Cognition and Docition in OFDMA-Based Femtocell Networks

Ana Galindo-Serrano, Lorenza Giupponi and Mischa Dohler
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,mischa.dohler}@cttc.es

Abstract—We address the coexistence problem between macro-cell and femtocell systems by controlling the aggregated interference generated by multiple femtocell base stations at the macro-cell receivers in a distributed fashion. We propose a solution based on intelligent and self-organized femtocells implementing a realtime multi-agent reinforcement learning technique, known as decentralized Q-learning. We compare this cognitive approach to a non-cognitive algorithm and to the well known iterative water-filling, showing the general superiority of our scheme in terms of (non-jeopardized) macrocell capacity. Furthermore, in distributed settings of such femtocell networks, the learning may be complex and slow due to mutually impacting decision making processes, which results in a non-stationary environment. We propose a timely solution — referred to as docition — to improve the learning process based on the concept of teaching and expert knowledge sharing in wireless environments. We demonstrate that such an approach improves the femtocells learning ability and accuracy. We evaluate the docitive paradigm in the context of a 3GPP compliant OFDMA (Orthogonal Frequency Division Multiple Access) femtocell network modeled as a multi-agent system. We propose different docitive algorithms and we show their superiority to the well known paradigm of independent learning in terms of speed of convergence and precision.

Index Terms—Femtocell system, interference management, multi-agent system, decentralized Q-learning.

I. INTRODUCTION

An effective way to increase the capacity of a wireless communication system is to reduce the distance between the transmitter and the receiver. Based on this consideration, femtocells [1] have recently been proposed as short-range, low-cost, low-power home base stations (BSs) installed by the consumer and designed to serve a very small area, such as a home or an office environment, providing radio coverage of a certain cellular network standard. Due to these characteristics, they can be deployed far more densely than macrocells, so that the spectrum can be reused more efficiently than in macro networks. Femtocells serve approximately one to four users; therefore, they can devote a larger portion of resources to their users compared to large coverage macrocells. Indoor users hence have high and more reliable data rates, while the operator reduces traffic in the macrocell since it focuses only on outdoor and mobile users.

Despite these technical advantages and the promised benefits from both the operator’s and the subscriber’s perspectives, femtocell technology as it stands today still faces several issues that urgently need to be solved. Femtocells can share their deployment frequency band with the existing macro network, or can operate in a dedicated frequency band. The interference management is more challenging in case of co-channel operation, but this option is more rewarding for the operator due to the increased spectral efficiency. We therefore focus on the latter type of operation. However, since femtocells are placed by end consumers, their positions and numbers are unknown to the network operator, so that the interference cannot be handled by means of a centralized frequency planning leading to a distributed interference management problem. This has to be handled in a self-organized fashion in order to reduce the signalling burden on the backhaul.

As a result of that, in this paper we focus on the autonomous management of closed access femtocells where a distributed machine learning based approach facilitates the coexistence in terms of interference generated to the macro system. The machine learning community has proposed a powerful learning technique known as reinforcement learning (RL), which is capable of finding optimal decision policies in dynamic scenarios. RL does not require environmental models and allows nodes to take actions while they learn. Among RL techniques, Q-learning [2] has been especially well studied, and possesses a firm foundation in the theory of Markov decision processes (MDPs). Q-learning can still be applied to the distributed femtocell setting, in the form of the so called decentralized Q-learning. Here, each node learns independently from the other nodes, which are assumed to be part of the surrounding environment. This paradigm, known as independent learning, will be shown to outperform, in terms of (non-jeopardized) macrocell capacity, a reference non-cognitive paradigm and the well known iterative water-filling (WF) approach where femto BSs selfishly maximize their capacity.

However, the environment is no longer stationary, since it consists of other nodes who are similarly adapting. This may generate oscillating behaviors that not always reach an equilibrium and that are not yet fully understood, even by machine learning experts. The dynamics of learning may thus be long and complex in terms of required operations and memory, with complexity increasing with an increasing observation space. A possible solution to mitigate this problem, to speed up the learning process and to create rules for unseen situations, is to facilitate expert knowledge exchange among learners [3][4].

Even as cognition and learning have received a considerable attention from various communities in the past, the process of knowledge transfer, i.e., teaching over the wireless medium
has received fairly little attention to date. We thus aim at introducing in this paper an emerging framework for femtocells, referred to as docincell [5], from “docere” = “to teach” in Latin, which relates to nodes teaching other nodes. The femto BSs are not (only) supposed to teach end-results, but rather elements of the methods of getting there. This concept perfectly fits a femtocell network scenario, where a femtocell is active only when the users are at home. When a femto BS is switched on, instead of starting a very energy expensive context awareness phase to sense the spectrum and learn the proper radio resource management (RRM) policy, it can take advantage of the decision policies learnt by the neighbor femtocells, which have been active during a longer time. This novel paradigm for femtocells will be shown to capitalize on the advantages but, most importantly, mitigate major parts of the drawbacks of purely self-organized and cognitive schemes, thus increasing their precision and accuracy and speeding up the learning process.

This paper is structured as follows. In Section II, we describe the system model. In Section III, we present the distributed Q-learning algorithm; then we modify this algorithm in order to introduce different docitive techniques. In Section IV, we describe the simulation scenario considered for performance evaluation, based on recent 3GPP specifications. In Section V, we describe relevant simulation results. Finally, in Section VI, we summarize and outline open research lines.

II. SYSTEM MODEL

We consider $L$ macrocells, with a BS located at the center of each hexagonal coverage area. $Q_m$ macrousers are randomly located inside the macro coverage area. The macrocells are deployed in an urban area and coexist with $N$ femtocells. Each femtocell provides service to its $Q_f$ associated femtousers. We consider that the total bandwidth $BW$ is divided into sub-channels of width $\Delta f$. Orthogonal frequency division multiplexing (OFDM) symbols are grouped into resource blocks (RBs), consisting of a certain number of sub-channels. Both macrocells and femtocells operate in the same frequency band and have the same amount $R$ of available RBs, which allows to increase the spectral efficiency per area through spatial frequency reuse. We focus only on the downlink operation.

We denote by $p_{i,F}^{r}$ and $p_{i,M}^{r}$ the transmission power vector of femtocell $i$ and macrocell $j$ respectively. With $p_{i,F}^{r}$ and $p_{i,M}^{r}$ denoting the downlink transmission power of femto and macro BS in RB $r$ respectively. We assume that the $R$ RBs in both the femto and macro systems are defined according to a proportional scheduling policy. The maximum transmission power for femto and macro BSs are $P_{F}^{\text{max}}$ and $P_{M}^{\text{max}}$ respectively, where $\sum_{r=1}^{R} p_{i,r}^{F} \leq P_{F}^{\text{max}}$ and $\sum_{r=1}^{R} p_{i,r}^{M} \leq P_{M}^{\text{max}}$.

The signal to interference noise ratio (SINR) at macrouser $q$ allocated in RB $r$ of macrocell $j$ is:

$$\gamma_{r}^{j,M} = \frac{p_{r}^{F,j,M} h_{r}^{M,j,r}}{\sum_{k=1,k\neq j}^{L} p_{r}^{F,k,M} h_{r}^{M,k,r} + \sum_{i=1}^{N} p_{r}^{F,i,M} h_{r}^{M,i,r} + \sigma^{2}}$$  

with $j = 1, \ldots, L$. Here, $h_{r}^{M,j,r}$ indicates the link gain between the transmitting macro BS $j$ and its macrouser $q$; $h_{r}^{M,k,r}$ indicates the link gain between the transmitting macro BS $k$ and macrouser $q$ in macro BS $j$; $h_{r}^{F,j,r}$ indicates the link gain between the transmitting femto BS $i$ and macrouser $q$ of macrocell $j$; finally, $\sigma^{2}$ is the noise power.

The capacity of macrocell $j$ is:

$$C^{j,M} = \sum_{r=0}^{R} \frac{BW}{R} \log_{2} \left( 1 + \gamma_{r}^{j,M} \right)$$  

with $j = 1, \ldots, L$.

III. LEARNING AND TEACHING TECHNIQUES

In the considered scenario the $N$ femto BSs have to distributively learn an optimal policy for each RB to achieve the common objective of maintaining the aggregated interference at macrousers in all RBs below a certain threshold. The multi-agent learning problem can be solved by means of distributed RL approaches, such as distributed Q-learning, when the probabilistic transition function cannot be deduced. The main challenge in this field is how to ensure that individual decisions of the nodes result in jointly optimal decisions for the entire network, considering that the standard convergence proof for Q-learning does not hold in this case as the transition model depends on the unknown policy of the other learning nodes.

The Nash Q-learning algorithm proposed by Hu and Wellman [6], extends the Q-learning to multi-agent domain, by taking into account the joint actions of the participants and is shown to converge to a Nash equilibrium with probability 1, under some conditions. However, this approach presents scalability issues since both the state and action spaces scale exponentially with the number of nodes. Alternatively, we can let each node learn its policy independently of the other nodes (i.e., independent learning), but then the transition model depends on the policy of the other learning nodes, which may result in oscillatory behaviors and in slow speed of convergence to prior set targets [7]. This introduces game-theoretic issues to the learning process, which are not yet fully understood.

As a solution to these problems, we propose a distributed approach where nodes share potentially differing amounts of intelligence acquired on the run. This is expected to sharpen and speed up the learning process. Any achieved gain, however, needs to be gauged against the overhead incurred due to the exchange of docitive information which is not focus of this paper.

A. Decentralized Q-Learning

It is assumed that the environment is a finite-state, discrete-time stochastic dynamical system. Let $S$ be the set of possible states $S = \{s_1, s_2, \ldots, s_m\}$, and $A$ be a set of possible actions $A = \{a_1^1, a_2^1, \ldots, a_m^1\}$ that each femto BS $i$ may choose with respect to RB $r$. The interactions between the multi-agent system and the environment at each time instant $t$ corresponding to RB $r$ consist of the following sequence.

- The agent $i$ senses the state $s_i^r = s \in S$.
Based on $s$, agent $i$ selects an action $a_i^{\pi} = a \in A$.
- As a result, the environment makes a transition to the new state $s_{t+1} = v \in S$.
- The transition to the state $v$ generates a cost $c^{i,r}_t = c \in \mathbb{R}$, for agent $i$.
- The cost $c$ is fed back to the agent and the process is repeated.

The objective of each agent is to find an optimal policy $\pi^*(s) \in A$ for each $s$, to minimize some cumulative measure of the cost $c^{i,r}_t = c(s,a)$ received over time. For each agent $i$ and learning process $r$, we define an evaluation function, denoted by $Q^{i,r}(s,a)$, as the expected total discount cost over an infinite time. To simplify the notation, in the following we refer to $Q^{i,r}(s,a)$ as $Q(s,a)$:

$$Q(s,a) = \mathbb{E}\left\{ \sum_{t=0}^{\infty} \gamma^t c(s_t, \pi(s)) | s_0 = s \right\} \tag{3}$$

where $0 \leq \gamma < 1$ is a discount factor. If the selected action $a$ at time $t$ following the policy $\pi(s)$ corresponds to the optimal policy $\pi^*(s)$, the Q-function is minimized with respect to the current state.

Let $P_{s,v}(a)$ be the transition probability from state $s$ to state $v$, when action $a$ is executed. Then, (3) can be expressed as:

$$Q(s,a) = \mathbb{E}\{c(s,a)\} + \gamma \sum_{v \in S} P_{s,v}(a)Q(v,b) \tag{4}$$

where $\mathbb{E}\{c(s,a)\}$ denotes the expected value of $c(s,a)$ and $b$ is the action to take in state $v$. Equation (4) indicates that the Q-function of the current state-action pair, for each agent $i$ and learning process $r$, can be represented in terms of the expected immediate cost of the current state-action pair and the Q-function of the next state-action pairs. The task of Q-learning is to determine an optimal stationary policy $\pi^*$ without knowing $\mathbb{E}\{c(s,a)\}$ and $P_{s,v}(a)$, which makes it well suited for learning a power allocation policy in a femtocell system.

The principle of Bellman’s optimality assures that, for single agent environments, there is at least one optimal stationary policy $\pi^*$ which is such that [2]:

$$V^*(s) = V^{\pi^*}(s) = \min_{a \in A} \mathbb{E}\{c(s,a)\} + \gamma \sum_{v \in S} P_{s,v}(a)V^*(v) \tag{5}$$

In multi-agent settings, where each agent learns independently from the other agents, we approximate the other agents as part of the environment, and we still can apply Bellman’s criterion. In this case, the convergence to optimality proof does not hold strictly, but such an independent learning approach has been shown to correctly converge in multiple applications [8]. Applying Bellman’s criterion, first we have to find an intermediate minimal of $Q(s,a)$, denoted by $Q^*(s,a)$, where the intermediate evaluation function for every possible next state-action pair $(v,b)$ is minimized, and the optimal action is performed with respect to each next state $v$. $Q^*(s,a)$ is:

$$Q^*(s,a) = \mathbb{E}\{c(s,a)\} + \gamma \sum_{v \in S} P_{s,v}(a)\min_{b \in A} Q^*(v,b) \tag{6}$$

This allows us to determine the optimal action $a^*$ with respect to the current state $s$. In other words, we can determine $\pi^*$. Therefore, $Q^*(s,a^*)$ is minimal, and can be expressed as:

$$Q^*(s,a^*) = \min_{a \in A} Q^*(s,a)$$

As a result, the Q-value $Q(s,a)$ represents the expected discounted cost for executing action $a$ at state $s$ and then following policy $\pi$ thereafter. The Q-learning process tries to find $Q^*(s,a)$ in a recursive manner using available information $(s,a,v,c)$, where $s$ and $v$ are the states at time $t$ and $t+1$, respectively; and $a$ and $c$ are the action taken at time $t$ and the immediate reward due to $a$ at $s$, respectively. The Q-learning rule to update the Q-values relative to agent $i$ and learning process $r$ is:

$$Q(s,a) = Q(s,a) + \alpha[c + \gamma \min_{a} Q(v,a) - Q(s,a)] \tag{7}$$

where $\alpha$ is the learning rate. A study on computational complexity is discussed in [9]. For more details about RL and Q-learning the reader is referred to [2].

### B. Docitive Femtocells

Some early contributions in machine learning literature [3][4] suggest that the performances of a multi-agent learning can be improved by using cooperation among learners in a variety of ways. In our scenario, for example, a femto BS which has been switched on, can take advantage of the exchange of information and expert knowledge from other femto BSs in the neighborhood [4], the so-called docitive femtocells. The agents select the most appropriate femto BS from which to learn, based on the level of expertise and the similar impact that their actions may have on the environment, which is captured by a gradient $\nabla_i$. Notice that in terms of signaling overhead the gradient is only a numerical value to exchange sporadically among femto BSs. This gradient, for femtocell $i$ and RB $r$, is defined as:

$$\nabla_i^r = \frac{\text{SINR}^r_i - \text{SINR}^{r-1}_i}{a^r_i - a^{r-1}_i}, \tag{8}$$

where $a^r_i$ and $a^{r-1}_i$ represent the actions taken for RB $r$ at time $t$ and $t-1$, respectively, and $\text{SINR}^r_i$ and $\text{SINR}^{r-1}_i$ represent the SINR at the macrouser in RB $r$ at time $t$ and $t-1$, respectively. The rationale behind the definition of this gradient is that nodes should learn from nodes in similar situations, e.g., a femto BS which is located close to a macrouser should learn the policies acquired by a femto BS operating under similar conditions. Depending on the degree of docition among nodes, we consider in this paper the following cases:

- **Startup Docition.** Docitive femto BSs teach their policies to any newcomers joining the network. In this case, each node learns independently; however, when a new femto-cell joins the network, instead of learning from scratch
radius block of apartments has two stripes, separated by a
We consider the macro and femto systems to be based on
also located outdoor and randomly between the two blocks of
randomly located inside the femtocell area. Macrousers are
operates at $1850$ MHz. The antenna patterns for macro BS, femto BS and
macro/femto users are omnidirectional, with $18$ dBi, $0$ dBi and
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. 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 $10$ dBm.

The considered path loss ($PL$) models are for urban scen-
arios and are summarized in Table I. Here, $d$ and $d_{\text{indoor}}$ are the
total and indoor distances between the macro/femto BS and
the macro/femto user, respectively. The factor $0.7d_{\text{indoor}}$ takes
into account the penetration losses due to the walls inside the
apartments. $WP_{\text{out}} = 15$ dB and $WP_{\text{in}} = 5$ dB are the
penetration losses of the building external walls and of the walls
separating the apartments, respectively. Finally, $w$ represents
the number of walls separating apartments. We also consider
the 3GPP implementation of frequency-selective fading model
specified in [11] (urban macro settings) for macro BS to user
propagation, and a spectral block fading model with coherence
bandwidth $750$ kHz for indoor propagation.

As for the decentralized Q-learning, the state, actions
and cost are defined as follows:

- **State:** At time $t$ for femtocell $i$ and RB $r$ the state is
defined as:
  \[ s = \{ I_i^t, Pow_i^t \} \]
  where $I_i^t$ specifies the level of aggregated interference
  generated by the femtocell system. The set of possible
  values is based on:
  \[ I_i^t = \begin{cases} 
  0 & \text{ if } SINR_i < (SINR_{Th} - 2 \text{dB}), \\
  1 & \text{ if } (SINR_{Th} - 2 \text{dB}) \leq SINR_i \leq (SINR_{Th} + 2 \text{dB}), \\
  2 & \text{ otherwise} 
  \end{cases} \]
  where $SINR_i$ is the instantaneous SINR measured at the
  macrourser for RB $r$ and $SINR_{Th} = 20$ dB represents
  the minimum value of SINR that can be perceived by the
  macrourcers.
  \[ Pow_i^t = \sum_{r=0}^{R} p_i^t \]
  denotes the total transmission power by the femtocell $i$ in all RBs at time $t$. The set of possible
  values is based on:
  \[ Pow_i^t = \begin{cases} 
  0 & \text{ if } Pow_{Th} - 6 \text{dBm} < Pow_i^t, \\
  1 & \text{ if } (Pow_{Th} - 6 \text{dBm}) \leq Pow_i^t \leq Pow_{Th}, \\
  2 & \text{ otherwise} 
  \end{cases} \]
  where $Pow_{Th} = 10$ dBm is the maximum transmission
  power that a femto BS can transmit.

- **Action:** The set of possible actions are the $l = 60$ power
  levels that femto BS can assign to RB $r$. Those power
  levels range from $-80$ to $10$ dBm effective radiated
  power (ERP), with $1$ dBm granularity between $10$ dBm and

Fig. 1. System layout.
TABLE I
PATH LOSS MODELS FOR URBAN DEPLOYMENT.

<table>
<thead>
<tr>
<th>Macro BS to macro/femto user</th>
<th>outdoors</th>
<th>( PL(\text{dB}) = 15.3 + 37.6 \log_{10} d )</th>
</tr>
</thead>
<tbody>
<tr>
<td>indoors</td>
<td></td>
<td>( PL(\text{dB}) = 15.3 + 37.6 \log_{10} d + WP_{\text{out}} )</td>
</tr>
<tr>
<td>User in the same apartment stripe</td>
<td>( PL(\text{dB}) = 38.46 + 20 \log_{10} d + 0.7d_{\text{indoor}} + w * WP_{\text{in}} )</td>
<td></td>
</tr>
<tr>
<td>User outside the apartment stripe</td>
<td>( PL(\text{dB}) = \max(15.3 + 37.6 \log_{10} d, 38.46 + 20 \log_{10} d) + 0.7d_{\text{indoor}} + w * WP_{\text{in}} + WP_{\text{out}} )</td>
<td></td>
</tr>
<tr>
<td>User inside a different apartment stripe</td>
<td>( PL(\text{dB}) = \max(15.3 + 37.6 \log_{10} d, 38.46 + 20 \log_{10} d) + 0.7d_{\text{indoor}} + w * WP_{\text{in}} + 2 * WP_{\text{out}} )</td>
<td></td>
</tr>
</tbody>
</table>

- 40 dBm and 4 dBm granularity between -40 dBm and -80 dBm.

- **Cost**: The cost \( c_t^a \) assesses the immediate return incurred due to the assignment of action \( a \) in state \( s \). The considered cost function is:

  \[
  c_t^a = \begin{cases} 
  K & \text{if } \text{Pow}_t^i > \text{Pow}_{\text{Th}}, \\
  (\text{SINR}_t - \text{SINR}_{\text{Th}})^2 & \text{otherwise}
  \end{cases}
  \] (12)

  where \( K = 500 \).

  The rational behind this cost function is that the total transmission power of each femtocell does not exceed the allowed \( \text{Pow}_{\text{Th}} \), and the SINR at the macrouser is below the selected threshold \( \text{SINR}_{\text{Th}} \).

  With respect to the Q-learning algorithm, the learning rate is \( \alpha = 0.5 \) and the discount factor is \( \gamma = 0.9 \). Also, we introduce a probability \( \varepsilon = 0.05 \) of visiting random states in the initial 80% of the Q-learning iterations.

V. SIMULATION RESULTS

The decentralized Q-learning scheme, following the independent learning or docitive paradigm, has been compared to two reference algorithms:

- **Distance-Based Non-Cognitive**: The rationale behind this reference algorithm is that femtocell \( i \) selects the transmission power of RB \( r \) based on its distance from the macrouser using that RB. The set of possible values of power to assign is the same as for the decentralized Q-learning. Notice that this reference algorithm is only proposed as a non-cognitive benchmark for comparison purposes, and for its implementation we make the hypothesis that the femto network has at least some approximate knowledge of the position of the macrousers, which is a quite difficult hypothesis in a realistic cellular network.

- **Iterative Water-Filling**: It is a non-cooperative game where agents are selfish and compete against each other by choosing their transmit power to maximize their own capacity, subject to a total power constraint, such that:

\[
\max_{p_r^{i,F}} \sum_{r=1}^{R} \log \left( 1 + \frac{p_r^{i,F} h_{i,r}^{FF}}{\sum_{k=1,k\neq i}^{N} p_r^{k,F} h_{k,r}^{FF} + \sigma^2} \right) \\
\text{s.t. } \sum_{r=1}^{R} p_r^{i,F} \leq P_{\text{max}}, \quad p_r^{i,F} \geq 0 \quad (13)
\]

The solutions to (13) are given by the iterative WF power allocation solutions [12]:

\[
p_r^{i,F} = \max \left( \frac{1}{\lambda^{i,F}} - \frac{\sum_{k=1,k\neq i}^{N} p_r^{k,F} h_{k,r}^{FF} + \sigma^2}{h_{i,r}^{FF}}, 0 \right) \quad (14)
\]

where \( \lambda^{i,F} \) is the Lagrangian multiplier chosen to satisfy the power constraint.

Figure 2 shows the macrocell capacity as a function of femtocell density. It can be observed that learning techniques do not jeopardize the macrocell capacity, maintaining it at a desired level independently of the number of femtocells. On the other hand, with the distance-based reference algorithm, the macrocell capacity decreases when the number of femtocells increases, since the reference algorithm does not adaptively consider the aggregated interference coming from the multiple femtocells in the power allocation process. Furthermore, the iterative WF algorithm dramatically reduces the macrocell capacity due to its selfish power allocation policy. Finally, with respect to the implementation, it is worth mentioning that the decentralized Q-learning approaches only need feedback from the macro network about the SINR at the macrousers, which can be provided through the wired backhaul (i.e., X2-interface). However, the reference approaches rely on stronger hypotheses, such as the positions of the macrousers.
As for the performance of docition, Figure 3 shows performances in terms of precision, i.e., oscillations around the target SINR. We assumed a 50 % femtocell occupancy ratio, composed of the probability that a femtocell is present and that it is switched on. In particular, it represents the complementary cumulative distribution function (CCDF) of the variance of the average SINR at the control point with respect to the set target of $SINR_{Th} = 20$ dB. It can be observed that due to the distribution of intelligence among interactive learners the paradigm of docition stabilizes the oscillations by reducing the variance of the SINR with respect to the specified target. More precisely, at a target outage of 1 %, we observe that the IQ driven docition outperforms the startup docition by a factor of two, and the independent learning algorithm by about an order of magnitude. Figure 4 shows the probability that the total power at a femtocell is higher than $Pow_{Th}$ as a function of the learning time. It can be observed that the docitive approaches better satisfy the constraint in terms of total transmission power.

VI. CONCLUSIONS

In this paper we have presented a decentralized Q-learning approach for interference management in a macro-femto network to improve the systems’ coexistence. However, the main drawback of the proposed scheme is the length of the learning process. As a result, we have focused on the novel paradigm of docition, with which a femto BS can learn the interference control policy already acquired by a neighboring femtocell which has been active during a longer time, and thus saving significant energy during the startup and learning process. Notably, we have shown in a 3GPP compliant scenario that, with respect to decentralized Q-learning, docition applied at startup as well as continuously on the run yields significant gains in terms of convergence speed and precision.

VII. ACKNOWLEDGMENT

This work has been partially supported by the Generalitat de Catalunya under grant 2009-SGR-940 and performed in the framework of the ICT project ICT-4-248523 BeFEMTO, which is partly funded by the EU. The authors would like to acknowledge the contributions of their colleagues from BeFEMTO consortium.

REFERENCES


