inputs to the intelligent decision process are different from node to node since they come from spatially distributed sources of information. The learning process of each femto node is characterized by $R$ parallel tasks, one per each RB. In this section we present the proposed QL and FQL learning algorithms. For the sake of nomenclature simplicity, in what follows we do not include the time indicator $t$.

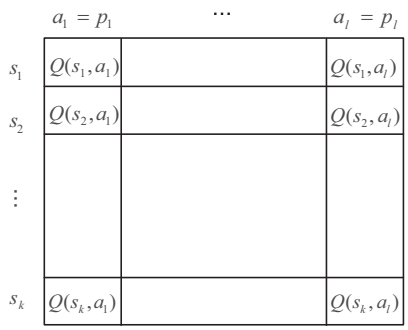


Fig. 1. Q-table for task $r$ of agent $f$.

### A. Q-Learning

In QL, the state of the surrounding environment and available actions to the agents are commonly represented by discrete sets. The QL algorithm is based on quantifying, by means of the Q-function, the quality of an action in a certain state. Therefore, to be able to learn from the past, the Q-values have to be stored in a representation mechanism. The lookup table, represented in Figure 1, is the most commonly used and the most direct method when the memory requirement is not a problem.

The state vector $\vec{x} \in \mathcal{S}$ is composed by values of representative variables capturing the surrounding environment, being $\mathcal{S} = \{s_1, \ldots, s_k\}$ the set of possible states the agent can perceive, for each RB $r$. The set of actions $\mathcal{A} = \{\mathsf{a}_1, \ldots, \mathsf{a}_l\}$ represent the decisions that the agent can make based on the state vector $\vec{x}$. Based on $\vec{x}$ and the corresponding Q-values, which are stored in the Q-table, the most appropriate action $a \in \mathcal{A}$ is selected. After the execution of $a$ in $\vec{x}$, the agent receives an immediate scalar cost $c$ and the corresponding Q-value $Q(\vec{x}, a)$ is updated according to the rule:

$$Q(\vec{x}, a) \leftarrow Q(\vec{x}, a) + \alpha[c + \delta \min_a Q(\vec{y}, a') - Q(\vec{x}, a)] \quad (4)$$

where $\alpha$ is the learning rate, $Q(\vec{y}, a')$ is the next state Q-value and $a'$ is the next state optimal action.

The state representation is given by the vector state:

$$\vec{x} = \{Pow^f, C_r^m, C_r^f\}$$

where $Pow^f = \sum_{r=0}^{r=R} p_r^f$ denotes the total transmission power of femtocell $f$ over all RBs. The considered cost for RB $r$ and femtocell $f$ is:

$$c = \begin{cases} K & \text{if} \quad Pow^f > P^F_{\max} \quad \text{or} \quad C_r^m < C^M_{Th}, \\ K \exp^{(-C_r^f C_r^m)} & \text{otherwise} \end{cases} \quad (5)$$

where $K$ is a constant initialized to a high value and $C^M_{Th}$ is the minimum capacity per RB that the macrocell has to fulfill. The rational behind this cost function is that if the macrocell capacity at RB $r$ is below the threshold $C^M_{Th}$ and the femtocell total transmission power is above the maximum allowed $P^F_{\max}$, the agent will receive a high cost value $K$. On the other hand, if both constraints are fulfilled the femtocell will focus on

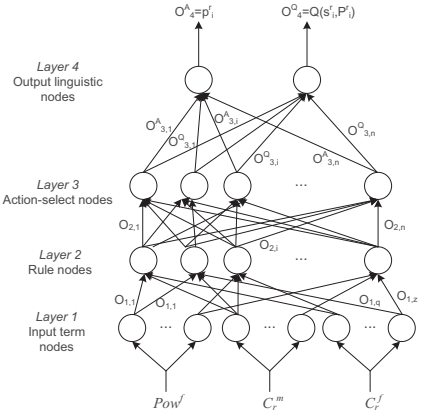


Fig. 2. FQL power, macro and femto capacity-based structure.

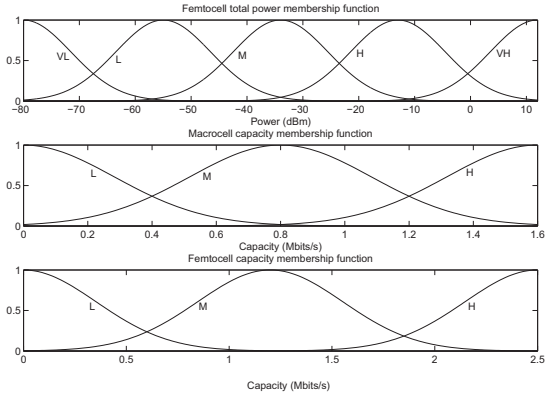


Fig. 3. Membership functions of the input linguistic variables.

maximizing its own capacity and the macrocell capacity at RB $r$.

### *B. Fuzzy Q-Learning*

When the number of state-action pairs is large or the input variables are continuous, the memory requirements may become infeasible. In addition, the selection of discrete sets for state and action definitions may highly affect the system performance. FQL scheme allows to solve those drawbacks, by combining the advantages of FIS and QL. In particular, the FIS allows to generalize the state space and to generate continuous actions [10], which is a big advantage in terms of precision, since femtocells will be able to better adjust their actions.

We consider an input state vector $\vec{x}$, represented by $L$ fuzzy linguistic variables. For each RB $r$ we denote $\bar{S} = \{\bar{s}_1, \ldots, \bar{s}_k\}$ the set of fuzzy state vectors of $L$ linguistic variables. For state $\bar{s}_i$, we denote $\mathcal{A} = \{\mathsf{a}_1, \ldots, \mathsf{a}_l\}$ the set of possible actions. The rule representation for state $\bar{s}_i$ is:

$$\begin{aligned} \textbf{If } \vec{x} \text{ is } \bar{s}_i, \ \textbf{Then } & \mathsf{a}_1 \text{ with } q(\bar{s}_i, \mathsf{a}_1) \\ & \cdots \\ \text{or } & \mathsf{a}_j \text{ with } q(\bar{s}_i, \mathsf{a}_j) \\ & \cdots \\ \text{or } & \mathsf{a}_l \text{ with } q(\bar{s}_i, \mathsf{a}_l) \end{aligned}$$

where $\mathsf{a}_j$ is the j-*th* action candidate which is possible to choose for state $\bar{s}_i$, and $q(\bar{s}_i, \mathsf{a}_j)$ is the fuzzy Q-value for each state-action pair $(\bar{s}_i, \mathsf{a}_j)$. The FQL has two outputs as a result of the defuzzification process. One corresponds to the inferred action after defuzzifying the $n$ rules and the other represents the Q-value for the state-action pair $(\vec{x}, a)$. They are given by:

$$a = \frac{\sum_{i=1}^{n} w_i \times \hat{\mathsf{a}}}{\sum_{i=1}^{n} w_i} \qquad Q(\vec{x}, a) = \frac{\sum_{i=1}^{n} w_i \times q(\bar{s}_i, \hat{\mathsf{a}})}{\sum_{i=1}^{n} w_i}$$

where $w_i$ represents the truth value (i.e., the fuzzy-AND operator) of the rule representation of FQL for $\bar{s}_i$, and $\hat{\mathsf{a}}$ is the action selected for state $\bar{s}_i$.

Q-values have to be updated after the action selection process. Since there is a fuzzy Q-value per each state-action pair, in each iteration, $n$ fuzzy Q-values have to be updated based on:

$$q(\bar{s}_i, \hat{\mathsf{a}}) = q(\bar{s}_i, \hat{\mathsf{a}}) + \alpha \Delta q(\bar{s}_i, \hat{\mathsf{a}})$$

where $\Delta q(\bar{s}_i, \hat{\mathsf{a}}) = [c + \delta(Q(\vec{y}, a') - Q(\vec{x}, a))] \times \frac{w_i}{\sum_{i=1}^{n} w_i}$. $c$ represents the cost obtained applying action $a$ in state vector $\vec{x}$ and $Q(\vec{y}, a')$ is the next-state optimal Q-value defined as:

$$Q(\vec{y}, a') = \frac{\sum_{i=1}^{n} w_i \times q(\bar{v}_i, \mathsf{a}^*)}{\sum_{i=1}^{n} w_i}$$

and $\mathsf{a}^* = \arg\min(q(\bar{v}_i, \mathsf{a}_j^*))$ with $j = (1, \ldots, l)$ is the optimal action for the next state $\bar{v}_i$, after the execution of action $\hat{\mathsf{a}}$ in the fuzzy state $\bar{s}_i$.

Figure 2 shows the FQL structure for the FQL power, macro and femto capacity-based (FQL-PMFB) algorithm as a four layer FIS. The functionalities of each layer are the following:

- *Layer 1*: This layer has as input three linguistic variables defined by the term sets: $T(Pow^f)$= {Very Low (VL), Low (L), Medium (M), High (H), Very High (VH)}, $T(C_r^m)$={L, M, H} and $T(C_r^f)$={L, M, H}. Therefore, considering the number of fuzzy sets in the three term sets in layer 1 we have $z = |T(Pow^f)|+|T(C_r^m)|+|T(C_r^f)| = 11$ term nodes. Every node is defined by a membership function with a bell shape form (Figure 3), so that the output $O_{1,h}$ for a generic component $x$ of $\vec{x}$ is given by:

  $$O_{1,h} = \exp^{-\frac{(x-e^h)^2}{(\rho^h)^2}} \qquad h = 1, \ldots, z$$

  where $e^h$ and $\rho^h$ are the mean and the variance of the bell shape function associated to node $h$, respectively. The membership function definitions for the proposed FQL are based on expert knowledge. Different mean and variance values, and the amount of fuzzy sets for the four term sets description, has been tested through simulations. Subjectivity in the term set definitions is absorbed by the learning precess in layer 3, due to the inherent adaptive capability of FQL algorithms.
- *Layer 2*: is the rule nodes layer. It is composed by $n = |T(Pow^f)| \times |T(C_r^m)| \times |T(C_r^f)| = 45$ nodes. Each node gives as output the truth value of the i-*th* fuzzy rule. Each node in layer 2 has three input values, one from one linguistic variable of each of the three components of the input state vector, so that the i-*th* rule node is represented by the fuzzy state vector $\bar{s}_i$. The layer 2 output $O_{2,i}$ is

the product of three membership values corresponding to the inputs. Truth values are represented as:

$$O_{2,i} = \prod_{h=1}^{3} O_{1,h} \qquad i = 1, \ldots, n$$

- *Layer 3*: This layer is composed by $n$ nodes each of which is an action-select node based on the $\varepsilon$-greedy policy. Here the set of possible actions for each layer 3 node are $l = 60$ power levels. Those power levels range from $-80$ dBm to 10 dBm effective radiated power (ERP). The $q(\bar{s}_i, \hat{a})$ values are initialized based on expert knowledge. The node $i$ generates two normalized outputs, which are computed as:

$$O^A_{3,i} = \frac{O_{2,i} \times \hat{a}}{\sum_{d=1}^{n} O_{2,d}} \qquad i = 1, \ldots, n$$

$$O^Q_{3,i} = \frac{O_{2,i} \times q(\bar{s}_i, \hat{a})}{\sum_{d=1}^{n} O_{2,d}} \qquad i = 1, \ldots, n$$

- *Layer 4*: This layer has two output nodes, action node $O^A_4$ and Q-value node $O^Q_4$, which represent the defuzzification method. The final outputs are given by:

$$O^A_4 = \sum_{i=1}^{n} O^A_{3,i} \qquad O^Q_4 = \sum_{i=1}^{n} O^Q_{3,i}$$

## V. Simulation Scenario

In this section we present the reference algorithms we use and some simulation results.

The scenario considered to validate the proposed approach is based on 3GPP TSG (Technical Specification Group) RAN (Radio Access Network) WG4 (Working Group 4) simulation assumptions and parameters [11]. Visualized in Figure 4, it is deployed in an urban area and operates at 1850 MHz. We consider $M = 1$ macrocells with radius $D = 500$ m and $BA = 1$ blocks of apartments, based on the dual stripes femto deployment model. We introduce an occupation ratio which determines whether inside an apartment there is a femtocell or not. Furthermore, each femtocell has a random activity pattern. Each femtocell provides service to its $U^F = 2$ associated femtousers, which are randomly located inside the femtocell area. Macrousers are also located randomly inside the femtocell block. We consider that macrousers are always outdoors and femtousers are always indoors. We consider the macro and femto systems to be based on Long Term Evolution (LTE), therefore, the frequency band is divided into RBs of width 180 kHz in the frequency domain and 0.5 ms in the time domain. Those RBs are composed of $N_{\mathrm{SC}} = 12$ subcarriers of width $\Delta f = 15$ kHz and 7 Orthogonal Frequency Division Multiplexing (OFDM) symbols. For simulations, we consider $R = 4$, which corresponds to an LTE implementation with a channel bandwidth of $BW = 0.7$ MHz. The antenna patterns for macro BS, femto BS and macro/femtousers are omnidirectional, with 18 dBi, 0 dBi and 0 dBi antenna gains, respectively. The shadowing standard deviation is 8 dB and 4 dB, for macro and femto systems, respectively. The macro and femto BS noise figures are 5 dB and 8 dB, respectively.

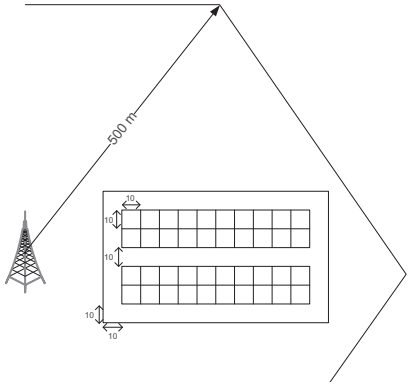


Fig. 4. System layout.

The transmission power of the macro BS is 46 dBm, whereas the femto BS adjusts its power through the learning scheme to a value of maximum $P^F_{\max} = 10$ dBm.

The considered Path Loss (PL) models are based on [11] for the case of urban scenarios and are described with details in [12]. We also consider the 3GPP implementation of frequency-selective fading model specified in [13] (urban macro settings) for macro BS to user propagation, and a spectral block fading model with coherence bandwidth of 750 kHz for indoor propagation.

With respect to the FQL algorithm, the variance $\rho^h$ of the membership functions in layer 1 are 12, 0.4 and 0.5 for total power, macrocell capacity and femtocell capacity inputs, respectively. The selected mean values $e^h$ for $Pow^f$ are -80, -55, -34 -13 and 12, for $C^m_r$ are 0, 0.8 and 1.6 and finally for $C^f_r$ are 0, 1.2 and 2.5. Also, we introduce a probability $\varepsilon = 0.07$ of visiting random states in the initial 50 % of the FQL iterations and this probability is decreased to $\varepsilon = 0.04$ until the 80 % of iterations. The benchmark algorithm is implemented considering $\eta = 0.35$ and $\beta = 0.8$.

## VI. Simulation Results

In this section we first present the reference algorithms we use to compare the proposed learning methods and then we show some simulation results.

### A. Reference algorithms

The proposed algorithms are compared to two reference FQL algorithms with lower grade of complexity than the proposed FQL-PMFB and a SPC based on interference measurements proposed in [7].

- **FQL Power-Based (FQL-PB):** The algorithm's objective is to maintain only the total femto BS transmission power below a given threshold. The vector state is represented as:

$$\vec{x} = \{Pow^f\}$$

And the cost equation is the following:

$$c = \begin{cases} K & \text{if} \quad Pow^f > P^F_{\max}, \\ 0 & \text{otherwise} \end{cases}$$

The rational behind this cost function is that the total transmission power of each femtocell does not exceed the allowed $P^F_{\max}$.
Since the structure of this algorithm is very similar to the FQL-PMFB we only highlight the different parameters.

- *Layer 1*: The term set of the input linguistic variables is defined by the following fuzzy set: $T(Pow^f) = \{\mathtt{VL}, \mathtt{L}, \mathtt{M}, \mathtt{H}, \mathtt{VH}\}$. The layer 1 of the fuzzy system is composed by $z = |T(Pow^f)| = 5$ term nodes.
- *Layer 2*: it is composed by $n = |T(Pow^f)| = 5$ nodes.

- **FQL power and macrocell capacity-based (FQL-PMB):** This algorithm objective is to maintain the total femto BS transmission power below the threshold and at the same time maximize the macrocell capacity. The input state variable is the following:

$$\vec{x} = \{Pow^f, C^m_r\}$$

and the cost equation:

$$c = \begin{cases} K & \text{if} \quad Pow^f > P^F_{\max} \quad \text{or} \quad C^m_r < C^M_{Th}, \\ 0 & \text{otherwise} \end{cases}$$

The rational behind this cost function is that, besides the total transmission power control of the femtocell, it guarantees that the macrocell capacity is above a desired threshold.
The layered structure is as for FQL-PMFB, but:

- *Layer 1*: The linguistic variables of the system are defined by the following fuzzy sets: $T(Pow^f) = \{\mathtt{VL}, \mathtt{L}, \mathtt{M}, \mathtt{H}, \mathtt{VH}\}$, $T(C^m_r) = \{\mathtt{L}, \mathtt{M}, \mathtt{H}\}$, so that layer 1 is composed by $z = |T(Pow^f)| + |T(C^m_r)| = 8$ term nodes, each one representing a fuzzy term of an input linguistic variable.
- *Layer 2*: it is composed by $n = |T(Pow^f)| \times |T(C^m_r)| = 15$ nodes.

- **Smart power control:** In this algorithm the femtocell BS adjusts its RBs transmission power based on the total received interference at the femtocell BS. The scheme is open loop, and it does not involve the femtousers and signaling between network nodes [7]. The femto BS adjusts the RBs maximum transmission power according to:

$$p^f_r = \max(\min(\eta \cdot (E_c + 10\log(R \cdot N_{\text{sc}})) + \beta, P^F_{\max}), P^F_{\min})$$

where $\eta$ is a linear scalar that allows altering the slope of power control mapping curve and $\beta$ is a parameter expressed in dB, both of which are femtocell configuration parameters. Furthermore, $N_{\text{sc}}$ is the number of subcarriers, $P^F_{\min}$ is the minimum femtocell transmit power, and $E_c$ is the reference signal received power per resource element present at the femto node.

### B. Simulation Results

In this subsection we evaluate the QL and fuzzy algorithms behavior in terms of macrocell, average femtocell, total system capacity and convergence speed. Figure 5 depicts the

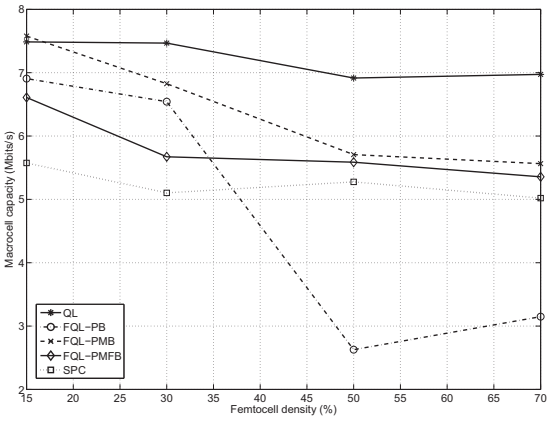


Fig. 5. Macrocell capacity.

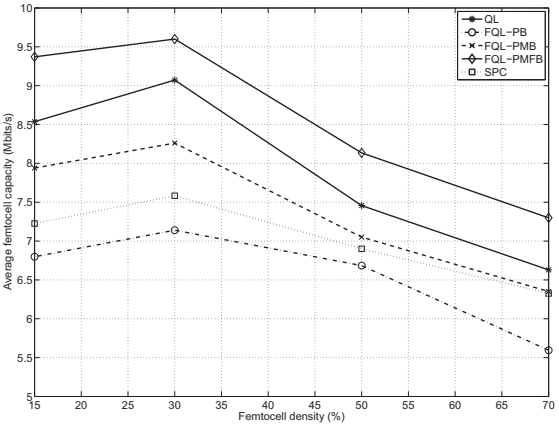


Fig. 6. Average femtocell capacity.

behavior in terms of macrocell capacity. Analyzing the fuzzy algorithms behavior, we can see that the FQL-PMB and FQL-PMFB algorithms, differently from the FQL-PB algorithm, are required in their cost equations to maintain the macrocell capacity above the $C^M_{Th}$, and so they behave, contrarily to FQL-PB, which is not able to maintain this target. In addition, it is worth mentioning that the FQL-PMFB obtains lower values of macrocell capacity than QL and FQL-PMB, since due to its cost function definition, it also aims at maximizing the femtocells capacity. QL maintains the macrocell capacity at higher values due to the discrete nature of its action space, while femtocells applying FQL, to fulfil the macrocell performance requirements, can transmit at higher power levels which increases the femto capacity but decreases the macro one. Finally, the QL, the FQL-PMB and FQL-PMFB better perform than the benchmark algorithm, since when increasing the density of femtocells, it does not adaptively operate to maintain the interference below a threshold.

Figure 6 shows the system behavior in terms of average femtocell capacity. It can be observed that average femtocell

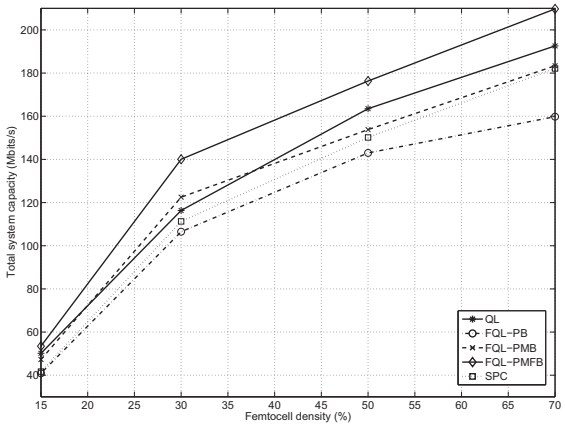


Fig. 7. Total system capacity.

capacity decreases with the femtocell density due to the increase of the femto-to-femto interference in all cases. Since QL and FQL-PMFB include in their cost equation (5) the maximization of the femtocell capacity, they are able to increase the average femtocell capacity up to more than 0.5 and 1 Mbits/s with respect to FQL-PMB and 1.5 and 2 Mbits/s with respect to FQL-PB, respectively. As mentioned before, the FQL-PMFB is able to choose more precise actions, which allow femtocells to reach higher performances maintaining the macrocell system requirements. In addition, similarly to the macrocell capacity case, the QL, FQL-PMB and FQL-PMFB algorithms outperform the SPC algorithm.

Figure 7 represents the system behavior in terms of total system capacity. As it was expected QL and FQL-PMFB have the higher system capacity, outperforming the FQL-PMB and the SPC algorithm of up to 20 Mbits/s and the FQL-PB of up to 40 Mbits/s. Total system capacity increases with the femtocell occupation ratio since there are more nodes in the network. On the other hand, femtocells capacity decreases as a function of femto nodes density due to the the interference between them.

Finally, to assess the speed of convergence, we compare the iterations before convergence required by the FQL algorithm and by the QL, in a scenario with 70 % of femtocell density. The FQL increases the speed of convergence up to 25 % with respect to the QL algorithm, due to the previous knowledge that can be embedded in the rules.

## VII. Conclusions

In this paper we have presented two multiagent RL approaches to control the aggregated interference generated by a femtocell system. The first one is the common QL algorithm, which presents some drawbacks in terms of accuracy and subjectivity. To solve this drawback we presented a learning algorithm which combines fuzzy logic with QL, allowing a more precise and adaptive system. We showed that both learning algorithms are able to learn a policy which allows them to correctly adapt to each situation. In order to present the behavior of the proposed FQL algorithm, we compare the obtained results with two less complex FQL approaches, with QL and with a benchmark proposed by 3GPP. Results show that the proposed learning methods better perform than the SPC allowing a more efficient use of the network resources.

## VIII. Acknowledgment

This work has been partially funded by the Spanish Research Council under SOFOCLES grant (TEC2010-21100), by European Union under ICT project ICT-4-248523 BeFEMTO, ACROPOLIS-NoE (ICT-2009.1.1), COST Action IC0902 and by the Spanish Ministry of Education (FPU grant AP2009-4993).

## References

[1] V. Chandrasekhar, J. G. Andrews, and A. Gatherer, "Femtocell networks: A survey," *IEEE Communication Magazine*, vol. 46, no. 9, pp. 59–67, Sept. 2008.

[2] K. P. Sycara, "Multiagent systems," *AI Magazine*, vol. 19, no. 2, pp. 79–92, 1998.

[3] M. E. Harmon and S. S. Harmon, "Reinforcement learning: A tutorial," 2000. [Online]. Available: http://www.nbu.bg/cogs/events/2000/Readings/Petrov/rltutorial.pdf

[4] X. Chen, Z. Zhao, H. Zhang, and T. Chen, "Applying multi-agent Q-learning scheme in cognitive wireless mesh networks for green communications," in *Personal, Indoor and Mobile Radio Communications Workshops (PIMRC Workshops), 2010 IEEE 21st International Symposium on*, Sept. 2010, pp. 336–340.

[5] H. Li, "Multi-agent Q-learning of channel selection in multi-user cognitive radio systems: a two by two case," in *Proceedings of the 2009 IEEE international conference on Systems, Man and Cybernetics*, ser. SMC'09. Piscataway, NJ, USA: IEEE Press, 2009, pp. 1893–1898. [Online]. Available: http://portal.acm.org/citation.cfm?id=1732003.1732028

[6] A. Galindo-Serrano, L. Giupponi, and G. Auer, "Distributed learning in multiuser OFDMA-based femtocell networks," in *Proc. of the IEEE 73th Vehicular Thechnology Conference, VTC Spring 2011*, Budapest, Hungary, May 2011.

[7] 3GPP, "3GPP TR 36.921 evolved universal terrestrial radio access (E-UTRA); FDD home eNode B (HeNB) radio frequency (RF) requirements analysis," 3GPP, Tech. Rep., March 2010.

[8] R. Bellman, *Dynamic Programming*. Princeton, NJ: Princeton Univ. Press, 1957.

[9] L. Panait and S. Luke, "Cooperative multi-agent learning: The state of the art," *Autonomous Agents and Multi-Agent Systems*, vol. 3, no. 11, pp. 383–434, Nov. 2005.

[10] Y.-H. Chen, C.-J. Chang, and C. Y. Huang, "Fuzzy Q-learning admission control for WCDMA/WLAN heterogeneous networks with multimedia traffic," *IEEE Transactions on Mobile Computing*, vol. 8, pp. 1469–1479, 2009.

[11] 3GPP, "3GPP TSG RAN WG4 (Radio) Meeting 51: Simulation assumptions and parameters for FDD HeNB RF requirements," Tech. Rep., 4-8 May 2009.

[12] A. Galindo-Serrano and L. Giupponi, "Distributed Q-learning for interference control in OFDMA-based femtocell networks," in *Proc. of the IEEE 71th Vehicular Thechnology Conference, VTC Spring 2010*, Taipei, Taiwan, May 2010.

[13] J. Salo, G. Del Galdo, J. Salmi, P. Kysti, M. Milojevic, D. Laselva, and C. Schneider, "MATLAB implementation of the 3GPP Spatial Channel Model (3GPP TR 25.996)," On-line, Jan. 2005, http://www.tkk.fi/Units/Radio/scm/.