Learning Coarse Correlated Equilibria in Two-Tier Wireless Networks

Mehdi Bennis, Samir Medina Perlaza, Mérouane Debbah

Learning Coarse Correlated Equilibria in Two-Tier Wireless Networks

M. Bennis*, S. M. Perlaza† and M. Debbah†
*Centre for Wireless Communications, University of Oulu, Finland
†Alcatel-Lucent Chair in Flexible Radio, SUPELEC, France
E-mail: bennis@ee.oulu.fi, {merouane.debbah, samir.medinaperlaza}@supelec.fr

***Abstract*—In this paper, we study the *strategic* coexistence between macro and femto cell tiers from a game theoretic *learning* perspective. A novel *regret-based* learning algorithm is proposed whereby *cognitive* femtocells mitigate their interference toward the macrocell tier, on the downlink. The proposed algorithm is fully decentralized relying only on the signal-to-interference-plus-noise ratio (SINR) feedback to the corresponding femtocell base stations. Based on these local observations, femto base stations learn the probability distribution of their transmission strategies (power levels and frequency band) by minimizing their regrets for using certain strategies, while adhering to the cross-tier interference constraint. The decentralized regret based learning algorithm is shown to converge to an ϵ-coarse correlated equilibrium (ϵ-CCE) which is a generalization of the classical Nash Equilibrium (NE). Finally, numerical results are shown to corroborate our findings where, quite remarkably, our learning algorithm achieves the same performance as the classical regret matching, but with *substantially* much less overhead.**

I. Introduction

Wireless data traffic has been increasing exponentially in recent years in which the emergence of novel wireless services such as social networking and gaming has introduced stringent quality-of-service (QoS) and data rate constraints on next-generation wireless cellular networks. This increase led mobile operators to explore new ways to achieve network coverage improvements, higher spectral efficiencies, and OPEX/CAPEX reductions. In view of this, femtocell technology (and small cells in general) represents a novel networking paradigm based on the idea of deploying low-power, low-cost base stations underlaying the legacy macrocell network [1].

The deployment of future heterogeneous cellular networks supporting macro, pico, and femtocells coexisting on the same spectrum and in the same geographical area entails new technical challenges for mobile operators. These challenges encompass *co-tier* and *cross-tier* interference, coverage holes due to large transmit power disparity across small cells, handover optimization, and heterogenous backhaul design. Furthermore, because femtocells are user-deployed, self-organizing network (SON) capabilities requiring innovative interference and mobility management are of vital importance. For this reason, both academia and standardization bodies are currently looking at SON aspects in which self-configuration and self-optimization are deemed instrumental to make the deployment of femtocell networks feasible.

The authors would like to thank the Finnish funding agency for technology and innovation, Elektrobit, Nokia and Nokia Siemens Networks for supporting this work. This work has been performed in the framework of the ICT project ICT-4-248523 BeFEMTO, which is partly funded by the EU.

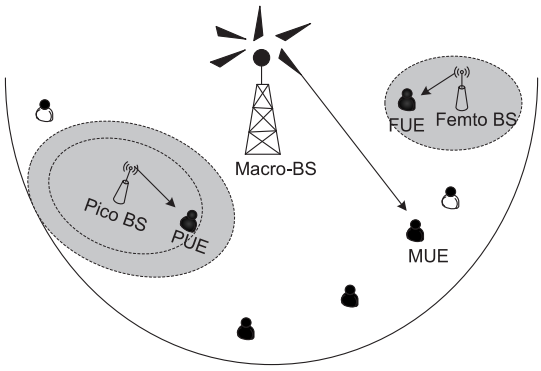

Fig. 1. Two-tier network topology including a mix of high-power (Macro-BS) and low-power base stations (Pico-BS and Femto-BS). MUE/FUE/PUE stand for macro/femto/pico user equipments.

Interference management for femtocell networks can be found in a number of works such as [2], [3], [4] among others where a variety of techniques including dynamic frequency planning, dynamic spectrum occupation, power control, closed/open/hybrid access group [3], context awareness capabilities were studied. Interference management solutions based on reinforcement learning (RL) techniques such as Q-learning, cooperative Q-learning, replication dynamics and their variants were studied in [4], [6].

In this paper, we propose a fully decentralized algorithm for interference mitigation from the femto-to-macrocell tier, on the downlink, as shown in Fig. 1. The basic idea is that each femtocell base station (FBS) *learns the regret* of its actions taken at every time instant, aiming at minimizing its average regret over time. At the same time and owing to the two-tier *hierarchy*, femtocells need to mitigate their interference brought onto the macrocell. A player's *regret* is defined as the difference between its average utility when playing the same action in all previous stages of the game, and its average utility obtained by constantly changing its actions. The underlying assumption in our work is that feedback messages from macrocell users (MUEs) to their serving macrocell base station (MBS) containing their instantaneous signal to interference

plus noise ratio (SINR) can be decoded by all FBSs. The repetitive observation of the SINR is used by the FBS to dynamically configure how often different frequency bands are used such that, a minimum time-average SINR level can be guaranteed at the macrocell tier. Finally, our proposed algorithm allows us to capture fundamental issues, namely, (1) joint estimation of players' own utility function due to perturbation and regret minimization, and (2) alleviate the setback of the classical regret matching procedure in terms of information requirement.

This paper is organized as follows. In Section II, the system and game model are presented. Section III describes the regret-based learning procedure carried out by femtocells to learn their optimal transmission strategies and mitigate their interference towards the macrocell tier. The distributed regret based learning algorithm is described in Section IV. Finally, numerical results are presented in Section V, and Section VI concludes this paper.

II. Models

A. Notations

Boldface upper and lower case symbols represent vectors and scalars. Given a random variable z, the expectation with respect to the probability distribution of z is denoted by $\mathbb{E}_z[.]$. The indicator function is denoted by $\mathbf{1}_{\{condition\}}$ and it equals 1 (resp. 0) when *condition* is true (resp. false). Given a finite set $\mathcal{A}$, $\triangle(\mathcal{A})$ represents the set of all probability distributions over the elements of the finite set $\mathcal{A}$. Let the vector $\boldsymbol{e}_s^{(S)} = \left(e_{s,1}^{(S)}, ..., e_{s,S}^{(S)}\right) \in \mathbb{R}^S$ denote the s-th vector of the canonical base spanning the space of real vectors of dimension S. Here, $\forall n \in \{1, ..., S\} \setminus \{s\}$, $e_{s,n}^{(S)} = 0$ and $e_{s,s}^{(S)} = 1$.

B. System Model

Let us assume $M = 1$ macrocell base station operating over a set $\mathcal{S} = \{1, \ldots, S\}$ of S frequency bands. Let $\boldsymbol{\Gamma_0} = (\Gamma_0^{(1)}, \ldots, \Gamma_0^{(S)})$ denote the minimum average SINR offered by the MBS to its macrocell user equipment over its corresponding spectrum band. Consider now a set $\mathcal{K} = \{1, \ldots, K\}$ of K femtocells underlaying the macrocell. Each femtocell can use any of the available frequency bands to serve its corresponding femto end-users (FUE) as long as it does not induce a lower average SINR than the minimum required by the MUE. At each time interval each FBS serves one FUE over one of the available channels following a time division multiple access (TDMA) policy.

Designate the MBS's transmit power on a given sub-carrier to be $p_0^{(s)}$ and let $|h_{0,0}^{(s)}|^2$ denote the channel gain between the MBS and its associated MUE in sub-carrier $s \in \mathcal{S}$. Likewise, $|h_{i,j}^{(s)}|^2$ denotes the channel gain between transmitter j and receiver i on sub-carrier s, and let ${\sigma_k^{(s)}}^2$ be the variance of the additive white Gaussian noise at receiver k, which is assumed to be constant over all sub-carriers for simplicity. Let $p_{k,\max}$ with $k \in \mathcal{K}$ be the maximum transmit power of FBS k. For all $k \in \mathcal{K}$, let the S-dimensional vector $\boldsymbol{p}_k(t) = \left(p_k^{(1)}(t), ..., p_k^{(S)}(t)\right)$ denote the power allocation (PA) vector of FBS $k \in \mathcal{K}$ at time t. Here $p_k^{(s)}(t)$ is the transmit power of femtocell k over frequency band s at time t. All FBSs are assumed to transmit over one frequency band only at each time t at a given power level not exceeding $p_{k,\max}$. Let $L_k \in \mathbb{N}$ be the number of discrete power levels of FBS k and denote by $\boldsymbol{q}_k^{(\ell,s)}$ its ℓ-th transmit power level when used over channel s, with $(\ell, s) \in \mathcal{L}_k \times \mathcal{S}$, with $\mathcal{L}_k = \{1, \ldots, L_K\}$. Denote also by $\boldsymbol{q}_k^{(0,0)}$, with $k \in \mathcal{K}$, the S-dimensional null vector, i.e., $\boldsymbol{q}^{(0,0)} = (0, \ldots, 0) \in \mathbb{R}^S$. Thus, FBS k has $N_k = L_k \cdot S + 1$ possible PA vectors and for all $t \in \mathbb{N}$, $\boldsymbol{p}_k(t) \in \mathcal{A}_k$, where

$$\mathcal{A}_k = \boldsymbol{q}^{(0,0)} \cup \left\{ \boldsymbol{q}_k^{(\ell,s)} : (\ell, s) \in \mathcal{L}_k \times \mathcal{S} \right\}. \tag{1}$$

The SINR of MUE on carrier s is:

$$\gamma_0^{(s)} = \frac{|h_{0,0}^{(s)}|^2 p_0^{(s)}}{{\sigma_0^{(s)}}^2 + \underbrace{\sum_{k \in \mathcal{K}} |h_{0,k}^{(s)}|^2 p_k^{(s)}}_{\text{femtocells}}}, \tag{2}$$

The SINR for FBS $k \in \mathcal{K}$ serving its femto-user FUE is given as follows:

$$\gamma_k^{(s)} = \frac{|h_{k,k}^{(s)}|^2 p_k^{(s)}}{{\sigma_k^{(s)}}^2 + \underbrace{|h_{k,0}^{(s)}|^2 p_0^{(s)}}_{\text{macrocell}} + \underbrace{\sum_{j \in \mathcal{K} \setminus \{k\}} |h_{k,j}^{(s)}|^2 p_j^{(s)}}_{\text{femtocells}}}. \tag{3}$$

Finally, all FBSs are interested in optimizing a given interference mitigation metric denoted by $\phi_k : \mathbb{R}^{S \cdot K} \to \mathbb{R}^+$, which determines at each instant t the impact of the interference on the macro system based on the observation of all the SINR levels $\gamma_k^{(s)}$ and $\gamma_0^{(s)}$, with $(k, s) \in \mathcal{K} \times \mathcal{S}$. The interference mitigation metric considered in this work is:

$$\phi_k\left(\boldsymbol{p}_k, \boldsymbol{p}_{-k}\right) = \sum_{s=1}^{S} \log_2\left(1 + \gamma_k^{(s)}\right) . \mathbb{1}_{\left\{\gamma_0^{(s)} > \Gamma_0^{(s)}\right\}}. \tag{4}$$

This metric at a given instant t is different from zero only if the macrocell satisfies at time t the minimum SINR level at least over one of the available channels. Hence, as long as the macrocell tier sees its QoS requirement satisfied, femtocells obtain a positive reward/payoff.

C. Game Theoretic Model

The cross-tier interference mitigation problem described in the previous section can be modeled by a normal-form game $\mathcal{G} = \left(\mathcal{K}, \{\mathcal{A}_k\}_{k \in \mathcal{K}}, \{\phi_k\}_{k \in \mathcal{K}}\right)$. Here, $\mathcal{K}$ represents the set of FBSs in the network and for all $k \in \mathcal{K}$, the set of actions of FBS k is the set of power allocation vectors $\mathcal{A}_k$ described in (1). We denote by $\mathcal{A} = \mathcal{A}_1 \times ... \times \mathcal{A}_K$ the space of actions, and $\phi_k : \mathcal{A}_k \to \mathbb{R}_+$ is the payoff function of femtocell k.

At each time t and for all $k \in \mathcal{K}$, FBS k chooses its action from the finite set $\mathcal{A}_k$ following a probability distribution $\boldsymbol{\pi}_k(t) = \left(\pi_{k,\boldsymbol{q}_k^{(0,0)}}(t), \pi_{k,\boldsymbol{q}_k^{(1,1)}}(t), ..., \pi_{k,\boldsymbol{q}_k^{(L_k,S_k)}}(t)\right)$ where

$\pi_{k,\boldsymbol{q}_k^{(l_k,s_k)}}$ is the probability that femtocell k plays action $\boldsymbol{q}_k^{(l_k,s_k)}$ at time t, i.e.,

$$\pi_{k,\boldsymbol{q}_k^{(l_k,s_k)}} = \Pr\left(\boldsymbol{p}_k(t) = \boldsymbol{q}_k^{(l_k,s_k)}\right). \quad (5)$$

where $(l_k, s_k) \in \{1, ..., L_K\} \times \mathcal{S} \cup \{(0,0)\}$.

In what follows, we give a definition of the equilibrium concept to be used in the sequel of this paper.

D. ϵ-Coarse Correlated Equilibrium (ϵ-CCE)

Since the action set $\mathcal{A}$ is discrete and finite, the game $\mathcal{G}$ admits at least one equilibrium in mixed strategies. In what follows, we formally define the ϵ-coarse correlated equilibrium (ϵ-CCE):

*Definition 1 (**ϵ-Coarse Correlated Equilibria**):* The probability distribution $\boldsymbol{\pi} \in \Delta(\mathcal{A})$ is a ϵ-coarse correlated equilibrium if $\forall k \in \mathcal{K}$ and $\forall a'_k \in \mathcal{A}_k$,

$$\sum_{\boldsymbol{a}_{-k} \in \mathcal{A}_{-k}} \phi_k(a'_k, \boldsymbol{a}_{-k}) \pi_{-k,\boldsymbol{a}_{-k}} - \sum_{\boldsymbol{a} \in \mathcal{A}} \phi_k(a_k, \boldsymbol{a}_{-k})\, \pi_a \leqslant \epsilon, \quad (6)$$

where $\pi_{-k,\,\boldsymbol{a}_{-k}} = \sum_{a_k \in \mathcal{A}_k} \pi(\boldsymbol{a}_k, \boldsymbol{a}_{-k})$ is the marginal probability distribution w.r.t. a_k.

Note that, by letting $\epsilon = 0$, the classical definition of coarse correlated equilibrium is obtained. An CCE is a probability distribution over the set of action profiles of the game from which no player has incentives to deviate. When a player observes the values of other players' actions, the classical regret matching procedure exhibits convergence which is described in the following theorem.

*Theorem 1 (**Hart and Mas-Colell** [5]):* If every player plays according to the adaptive procedure of regret matching, then the empirical distribution of play converges *almost surely* as time goes infinity to the set of correlated equilibrium distributions of the game $\mathcal{G}$.

Remark 1: It is worth noting that correlated equilibria which are a generalization of Nash equilibria are more relevant within the context of decentralized and dense networks (such as femtocells), as it induces players to coordinate their actions, and hence reach better overall performance than the Nash approach (with no coordination). The following section describes how femtocells learn these equilibria in a totally decentralized manner.

III. Regret-Based Learning Procedure for Cross-Tier Interference Mitigation

Let us assume that a given FBS $k \in \mathcal{K}$ compares the time-average of its utility observations $\tilde{\phi}_k(n)$ obtained by constantly changing its actions following a particular strategy $\boldsymbol{\pi}_k$, with the case where it would have played the same action in all previous stages of the game, while other players use their current strategies $\boldsymbol{\pi}_{-k}$. Our behavioral assumption is that all femtocells are interested in choosing the probability distribution $\boldsymbol{\pi}^* \in \triangle(\mathcal{A})$ that minimizes the regret, where the regret of player k for not having played action $q_k^{(\ell_k,s_k)}$ from $n = 1$ up to time t is calculated as follows:

$$r_{k,q_k^{(\ell,s)}}(t) = \frac{1}{t}\sum_{n=1}^{t} \phi\Big(q_k^{(\ell,s)}, \boldsymbol{p}_{-k}(n)\Big) - \tilde{\phi}_k(n), \quad (7)$$

Note that if $r_{k,q_k^{(l_k,s_k)}}(t) > 0$, player $k \in \mathcal{K}$ would have obtained a higher average utility by playing action $q_k^{(\ell_k,s_k)}$ during all the previous stages. Thus, player k regrets for not having done it. On the contrary, if $r_{k,q_k^{(l_k,s_k)}}(t) \leqslant 0$, player k does not regret at all. Hence, given a vector of regrets up to time t, where $\boldsymbol{r}_k(t) = \Big(r_{k,q_k^{(0,0)}}(t), \ldots, r_{k,q_k^{(L_k,S_k)}}(t)\Big)$, player k would be inclined towards taking actions with the highest regret, that is:

$$\pi_{k,\boldsymbol{q}_k^{(l_k,s_k)}}(t) = \frac{\max\Big(0, r_{k,q_k^{(l_k,s_k)}}(t)\Big)}{\sum_{(l,s)\in\mathcal{A}_k} \max\Big(0, r_{k,q_k^{(l,s)}}(t)\Big)}. \quad (8)$$

Remark 2: At each time $t > 0$, FBS k calculates its regret vector $\boldsymbol{r}_k(t) = \Big(r_{k,q_k^{(0,0)}}(t), \ldots, r_{k,q_k^{(L_k,S_k)}}(t)\Big)$. This hinges on the fact that: (1) each FBS knows the explicit expression of its own utility function $\phi_k(.,.)$, and (2) each FBS observes the actions of all the other players at each time t, $\boldsymbol{p}_{-k}(t)$. Clearly, these assumptions are unrealistic in practice due to the distributed nature of femtocell networks, and thus techniques for relaxing the information conditions are paramount. The following section shows how to deal with such a case. Surprisingly as will be shown, one can design variants of the regret matching procedure which requires no knowledge whatsoever of other players' actions, and yet yields similar performance.

IV. Learning the ϵ-Coarse Correlated Equilibrium (ϵ-CCE)

As previously noted, the classical regret matching learning approach is unsuitable for solving the cross-tier interference mitigation problem due to the amount of required information at every FBS k. Here, we describe our novel distributed learning algorithm which significantly relaxes this assumption. To do that, femtocells face a trade-off between minimizing their regret and estimating their achieved time-average utility by playing a particular action at each time t. Hence, a suitable *behavioral rule* for each femtocell would be choosing the actions which yield high regrets more likely than those yielding lower regrets, but in any case always letting a *non-zero* probability of playing any of the actions. Formally speaking, the behavioral rule described above can be modeled by the probability distribution $\beta_k(\boldsymbol{r}_k^+(t))$ satisfying:

$$\beta_k(\boldsymbol{r}_k^+(t)) \in \arg\max_{\boldsymbol{\pi}_k \in \triangle(\mathcal{A}_k)} \Big[\sum_{\boldsymbol{p}_k \in \mathcal{A}_k} \pi_{k,\boldsymbol{p}_k} r_{k,\boldsymbol{p}_k}(t) + \kappa_k H(\boldsymbol{\pi}_k)\Big], \quad (9)$$

where, we denote by $\boldsymbol{r}_k^+(t)$ the vector of positive regrets: $\boldsymbol{r}_k^+(t) = \max(0, \boldsymbol{r}_k(t))$.

The *temperature* parameter $\kappa_k > 0$ represents the interest of FBS k to choose other actions rather than those maximizing the regret in order to improve the estimations of the vectors of regrets (7). The unique solution to the right hand side of the continuous and strictly concave optimization problem in (9) is written as:

$$\beta_k(\boldsymbol{r}_k^+(t)) = \left(\beta_{k,q_k^{(0,0)}}(\boldsymbol{r}_k^+(t)), \beta_{k,q_k^{(1,1)}}(\boldsymbol{r}_k^+(t)), ..., \beta_{k,q_k^{(L_k,A_k)}}(\boldsymbol{r}_k^+(t))\right) \tag{10}$$

where for all $k \in \mathcal{K}$ and for all $(l_k, s_k) \in \mathcal{L}_k \times \mathcal{S}$:

$$\beta_{k,q_k^{(l_k,s_k)}}(\boldsymbol{r}_k^+(t)) = \frac{\exp\left(\kappa_k r^+_{k,q_k^{(l_k,s_k)}}(t)\right)}{\sum\limits_{\boldsymbol{p}_k \in \mathcal{A}_k} \exp\left(\kappa_k r^+_{k,\boldsymbol{p}_k}(t)\right)}, \tag{11}$$

where $\beta_{k,q_k^{(l_k,s_k)}}(\boldsymbol{r}_k^+(t)) > 0$ holds with strict inequality regardless of the regret vector $\boldsymbol{r}_k(t)$. In what follows, the distributed regret based learning algorithm for cross-tier interference mitigation between macro and femtocell tier is formally described.

A. Distributed No-Regret Learning Algorithm

The distributed learning procedure carried out independently by every FBS $k \in \mathcal{K}$ is described below where $\forall k \in \mathcal{K}$ and $\forall (\ell_k, s_k) \in \{1, \ldots, L_k\} \times \mathcal{S} \cup \{(0,0)\}$:

$$\begin{cases} \hat{\phi}_{k,\boldsymbol{q}_k^{(\ell_k,s_k)}}(t) = \hat{\phi}_{k,\boldsymbol{q}_k^{(\ell_k,s_k)}}(t-1) + \\ \qquad \nu_k(t)\mathbb{1}_{\left\{\boldsymbol{p}_k(t)=\boldsymbol{q}_k^{(\ell_k,s_k)}\right\}}\left(\tilde{\phi}_k(t) - \hat{\phi}_{k,\boldsymbol{q}_k^{(\ell_k,s_k)}}(t-1)\right) \\ r_{k,q_k^{(l_k,s_k)}}(t) = r_{k,q_k^{(l_k,s_k)}}(t-1) + \lambda_k(t)\Big(\hat{\phi}_{k,\boldsymbol{q}_k^{(\ell_k,s_k)}}(t-1) - \\ \qquad r_{k,q_k^{(l_k,s_k)}}(t-1) - \tilde{\phi}_k(t)\Big), \\ \pi_{k,q_k^{(l_k,s_k)}}(t) = \pi_{k,q_k^{(l_k,s_k)}}(t-1) + \\ \qquad \alpha_k(t)\left(\beta_{k,q_k^{(l_k,s_k)}}(\boldsymbol{r}_k(t)) - \pi_{k,q_k^{(l_k,s_k)}}(t-1)\right) \\ \kappa_k(t) = \kappa_k(t-1) + \epsilon_k(t).\Psi(t) \end{cases} \tag{12}$$

where $\pi_k(0)$ and $r_k(0)$ are arbitrary initial actions and regrets, Ψ is a non-decreasing function, ν_k, α_k, λ_k and ϵ_k are learning rates[1] chosen such that:

$$\lim_{T\to\infty} \sum_{t=0}^{T} \alpha(t) + \lambda(t) = +\infty \tag{13}$$

$$\lim_{T\to\infty} \sum_{t=0}^{T} \alpha(t)^2 + \lambda(t)^2 < +\infty, \tag{14}$$

$$\lim_{t\to\infty} \frac{\lambda(t)}{\nu(t)} = 0, \tag{15}$$

$$\lim_{t\to\infty} \frac{\alpha(t)}{\lambda(t)} = 0 \tag{16}$$

$$\lim_{t\to\infty} \frac{\epsilon(t)}{\alpha(t)} = 0 \tag{17}$$

Remark 3: We would like to stress that in contrast to the classical regret matching approach where each player k perfectly knows its own utility function $\phi(q_k^{(l,s)}, \boldsymbol{p}_{-k}(t))$ as well as other players' chosen strategies, our proposed algorithm is totally decentralized and is composed of three phases. First, using its own instantaneous observed utility, each player is able to estimate its expected utility with each of its actions. Second, the estimated utility function allows players to compute their regrets of playing a given action. Third, players update the probability distribution of their transmission strategies. Moreover, conditions (13)-(17) are necessary to guarantee that the regret processes $\boldsymbol{r}_k(t)$ always see the process $\hat{\phi}_k(t)$ as fast processes always calibrated to the current value of the regret, and likewise, for the strategy distribution processes $\boldsymbol{\pi}_k(t)$ which sees the regret processes as fast processes always calibrated to the current values, as per (15). Finally, it is important to note that the temperature parameter κ_k is time-dependent in order to account for the fact in the beginning of the learning process femtocells may decide to explore all their actions (small κ) to favorite the estimation of the expected utility and corresponding regrets. Then as κ_k increases, the behavioral rule approaches the classical regret learning. That is, the ϵ (of the ϵ-CCE) is made smaller as time increases which ensures the asymptotic convergence to CCE. A formal proof of the convergence of this algorithm relies on stochastic approximation theory using the notions of multiple time scale and is not included here for space limitation.

V. Simulation Results

Let us consider one macrocell with radius $R_m = 500$ underlaid with K femtocells each of radius $R_f = 20$ m, transmitting over an arbitrary number of carriers S, with L transmit power levels. The minimum SINR of the macrocell UEs is given by $\boldsymbol{\Gamma_0} = \left(\Gamma_0^{(1)}, \ldots, \Gamma_0^{(S)}\right)$ where $\Gamma_0^{(1)} = \cdots = \Gamma_0^{(N)} = 3$ dB is assumed. The transmission power of the macro BS is set to 43 dBm, whereas the FBS adjusts its power through the various learning schemes to a value of maximum 10 dBm. The channel is represented as a combination of path-loss fading and log-normal shadowing with standard deviation of 8 and 4 dBm for outdoor and indoor communications, as per [8].

In Figure 2, we plot the average femtocell sum-rate for $K = 2$ FBSs underlaying one macrocell over $s = 2$ sub-

[1]In this work, femtocells are assumed to have similar learning rates, i.e., $\forall k \in \mathcal{K}, \lambda_k = \lambda, \alpha_k = \alpha$. Different learning rates across players is left for future work.

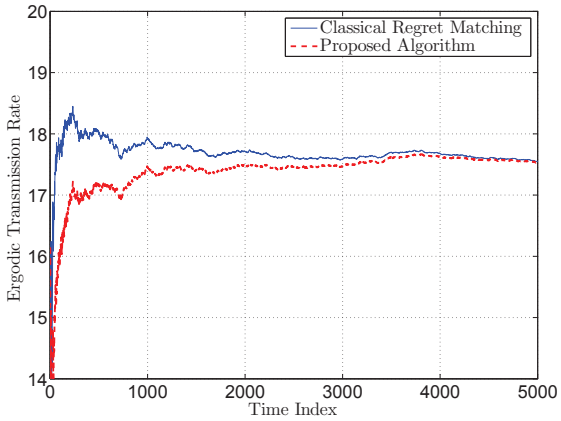

Fig. 2. Convergence of the proposed learning algorithm as well as the classical regret matching in terms of Ergodic transmission rate, for $K = 2$ femtocells, $S = 2$ channels, and $L = 2$ power levels.

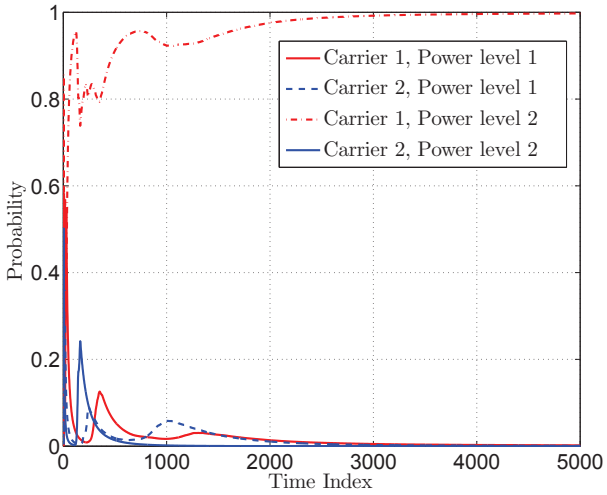

Fig. 3. Probability distribution over the set of actions of femtocell 1.

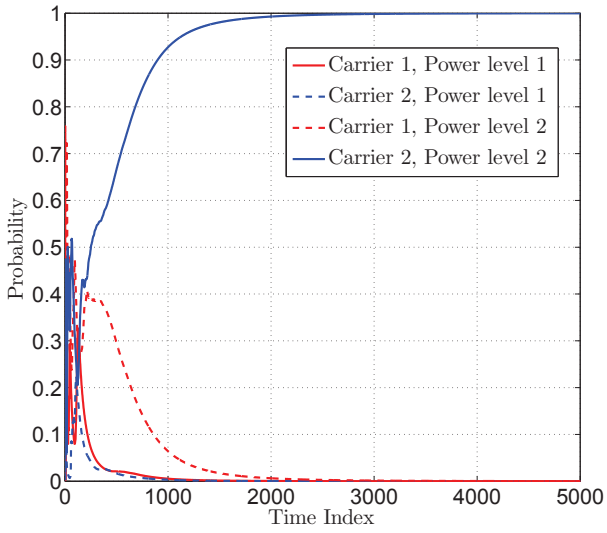

Fig. 4. Probability distribution over the set of actions of femtocell 2.

carriers, highlighting the convergence behavior of our proposed algorithm as well as the benchmark case (classical regret matching). We observe that surprisingly the proposed learning approach yields the same performance as the RM in the long-term. This is quite significant given the fact that unlike the RM approach, our algorithm is totally decentralized and does not require any information about other players' strategies.

Figures 3-4 plot the convergence of the probability distribution over the set of actions taken by both FBSs. As time goes by FBS $k = 1$ increases the probability to transmit with the maximum power level on frequency band $s = 1$, while the probability of transmitting on the other bands decreases. On the other hand, the probability that FBS $k = 2$ transmits with maximum transmit power level on carrier $s = 1$ decreases with time, whereas the probabilities of transmitting over carrier $s = 2$ with maximum transmit level increases. It can also be seen that although femtocells do not communicate with each other, they *implicitly* coordinate their transmissions by using different frequency bands with very high probability.

VI. Conclusions

In this paper, the problem of cross-tier interference mitigation was studied from a game theoretic *learning* perspective. Owing to their implicit coordination, a *regret-based* learning algorithm was proposed where femtocells jointly estimate their own utility function and learn their transmission strategies in a purely decentralized manner, relying only on local information. The considered behavioral rule was shown to converge to an ϵ-CCE where femtocells optimize their exploration and exploitation tradeoff. Remarkably, it was shown that the proposed algorithm achieves the same performance of the classical regret matching approach with much less overhead.

References

[1] V. Chandrasekhar and J. G. Andrews, "Femtocell networks: A survey," IEEE Commun. Magaz., 46(9): 59-67, September 2008.

[2] D. Lopez-Perez, G. d. l. Roche, A. Valcarce, A. Juttner, and J. Zhang, "Interference avoidance and dynamic frequency planning for WiMAX femtocells network," ICCS 2008, pp.1579-1584, 19-21 Nov. 2008.

[3] D. Choi, P. Monajemi, S. Kang, and J. Villasenor, "Dealing with loud neighbors: the benefits and tradeoffs of adaptive femtocell access," in Proc. IEEE GLOBECOM, 2008, p. 15.

[4] M. Bennis, S. Guruacharya, and D. Niyato "Distributed learning strategies for interference mitigation in femtocell networks," IEEE GLOBECOM, Houston, USA, 5-9 Dec. 2011.

[5] S. Hart and A. Mas-Colell, "A simple adaptive procedure leading to correlated equilibrium," Econometrica, vol. 68, no. 5, pp. 1127-1150, Sep. 2000.

[6] M. Bennis and S. M. Perlaza, "Decentralized Cross-Tier Interference Mitigation in Cognitive Femtocell Networks," in Proc. IEEE ICC, pp.1-5, 5-9 June 2011.

[7] J. F. Nash, "Equilibrium points in n-person games," National Academy of Sciences of the United States of America, 1950.

[8] 3GPP TR 25.820, "3G Home NodeB Study Item Technical Report (Release 8)," March 2008.