CO-TIER AND CROSS-TIER UPLINK INTERFERENCE MITIGATION USING Q-LEARNING

ABDULLAH ALQHTANI, YUHE ZHANG

Abstract. Efficient resource allocation in femtocell networks has become necessary owing to the enormous advantages of having femtocells deployed in a heterogeneous network. However, the interference arising from this deployment necessitate a mechanism for mitigation and optimal control of resource allocation. Q-learning, as an example of a reinforcement learning technique has been widely used for this purpose with more emphasis on downlink interference problems. Using a simplified heterogeneous model comprising of one macrocell and two femtocells, we extend the use of Q-learning to specifically model and address the uplink interference problem. We show by means of controlled experiments, that the proximity of the users in the network to their respective base stations, and the available power transmission levels plays a significant role in the total capacity of the network. This has the potential to enable networks to act in an improved manner. Results from the simulation can be used to configure any realistic network model.

Date: April 8, 2015.

Contents

1. Introduction

The need to enhance the capacity and high coverage of mobile users has caused network service providers to deploy femtocells while trying to mitigate the uplink and downlink interferences caused by their deployment [10]. The interference is more common when the existing macrocells and the deployed femtocells operate in a dedicated frequency band. This sort of deployment is common because of the advantages it offers to network service providers. However, this leads to a complicated type of interference. When these femtocells are set up by the end user at random positions away from the macrocells, the interference problem becomes complex which then requires an efficient interference mitigation strategy. In addition, as the number of femtocells in a network increases, managing the interference becomes more crucial in order to satisfy the quality of service of not just the macrousers, but also guaranteeing a certain measure of service for the femtousers.

1.1. **Literature Review.** While several interference mitigation techniques have been proposed to address the downlink interference problem in cellular networks, very few attempt have been focused on addressing the uplink interference problem where a femtocell base station (FBS) suffers from a co-tier interference with another femtocell user equipment (FUE) or from a cross-tier interference with a macrocell user equipment (MUE). In managing the interference, one common approach is to allow the agents—femtocells and macrocells—learn from the dynamic environment created by the coexistence of both cells. By learning from their environment, these devices can adjust their parameters such as power transmission level to satisfy the quality of service of their respective users. In literature, one common tool that has been used widely to achieve this learning is a reinforcement learning technique called Q-learning [2]. An advantage of Q-learning is that it does not need any prior information on the state of the environment. This means that we do not need to be concerned about the number of femtocells in the system or their spatial locations. Furthermore, the learning can be independently performed by each of the agents or it could be cooperative learning [25] where, for instance the femtocells share their information with each other. The actions taken by any agent in this setting affects the state or environment and also affects the learning process. One merit of Q-learning is that the agents can take actions while still learning from the actions of other agents. This learning leads to network acting in an improved manner [5]. This improved performance of the network and the easiness of the algorithm motivates us to adopt Q-learning in this work.

1.2. **Contribution.** To simplify the mathematical analysis of the use of Q-learning , our system model comprises two femtocells and one macrocell. For each of the femtocells, there is an associated femtouser and a macrouser associated with the macrocell. This simplified model of what an heterogeneous network looks like might impact the accuracy of the learning algorithm, however it is a proof of concept. We consider the uplink co-tier and cross tier interference in the network. By designing the power transmission levels in the network from which the users can select, a mathematical relation is used to estimate the signal to interference noise ratio (SINR) for each of the users and also compute the capacity of the macrocell and the femtocells in the system. Because the choice of the power transmission level affects the SINR, we investigate more than one range of power transmission levels. We also investigate the dependence on proximity between femtocells and their base stations relative to other femtocells and macrocells. Since most of the related work has focused on the use of Q-learning to address the downlink interference problem (e.g., [17], [25]), our main contribution in this work is to use Q-learning to address the uplink interference problem. This sets our work apart from those of [14] who does not consider the Q-learning method in their analysis of the uplink capacity in a similar network architecture.

1.3. **Report Structure.** We start by giving a brief overview of the network interference problem in Section 2. We briefly discuss some interference management approaches and methods, including examples of work that has been done using these approaches. We discuss and justify our adoption of Q-learning in line with related works. In Section 3, we set up our system model, describe the dynamics in the system, and the resulting interference problem. Finally, an implementation of Q-learning on our system model is shown with stylized experiments in Section 4. In Section 5, a summary of the report is given and future work is discussed.

2. Overview of Network Interference and Q-learning

There are different designs for femtocell/macrocell networks based on the mode of access and access control strategies. However, two common ones are the Code-Division Multiple Access (CDMA) based and the Orthogonal Frequency-Division Multiple Access (OFDMA) based networks. In our study, we shall be focusing on the OFDMA systems simply because it allows for more efficient utilization of frequency spectrum resources [10]. While femtocells have been very efficient for cellular network coverage, the deployment and positioning of femtocells gives rise to several technical challenges. One of the most significant challenges is the interference management between femtocells and macrocells. Unlike macro base stations (MBS), femto access points (femtocells) are usually deployed without network consideration. Therefore, the distribution of the femtocells causes interference among the cells and to/from other cells, including the macrocell. When there is an interference in a network, a mitigation or control technique is required in order to enhance network coverage and improve the signal to interference noise ratio at the user's equipment.

2.1. **Network Interference.** Depending on the cells interfering in the network, two main types of interference are observed. For example, in a two-tier femtocell network architecture,– a network with one macrocell and at least one femtocell–, we can have a co-tier interference or/and a cross-tier interference. Co-tier interference is the interaction between femtocells . Usually when there is a cluster of femtocells, there exist a co-tier interference between them (see e.g., [7],[29]). On the other hand, cross-tier interference describes the interaction between the macro cell or base station and a femtocell or /and a femto base station (see e.g., [8],[21]). The transmitter agent and the receivers determine the type of interference and transmission mode observed in a heterogeneous network. For example, if a femtocell user equipment is transmitting a signal and a femto base station (not associated with the femtocell user) is also active, the femto base station will suffer an uplink co-tier interference.

In general, there are three types of downlink interference and three types of uplink interference, with four cross-tier and two co-tier interference scenarios in the OFDMA-based femtocell network. The co-tier or co-channel interference problem is a traditional problem in wireless network. Although it has been investigated intensively in traditional cellular network, in heterogeneous network (HetNet), we face the same problem with new scenarios and challenges. Orthogonal scheduling is performed separately by the macro base stations and each femto base station (FAP), so the mobile users associated with the same FAP or MBS will surely not interfere with each other [18]. But due to high density of FAPs and FUEs, it is very likely that multiple FUEs, each associated with a different FAP, are using the same sub-channel in a nearby area, potentially causing a co-tier uplink interference [12]. Also in HetNet, FAP and MBS are using the same spectrum. Therefore, if a MUE is nearby in the same area of FAPs, co-tier and cross-tier interferences will happen together. This case is common in an office building scenario. The problem with this setting is that there might be a state of instability in the peer transmission levels and destructive interference in the network which affects the signal to noise transmission level of the users in the network. Hence, solving the power equilibrium problem in this kind of scenario is crucial. To address this problem, there are several approaches that have been proposed in the literature. This leads us to the subject of interference management.

2.2. **Interference Management.** Femto base stations can experience uplink interference in a heterogeneous network. The interference can be either from a nearby femto user equipment or a macro user equipment. In this case, the FBS is the victim. Similarly, the femto user equipment can cause uplink interference to nearby femto user equipments. The interference problem is a major one in networking paradigm and has drawn attention by many researchers. In [15], the authors presented an analytic study of the uplink interference and metrics that help characterize this type of interference, such as the outage probability and the signal to interference ratio. To mitigate the uplink interference at the FBS and limit the uplink interference created by the femto user equipment, several methods have been studied. For example, [13] presented an interference management solution for coexisting two-tier networks. In their work, they exploited the cognition and coordination between the tiers via the use of agile radios. Another work by [1] considers how to limit the interference by controlling the distance between the femtocell and other users in the network. A performance analysis of the uplink inter-carrier (cross-tier) interference was shown by [12]. In their work, the authors studied the timing misalignment of macro user signals in OFDMA network. They showed how the signal to interference ratio is degraded in the presence of a cross-tier interference. One common approach to manage interference and the resources used by agents in a network, is to study and control the power allocation for the system. Several methods have been applied in this regard. For instance, [11] used a game theory based method for modelling the power control and resource allocation problem in two-tier femtocell networks. The goal was also to maximize the energy efficiency of the femtocell users. They showed that their method can improve the utility of users significantly, compared with other power control and resource allocation methods. Furthermore, in [16], the authors described the advantage of a two-tier network by designing a multitiered wireless network which highlights increased stability in the system cost and showed a degradation in the system performance as more users are added to the network.

2.3. **Q-Learning.** Reinforcement learning (RL) is a machine learning (ML) algorithm where an agent learns from a dynamic environment through trial-and-error interactions. In Q-learning (QL), an example of RL, an agent can find the optimal decision policies through interaction with the environment. These interactions can be modelled as a Markov Decision Process (MDP) [3],[4]. The Q learning model can be defined as a tuple $(S, A, P_{s,s'}, R(s,a))$ where S $= (s_1, s_2, ..., s_n)$ is a finite set of environment states the agent can enter, A $= (a_1, a_2, ..., a_m)$ is a finite set of actions the agent may choose, $P_{s,s'}(a)$ is the state transition probability function from state s to the new state s' after taking action a, and $R(s,a)$ is the reward function that determines the reward for the agent when performing action a in state s. In a system of more than one agent, the agents may decide to share their actions with other agents or may not. The objective of each agent is to find an optimal policy π^*(s), which tells the agent the optimal action to choose in each state $s_t \in S$ in order to maximize the the total expected discounted reward. The expected discounted reward over infinite time given the policy π and initial state s can be written as:

$$Q(s,a) = E\left\{\sum_{t=0}^{\infty} \gamma^t R(s_t, \pi(s_t), s_{t+1}) | s_0 = s\right\} \tag{1}$$

where $\gamma \in [0, 1)$ is called the discount factor which determines the significance of future rewards. If $\gamma = 0$, the agent will ignore all future rewards. If $\gamma = 0.9$, the agent will put emphasis on future rewards. In QL, $Q(s, a)$ is called the Q value. When the agent selects action a according to the optimal policy $\pi^*(s)$, then the Q value (expected discounted reward) is maximized. The optimal Q value $Q^*(s, a)$ to find the optimal policy $\pi^*(s)$ is defined as:

$$Q^*(s,a) = E\{R(s,a)\} + \gamma \sum_{s' \in S} P_{s,s'}(a) max_{b \in A} Q^*(s', b) \tag{2}$$

where $Q^*(s', b)$ is the optimal Q value from the the next state s' after choosing the next action b. QL process finds $Q^*(s, a)$ in a recursive manner by using the following update rule:

$$Q(s,a) = Q(s,a) + \alpha[r + \gamma max_{b \in A} Q(s', b) - Q(s,a)] \tag{3}$$

where α is the learning rate which determine how much the new information will override the old information. It was proved that the update rule in Equation(3) converges to the optimal Q-value under certain conditions [2][6]. One of the conditions is that the agent has to try all the possible actions and visit all the possible states infinitely. To achieve this condition, the ϵ greedy exploration is introduced in the Q-learning algorithm.That is, in each iteration, the agent chooses a random action with probability $\epsilon \in$ (0,1) and chooses the greedy action that will maximize the Q-value with probability $1-\epsilon$.

Related works that have used Q-learning to address interference problems and resource allocation in HetNET include [17],[19],[20]. The main idea in these works is to model the femto network as a multi-agent system where the femto user equipments are the agents. The interaction between the agents and the surrounding environment leads to a learning of an optimal policy to solve the interference problem. The vast majority of the literature on Q-learning proposes algorithms that attempt to solve the power control problem, while showing improved performance in terms of convergence of the capacity, e.g. as in [19],[20]. The role of the choice of the reward function and how it can greatly affect the performance of the Q-learning algorithm was presented by [1]. The authors showed how the distance between the interacting agents and the total number of agents in the network can affect the capacity and the SINR. In addition, they affirmed the learning algorithm as an efficient tool for resource allocation and management in femtocell networks. Another significant work by [25] was aimed at solving the power control problem with Q-learning to manage the interference caused by the femtocells on the macrocells in the downlink. The authors showed that, using Q-learning, the femtocells can either learn independently or cooperatively, i.e. they share their ”Q-table”, in order to enhance their performance.

The number of related work described above and the opportunity to adopt the powerful Q-learning tool for resource allocation in a HetNET motivates us to explore the role of this algorithm to manage and control the uplink interference in a two-tier network. In the next section, we set up the system model which depicts the various agents and the associated interference patterns. This system model will lead to the problem formulation which enables us to design controlled experiments that will address the problems.

3. System Model and Implementation

3.1. **System Model.** The system model (Figure 1) consists of one MBS underlaid with 2 FAPs. The MUE is randomly located in the coverage area of the MBS, while the FUEs are randomly located in their associated FAPs. In our model, there is only one active FUE in each femtocell at every given time slot. Hence, we are only interested in the interference caused by the neighbouring active FUE and the neighbouring active MUE at the designated FAP.

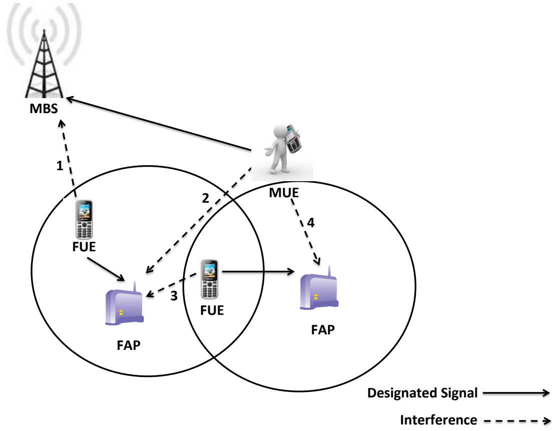

Figure 1. System model. The solid arrows are the designated communication links. The dashed arrows are the uplink interferences to the base stations. Link 1,2,4 are cross-tier interferences and link 3 is co-tier interference.

In this system, we consider the uplink co-tier and cross-tier interferences between femtocell and macrocell contrary to the downlink interference studied by [17]. To simplify our study, we assume all the MUE and FUEs are on the same sub-channel and have the same amount of available resource blocks (RB), which allows to increase the spectral efficiency per area through spatial frequency re-use [25][17]. In our model, the length of the RB is the length of the power transmission levels which we will revisit in the next section. The performance of our system is analyzed using the common parameter of signal interference noise ratio (SINR) and capacity (C). These parameters are as defined in related literature such as [1], [17], [20], [22], and [26].

The instantaneous SINR of FUE i associated with its designated FAP k is defined as:

$$\gamma_i^{k,f} = \frac{p_i^f g_i^{k,f}}{\sum_{j \in I_m} p_j^m g_j^m + \sum_{r \in I_f} p_r^f g_r^f + \sigma^2} \tag{4}$$

where p_i^f indicates the uplink transmission power of FUE i; $g_i^{k,f}$ indicates the channel gain or link gain between FUE i and its designated FAP k; I_m is the set of the MUEs that are interfering with the FUE i in the same sub-channel; I_f is the set of FUEs that are interfering with the FUE i in the same sub-channel. p_j^m and p_r^f indicate the transmission powers of MUE j and FUE r respectively that are on the same sub-channel with FUE i; g_j^m indicates the channel gain between MUE j and the FAP k; g_r^f indicates the channel gain between the FUE r and FAP k, and σ^2 is the channel noise power which denotes the variance of an additive white gaussian noise in our simulations. We assume this to be constant over all the sub-channels for simplicity.

Similarly, the SINR of MUE i, associated with its designated MBS k can be written as

$$\gamma_i^{k,m} = \frac{p_i^m g_i^{k,m}}{\sum_{j \in I_m} p_j^m g_j^m + \sum_{r \in I_f} p_r^f g_r^f + \sigma^2} \tag{5}$$

Here, p_i^m indicates the transmission power of MUE i; $g_i^{k,m}$ indicates the channel gain between MUE i and its designated MBS k. All other notations are as described previously and also with reference to Equation 4.

The capacities of the femtocell f, macrocell m and the total system capacity are :

$$C^{i,f} = \log_2(1 + \gamma_i^{k,f}) \tag{6}$$

$$C^{i,m} = \log_2(1 + \gamma_i^{k,m}) \tag{7}$$

$$C_{System} = C^{1,m} + C^{1,f} + C^{2,f} \tag{8}$$

The main objective of these parameters is to to guarantee, at every time-instant, that the MUE has a certain Quality-of-Service (QoS) requirement that is above a defined threshold. While ensuring this condition, another goal is to ensure that the FUEs equally have a reasonable QoS. This criteria will be discussed further when we discuss the concept of "rewards" in Q-learning.

3.2. **Problem Formulation.** To illustrate the interference problem in a heterogeneous network, it is imperative to understand the dynamics of the agents in the network. In addition, it is necessary to have a tool that can model the dynamics and enable us to analyze the result of the interaction. In a previous section, we introduced the Q-learning model which consists of a set of states and actions and whose ultimate goal is to find an optimal policy that maximizes the observed rewards over the interaction time of the agents (femtocells/femtousers or macrocells/macrousers). In general, the goal of the agent is to find an optimal policy for each state, using the Q-learning table, which maximizes a cumulative measure of the rewards over time . However, in our case, the agents in our system model are the FUEs and MUE. These agents are competing for the limited spectrum resource. The resulting effect of this interaction and competition is the co-channel interference. It has been reported in literature ([1], [20], [25]) that one of the benefits of the Q-learning table is that the learned information can be used to converge more quickly. In the long run, we would like to mitigate this interference after learning the optimal policy for each agent.

Each agent explores its environment, learns independently, the agents do not communicate with each other, i.e. they do not cooperate. This results in the agents not knowing the other agents' strategies. Because we would like the MUE to have high QoS, and would like to minimize the interference at the MUE, we assume all FUEs have the knowledge of the SINR of the MUEs. Otherwise most likely the MUEs will be left in a low SINR state. We incorporate this condition/assumption into the design of the reward function that is needed in the Q-learning model. Several authors -[1][17]- have proposed different reward functions depending on the problem. The choice of the reward function is critical as it defines what the system learns and in essence defines the objective of the system. While some reward functions [17] depend on the SINR of the MUE and FUE, others such as [1] depends on the proximity of the MUE and the FAP. The latter reward function is a proximity-based reward function. One common feature of all the reward functions is that they include a constant which is non-deterministic. For example, [17] proposed the following reward function :

$$r_t^i = K - (SINR_{MUE_t} - SINR_{th})^2 \tag{9}$$

where K is the constant; $SINR_{MUE_t}$ is the MUE SINR and $SINR_{th}$ is the MUE threshold SINR.

3.3. **State-Action Pairs.** We design different states and different reward functions for the MUE and FUEs respectively. The learning agents, actions, states and reward functions are designed and explained as follows:

(1) Agents: The learning agents are the FUEs and the MUE associated with the only MBS in the considered channel. If there are Nf FAPs, there are Nf + 1 learning agents in the system. In our example, Nf = 2, hence there are 3 learning agents.

(2) Actions: The actions of the learning agents are the predefined transmission power levels. For illustration purposes, we consider actions with an uniform step size, e.g., consider the actions: -2dB, 0dB, 2dB,$\cdots$,16dB with step size of 2dB.

(3) States: The state for FUE i at a particular time step is represented as a tuple of three indicators: $s_i^{FUE} = \{I_{\gamma_m},\ I_{\gamma_i},\ I_r\ \}$. Here I_{γ_m} represents the instantaneous SINR condition of the MUE and I_{γ_i} indicates the SINR condition of FUE i. $I_r = \gamma_i/p_i$ is the SINR and energy ratio of FUE i, used to measure the energy efficiency of the FUE i. These three indicators $\{I_{\gamma_m},\ I_{\gamma_i},\ I_r\ \}$ are defined in equation (10). We can see that FUE i states not only consider the SINR of the FUE i and the energy efficiency of the FUE i but also considers the SINR of the MUE. This is because our aim is to let the FUEs achieve as high efficiency as possible and protect the MUE at the same time. However, the MUE only needs to care about its own SINR and try to achieve high energy efficiency. The state of MUE at a particular time step is represented as a tuple of two indicators: s^{MUE} $= \{I_{\gamma_m},\ I_r\ \}$.

$$(10)\qquad I_{\gamma_m} = \begin{cases} 1 & \text{if } \gamma_m \geq \gamma_T \\ 0 & \text{if } \gamma_m < \gamma_T \end{cases} \qquad I_{\gamma_i} = \begin{cases} 1 & \text{if } \gamma_i \geq \gamma_T \\ 0 & \text{if } \gamma_i < \gamma_T \end{cases} \qquad I_r = \begin{cases} 0 & \text{if } \gamma_i/p_i \leq T_a \\ 1 & \text{if } T_a < \gamma_i/p_i < T_b \\ 2 & \text{if } \gamma_i/p_i \geq T_b \end{cases}$$

where γ_T is the minimum SINR for reliable communication, γ_m and γ_i are the instantaneous SINR of the MUE and FUE respectively, and p_i is the transmission power of the FUE. T_a and T_b are two predetermined thresholds. If the SINR energy ratio is below T_a, the FUE is in a low energy efficiency mode. If the ratio is above T_b, then the user is in high efficiency mode. Our aim is to let the FUEs achieve as high energy efficiency as possible without subjecting the MUE to lower energy efficiency. Basically, there are 12 possible states for each FUE and 6 possible states for each MUE. Table 1 shows the possible states for both the FUE and the MUE.

FUE States	MUE States
$s_0(0,0,0)$	$s_0(0,0)$
$s_1(0,0,1)$	$s_1(0,1)$
$s_2(0,0,2)$	$s_2(0,2)$
$s_3(1,0,0)$	$s_3(1,0)$
$s_4(1,0,1)$	$s_4(1,1)$
$s_5(1,0,2)$	$s_5(1,2)$
$s_6(0,1,0)$	
$s_7(0,1,1)$	
$s_8(0,1,2)$	
$s_9(1,1,0)$	
$s_{10}(1,1,1)$	
$s_{11}(1,1,2)$	

Table 1. Possible States for the the FUE and the MUE

(4) Rewards: In choosing a reward function, we would like to keep the MUE above a pre-defined QoS signal-to-interference and noise ratio or capacity threshold. In doing so, we would also like to maximize the SINR or capacity of each FUE. Contrary to conventional reward functions that are either proximity-based [1] or SINR dependent [17], we choose an arbitrary reward for the MUE and FUE. This choice of reward function is amenable to modification and can be tuned by the user to simulate a different mode of operation. A more realistic reward function would depend on the spatial location of the FUE and macrocell.

The choice of reward function for the MUE and FUE are

$$r_i^m = \begin{cases} 100 & \text{if } \gamma_m \geq \gamma_T \\ -1 & \text{otherwise} \end{cases} \qquad r_i^f = \begin{cases} 100 & \text{if } \frac{\gamma_m}{\gamma_T} \geq 1, \frac{\gamma_i}{\gamma_T} \geq 1 \\ -1 & \text{if } \frac{\gamma_m}{\gamma_T} \geq 1, \frac{\gamma_i}{\gamma_T} < 1 \\ -1 & \text{if } \frac{\gamma_m}{\gamma_T} < 1, \frac{\gamma_i}{\gamma_T} < 1 \end{cases} \tag{11}$$

The rationale behind these reward functions is to maintain high capacity and protection for the MUE and let the FUEs utilize the same spectrum as much as possible without interfering with the MUE at the same time. The MUE will be rewarded 100 if its SINR is higher than the threshold and punished -1 if its SINR is below the threshold because we want to give the MUE higher SINR and at the same time obtain a reasonable SINR for the FUE. Having a negative reward such as -1 would result in a low Q-value. For the FUE, there are three different cases listed above. The FUE will only be rewarded 100 if both its own SINR and the MUE's SINR are above the thresholds. The FUE will be punished -1 when either its own SINR is below the threshold or the MUE's SINR is below the threshold. Since the reward function determines the Q-value which is constantly updated as iteration progresses, this choice of values for the rewards might further bridge the gap between the rewards for the FUE and MUE. Through this design, the FUEs will have to consider the MUE when selecting transmission powers. The agents will choose the actions which have the highest Q-value at every state, after a specified number of iterations.

3.4. **Dynamics of Multi-Agents.** Figure 2 illustrates one FUE decision process modelled as MDP with three states and one possible action associated with each state. The Figure shows a simplified example of how the interaction process of the FUE can be modelled as MDP. However, in practice, we have multiple agents and they have multiple states and actions. The criteria for switching between states and actions can be described as follows:

(1) All the agents choose a random action simultaneously.
(2) Each agent learns the actions of other agents and uses this information to estimate its instantaneous SINR based on Equation 4 for the FUE and based on Equation 5 for the MUE.
(3) The SINR and the predefined threshold values determine the next state of the agent.

To demonstrate how an agent moves between the possible states, we consider the following example: Let's assume the agent FUE_i is at initial random state s_0(0,0,0) and chooses a random action a which corresponds to a predefined transmission power level of 10dB (see Figure 2). Since the other agents simultaneously choose their respective random actions, the agent FUE_i can now estimate its SINR based on the other agents' actions. This is the only information shared with the agents via the designated base station. This random process gives the FUE_i an instantaneous SINR(γ_i) of 8dB which is calculated using Equation 4. We also assume the following values: γ_T = -2dB, $\gamma_m = 9$dB, $T_b = 7$dB, $T_a = 4$dB and $p_i = 1$dB. We note that these constants are deterministic.

From the chosen actions and computed values, since $\gamma_m \geq \gamma_T$, set $I_{\gamma_m} = 1$. Since $\gamma_i \geq \gamma_T$, set $I_{\gamma_i} = 1$. Since $\gamma_i/p_i \geq T_b$, set $I_r = 2$. Based on the predefined values, the agent FUE_i will move to the new state $s_1 = \{1,1,2\}$ corresponding to the computed values of (I_{γ_m}, I_{γ_i}, I_m). This is because the instantaneous SINR of the FUE_i (γ_i) is above the minimum SINR for reliable communication (γ_T), the instantaneous SINR of the MUE is also above γ_T, and the energy ratio of the FUE_i (γ_i/p_i) is above the high efficiency threshold (T_b).

Similarly, when the agent is at state $s_1 = \{1,1,2\}$, it chooses another random action (transmission power level of 16dB), while learning the action of other agents, which reduces its SINR to a value below the low efficiency threshold T_a. Consequently, $\gamma_i \geq \gamma_T$, $\gamma_m \geq \gamma_T$ but $\gamma_i/p_i \leq T_a$. Therefore the agent moves to the next state s_2(1,1,0).

Finally, while in state s_2(1,1,0), suppose the agent chooses an action (transmission power level of 14dB) which not only reduces its SINR to a value below γ_T, and learns that the SINR of the MUE is also below γ_T, this process transitions the agent to the state s_0(0,0,0). We note that this is only an instance of several possible decisions the agent may take and the chain in Figure 2 may take different paths. It is this sort of dynamics that we attempt to model in the next section by means of controlled experiments.

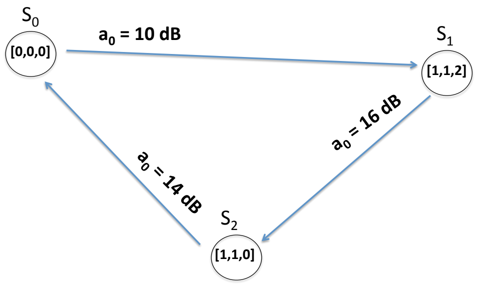

FIGURE 2. MDP for one FUE with three states and one possible action associated with each state. The actions are represented by the blue arrows which are the predefined transmission power level.

4. Numerical Experiments

Using the system model we described in the previous section, we apply a markov decision process to simulate the co-tier and cross tier interference observed at the femtocells and macrocells respectively. Q-learning is introduced during simulation to address the learning mechanism of the active agents and the interfering agents.

4.1. **Simulation Scenario and Parameters.** The wireless network we consider consist of one macrocell underlaid with two femtocells. Each of the cells have an associated macro user and femtousers respectively. The channel gain between the transmitter i and receiver j depends on the distance between the transmitter and receiver. This is modeled as $g_{ij} = d_{ij}^{-k}$, where d_{ij} is the physical distance between the transmitter i and receiver j and k is the path-loss exponent. In our simulations, we choose $k = 4$ just like [1]. In our network, we are interested in studying the uplink interference from the femtouser to the macro user and the interference between the two femtousers. Therefore, the signal to interference noise ratio of one FUE depends on the strength of the interference from the nearby MUE and the nearby FUE. Furthermore only one FUE is allowed to cause interference at the MUE while the MUE can cause interference at both FUEs.

Because of the infinite range of possible distances between the transmitters and receivers, even in such a simplified network model, we tried several values for d in a bid to attain specific deterministic values for each pair of d_{ij}. Nevertheless, we initialized our simulations by setting $d_{ij} = 1$ to verify the usability of the simulation algorithm. The effect of this distance value will be explored in another related experiment.

The Q-learning aspect of our simulation considers two main factors - the reward function and the thresholds that were chosen adaptively. The reward function we chose has already been discussed in the previous section. A range of possible threshold levels (T_a, T_b, γ_T) were also investigated in the simulation. Computation of the SINR followed the conventional approach in literature, so we set the noise power $\sigma^2 = 0.1$ in line with [1]. In addition, the learning rate $\alpha = 0.5$ and the discounted factor used in the Q-learning algorithm is set to 0.9. Finally, the simulation is implemented in MATLAB on a desktop.

4.2. **Experimental Setup.** In the setup of our experiments, one major factor was considered namely, the position of the agents in the network. Contrary to other robust setups where the agents can be at random locations, choosing random actions at every instance [13], we choose to fix the position of the agents at every iteration for simplicity. This design may affect the accuracy of our model and we leave the extension for future work. The transition between states should not be misconstrued with the spatial locations of the agents in the network. The transition from state to state by the agents gives them a reward function based on their chosen actions and transmission levels. The role of Q-learning is to update the Q-table while this transition is in progress. After half of the iterations has been reached, the agents would have learnt the optimal action or optimal policy to take at each state to maximize their reward. On reaching this stage, the agents does not choose any random actions, instead it chooses the action based on the Q-table value. Throughout the experiment, there is no shared learning involved which has a potential to impact the network ability to efficiently allocate resources.

4.2.1. *Independent Q-learning with Constant Gain.* At the beginning of our simulation, we assume all the agents' distance is exactly the same, so we fix the distance to a unit of 1. This implies that the path loss is the same and this constant path loss is used in the Q-learning algorithm. Other parameters are the same as those used in the illustration in section 3.4 unless otherwise stated. This fixed condition on the distance means that the gains at each base station will be exactly the same and the SINR will simply be dominated by the choice of transmission level denoted by the random action the agents take at every iteration. Setting the threshold parameters as $(T_b = 7, T_a = 4, \gamma_T = 2)$, we performed 1000 Q-learning iterations where at each iteration, the agents choose a random action corresponding to one of the transmission power levels in the range $2dB, 4dB, \cdots 18dB$. In the first half of the iterations, the agents choose random actions, resulting in oscillating capacities as shown in Figure 3 while the agents continue to learn the actions of other agents. Finally information from the Q-table of each agent is used from the second half of the iteration leading to a convergence in the system for each of the FUEs and the MUE. As seen from the Figure, the capacity of the macrocell (labelled as MUE) is higher than the respective femtocells (labelled as FUEs), a condition we prefer for high QoS especially at the macrocell. In terms of scaling, we also see that the closer the agents, the stronger the interference leading to high magnitudes for the capacities. On the other hand, a visualization of the cumulative capacity as the iteration progressed shows very interesting information. Figure 4 shows that the FUE that is not experiencing a cross-tier interference from the other FUE attains a capacity that approaches the capacity of the MUE. This phenomenon is probably attributed to the fact that the two FUEs did not share their information, calling for the prospect of a distributed learning or cooperative learning idea. Finally, in this first phase of experiments, we computed the aggregate capacity of the femtocells and the macrocells. This is necessary to determine the load or throughput in the network. By taking the sum of the individual femtocell capacities at each iteration and plotting it against the aggregate or sum capacity of the macrocell, we observe that the total FUE capacity is more than that of the macrocell as shown in Figure 5. This result is expected given that the agents (FUEs) do not cooperate in the learning process. Another possible explanation is the fact that we have assumed a constant gain which relates to the spatial location of the FUEs and MUEs from their associated access points or/and the stations they interfere with. This necessitated the run of another experiment that addresses the question of distance related convergence.

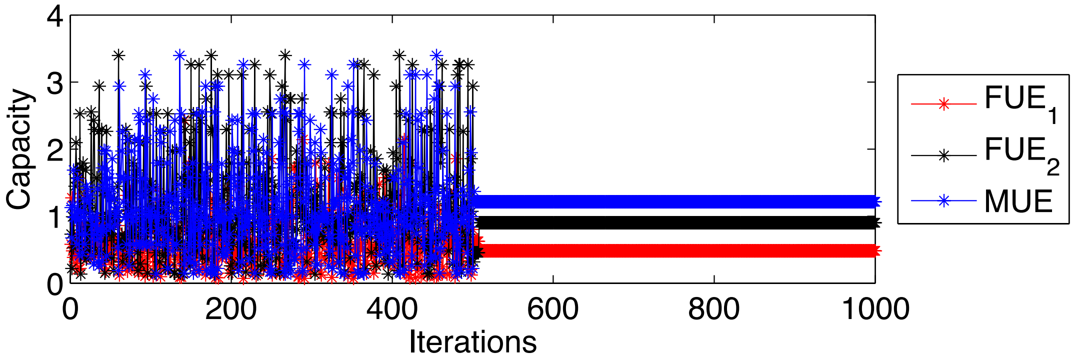

FIGURE 3. Macrocell and femtocells' capacity as a function of iterations. Independent Q-learning with agents at common distance

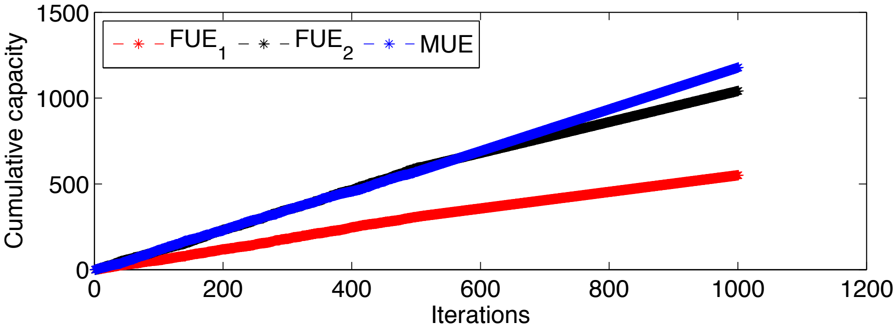

FIGURE 4. Cumulative capacity at the base stations as a function of iterations. Agents at common distance

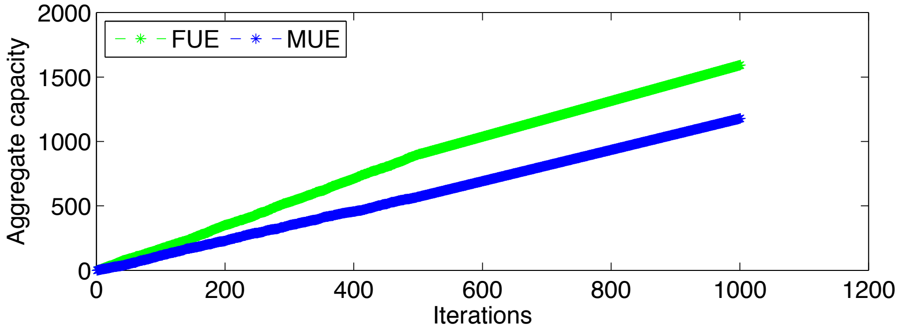

FIGURE 5. Total capacity of the FUEs and the MUE in the network as a function of iterations using independent Q-learning of agents at common distance

4.2.2. *Independent Q-learning with Varying Gain.* In this section, we repeated the sequence of experiment described in the previous section, However, we altered the distances for each of the agents. Therefore, for the FUEs and MUEs, we fixed their distances from their respective base stations to $d = 100$. This ensures they have a constant gain from their dedicated base stations. On the other hand, their locations from the interfering stations was set at $d = 10$. So, this changes the problem slightly from the previous experiments where $d = 1$ in all cases. After performing 1000 iterations, and invoking Q-learning in the process, the observed results are shown in Figures 6, 7 and 8. Comparison of these results with the Figures shown previously follows.

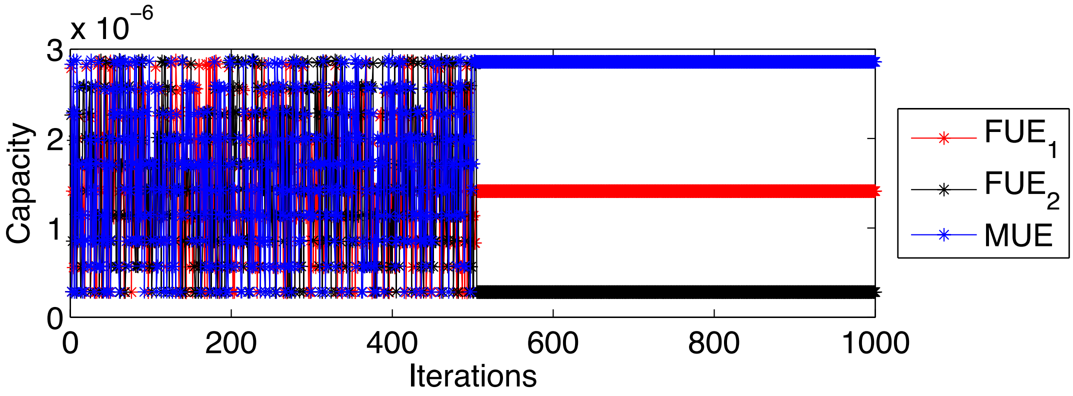

FIGURE 6. Macrocell and femtocells' capacity as a function of iterations. Independent Q-learning with agents at different distances from base stations

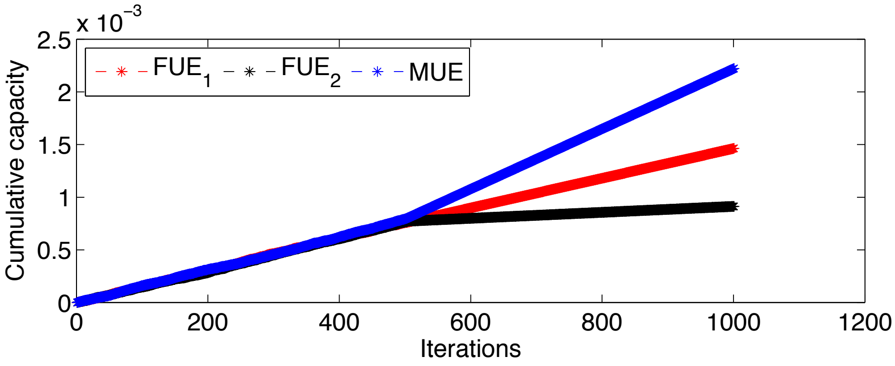

FIGURE 7. Cumulative capacity at the base stations as a function of iterations. Agents at different distances from base stations

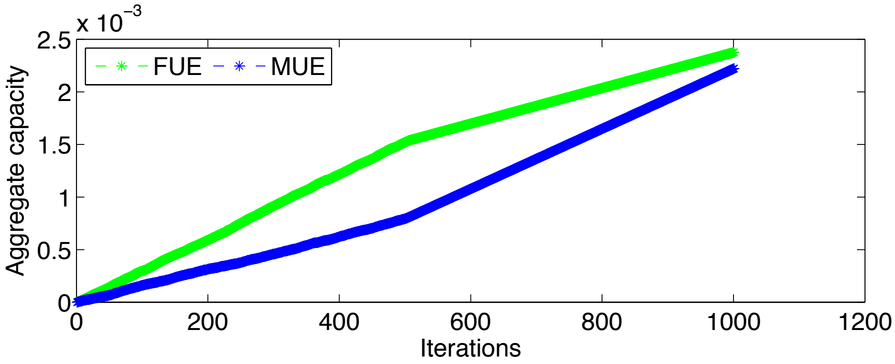

FIGURE 8. Total capacity of the FUEs and the MUE in the network as a function of iterations using independent Q-learning of agents at different distances from base stations

In this second set of experiments, our first observation was the drop in magnitude of the capacity. This decrease is related to the distance, hence the varying path loss in the network between the femtocells and macrocells. Further, the capacity of the macrocell after half of the iterations is still higher than the femtocells. Then, we observe reversal in the role of the two FUEs - the FUE experiencing both cross-tier and co-tier interference subsequently has a higher capacity than the one experiencing only a cross-tier interference. Perhaps, this is related to the change in the distance between the two FUEs. We also noticed that we no longer see the convergence between the FUE_2 and the MUE we saw previously. There is a distinct difference in the convergence of the cumulative capacities for all the agents. These observations shows the robustness of the Q-learning algorithm as a tool to efficiently allocate resources when the positions of the agents can be controlled. A cooperative learning between the agents might be a way to improve the network ability to manage the interference more efficiently.

5. Conclusions and Future Work

By means of numerical simulations and results, we have evaluated the performance of Q-learning algorithm. Our system comprising of one macrocell and two femtocells can be extended to multiple macrocells and femtocells. We showed how convergence can be attained using the Q-learning algorithm. We also illustrated how we can achieve a high quality-of-service for the macro user equipment while dealing with the uplink co-tier interference from the femtouser(s). The closer the interfering agent is to the base station, the stronger the interference. In another experiment, we investigated the effect of proximity in this network and how Q-learning helps to address it. Here, the idea was to keep the users closer to their respective base stations while keeping them far apart from each other to minimize the interference. We observe how the cumulative capacity graph gets closer for the FUE and MUE. We postulate that more experiments may need to be performed in order to draw a firm conclusion from this behaviour. We have shown how to control the learning using the Q-table to store the updated Q-values after a fixed number of iterations, before using the optimal policy for subsequent actions taken in the network. We also observe how the capacity of the macrocells and femtocells coincide depending on the spatial distribution of the cells. This confirms how the dynamics of the agents in an heterogeneous network can adversely affect the quality of service received by each user. In summary, these experiments confirm the power of Q-learning to enable us evaluate a reward function and the rate at which convergence is attained in a heterogeneous network under different scenarios. These scenarios range from the spatial distribution of the agents in the network, the available transmission levels, and the number of agents in the network.

It will be interesting to observe how these results are affected when we consider cooperative Q-learning. It will also be worth investigating how other rewards functions will affect the convergence graph and quantify the rate of convergence. While we have allowed every agent to interfere with others, a constrained interference, where we adaptively select how these agents are allowed to interfere, is also an area of research worth investigating. Another potential future work will be to investigate other mathematical approaches to deal with network interference problems such as game theory.

6. Group Contribution Statement

Introduction (Yuhe):

Yuhe - Gather related resources for project preparation and make introductions.

Literature Review(Abdullah):

Abdullah - Gather related resources for review purpose.

System Model (Abdullah, Yuhe):

Abdullah - Propose the system model based on related works.

Yuhe - Verify the system model based on data analysis.

Problem Formulation (Abdullah, Yuhe):

Abdullah - Propose the problem formulation based on the system model.

Yuhe - Modify and verify the problem fomrulation.

Numerical experiments and results (Abdullah, Yuhe):

Abdullah - Construct the main program for Q-Leaning algorithm, and provide information for reward functions.

Yuhe - Construct and modify the reward functions to satisfy different experimental needs, and help improving the main program.

Discussion and Conclusion (Abdullah, Yuhe):

Abdullah, Yuhe - Discuss in group to conclude the experiment results and make further discussion.

References

[1] The Japan Reader *“A proximity-based Q-learning reward,” function for femtocell networks”* in IEEE Vehicular Technology Conference, Sept. 2013

[2] C. J. C. H. Watkins and P. Dayan *“Technical note Q-learning,” Journal of Machine Learning, vol.* 8, pp. 279-292, 1992.

[3] Hu, Junling, and Michael P. Wellman. ”Multiagent reinforcement learning: theoretical framework and an algorithm.” ICML. Vol. 98. 1998.

[4] Michael, T., and I. Jordan. ”Reinforcement learning algorithm for partially observable Markov decision problems.” Proceedings of the Advances in Neural Information Processing Systems (1995): 345-352.

[5] Giupponi, Lorenza, Ana Galindo-Serrano, Pol Blasco, and Mischa Dohler. ”Docitive networks: an emerging paradigm for dynamic spectrum management [dynamic spectrum management].” Wireless Communications, IEEE 17, no. 4 (2010): 47-54.

[6] E. H. Norman *“Convergence of Q-learning: A simple proof,”* Institute Of Systems and Robotics, Tech. Rep. Relations.

[7] Radaydeh, Redha M., and M-S. Alouini. ”Low-complexity co-tier interference reduction scheme in open-access overlaid cellular networks.” Global Telecommunications Conference (GLOBECOM 2011), 2011 IEEE. IEEE, 2011.

[8] Kim, Min-Sung, Hui Je, and Fouad A. Tobagi. ”Cross-tier interference mitigation for two-tier OFDMA femtocell networks with limited macrocell information.” Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE. IEEE, 2010.

[9] Yavuz, Mehmet, Farhad Meshkati, Sanjiv Nanda, Akhilesh Pokhariyal, Nick Johnson, Balaji Raghothaman, and Andy Richardson. ”Interference management and performance analysis of UMTS/HSPA+ femtocells.” Communications Magazine, IEEE 47, no. 9 (2009): 102-109.

[10] Saquib, Nazmus, Ekram Hossain, Long Bao Le, and Dong In Kim. ”Interference management in OFDMA femtocell networks: Issues and approaches.” Wireless Communications, IEEE 19, no. 3 (2012): 86-95.

[11] Zhao, Jun, et al. *”Game Theory Based Energy-Aware Uplink Resource Allocation in OFDMA Femtocell Networks.” International Journal of Distributed Sensor Networks* 2014 (2014).

[12] Wang, Hong, Rongfang Song, and S. Leung. ”Analysis of Uplink Inter-Carrier-Interference in OFDMA Femtocell Networks.” (2013): 1-1.

[13] Guler, Basak, and Aylin Yener. ”Selective interference alignment for MIMO cognitive femtocell networks.” Selected Areas in Communications, IEEE Journal on 32.3 (2014): 439-450.

[14] Chandrasekhar, Vikram, and Jeffrey G. Andrews. ”Uplink capacity and interference avoidance for two-tier femtocell networks.” Wireless Communications, IEEE Transactions on 8.7 (2009): 3498-3509.

[15] Chakchouk, Nesrine, and Bechir Hamdaoui. ”Statistical characterization of uplink interference in two-tier co-channel femtocell networks.” Wireless Communications and Mobile Computing Conference (IWCMC), 2012 8th International. IEEE, 2012.

[16] Ganz, Aura, C. M. Krishna, Dingyi Tang, and Zygmunt J. Haas. ”On optimal design of multitier wireless cellular systems.” Communications Magazine, IEEE 35, no. 2 (1997): 88-93.

[17] Galindo-Serrano, Ana, and Lorenza Giupponi. ”Distributed Q-learning for interference control in OFDMA-based femtocell networks.” Vehicular Technology Conference (VTC 2010-Spring), 2010 IEEE 71st. IEEE, 2010.

[18] Mhiri, Fadoua, Kaouthar Sethom, and Ridha Bouallegue. ”A survey on interference management techniques in femtocell self-organizing networks.” Journal of Network and Computer Applications 36.1 (2013): 58-65.

[19] Simsek, Meryem, Andreas Czylwik, Ana Galindo-Serrano, and Lorenza Giupponi. ”Improved decentralized Q-learning algorithm for interference reduction in LTE-femtocells.” In Wireless Advanced (WiAd), 2011, pp. 138-143. IEEE, 2011.

[20] Bennis, Mehdi, and Dusit Niyato. ”A Q-learning based approach to interference avoidance in self-organized femtocell networks.” GLOBECOM Workshops (GC Wkshps), 2010 IEEE. IEEE, 2010.

[21] Bennis, Mehdi, and Samir Medina Perlaza. ”Decentralized cross-tier interference mitigation in cognitive femtocell networks.” Communications (ICC), 2011 IEEE International Conference on. IEEE, 2011.

[22] Dhahri, Chaima, and Tomoaki Ohtsuki. ”Learning-based cell selection method for femtocell networks.” Vehicular Technology Conference (VTC Spring), 2012 IEEE 75th. IEEE, 2012.

[23] Bennis, Mehdi, Sudarshan Guruacharya, and Dusit Niyato. ”Distributed learning strategies for interference mitigation in femtocell networks.” Global Telecommunications Conference (GLOBECOM 2011), 2011 IEEE. IEEE, 2011.

[24] Busoniu, Lucian, Robert Babuska, and Bart De Schutter. ”A comprehensive survey of multiagent reinforcement learning.” Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on 38.2 (2008): 156-172.

[25] Saad, Hussein, Amr Mohamed, and Tamer ElBatt. ”Distributed cooperative Q-learning for power allocation in cognitive femtocell networks.” Vehicular Technology Conference (VTC Fall), 2012 IEEE. IEEE, 2012.

[26] Saad, Hussein, Amr Mohamed, and Tamer ElBatt. ”A cooperative Q-learning approach for distributed resource allocation in multi-user femtocell networks.” Wireless Communications and Networking Conference (WCNC), 2014 IEEE. IEEE, 2014.

[27] Lopez-Perez, David, Alvaro Valcarce, Guillaume De La Roche, and Jie Zhang. ”OFDMA femtocells: A roadmap on interference avoidance.” Communications Magazine, IEEE 47, no. 9 (2009): 41-48.

[28] Kivanc, Didem, Guoqing Li, and Hui Liu. ”Computationally efficient bandwidth allocation and power control for OFDMA.” Wireless Communications, IEEE Transactions on 2.6 (2003): 1150-1158.

[29] Pateromichelakis, Emmanouil, Mehrdad Shariat, and R. Tafazolli. ”On the analysis of co-tier interference in femtocells.” Personal Indoor and Mobile Radio Communications (PIMRC), 2011 IEEE 22nd International Symposium on. IEEE, 2011.

[30] Seong, Kibeom, Mehdi Mohseni, and John M. Cioffi. ”Optimal resource allocation for OFDMA downlink systems.” Information Theory, 2006 IEEE International Symposium on. IEEE, 2006.