This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE Globecom 2011 proceedings.

Distributed Learning Strategies for Interference Mitigation in Femtocell Networks

M. Bennis†, S. Guruacharya‡, and D. Niyato‡
† Centre for Wireless Communications, University of Oulu
‡ School of Computer Engineering, Nanyang Technological University, Singapore
E-mail: bennis@ee.oulu.fi, {suda0002,dniyato}@ntu.edu.sg

***Abstract*—In this paper, the *strategic* coexistence between macro and femtocell tiers is studied using tools from evolutionary game theory and reinforcement learning. In the first case, femto base stations (FBSs) exchange information through a central controller, and adapt their strategies based on their instantaneous payoffs and average payoffs of the femtocell population. A fictitious play formulation is also examined where FBSs maximize their payoffs given the empirical frequency of other femtocells' actions. In the second case, when information exchange among femtocells is no longer possible, each femtocell gradually learns by interacting with its local environment through trials-and-errors, and adapt its strategies. Variant of the evolutionary game approach (referred to as replication by imitation) is also investigated where femtocells probabilistically review their strategies and imitate other femtocells in the network. Finally, the overall performance of the network in terms of spectral efficiency and convergence is shown to be adamantly driven by the type of information available at femtocells.**

I. Introduction

Recently, a new type of indoor Base Station (BS), called femtocell, has gained the attention of the industry [1], [2] due to the enormous benefits that it brings to both end-users and network operators. For instance, end-users can enjoy better signal qualities due to the reduced distance between the transmitter and the receiver, resulting in higher throughputs, power and battery savings. From the operator's point of view, femtocells will extend the indoor coverage, enhance system capacity, and share the spectrum in a more efficient manner. However, these benefits are hard to reap where a gamut of technical challenges such as cross-tier and co-tier interference need to be tackled. Therefore, fully distributed, scalable and efficient strategies need to be designed to make the deployment of femtocell networks feasible. Many results exist along this direction, e.g., see [5], [10] among others where a Q-learning based algorithm was investigated in the context of network selection for heterogeneous wireless networks, and channel selection in multiuser cognitive radios, respectively. Recently, in [6], a reinforcement learning framework based on Q-learning was studied for interference mitigation among femtocells. In [8],

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.

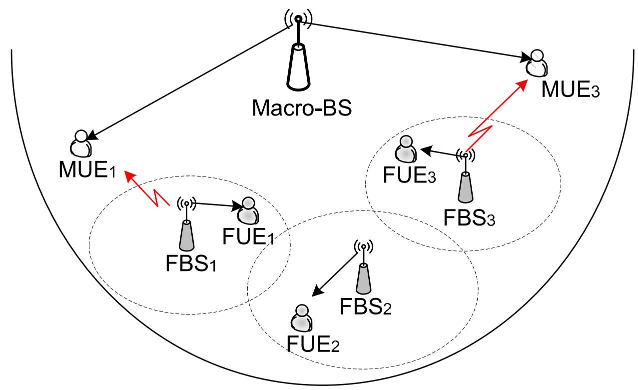

Fig. 1. Network topology with one macrocell underlaid with three femtocell networks. MUE and FUE stand for macro/femtocell user equipment, respectively. MBS and FBS stand for macro and femtocell base station, respectively.

the macro and femtocell coexistence was modeled as a Stackelberg game where macrocell base stations are assumed to have enough information to predict the response of the femtocells for given macrocell power profile. Nevertheless, the existing works often require information exchange among transmitters, which represents a non-affordable increment of signaling messages, and falls short of examining the non-trivial tradeoff between information exchange and performance improvements.

In this paper, we analyze the strategic coexistence between macrocell and femtocell tiers using tools from machine learning and evolutionary game theory, under different information exchange scenarios. Fully distributed algorithms are devised relying on local information aiming at mitigating interference towards the macrocell tier. A comparison in terms of spectral efficiency, convergence and inherent tradeoffs is provided for the different algorithms.

This paper is organized as follows. In Section II, the system model is presented. Section III deals with the evolutionary based approach with both replication dynamics and by imitation, and fictitious play formulation. Section IV studies the reinforcement learning formulation. Section V presents numerical results, and finally conclusions are drawn in Section VI.

II. System Model

Let us assume that there exists $M = 1$ macrocell base station (MBS) operating over a set $\mathcal{N} = \{1, \ldots, N\}$ of N frequency bands. Let $\mathbf{\Gamma_0} = (\Gamma_0^{(1)}, \ldots, \Gamma_0^{(N)})$ denote the minimum time-average signal-to-interference-plus-noise ratio (SINR) offered by the MBS to its macrocell user equipment (MUE) 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 time-average SINR than the minimum required by the MUE. At each time, each FBS serves one FUE over one or a subset of the available channels following a time division multiple access (TDMA) policy.

Let $t \in \{1, \ldots, \infty\}$ be a discrete time index and let a MUE denote the scheduled macro-user connected to its serving MBS. Designate MBS's transmit power on a given sub-carrier to be $p_0^{(n)}$. Let $|h_{0,0}^{(n)}|^2$ denote the channel gain between the MBS and its associated MUE in sub-carrier $n \in \mathcal{N}$. Likewise, $|h_{i,j}^{(n)}|^2$ denotes the channel gain between transmitter i and receiver j on sub-carrier n. Moreover, let σ_n^2 be the variance of Additive White Gaussian Noise at MUE, which is assumed to be constant over all sub-carriers for simplicity. The transmit power of FBS $k \in \mathcal{K}$ on sub-carrier n is denoted by $p_k^{(n)}$.

The signal to interference plus noise ratio at the MUE (assuming Gaussian signalling) is:

$$\gamma_0^{(n)} = \frac{|h_{0,0}^{(n)}|^2 p_0^{(n)}}{\sigma^2 + \underbrace{\sum_{k\in\mathcal{K}} |h_{k,0}^{(n)}|^2 p_k^{(n)}}_{\text{femtocells}}}, \tag{1}$$

Likewise, the SINR at the FUE served by FBS $k \in \mathcal{K}$ is:

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

III. Macro-Femtocell Coexistence With Information Exchange Among Femtocells

In this section, focus is on cross-tier interference mitigation algorithms with information exchange among femtocells.

A. Evolutionary game: replication dynamics

This interference mitigation mechanism is based on the concept of evolutionary game theory (EGT) [4], [5], where each FBS chooses its strategy against other FBSs within the same network (referred to as population). Femtocells observe the behavior of other interfering femtocells, learn from these observations, and thus make the best decision based on their instantaneous payoff, as well as the *average* payoff of all other femtocells. Let us denote by $\mathcal{G}^{ev} = \left(\mathcal{K}, \{\mathcal{A}_k\}_{k\in\mathcal{K}}, \{u_k\}_{k\in\mathcal{K}}\right)$ the EGT model. Here, $\mathcal{K}$ represents the set of deployed femtocells in the network. For all $k \in \mathcal{K}$, the set of actions of FBS k is the set of power allocation vectors, $\mathcal{A}_k = \{q_k^{(l,n)} : l \in \{0, \ldots, L_k\}$, and $n \in \mathcal{N}\}$, where $\mathcal{L}_k = \{1, \ldots, L_k\}$ and $L_k \in \mathbb{N}$ is the number of discrete power levels of FBS k, i.e., $\frac{p_{k,max}}{L_k}, \ldots, p_{k,max}$ and $l \in \mathcal{L}_k$ is the transmit power level. The power allocation vector when FBS k transmits over sub-carrier n with power level l is:

$$q_k^{(l,n)} = \frac{l}{L} p_{k,max}, \tag{3}$$

and, hence, FBS k has $N'_k = (L_k + 1)N$ possible power allocation vectors $q_k^{(0,0)}, q_k^{(1,1)}, \ldots, q_k^{(L_k,N)}$. Finally, $u_k : \mathcal{A} \to \mathbb{R}$ is the payoff or interference mitigation metric of femtocell $k \in \mathcal{K}$ and the *average* payoff of the entire femtocell population at time t is defined as $\bar{u}(t) = \frac{\sum_{k\in\mathcal{K}} u_k(t)}{K}$.

In this scenario, an entity which is referred to as HeNB-gateway [9], collects the payoffs for all femtocells and calculates the average payoff of the entire femtocell network. The payoff $u_k(t)$ of FBS k is then compared with the average payoffs $\bar{u}(t)$ and in the case when it is less than the average payoff of the femtocell network, a random strategy is chosen and the whole process is repeated again. Let $\zeta_a^{(l,n)}(t) = \sum_{k=1}^{K} \mathbf{1}_{\left\{p_k(t)=q_k^{(l,n)}=a\right\}}$ represent the total number of femtocells choosing strategy $q_k^{(l,n)}$, and $x_a(t) = \frac{\zeta_a^{(l,n)}(t)}{\sum_a \zeta_a^{(l,n)}(t)}$ is the proportion of femtocells using strategy $a \in \mathcal{A}$. Hence the replication dynamic equation can be defined as follows, $\forall (l,n) \in (\mathcal{L} \times \mathcal{N})$:

$$\dot{x}_a(t) = x_a(t) \left(u_a(t) - \bar{u}(t)\right), \tag{4}$$

where $u_a(t)$ is the payoff at time t of femtocell k when using action a, and $\bar{u}(t)$ is the corresponding average payoff of the entire femtocell population over all strategies $a \in \mathcal{A}$ where $\bar{u}(t) = \sum_{a\in\mathcal{A}} x_a u_a(t)$.

B. Evolutionary game: replication by imitation

Here, each player sticks to some pure strategy for some time period, and every now and then reviews her strategy, sometimes resulting in a change of strategy. First, a specification is defined for the time rate at which femtocells review their strategy depending on the current performance and other aspects of the current population state. Let $r_k^i(x)$ denote the average review rate of a FBS $k \in \mathcal{K}$ using strategy $i \in \mathcal{A}_k$. Second, the probability that FBS k in population x will change her strategy from i to j is denoted by $pr_i^j(x)$, which depends on the current performance of these strategies and other aspects of the current population state. Let $\boldsymbol{pr}_i(x) = \left(pr_i^{(1)}(x), \ldots, pr_i^{(N_k)}(x)\right)$ denote the probability distribution vector over the set of $\mathcal{A}_k$ of pure strategies; $pr_i^i(x)$ means that player k does not change her strategy i. Assuming that all players' Poisson processes are statistically independent, the aggregate of reviewing times in population i is itself a Poisson process with arrival rate $x_i r_i(x)$. Assuming that strategy switches are statistically independent random variables across players, the arrival rate of the aggregate Poisson process of switches from strategy i to strategy j is $x_i r_i(x) pr_i^j(x)$. Finally, the dynamic of the replication by imitation is given as follows:

$$\dot{x_i} = \sum_{j\in\mathcal{A}_k} (1 - x_i) r_i(x) pr_j^i(x) - \sum_{j\in\mathcal{A}_k} x_i pr_i^j . r_i(x), \tag{5}$$

where the first term in the right hand side of (5) regards the share of the femtocell population that randomly meets another femtocell and changes its strategy from j to i. The second term is the remaining population which changes its strategy i to other strategies. Note that in (5), there is no need to know the average payoff of the whole population as in the replication dynamics. Moreover, to guarantee that (5) induces a well-defined dynamics on the state space, it is assumed that $r_i : X \rightarrow R_+$ and $pr_i : X \rightarrow \Delta$ are Lipschitz continuous functions. Hence, by the Picard-Lindelof theorem, (5) has a unique solution through any initial state.

C. Fictitious play with complete information

The fictitious play (FP) formulation [4] can be defined as follows: Let us first assume that femtocells have *complete* and *perfect* information, i.e., they know the structure of the game and observe at each time t the power allocation vector taken by all other femtocells. Each FBS $k \in \mathcal{K}$ assumes that all of its counterparts play independent and stationary (*time-invariant*) mixed strategies π_j, $\forall j \in \mathcal{K} \setminus \{k\}$, where $\pi_j = \left(\pi_{j,q_j^{(1,1)}}, \ldots, \pi_{j,q_j^{(L_j,N)}}\right)$ and $\pi_{j,q_j^{(l,n)}} = Pr\left(p_j(t) = q_j^{(l,n)}\right)$. Under these conditions, femtocell k is able to build an empirical probability distribution over each set $\mathcal{A}_j$, for all $j \in \mathcal{K}\setminus\{k\}$. Let $f_{k,p_k(t)} = \frac{1}{t}\sum_{s=1}^{t} \mathbf{1}_{\left\{p_k(t)=q_k^{(l,n)}\right\}}$ be the empirical probability with which players $j \in \mathcal{K} \setminus \{k\}$ observe that player k plays action $q_k^{(l,n)} \in \mathcal{A}_k$. Hence, $\forall k \in \mathcal{K}$ and $\forall p_k \in \mathcal{A}_k$, the following recursive expression holds:

$$f_{k,p_k}(t+1) = f_{k,p_k}(t) + \frac{1}{t+1}\left(\mathbf{1}_{\left\{p_k(t)=q_k^{(l,n)}\right\}} - f_{k,p_k}(t)\right), \quad (6)$$

Let $\bar{f}_{k,p_{-k}}(t) = \prod_{j\neq k} f_{j,p_j}(t)$ be the probability with which player k observes the action profile $p_{-k} \in \mathcal{A}_{-k}$ at time t, for all $k \in \mathcal{K}$. Let the $|\mathcal{A}_{-k}|$ dimensional vector $\boldsymbol{f}_k(t) = (\bar{f}_{k,p_{-k}})_{\forall p_{-k}\in\mathcal{A}_{-k}}$ be the empirical probability distribution over the set $\mathcal{A}_{-k}$ observed by player k. In what follows, the vector $\boldsymbol{f}_k(t)$ represents the beliefs of player k over the strategies of all its corresponding counterparts. Hence, at each time t and based on its own beliefs $\boldsymbol{f}_k(t)$, each FBS k chooses its action $p_k(t) = q_k^{(l,n)}$, i.e.,

$$(l,n) \in \arg\max_{(l,n)\in(\mathcal{L}\times\mathcal{N})} \bar{u}_k\left(p_k(t), \boldsymbol{f}_k(t)\right). \quad (7)$$

where for all $k \in \mathcal{K}$, $\bar{u}_k(\pi) = \mathbb{E}_\pi\left[u_k(p_k, p_{-k})\right]$. From (6), it can be implied that by playing FP, players become myopic, by building beliefs of strategies used by all the other players, and at each time t, players choose the action that maximizes their instantaneous expected utility.

IV. Macro-Femtocell Coexistence With No Information Exchange

Unlike Section-III, information exchange among femtocells is no longer possible. In the following, we investigate two learning mechanisms adopted by femtocells to mitigate interference towards the macrocell tier.

A. Q-learning

The Q-learning model consists of a set of states S and actions A aiming at finding a policy that maximizes the observed rewards over the interaction time of the players (i.e., femtocells). Every FBS $k \in \mathcal{K}$ explores its environment, observes its current state s, and takes a subsequent action a, according to a decision policy $\pi : s \rightarrow a$. With their ability to learn, the knowledge about other players' strategies is not needed. Instead, a Q-function maintains the knowledge about other players in the network, based on which decisions can be made accordingly. Let us denote by $\mathcal{G}^Q = \left(\mathcal{K}, \{\mathcal{P}_k\}_{k\in\mathcal{K}}, \{u_k\}_{k\in\mathcal{K}}\right)$ the Q-learning game. Here, the players are the FBSs $k \in \mathcal{K}$, $\boldsymbol{s}_k(t) = \left(s_k^{(1)}(t), \ldots, s_k^{(N)}(t)\right)$ is the composite state of FBS k at time t in sub-carrier $n \in \mathcal{N}$. $s_k^{(n)}(t) \in \{0,1\}$ indicates whether FBS k generates interference towards MUE or not, in carrier n where the QoS of MUE is violated. $\boldsymbol{a}_k(t) = \left(a_k^{(1)}(t), \ldots, a_k^{(N)}(t)\right)$ is the action vector of FBS k, where $a_k^{(n)}(t) \in \{0,1\}, \forall n \in \mathcal{N}$. Finally, $\mathbf{u}_k(t) = \left(u_k^{(1)}(t), \ldots, u_k^{(N)}(t)\right)$ is the utility function or payoff vector of FBS k at time-instant t. The *expected* discounted reward over an infinite horizon is given by:

$$V^\pi(s) = \mathbb{E}\left[\sum \gamma^t \times r(s_t, \pi^*(s_t)) | s_0 = s\right], \quad (8)$$

where $0 \leq \gamma \leq 1$ is a discount factor and r is the agent's reward at time t. Equation (8) can be rewritten as follows:

$$V^\pi(s) = R(s, \pi^*(s)) + \gamma \sum_{v\in S} P_{s,v}(\pi(s))V^\pi(v), \quad (9)$$

where $R(s,\pi^*(s)) = \mathbb{E}\{r(s,\pi(s))\}$ is the mean value of reward $r(s,\pi(s))$, and $P_{s,v}$ is the transition probability from state s to v. Moreover, the optimal policy π^* satisfies the optimality criterion:

$$V^*(s) = V^{\pi^*}(s) = \max_{a\in\mathcal{A}}\left(R(s,a) + \gamma \sum_{v\in\mathcal{S}} P_{s,v}(a)V^*(v)\right). \quad (10)$$

It is generally difficult to explicitly calculate the reward $R(s,a)$ and transition probability $P_{s,v}(a)$. However, through Q-learning, the knowledge of these values can be gradually learnt and reinforced with time. For a given policy π, define a Q-value as follows:

$$Q^*(s,a) = R(s,a) + \gamma \sum_{v\in S} P_{s,v}(a)V^\pi(v), \quad (11)$$

which is the expected discounted reward when executing action a at state s and then following policy π thereafter.

Here, we use the Q-learning algorithm to iteratively approximate the state-action value function $Q(s,a)$. The agent keeps trying all actions in all states with non-zero probability and must sometimes explore by choosing at each step a random action with probability $\epsilon \in (0,1)$, and the greedy action with probability $(1-\epsilon)$. This is referred to as ϵ-greedy exploration [3].

The Q-learning process aims at finding $Q(\mathbf{s},\mathbf{a})$ in a recursive manner where the update equation is given as [3]:

$$Q_t(\mathbf{s},\mathbf{a}) = (1-\alpha)Q_{t-1}(\mathbf{s},\mathbf{a}) + \alpha\left[r_t(\mathbf{s},\mathbf{a}) + \gamma V_{t-1}(v_t)\right] \quad (12)$$

where $V_{t-1}(v_t) = \max_{b \neq a} Q_{t-1}(v,b)$ and α is the learning rate.

B. Cooperative Q-learning

Unlike the classical case where femtocells build their Q-tables in a non-cooperative manner, femtocells have the possibility to adopt the learning-by-*expert* scheme, where cooperation between femtocells takes place based on temporal difference of their utility function $d_k(t) = u\left(p_k(t), p_{-k}(t)\right) - u\left(p_k(t-1), p_{-k}(t-1)\right)$ [12]. Thus if a femtocell i finds a femtocell j to have similar difference such that $|d_i(t) - d_j(t)| \leq \zeta$, then it considers FBS j as an expert to which an appropriate weight is assigned. As a result, a given femtocell modifies accordingly its Q-table by learning from a small group of femtocells considered as experts. Finally, this information exchange among femtocells takes place periodically.

V. Simulation Results

To evaluate the performance of the interference mitigation algorithms, we consider the case where one macrocell with radius $R_m = 500$ is underlaid with K femtocells each of radius $R_f = 20$ m, transmitting over $N = 8$ sub-carriers. We assume that femtocells have $L = 3$ transmit power levels. The minimum SINR of the MUEs is given by $\mathbf{\Gamma_0} = \left(\Gamma_0^{(1)}, \ldots, \Gamma_0^{(N)}\right)$. It is assumed that $\Gamma_0^{(1)} = \cdots = \Gamma_0^{(N)} = 3$ dB. The transmission power of the MBS is 43 dBm, whereas the FBS adjusts its power through the various learning schemes to a value of maximum 10 dBm. The channel is inline with those of 3GPP [9] composed of path-loss fading and log-normal shadowing with standard deviation of 8 and 4 dBm for outdoor and indoor communications, respectively. For learning through replication by imitation, we set the intensity of review (average review rate) to be $r = 0.01$, for all players. For Q-learning and cooperative Q-learning approaches, we set the discount factor $\gamma = 0.95$, exploration probability $\epsilon = 0.1$, and impressionability $\beta = 0.3$. For cooperative Q-learning, cooperation was set to occur after every 4 intervals of time.

The interference mitigation metric considered in this work is:

$$u\left(p_k(t), p_{-k}(t)\right) = \sum_{n=1}^{N} \log_2\left(1 + \gamma_k^{(n)}\right) . \mathbf{1}_{\left(\gamma_0^{(n)} > \Gamma_0^{(n)}\right)}. \quad (13)$$

This metric at a given time t is different from zero only if the macrocell satisfies at time t the minimum SINR level required for their own communications. Hence, as long as the macrocell sees its QoS requirement satisfied, femtocells obtain a positive reward. This models a certain altruism from the behavior of the FBSs which sacrifice their performance to guarantee the QoS of the macrocell system.

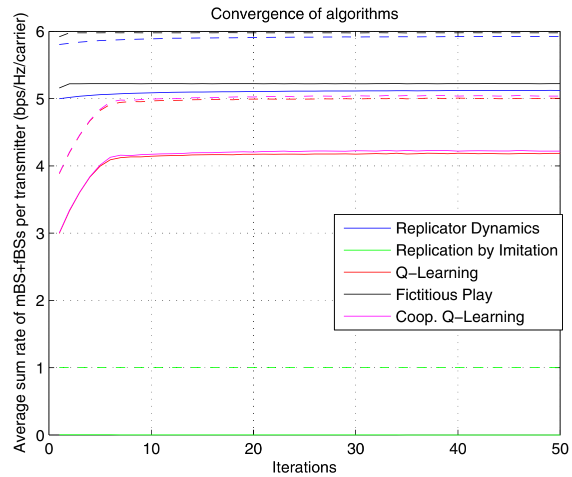

Fig. 2. Convergence of various learning algorithms and their impact on the average femtocell sum-rate and total average system sum-rate for $K = 50$ femtocells and $N = 8$ sub-carriers.

A. Total sum-rate of the network vs. time for all learning algorithms

In Figure 2, we plot the average femtocell sum-rate for $K = 50$ FBSs underlaying one macrocell over $N = 8$ sub-carriers (solid line), highlighting the convergence behavior for different learning algorithms. From this plot, we notice that replicator dynamics, fictitious play, Q-learning, and cooperative Q-learning schemes eventually converge to some steady state, whereas the replication by imitation converges to zero. This is attributed to the fact that the reviewing femtocell imitates a random femtocell based on the population distribution of femtocells over the strategy space, and not necessarily the femtocell with better payoff. Thus the reviewing femtocell has a higher probability of choosing the more "popular" strategy, instead of a performance enhancing strategy. This leads femtocells to congregate around few strategies, causing higher interference to the macrocell, yielding forced femtocell shutdowns.

In addition, Figure 2 depicts the total network sum-rate (dashed-line). We note that even though the performance of macrocell is degraded over time by the femtocells, the number of femtocells operating within the macrocell area tends to boost the overall network performance. However, we note that for replication by imitation case, since only the macrocell contributes towards the network performance, the overall system has a poorer performance compared to the other methods.

B. Effect of femtocell density on the performance of learning algorithms

Figures 3-4 plot the impact of the femtocell density on the average femtocell sum-rate and average total system sum-rate for different learning algorithms, for $K = \{50, 100, 150, 200, 250\}$ femtocells. A general decline in performance is seen as the number of femtocells increases. Nonetheless, we see that the rate of decrease in performance is not the same for all algorithms. In particular, for $K = 250$, the femtocell average sum-rate of around 5 bps/Hz is obtained using replicator dynamics

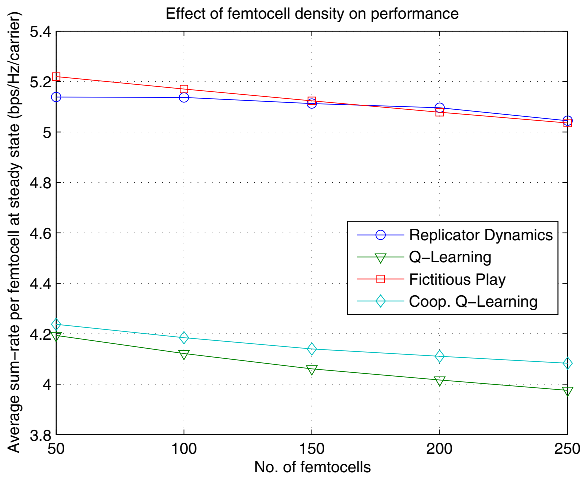

Fig. 3. Effect of femtocell density on the average femtocell sum-rate, for different learning algorithms.

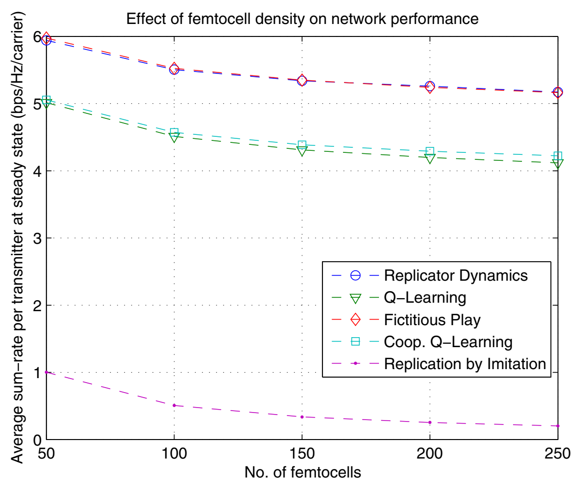

Fig. 4. Effect of femtocell density on the average network sum-rate, for different learning algorithms.

and fictitious play, whereas approximately 4 bps/Hz is obtained via Q-learning and cooperative Q-learning, respectively. This discrepancy is due to the fact that both Q-learning algorithms are essentially ON/OFF algorithms. Finally, it is worth noting that the cooperative Q-learning outperforms the traditional Q-learning algorithm.

C. Dedicated vs. Shared Spectrum Between Macro and Femtocell Tiers

Figure 5 plots the CDF of the average femtocell sum-rate for both dedicated and shared bands. It can be noted that the cooperative Q-learning approach outperforms the classical Q-learning approach and the fictitious play outperforms the replication dynamics approach. Interestingly, the dedicated case of FP and replication dynamics outperforms the shared case, whereas the opposite holds true for both classical and cooperative Q-learning schemes.

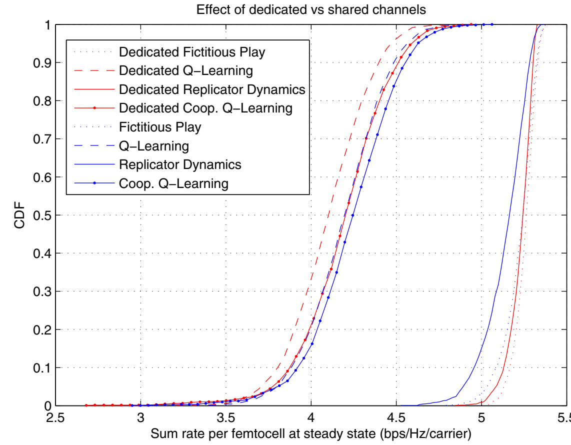

Fig. 5. CDF of average sum rates of $K = 50$ femtocells at steady state.

VI. Conclusions and Future Work

In this paper, distributed interference mitigation algorithms pertaining to different information exchange scenarios among femtocells have been examined for multi-tier networks. Better overall performance and faster convergence are achieved at the expense of more information exchange among femtocells. In future work, we will extend the learning framework for channel selection to the case where femtocells transmit over several sub-carriers.

References

[1] V. Chandrasekhar and J. G. Andrews, “Femtocell networks: A survey,” IEEE Communication Magazine, 46(9):5967, September 2008.

[2] H. Claussen, L. T. W. Ho, and L. G. Samuel, “An overview of the femtocell concept. Bell Labs Technical Journal,” Wiley, 3(1): 221-245, May 2008.

[3] M. E. Harmon and S. S. Harmon, “Reinforcement learning: A tutorial,” 2000.

[4] D. Fudenberg and D. K. Levine, “The Theory of Learning in Games,” The MIT Press, Cambridge, MA, 1998.

[5] D. Niyato and E. Hossain, “Dynamic of network selection in heterogeneous wireless networks: an evolutionary game approach,” IEEE Transactions on Vehicular Communications, vol. 58, no. 4, 2651-2660, May 2009.

[6] M. Bennis and D. Niyato, “A Q-learning based approach to interference avoidance in self-organized femtocell networks,” 1st IEEE International Workshop on Femtocell Networks (FEMnet) in conjunction with IEEE GLOBECOM 2010, Miami, FL, USA.

[7] M. Bennis and S. M. Perlaza, “Decentralized Cross-Tier Interference Mitigation in Cognitive Femtocell Networks,” IEEE International Conference on Communications (ICC), Kioto, Japan, June, 2011.

[8] S. Guruacharya, D. Niyato, E. Hossain, and D. I. Kim, “Hierarchical competition in femtocell-based cellular networks,” to be presented in IEEE GLOBECOM’10, Miami FL USA, 6-10 December 2010.

[9] 3GPP TR 25.820, “3G Home NodeB Study Item Technical Report (Release 8),” March 2008.

[10] H. Li, ”Multi-agent Q-Learning of Channel Selection in Multi-user Cognitive Radio Systems: A Two by Two Case,” IEEE Conference on System, Man and Cybernetics (SMC), 2009.

[11] M. N. Ahmadabadi, M. Asadpour, “Expertness based cooperative Q-learning,” Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on , vol.32, no.1, pp.66-76, Feb 2002.

[12] Angelia Nedic, Asuman Ozdaglar, “Distributed Subgradient Methods for Multi-Agent Optimization,” Automatic Control, IEEE Transactions on , vol. 54, no. 1, pp.48-61, Jan 2009