Temporal and Spatial Spectrum Assignment in Next Generation OFDMA Networks through Reinforcement Learning

Francisco Bernardo, Ramón Agustí, Jordi Pérez-Romero and Oriol Sallent
Signal Theory and Communications Department
Universitat Politècnica de Catalunya (UPC)
08034 Barcelona, Spain
Email: [fbernardo, ramon, jorperez, sallent]@tsc.upc.edu

Abstract—**This paper proposes a Dynamic Spectrum Assignment strategy in the context of next generation multicell Orthogonal Frequency Division Multiple Access networks. The proposed strategy is able to dynamically find spectrum assignments per cell depending on the spatial and temporal distribution of the users over the scenario. Reinforcement Learning methodology has been employed to implement the strategy, which compared with other fixed and dynamic spectrum assignment strategies shows the best tradeoff between spectral efficiency and Quality-of-Service.**

Index Terms—**Dynamic Spectrum Assignment, Multicell OFDMA, Reinforcement Learning.**

I. Introduction

The detected spectrum scarcity and its underutilization in current networks [1] claim for a new paradigm of spectrum access that overcomes current regulatory and technological barriers and promotes the usage of the spectrum dynamically and opportunistically by accounting for the different temporal and spatial spectrum demands [2]. OFDMA (Orthogonal Frequency Division Multiple Access) is a multiple access technique that is in the main stream of current proposed next generation broadband wireless systems (3G LTE, WiMax). It provides an extremely flexible radio interface, where the operation bandwidth is divided into small flat frequency response subchannels. Hence, this interface is suitable for current spectrum needs where the objective is to find a spectrum assignment over the radio interface that (a) improves the spectral efficiency and (b) adapts system spectrum to users' QoS requirements, taking into account the spatial and temporal variations of the network load.

Several approaches have been proposed so far to this problem. On the one hand, frequency reuse schemes based on a certain Frequency Reuse Factors (FRF) [3] try to mitigate intercell interference and in this way improve spectral usage at the cells' edge. Universal reuse (FRF1), where all cells share the same set of subchannels, increases system capacity at the cost that some users at the edge cannot be served due to excessive intercell interference. Thus, higher FRFs such as FRF3 or FRF7, where the entire bandwidth is distributed among clusters of 3 and 7 cells respectively, have been proposed. However, FRF3 and FRF7 schemes considerably reduce cell capacity. Additionally, the number of subchannels devoted to each FRF in each cell must be equal, which is not optimal for heterogeneous spatial distributions of the users [4]. In these heterogeneous scenarios, high flexibility of the spectrum management is preferred in order to properly distribute the subchannels over cells depending on each cell demands [5].

On the other hand, radio resource allocation strategies have been proposed with the objective of minimizing the total power consumption [6], maximizing overall system throughput [7], mitigate intercell interference [8] or guarantee users' QoS requirements [9]. However, some of these schemes [6][7] are proposed for single cell scenarios and they need high computational requirements and information exchange among users and cells to perform the subchannel/power/modulation/cell/user assignment in the short-term (due to instantaneous variations of subchannel conditions in multipath propagation environments). Others [8][9] introduce hierarchical schemes to circumvent these drawbacks. Thus, the controller of a cluster of cells decides which subchannels should be used by each cell under control, whereas in the short-term each cell independently decides how to schedule users' transmissions into available subcarriers regarding the channel status reported by the users. Nevertheless, users' QoS requirements [8] and intercell interference [9] are not considered in the optimization problem.

This paper presents a Dynamic Spectrum Assignment (DSA) strategy in a multicell OFDMA system that embraces the DSA problem not just at a cell level but at a network level by sharing the whole spectrum among different cells by means of a hierarchical architecture. Cognitive network functionalities like network observation, analysis, learning and decision making are introduced to provide the network

This work has been performed in the framework of the project E^3, which has received research funding from the Community's Seventh Framework programme. Also, the Spanish Research Council and FEDER funds under COGNOS grant (ref. TEC2007-60985) have supported this work. This paper reflects only the authors' views and the Community is not liable for any use that may be made of the information contained therein. The contributions of colleagues from E^3 consortium and the support of the Spanish Ministry of Science and Innovation via FPU grant AP20051165 are hereby acknowledged.

with the ability to automatically detect the instants when the current spectrum assignment is no longer valid and then dynamically find spectrum assignments per cell depending on the spatial and temporal distribution of the users over the scenario. To this end, we propose a DSA strategy based on Reinforcement Learning (RL) denoted here as DSA. The ability of RL to learn from interaction with the network is exploited to discover proper dynamic spectrum assignments of groups of contiguous OFDMA subcarriers or *chunks* to cells. Compared with classical fixed frequency assignments for cellular OFDMA planning, RL-DSA (*i*) improves spectral efficiency, (*ii*) maintains users' QoS satisfaction and, (*iii*) adapts spectrum to temporal and spatial variations of the network load. Thus the proposed algorithm demonstrates the best tradeoff between spectral efficiency and QoS.

Sections II and III describe the DSA system model and the RL-DSA algorithm, respectively. Section IV is devoted to detail the simulation model and section V to analyze obtained results. Finally, section VI presents final concluding remarks.

II. Dynamic Spectrum Assignment System Model

A hierarchical framework is proposed to manage spectrum dynamically and is depicted on Fig. 1. A generalized OFDMA radio interface is supposed where time is divided into frames and frequency into chunks. Hence, the minimum radio Resource Block (RB) assignable to users is a specific chunk into a frame. The cluster controller performs long-term cell-by-cell spectrum assignments that adapt to temporal and spatial variations of the spectrum demands depending on users' QoS and intercell interference. Additionally, the Short-Term Scheduler (STS) independently provides the short-term exploitation of multiuser diversity at each cell by dynamically assigning available radio resources to users.

As shown in Fig. 1(a), there is a *DSA controller* that is located in a network node with the ability to control a set of cells. The *RL Trigger* entity is in charge of executing the RL-DSA algorithm. The network variable status is observed and analyzed by this entity to detect the instants when the current spectrum assignment is no longer valid regarding the users' QoS performance as in the following.

Let us define $P_k^{T_{th}}$ as the average user dissatisfaction in cell k that reflects the percentage of time that the received throughput per user during one second is below a target throughput T_{th}.Then, the RL Trigger entity checks the following condition each L seconds period.

Condition 1:

```
if (($P_k^{T_{th}} > \delta^{up}$) OR ($P_k^{T_{th}} < \delta^{down}$)) then
    EXECUTE $RL\_DSA$ ALGORITHM
endif
```

Parameters δ^{up} and δ^{down} are thresholds to determine whether current assigned resources are insufficient or over-provisioned respectively.

Once the RL-DSA algorithm is executed, its intermediate actions (i.e., possible spectrum assignments) are applied to a *Network Characterization Entity* (NCE) that mimics the behavior of the real network based on real measurements. Inputs to NCE are the deployment and powers of base stations, the load per cell and the average pathloss of the users from serving and neighboring cells. These load and average pathloss measurements can actually be obtained from real networks and are acquired when the RL-DSA algorithm has to be run. As it is discussed in next section, RL-DSA converges to a solution based on the interaction with the NCE that returns a reward for a given action.

Finally, a *Decision Maker* analyses the status of the algorithm to stop the RL-DSA algorithm when it has converged and to decide the final chunk-to-cell assignment. Moreover, it implements the procedures to redeploy the new assignment in the real system. Details about the decision process are given in next section.

III. RL-DSA Algorithm

Consider N available chunks in a downlink OFDMA cellular system to distribute over K cells. Chunks are numbered

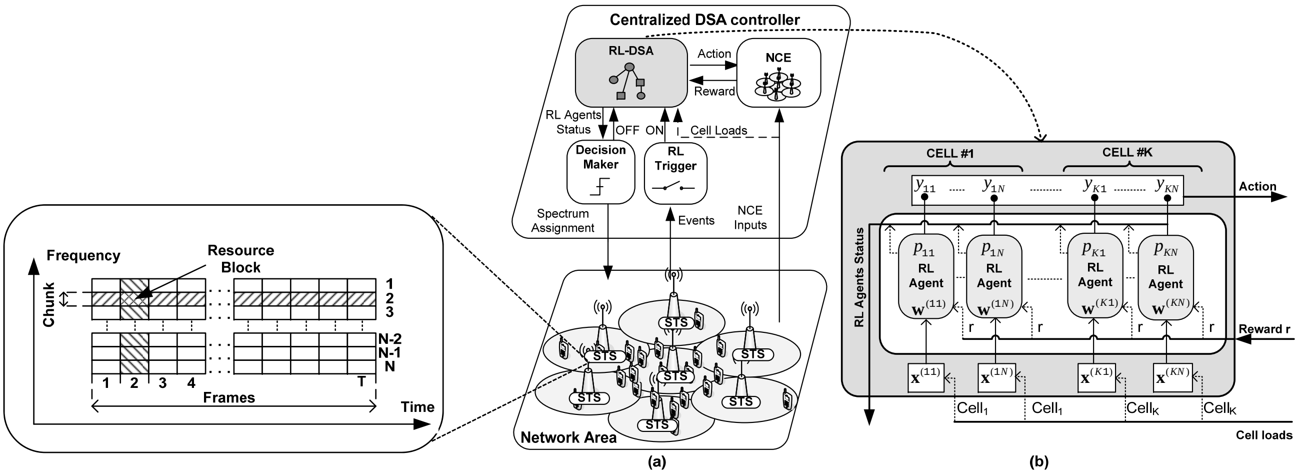

Fig. 1. Proposed DSM Model based on RL. (a) System Model, (b) RL-DSA Model.

from 1 to N and cells are numbered from 1 to K. The RL-DSA algorithm is based on the RL REINFORCE methods [10]. We propose a feed-forward network composed of KN agents to implement the RL-DSA algorithm (Fig. 1(b)) that interacts with the NCE in a step by step basis until convergence is reached. Notice that N RL agents are devoted to the k-th cell where each kn-th RL agent is devoted to learn whether the n-th chunk should be assigned to the k-th cell. For each single RL agent the output $y_{kn}\in\{0,1\}$ is a Bernoulli random variable where the agent's internal parameter p_{kn} contains its knowledge since represents the probability that the output y_{kn} is 1. Probability p_{kn} depends on an input x_{kn} and a weighting value w_{kn} as [10]

$$p_{kn}=\max\left\{\min\left\{\left(1+e^{-x_{kn}w_{kn}}\right)^{-1},1-p_{\exp}\right\},p_{\exp}\right\} \quad (1)$$

where probability $p_{\exp}$ is introduced as a small bias in order to enforce some exploratory behavior in the agent even if its internal probability is very near to 1 or 0.

The RL-DSA algorithm works as follows:

1) For each RL step t, it is considered that the n-th chunk is assigned to the k-th cell if the output $y_{kn}(t)$ is 1.
2) In next step, for each assignment the NCE returns a reward $r(t{+}1)$ (i.e, a numerical representation of the assignment suitability detailed below) that is used by the respective RL agents to update its internal weight as [10]

$$w_{kn}(t{+}1)=w_{kn}(t){+}\Delta w_{kn}(t{+}1), \quad (2)$$

$$\begin{aligned}\Delta w_{kn}(t{+}1) = \alpha(t{+}1)\cdot[r(t{+}1)-\bar{r}(t)]\cdot\\ \cdot[y_{kn}(t)-p_{kn}(t)]\cdot x_{kn}(t). \quad (3)\end{aligned}$$

   Notice that the learning from reward is enforced in the weighting value. Parameter $\alpha(t)$ is called the learning rate and, to improve convergence [11], it is linearly decreased as $\alpha(t){=}\alpha(t{-}1){-}\Delta$, where Δ is a factor that should be small enough to assure a smooth transition between steps. Moreover, the maximum number of steps (MAX_STEPS) must be set so that negative values of α are avoided. $\bar{r}(t)$ is the reinforcement baseline or average reward calculated using a exponential moving average with parameter β.
3) The agents capture the inputs $x_{kn}(t{+}1)$ that reflect the percentage of users in the kn-th cell so that the algorithm is able to adapt to homogeneous and heterogeneous spatial distributions of the users.
4) Probabilities $p_{kn}(t{+}1)$ and outputs $y_{kn}(t{+}1)$ are obtained for next assignment. If the decision maker detects that the variation of all p_{kn} between two successive steps is below ε in S steps, or $t > MAX\_STEPS$, then step 5) is executed. If not, next assignment is tested from 1).
5) Decision maker stops the RL-DSA algorithm. It decides that a chunk is assigned (not assigned) to a cell if p_{kn} is greater (lower) than 0.5.

The first time that RL-DSA is run full assignment is set, i.e.,$y_{kn}(0){=}1$ $\forall n,k$, and in subsequent executions RL-DSA algorithm begins from the assignment learnt in the previous execution so that the knowledge acquired until that moment is retained.

REINFORCE methods are characterized by low complexity and optimal behavior in the sense that a maximum reward is guaranteed in the long-term. Thus, a reward signal that captures the objective of maximizing the spectral efficiency per cell η_k and guarantees the satisfaction of the users has to be defined to assure the success of the proposed strategy. The reward signal r_k per cell k is defined as

$$r_k(t)=\begin{cases}0, & \text{if } th_k(t) < T_{th}\\ \lambda\eta_k(t), & \text{otherwise}\end{cases} \quad (4)$$

where $th_k(t)$ is the average user throughput for cell k at step t, $\eta_k(t)$ is the average spectral efficiency, λ is a positive weighting constant whose value should be high enough (around 100) in order to distinguish very similar spectral efficiencies, and T_{th} is the user satisfaction throughput threshold.

With the above definition, r_k reflects the spectral efficiency obtained in a cell if average user throughput is greater than the target. Otherwise RL-DSA does not receive any reward. Thus, it is assured that spectral efficiency is maximized only if QoS is fulfilled.

IV. Simulation Model

Results were obtained by means of dynamic simulations over a 19 hexagonal cells scenario representing a simulated time of 1 hour. At the beginning 190 users equally distributed among cells are considered moving at 3Km/h with a random walk model [12] (i.e., 10 active sessions per cell). Users always remain within their cell (i.e., handovers are not considered) and a full-buffer traffic model is considered. During a 10 minutes period between the minutes 25 to 35, 3 new sessions per minute are started in one of the cells and one session per minute is stopped in another cell (see Fig. 2). In this way, simulations consider both spatial and temporal variation of the traffic.

SINR for m-th user is computed for each chunk n (denoted as $\gamma_{m,n}$) considering distance dependant pathloss, shadowing

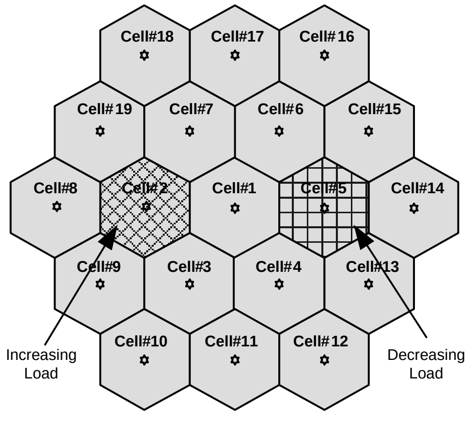

Fig. 2. Scenario layout

and frequency selective fast fading for both serving cell and interfering cell. Chunk power is constant and users' transmission bit rate is variable by means of Adaptive Coding and Modulation (ACM). The detailed SINR thresholds for each modulation and coding rate considered are given in Table I [13]. Then, the m-th user achievable bit rate for each chunk n, $(R_{m,n})$ is computed as

$$R_{m,n} = Bq(\gamma_{m,n}) \qquad (5)$$

where, B is the chunk bandwidth in Hz and $q(\gamma_{m,n})$ stands for the available spectral efficiency for a given SINR threshold obtained from Table I. Notice that the maximum net spectral efficiency is limited to a maximum value of $\eta_{\max} = 5$ bits/s/Hz corresponding to 64QAM modulation with a coding rate of 5/6. Finally, users transmissions are scheduled by STS following a Proportional Fair strategy [14] with an averaging window of 50 frames. Other simulation parameters are provided in Table II.

RL-DSA is triggered each time **Condition 1** holds for δ^{up}=10% and δ^{down}=0.1% for any cell, being T_{th}=256 kbps the user satisfaction throughput. RL-DSA performance is compared with classical Frequency Reuse Factors (FRF), such as FRF1 (full assignment) and FRF3 (1/3 of the band assigned to each cell), and a fully dynamic heuristic strategy [5] whose configuration parameters are also included in Table II. This strategy heuristically decides the number of chunks per cell based on its respective number of users.

TABLE I
MODULATION AND CODING SCHEMES

Modulation m (bits/s/Hz)	Coding Rate r (bits/s/Hz)	Spectral efficiency q (bits/s/Hz)	SINR threshold (dB)
2 (QPSK)	1/3	0.66	≥ 0.9
2 (QPSK)	1/2	1	≥ 2.1
2 (QPSK)	2/3	1.33	≥ 3.8
4 (16QAM)	1/2	2	≥ 7.7
4 (16QAM)	2/3	2.66	≥ 9.8
4 (16QAM)	5/6	3.33	≥ 12.6
6 (64QAM)	2/3	4	≥ 15.0
6 (64QAM)	5/6	5	≥ 18.2

TABLE II
SIMULATION PARAMETERS

Number of cells	$K = 19$
Cell Radius	$R = 500$ meters
Antenna patterns	Omnidirectional
Frame time	2 ms
Number of chunks	$N = 12$
Chunk bandwidth	$B = 375$ KHz
Power per chunk	$P = 33$ dBm
Path loss in dB at d km	$128.1 + 37.6log10(d)$ [12]
Shadowing standard deviation	8 dB [12]
Shadowing decorrelation distance	5 m [12]
Small Scale Fading model	ITU Ped. A [12]
UE thermal noise	-174 dBm/Hz
UE noise factor	9 dB
RL parameters $[\alpha, \beta, \Delta]$	$\left[10, 0.01, 10^{-5}\right]$
Exploratory probability	$p_{exp} = 0.1\%$
Reward constant λ	100
RL convergence criterion $[\varepsilon, S]$	$\left[10^{-4}, 5000\right]$
MAX_STEPS	100000
RL-Trigger period	$L = 30$ seconds

V. RESULTS

Results are presented in terms of spectrum utilization, dissatisfaction probability and spectral efficiency in Fig. 3, Fig. 4 and Fig. 5 respectively.

Fig. 3 illustrates the number of assigned chunks for each one of the studied strategies. Average number of assigned chunks over all cells in the scenario as well as in the most and least loaded cell is represented. As expected, RL-DSA and DSA-heuristic as dynamic strategies adapt the number of chunks per cell depending on traffic demands. However, RL-DSA adapts more progressively the spectrum proving the validity of the proposed system architecture for the RL-DSA algorithm by detecting the instants when RL-DSA has to be run during temporal changes in the scenario. The instants where RL-DSA is executed are plotted in the figure as vertical arrows. Notice that, after each execution a spectrum modification (i.e., variation of the number of assigned chunks to cells) is enforced. In this way, increasing load cells have also an increasing number of chunks and vice versa. Finally, in all cases, RL-DSA is the strategy that minimizes the number of chunks per cell. This feature translates in an improved spectral efficiency without compromising dissatisfaction probability as it is shown in the following.

Fig. 4 shows the average dissatisfaction probability per cell for the studied strategies. Again, this performance is represented for all cells in the scenario (average) and for the most and least loaded cells. It can be seen that RL-

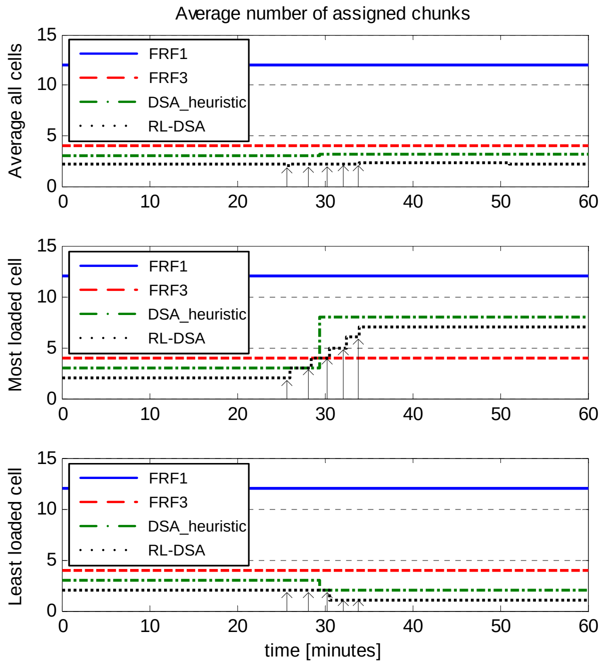

Fig. 3. Average number of assigned chunks

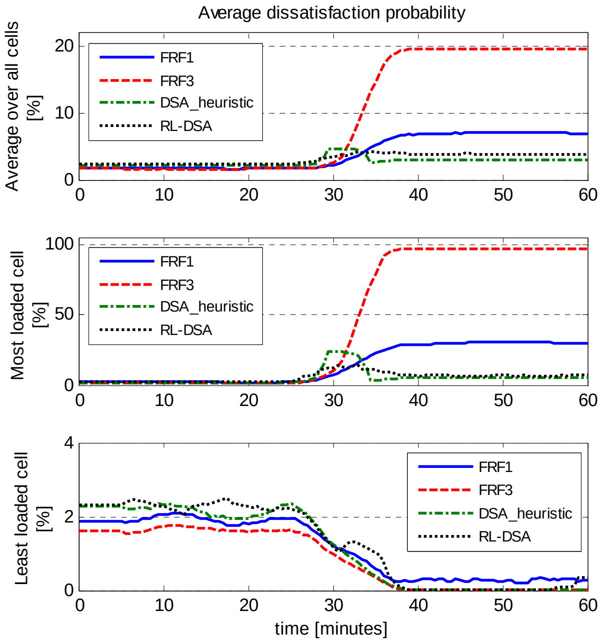

Fig. 4. Average dissatisfaction probability per cell

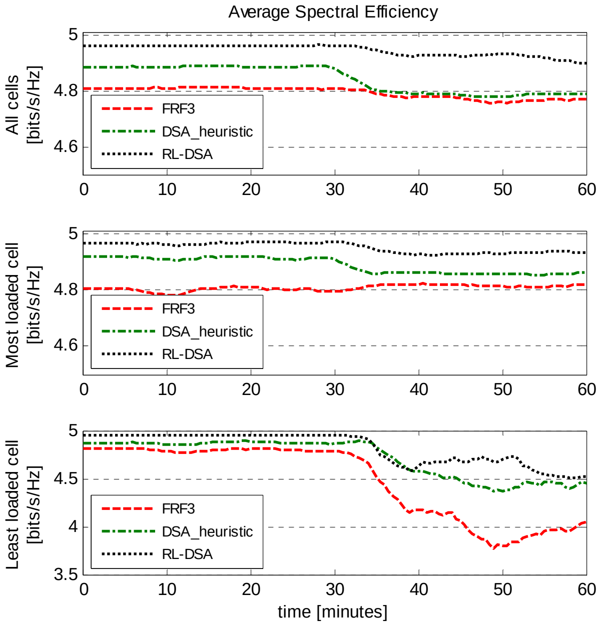

Fig. 5. Average spectral efficiency per cell

DSA maintains dissatisfaction probability below 5% in all cases and improvements near to 95% can be obtained with respect to FRF3 in the most loaded cell. Fixed strategies demonstrate poor performance for heterogeneous distribution of the load (i.e., after minute 35) since they are not capable of re-adapting their spectrum whereas DSA-heuristic strategy also has a reasonable dissatisfaction probability but performs worse than proposed RL-DSA algorithm in terms of spectral efficiency as shown in Fig. 5.

RL-DSA obtains the best spectral efficiency results as depicted in Fig. 5. FRF1 performance is not represented in those plots since its spectral efficiency maintains below 4 bits/s/Hz, showing that, due to excessive intercell interference, spectrum cannot be fully exploited by users. RL-DSA achieves a spectral efficiency near the maximum of 5 bits/s/Hz. Notice that for the least loaded cell a reduction of the spectral efficiency is reported after minute 35 for all strategies. This is because only one session is active in this cell and, then, multiuser diversity cannot be exploited. Also in this situation RL-DSA achieves the best spectral efficiency since it is the strategy that assigns only one chunk to this cell.

VI. Conclusion

In this paper a framework for Dynamic Spectrum Assignment (DSA) in next generation OFDMA-based networks has been proposed, as well as a RL-based algorithm whose foundations reside on an optimal RL methodology called REINFORCE. Compared with other fixed spectrum planning and dynamic strategies, the proposed algorithm demonstrates the best tradeoff between spectral efficiency and QoS fulfillment thanks to an adequate adaptability to temporal and spatial variations of the spectrum demands.

References

[1] J. A. Hoffmeyer, “Regulatory and standardization aspects of DSA technologies - global requirements and perspective,” in *IEEE New Frontiers in Dynamic Spectrum Access Networks (DySPAN)*, 2005, pp. 700–705.

[2] M. M. Buddhikot, “Understanding dynamic spectrum access: Models, taxonomy and challenges,” in *IEEE New Frontiers in Dynamic Spectrum Access Networks (DySPAN)*, 2007, pp. 649–663.

[3] W. Zhaocheng and R. Stirling-Gallacher, “Frequency reuse scheme for cellular OFDM systems,” *Electronics Letters*, vol. 38, no. 8, pp. 387–388, 2002.

[4] H. Kim, Y. Han, and J. Koo, “Optimal subchannel allocation scheme in multicell OFDMA systems,” in *IEEE 59th Vehicular Technology Conference VTC 2004-Spring*, vol. 3, 2004, pp. 1821–1825.

[5] F. Bernardo, R. Agustí, J. Pérez-Romero, and O. Sallent, “Dynamic spectrum assignment in multicell OFDMA networks enabling a secondary spectrum usage,” *Accepted to Wireless Communications and Mobile Computing (WCMC)*, 2009.

[6] S. Pietrzyk, *OFDMA for broadband wireless access.* Artech House, 2006.

[7] K. Kim, H. Kim, and Y. Han, “Subcarrier and power allocation in OFDMA systems,” in *60th IEEE Vehicular Technology Conference-Fall*, vol. 7, 2004, pp. 1058–1062.

[8] G. Li and H. Liu, “Downlink radio resource allocation for multi-cell OFDMA system,” *IEEE Transactions on Wireless Communications*, vol. 5, no. 12, pp. 3451–3459, December 2006.

[9] H. Kwon, W.-I. Lee, and B. G. Lee, “Low-overhead resource allocation with load balancing in multi-cell OFDMA systems,” in *IEEE 61st Vehicular Technology Conference VTC 2005-Spring*, vol. 5, May-1 June 2005, pp. 3063–3067.

[10] R. J. Williams, “Simple statistical gradient-following algorithms for connectionist reinforcement learning,” *Machine Learning*, vol. 8, no. 3, pp. 229–256, May 1992.

[11] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. The MIT Press, March 1998.

[12] 3GPP, “Physical layer aspects for evolved universal terrestrial radio access (UTRA),” 3GPP, Tech. Rep. TR 25.814 v7.1.0, September 2006, release 7.

[13] R. Schoenen, R. Halfmann, and B. Walke, “MAC performance of a 3GPP-LTE multihop cellular network,” in *IEEE International Conference on Communications ICC’08*, May 2008, pp. 4819–4824.

[14] C. Wengerter, J. Ohlhorst, and A. v. Elbwart, “Fairness and throughput analysis for generalized proportional fair frequency scheduling in OFDMA,” in *IEEE 61st Vehicular Technology Conference 2005-Spring*, vol. 3, 2005, pp. 1903–1907.