Learning-based Cell Selection Method for Femtocell Networks

Chaima Dhahri and Tomoaki Ohtsuki
Department of Computer and Information Science, Keio University
Yokohama, Kanagawa, Japan

***Abstract*—In open-access non-stationary femtocell networks, cellular users (also known as macro users or MU) may join, through a handover procedure, one of the neighboring femtocells so as to enhance their communications/increase their respective channel capacities. To avoid frequent communication disruptions owing to effects such as the ping-pong effect, it is necessary to ensure the effectiveness of the cell selection method. Traditionally, such selection method is usually a measured channel/cell quality metric such as the channel capacity, the load of the candidate cell, the received signal strength (RSS), etc. However, one problem with such approaches is that present measured performance does not necessarily reflect the future performance, thus the need for novel cell selection that can predict the *horizon*. Subsequently, we present in this paper a reinforcement learning (RL), i.e, Q-learning algorithm, as a generic solution for the cell selection problem in a non-stationary femtocell network. After comparing our solution for cell selection with different methods in the literature (least loaded (LL), random and capacity-based), simulation results demonstrate the benefits of using learning in terms of the gained capacity and the number of handovers.**

I. Introduction

Nowadays, femtocell networks offer an interesting solution for indoor communications with short-range, low-cost, low-power base station (BS) that maintains good indoor coverage and capacity [1]. For an open-access femtocell network, Macro User (MU) may leave its serving Macro Base Station (MBS) and switch to one of its neighboring Femto Base Stations (FBSs). In general, for large number of neighboring femtocells, a user has multiple choices for the next step and some of the choices are better than others. Thus, it may be worthwhile to investigate a method for cell selection during a handover procedure in order to maximize the user benefits in terms of gained capacity and to eliminate redundant handovers.

Different handover metrics were proposed in the literature. Some of these metrics are signal to noise ratio (SNR) [2], signal strength [3] and bit error rate [4]. Considering a handover decision in femtocell networks, previous works like [5] evaluate the performance of received signal strength (RSS)-based and capacity-based cell selection methods. It has been shown that the capacity-based cell selection metric outperforms the RSS-based cell selection metric.

From a capacity point of view, the problem with previous work is that it guarantees capacity at present time t and can not ensure that this capacity will be maintained in the future, because the measure of the gained capacity at time t does not reflect the behavior of the target cell at time $t + \Delta t$. Indeed, in some situations the capacity may degrade abruptly either because of channel condition, i.e. owing to propagation effects, or receiver location (cell edge receiver). Such situations may cause undesirable effects such as the ping-pong effect. However, in our view, handover is a lengthy problem signaling-and-bandwidth expensive. Thus, it is necessary that a handover procedure be as less frequently as possible. Ensuring an efficient handover decision that guarantees better capacity for sufficiently long period of time becomes of the essence. Towards this end, we propose an Reinforcement Learning (RL)-based cell selection method for an open-access femtocell network to ensure a reliable cell selection based on learning the behavior of different cells in the past and predicting their behaviors in the future.

In the RL framework [6], an agent learns optimal actions/decisions through trial-and-error interaction with its dynamic environment. On each step, the agent chooses an action that changes the state of the environment through a transition phase then receives a reward representing how good or bad the action was. The agent's goal is to maximize this reward by finding the optimal policy defining the best action for each state of the environment.

The solution for open-access mode in femtocell networks is presented as a model-free RL framework, because it requires no explicit knowledge of the transition probability and the reward function. The MU is modelled as a *selfish* agent that wants to maximize its capacity. The MU takes advantage of the RL algorithm to estimate the efficiency of neighboring femtocells based on their past behavior: RL relies on exploration of the environment to converge toward a policy that maximizes MU capacity.

II. System Model and Problem Statement

A. System Model

Our system model is depicted in Fig. 1. We consider one MBS deployed in a non-stationary environment (considering user mobility) where N_{MU} MUs are randomly located inside the macrocell coverage area. This MBS underlaid with F FBSs providing service to N_{FU} Femto Users (FUs). All base stations (MBS and FBSs) and user equipment (UE) are equipped with one antenna. We consider a Time Division Duplex (TDD) Transmission mode for femto and macro links. TDD time-synchronization among all the cells is assumed to be perfect. Finally, we assume a co-channel open-access deployment of femtocells to enable efficient utilization of the available spectrum.

978-1-4673-0990-5/12/$31.00 ©2012 IEEE

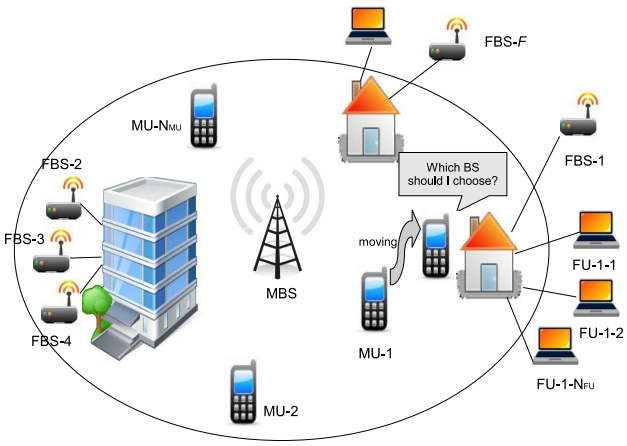


Fig. 1. System model

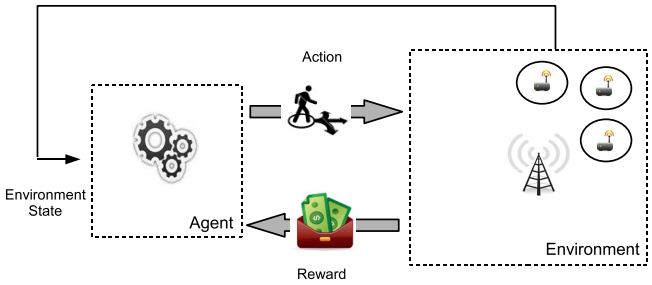


Fig. 2. RL framework

Let P_{MBS} be the MBS transmit power and $h_{MBS,k}$ the channel gain between the MBS and its serving k^{th} MU. Likewise, $h_{i,j}$ denotes the channel gain between the i^{th} FBS and the j^{th} FU. Finally, P_i denotes the transmit power of the i^{th} FBS. We assume an additive white Gaussian noise (AWGN) at MUs with power σ^2. The capacity at MU k from its serving MBS is given by:

$$C_k = \frac{B}{N_{MU}} \log_2 \left(1 + \frac{|h_{MBS,k}|^2 P_{MBS}}{\sigma^2 + I}\right) \quad (1)$$

where B is the available bandwidth, $I = \sum_{i=1}^{N_{NCL}} |h_{i,k}|^2 P_i$ is the interference from neighboring FBSs and N_{NCL} is the number of neighboring FBSs. We assume that the bandwidth is equally allocated to all users (FUs and MUs). The capacity at FU j from FBS i is given by:

$$C_j = \frac{B}{N_{FU}} \log_2 \left(1 + \frac{|h_{i,j}|^2 P_i}{\sigma^2 + I_{MBS} + I_{FBS}}\right) \quad (2)$$

where $I_{MBS} = |h_{MBS,j}|^2 P_{MBS}$ is the interference from MBS, $h_{MBS,j}$ is the channel gain between the MBS and user j. Also, $I_{FBS} = \sum_{l \neq i} |h_{l,j}|^2 P_l$ is the interference from other FBSs and $h_{l,j}$ is the channel gain between FBS l, transmitting with power P_l, and user j.

B. Problem Statement

In a large-scale, highly dense femtocell network, we face the problem of cell selection during a handover procedure. The most critical handover step is the selection decision because a bad decision in cell selection may cause undesirable effects, i.e. ping-pong effect. This problem can be regarded as an optimization problem where the past behavior of different cells has an impact on the future decisions. Thus, the need of a learning algorithm that can predict, based on the past, the best performing cell in the future. RL is a learning algorithm that tries to find a policy (mapping situations to actions) in a recursive fashion in order to maximize a received reward. A typical RL algorithm used in the proposed solution is the Q-learning: an online learning algorithm that needs no explicit information (transition distribution and the reward function) during its execution. These features, that RL (Q-learning) enjoys, make it an attractive solution for cell selection in a femtocell networks.

III. Proposed Solution

A. General View

Nowadays, a user in a femtocell network can support a searching function for the neighboring cells and distinguish accessible FBS from unauthorized ones [7]. A mobile MU can be *selfish* and switch from a serving cell (MBS or FBS) to another cell in order to increase its capacity. To do so, MU must make a selection decision: choose the best FBS from its Neighboring Cell List (NCL) or stay connected to the MBS. The solution breaks down into three steps:

- MU j collects information about its neighboring cells: estimation of the channel gain $\hat{h}_{i,j}$ between itself and its neighboring cell i using LMMSE estimator like in [8].
- MU predicts the best cell through learning (Q-learning).
- MU joins the new serving cell.

B. Q-learning Algorithm

Definition: With Q-learning algorithm, an agent aims to find the policy that maximizes the Q-value function that gives the expected utility of taking an action a in a current state s (see Fig. 2). Here, we define all parameters related to the Q-learning algorithm:

- Environment: represents all elements outside the agent. In our case, it includes the MBS and all FBSs in the MU's NCL. We assume that the environment is a finite-state, discrete-time, stochastic dynamical system.
- Agent: is the decision maker. In our framework, the agent represents the MU who wants to perform a handover procedure from its serving cell (MBS or FBS) to a better neighboring cell.
- State: is the current state of the environment. In our case, it represents the current MU serving cell (MBS or FBS). The state set S is defined as $S = \{s = 1, 2, ..., N_{NCL} + 1\}$ where N_{NCL} is the number of neighboring femtocells. The initial state corresponds to the case where MU is connected to MBS $(s = 1)$.

- Action: is the decision taken by the agent. In our case, it represents the identifier of the target cell: the MU can stay connected to the serving MBS (action 1) or choose one of the FBSs from its NCL (action 2,..., action N_{NCL}). The action set A is defined as $A = \{a = 1, 2, ..., N_{NCL} + 1\}$ where N_{NCL} is the number of neighboring femtocells.
- Reward: is a utility function, r, representing how good the action was. In our case, the reward is the gained capacity after joining the target cell (MBS or FBS).

Formulation: The aim of a Q-learning algorithm is to find an optimal policy Π_{opt} that maximizes the discounted cumulative reward function over an infinite horizon:

$$\max_{\Pi} \sum_{t=0}^{\infty} \gamma^t r_t \tag{3}$$

where γ is the discount factor $(0 \leq \gamma \leq 1)$ and r_t is the received reward at time t. For $\gamma = 0$ future rewards have no impact on the state value, while for γ close to 1, future actions are considered as important as the immediate rewards. For a given policy Π, we define a Q-value as:

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

where:

- $R(s, a)$ is the expected reward when the environment is in the state s and the agent executes the action a: expected reward of the current pair state-action
- $P_{s,v}(a)$ is the transition probability from the initial state s to the new state v as a result of the action a.
- $Q(v, b)$ is the Q-function of the next state-action pairs.

Applying Bellman's optimality for a single agent environment, we can guarantee at least one optimal policy Π^* [9]. Let $Q^*(s, a)$ be the maximum of the Q-function that determines the optimal action b for every possible next pair (v, b).

$$Q^*(s, a) = R(s, a) + \gamma \sum_{v \in S} P_{s,v}(a) \max_{b \in A} Q^*(v, b) \tag{5}$$

Q-learning determines this optimal $Q^*(s, a)$ in an iterative process. At each step during the learning process, the Q-value function must be updated using the following equation:

$$\begin{aligned} Q_t(s, a) &= (1 - \alpha)\, Q_{t-1}(s, a) \\ &+ \alpha \left(R_t(s, a) + \gamma \max_b Q_{t-1}(v, b) \right) \end{aligned} \tag{6}$$

where α is the learning rate.

C. Proposed Solution

Single-Agent Multiple-State RL: We consider one MU performing a handover procedure (single-agent) and multiple choices for the serving cell (multiple-state). Therefore, MU has to decide which cell to join in order to maximize its capacity and maintain it for long period of time. The selection problem can be modelled as a Single-Agent, Multiple-State RL (SAMSRL) framework where we need to find a policy that gives the best action for each state. In our system, $Q(s, a)$ is the expected reward $r(s, a)$ of each state-action pair (s, a).

$$Q(s, a) = E[r(s, a)] \tag{7}$$

Because of the dynamic nature of our environment, we can not have an exact knowledge about the reward received from the new selected cell (it can be either one of the neighboring FBSs or the MBS). Thus, we define $Q_t(s, a)$ as the estimation of $Q(s, a)$ at time t. It is an average of all received rewards when action a was chosen for the state s before time t [10]. At $t = 0$, we set $Q_0(s, a) = 0$ for every possible pair (s, a).

Reward function: Let us define the utility function corresponding to the reward r in the proposed algorithm. Our goal is to maximize and maintain the capacity of MU joining a new cell after a handover procedure. Therefore, the utility function r is the perceived reward (capacity) of the target cell and it is expressed as (1) or (2), i.e.

- if MU chooses the MBS as a serving cell, r is expressed as (1).
- if MU chooses to join one of the FBSs in its NCL, r is expressed as (2).

Action Selection Strategy: Different strategies might be taken to learn $Q(s, a)$. In our proposal, we consider ϵ-Greedy method. In this method, the best cell is selected for a proportion $1 - \epsilon$ of the trials (exploitation phase), and another cell is randomly selected, with uniform probability, for a proportion ϵ of the trials (exploration phase) where ϵ is between 0 and 1. In this case, as the number of trials gets larger $(n \rightarrow \infty)$, we will guarantee the proper estimates of $Q(s, a)$ for all $s \in S$ and $a \in A$ [10].

IV. Performance Evaluation

A. Simulation Parameters

In this section, we evaluate the performance of the proposed method for one MU performing a handover procedure and trying to select an optimal cell in open-access non-stationary femtocell network where the femtocells are distributed randomly. We consider a pedestrian MU moving, in a limited region, at a speed of 1 m/s following straight line. As initial state, the MU is served by the MBS. The channel is designed as a multipath fading Rayleigh with Doppler spread. The considered path loss (PL) models are similar to [11]. The simulation parameters considered to validate the proposed solution are defined in TABLE I. To simulate different MUs performing handover simultaneously, we have to consider Multiple-Agent Multiple-State RL where each MU represents an agent. This scenario is beyond the scope of this paper.

B. Simulation Results

To evaluate the performance of the proposed SAMSRL method from a capacity point of view, we compare it with different selection methods found in the literature such as random, least loaded (LL) and capacity-based [5] methods. In Fig. 3, we compare the performance of the proposed scheme (SAMSRL), in terms of MU gained capacity with Random and Least Loaded (LL) algorithms. First, we define the load as the number of users in a cell. LL algorithm will select a cell with the minimum number of users. For each algorithm, we call the program 2000 times then we plot the capacity received by MU, after performing a handover, averaged over 2000. We

TABLE I
SIMULATION PARAMETERS

Parameter	Value
Carrier Frequency	2000 MHz
System Bandwidth	10 MHz
Number of Paths	6
Time Sampling	1/1000 s
Macrocell Size	1000 m
Femtocell Size	30 m
Transmit Power at MBS	46 dBm
Transmit Power at FBS	20 dBm
FBSs Distribution	random
Transmit Power at User Equipment(UE)	23 dBm
Noise Power at each UE	-174 dBm
Noise Figure at UE	9 dB
Total Number of MUs: N_{MU}	10
Number of MUs Performing a Handover	1
Number of FUs at each FBS: N_{FU}	Poisson distribution ($\lambda = 2$)
Total Number of FBSs: F	30
Outer Wall Loss	15 dB
α	0.5
γ	0.9
ϵ	0.1

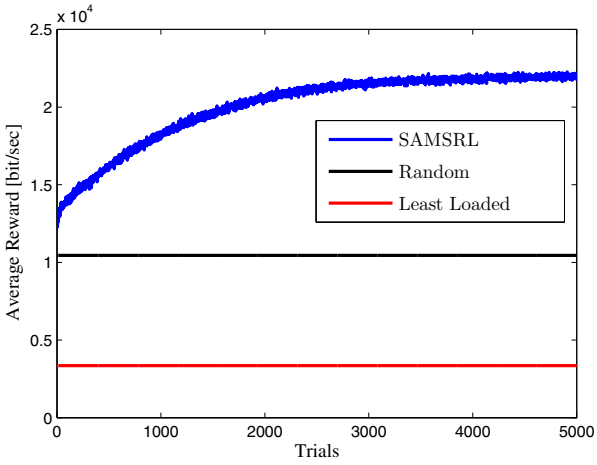


Fig. 3. Comparison of SAMSRL, random and Least Loaded

observe that, the proposed scheme (SAMSRL) allows higher capacity than a random and LL selection algorithms: after a certain number of trials, the Q-learning algorithm (blue curve) converges to a value (above 2×10^4 bit/sec) that compares very well with the capacity gained from random (black line) and LL (red line) selection algorithm. Also, we can see that random algorithm gives better result (higher capacity) than LL algorithm. In fact, with a random strategy, all actions are equiprobable, hence the best action will be selected with the same probability as other actions while in LL algorithm, the least loaded cell is always selected. However, a cell with minimum number of users is not necessary the best one because other parameters like interference, channel, etc have an impact on the received capacity.

When only one handover is carried out: Fig. 4 illustrates the instantaneous capacity gains of the proposed scheme (in blue) with respect to the conventional capacity-based in [5] (in red),

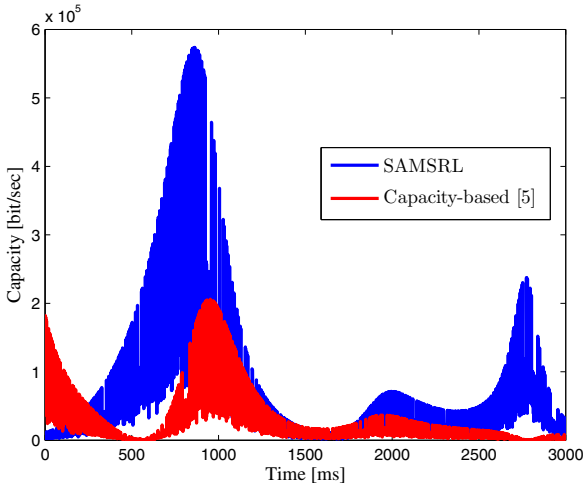


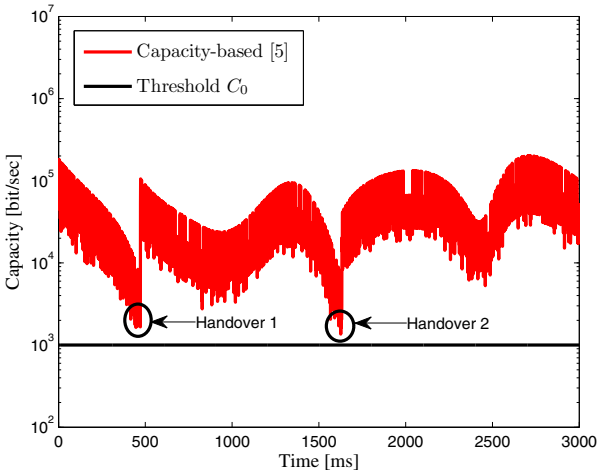
Fig. 4. Comparison in terms of the gained capacity

after performing a single handover. We observe two phases:

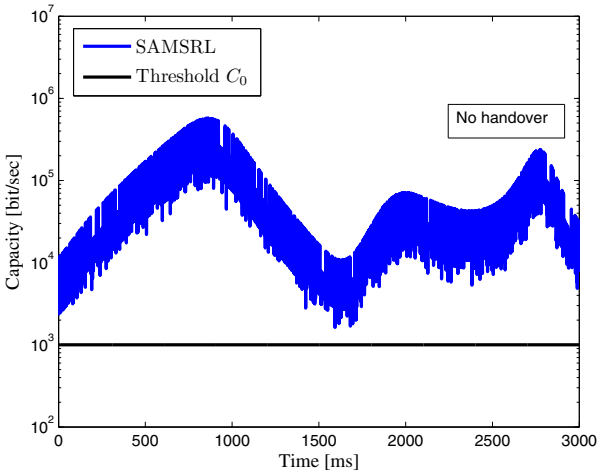
- At the beginning of the simulation, the conventional scheme [5] outperforms the learning method (SAMSRL). Indeed, given such short time frame, it is unlikely for the quality of the current best cell to degrade.
- However, as time elapses, the quality of the best cell selected by the conventional scheme [5] becomes more likely to degrade. Quite the opposite, the proposed scheme (SAMSRL) picks a cell whose channel quality is likely to remain good over an extended period of time owing to the learning method. Thus, we observe that the proposed scheme (blue line) outperforms the conventional scheme (red line) in the remainder of the simulation time.

This confirms our intuition that Q-learning helps predict the future channel quality (*horizon*) of the selected cell.

When more than one handover is carried out: In Fig. 5, we extend the results of Fig. 4 to the case where MU is allowed to carry out more than one handover. For clarity, the capacity gains relative to the conventional (capacity-based [5]) and the proposed (learning-based SAMSRL) methods are reported in different figures. In our simulations, we required the MU to trigger the handover procedure whenever the capacity, given by (1) or (2), fell below a given threshold $C_0 = 10^3$ bit/sec. In the conventional scheme (Fig. 5(a)), we observe that the handover procedure was triggered twice, precisely at $t_1 = 472$ ms and $t_2 = 1633$ ms. This is owing to the fact that the best cell picked by the conventional scheme had its channel quality degrade after a certain period of time. On the contrary, the proposed scheme (Fig. 5(b)) ensures that the picked cell will have its channel quality sufficiently good over an extended period of time. Thus, we observe in this figure that no handover procedure was triggered owing to the proposed learning method. To confirm such observation, we evaluate in Fig. 6 the average number of handovers over 40 runs for the conventional (capacity-based [5]) and the proposed (SAMSRL) methods considering different durations. We observe that, for both methods, the average number of handovers increases



(a) capacity-based



(b) learning-based

Fig. 5. Comparison in terms of the number of handovers

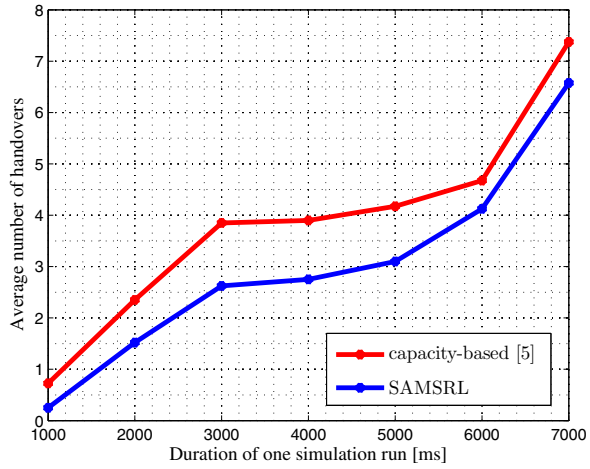


Fig. 6. Average number of handovers for different durations of simulation

when the duration of the simulation increases. However, in the proposed scheme (SAMSRL) it is always less than that of the conventional scheme. Such result is in agreement with the observation in Fig. 5 that the proposed scheme is less likely to incur situations where a handover may be required even for long duration of observation.

Summary of findings: From these figures, we conclude that with respect to conventional methods (random, LL, capacity-based), the merits of the proposed cell selection (SAMSRL) are twofold:

- A capacity gain (Fig. 3 and Fig. 4) owing to the stability of the channel quality of the target cell chosen by the proposed method (SAMSRL).
- A signaling cost reduction by reducing the number of expected handovers for a given mobility pattern (Fig. 5 and Fig. 6).

V. Conclusion

In this paper, we propose a learning-based cell selection method for an open-access femtocell network. This method is based on Single-Agent Multiple-State RL (SAMSRL) framework for cell selection to guarantee higher capacity for a *selfish* MU that wants to proceed a handover from its serving cell (MBS or FBS) to one of its neighboring cells. First, we justify the need of learning and we explain, in detail, the Q-learning algorithm used for cell selection. Then, we compare the proposed solution with random, Least Loaded (LL) and capacity-based selection methods. Simulation results show that when the MBS is underlaid with many FBSs, the proposed method could significantly improve the user (MU) experience, when performing a handover, in terms of the gained capacity and the number of handovers. The scenario where multiple mobile users decide to handover simultaneously (group handover) can be considered as a future work and can be modelled as Multiple-Agent, Multiple-State RL framework.

References

[1] V. Chandrasekhar, J. Andrews, and A. Gatherer, "Femtocell networks: a survey," *IEEE Commun. Mag.*, vol. 46, no. 9, pp. 59–67, Sept. 2008.
[2] E. Frech and C. Mesquida, "Cellular Models and Hand-Off Criteria," in *VETEC 1989*, vol.1, pp. 128-135, May 1989.
[3] J. Holtzman and A. Sampath, "Adaptive Averaging Methodology for Handoffs in Cellular Systems," *IEEE Trans. Veh. Technol.*, vol. 44, no. 1, pp. 59–66, Feb. 1995.
[4] K. Cornett and S. Wicker, "Bit Error Rate Estimation Techniques for Digital Land Mobile Radios," in *VETEC 1991*, pp. 543-548, May 1991.
[5] H. Mahmoud, I. Gu andvenc, and F. Watanabe, "Performance of Open Access Femtocell Networks with Different Cell-Selection Methods," in *VTC Spring 2010*, pp. 1-5, May 2010.
[6] L. Pack Kaelbling, M. Littman, and A. Moore, "Reinforcement Learning: A Survey," *Journal of Artificial Intelligence Research*, vol. 4, pp. 237–285, May 1996.
[7] J. Zhang and G. de la Roche, *Femtocells: Technologies and Deployment*. Wiley, 2010.
[8] M. Noh, Y. Lee, and H. Park, "Low complexity LMMSE channel estimation for OFDM," *Communications, IEE Proceedings*, vol. 153, no. 5, pp. 645 –650, Oct. 2006.
[9] J. C. H. Watkins and P. Dayan, "Technical note: Q-learning," *Machine Learning*, vol. 8, no. 279-292, 1992.
[10] D. R. Hundley, "Case Study on Learning," 2011. [Online]. Available: http://people.whitman.edu/~hundledr/courses/M472/
[11] FemtoForum, "Interference Management in OFDMA Femtocells," Mar. 2010. [Online]. Available: http://www.femtoforum.org