Technical Correspondence

Intercell Interference Management in OFDMA Networks: A Decentralized Approach Based on Reinforcement Learning

F. Bernardo, *Student Member, IEEE*, R. Agustí, *Member, IEEE*, J. Pérez-Romero, *Member, IEEE*, and O. Sallent

***Abstract*—This paper presents a decentralized framework for dynamic spectrum assignment in multicell orthogonal frequency division multiple access (OFDMA) networks. The proposed framework allows each cell to autonomously decide the frequency resources it should use through a procedure that incorporates concepts from self-organization and machine learning in multiagent systems (MASs). Simulation results have been obtained for several scenarios, including both macrocells (MCs) and femtocells (FCs), revealing important improvements in terms of spectral efficiency and intercell interference mitigation over reference approaches, and close performance with the one obtained by a centralized strategy. Results also suggest that the framework would be practical for future FC cellular deployments where a high degree of independence of the network nodes is expected to reduce operational costs.**

***Index Terms*—Cellular mobile networks, intercell interference, multiagent systems (MASs), orthogonal frequency division multiple access (OFDMA), reinforcement learning.**

I. Introduction

Current mobile cellular networks are difficult to manage and require a lot of human interaction. For example, tasks such as assigning spectrum resources to macrocells (MCs), i.e., operator-deployed cells of wide-range coverage areas, are carried out off-line during network deployment. Then, the spectrum assignment remains unaltered until new infrastructure is added to the system, and a tedious manual frequency planning is repeated. This is also the case for next-generation mobile cellular networks such as Third-Generation Partnership Plan (3GPP) long-term evolution (LTE) or IEEE 802.16e (Mobile WiMax), whose downlink radio access network is based on orthogonal frequency division multiple access (OFDMA). Such an interface divides a broad frequency band into small bandwidth frequency resources named *chunks* [1] that have to be allocated to the different cells.

When performing frequency planning, the network operator aims to cover the maximum demand anyplace within the service area while reducing the *intercell interference* (i.e., the interference that two or more neighboring cells using the same frequency resource cause each other). Usually, different frequency reuse factors (FRF) [2] are deployed, where the spectrum is distributed among MCs following a static regular pattern. Additionally, it has been showed [3] that global network performance can be improved by fragmenting spectrum according to the distance of the users to MCs. Then, several proposals to arrange the division between a central and an edge subband have been proposed, where one of the most representatives is called partial-frequency reuse (PR) [4]. However, either FRF or PR spectrum assignment strategies can be clearly inefficient with variable traffic demands [5], leading to underutilization of the spectrum in some cells and lack of spectrum resources in others.

Moreover, current trend in cellular networks is to decentralize the spectrum assignment tasks [6], [7]. This has been mainly motivated by 1) the tendency to enhance the radio resource management capabilities in MCs in order to reduce operational costs of the network and 2) the advent of new cellular deployments based on femtocells (FCs). These are small range user-deployed base stations introduced at a considerable amount of random locations [8] that require a high degree of autonomy due to their independent nature. Then, spectrum management in OFDMA FCs is a challenging task where different possibilities are open [9], especially when FCs are operating under the coverage of an MC deployment. Therefore, orthogonal division of the spectrum can be made to avoid cross-layer interference (i.e., that interference between MCs and FCs). In that arrangement, MCs and FCs operate in different parts of the spectrum to avoid cross-layer interference but at the cost of reducing the available capacity as well. On the other hand, the authors of [10] proposed a hybrid scheme where, depending on the position of the FC, it accesses to an orthogonal or cochannel part of the spectrum (i.e., the MC layer and the FC layer could share the same spectrum band, so that cross-layer intercell interference could arise). However, this scheme needs to locate the FCs within the MC and from explicit communication between FCs and MCs. Finally, FC spectrum management with regard to other FCs is also an important task because in scenarios with a high density of FCs, interference between them cannot be neglected due to their proximity. In this case, a reference approach introduced in [9], called FRS_x, consists of dividing the available spectrum into x equal portions so that the FCs randomly select one of them to operate after switch on. Then, the spectrum management is simple and autonomous, but it can be far from optimal.

Centralized spectrum assignment strategies concentrate on the decision tasks in a centralized controller with global knowledge of the status of the cellular network. On the other hand, decentralized schemes provide 1) flexibility because they can adapt to a vast variety of scenarios without having to deploy a centralized architecture; 2) scalability, maintaining constant the signaling and computational requirements with an increment of the number of cells; and 3) robustness, since the decentralized strategy avoids the need of a centralized controller which constitutes a single point of failure. However, there are open questions regarding decentralized management such as how close the performance of the decentralized system can be to the one in the centralized counterpart (having in mind that each entity has only a partial knowledge of the status of the cellular network) or how the different entities self-organize to achieve a stable solution.

From an operator's perspective, it would be desirable that a decentralized strategy for spectrum assignment includes automatic mechanisms in the network procedures. Then, the spectrum assignment would be dynamically adapted to temporal and spatial traffic load needs at the same time that operational and capital expenditures reductions are achieved [11]. In this context, machine learning appears as a potential

Manuscript received May 17, 2010; revised August 11, 2010, and November 23, 2010; accepted December 1, 2010. Date of publication January 13, 2011; date of current version October 19, 2011. This work was supported in part by the Community's Seventh Framework Programme under Project E^3, in part by the Spanish Research Council under Cognitive Management of Radio Resources and Spectrum in Heterogeneous Mobile Networks with End to End Quality of Service Provision (COGNOS) Grant TEC2007-60985, and in part by the European Regional Development Funds. This paper was recommended by Associate Editor B. Chaib-Draa.

F. Bernardo is with the Department of Signal Theory and Communications, Universidad de Sevilla, Seville 41092, Spain (e-mail: fbernardo@tsc.upc.edu).

R. Agustí, J. Pérez-Romero, and O. Sallent are with the Universitat Politècnica de Catalunya, Barcelona 08034, Spain (e-mail: ramon@tsc.upc.edu; jorperez@tsc.upc.edu; sallent@tsc.upc.edu).

Digital Object Identifier 10.1109/TSMCC.2010.2099654

approach to implement cognitive and autonomic procedures in each of the entities of the system [12].

In this paper, we approach the decentralized spectrum assignment problem as a learning task in a multiagent system (MAS) [13]. This approach is useful in scenarios where several entities pursue different, even contradictory goals. This is the case of the spectrum assignment problem in cellular systems, where intercell interference supposes a conflict between cells. Here, each cell behaves as an autonomous agent that executes a reinforcement learning (RL) algorithm to decide its best spectrum assignment based on context information obtained from their surroundings. This context information includes partial knowledge of the spectrum assignment choice taken by other cells so that, although the learning is localized in each cell, it is influenced by other agents' decisions, constituting a form of learning in the MAS [14]. Other approaches consider a global common learning for the agents in the MAS [15] or the existence of a coordination entity that entails the learning task for a set of agents [16], but it is an open question whether local or global learning is better depending on the specific problem to be solved [14].

Work in this paper extends our previous work in [17] that presented a preliminary version of the decentralized approach. However, work in [17] did not consider future advances in mobile networks such as the possibility of intercell signaling interfaces between cells or the channel-aware packet scheduling strategies. On the other hand, two novel strategies for the decentralized spectrum assignment are presented in this paper, named *communicative* and *noncommunicative spectrum assignment schemes*, respectively, depending on whether there exists communication between adjacent cells or not. Moreover, the reinforcement learning-dynamic spectrum assignment (RL-DSA) strategy presented in [18] has been adapted to be independently executed by each cell to determine its used spectrum. In fact, work in [18] focused on a centralized approach where a single node entails the spectrum assignment procedure for a set of cells. Clearly, this approach could become intractable with the advent of new cellular deployments based on FCs. Finally, simulation results have been extended in this paper to be obtained in three case studies including a typical 19 MCs scenario, a scenario with up to 100 FCs, and a combined MC and FC scenario, where finding a spectrum assignment to reduce intercell interference is quite challenging (i.e., intercell interference could appear between MCs, between FCs, and between MCs and FCs). The proposed scheme exhibits a superior performance to classical spectrum assignment strategies, and a close performance to its centralized version taking into account that cells in the decentralized approach handle limited information compared with the centralized controller. Also, the better scalability of the decentralized approach with respect to the centralized counterpart has been qualitatively analyzed.

In the following, Section II presents the operation of the proposed framework. Then, Section III presents the *Cell DSA controller*, which is the functional block in each cell where the RL-DSA is executed. There, a detailed description of internal procedures of the controller is given. Section IV is devoted to present the simulation model and the results for three case studies in this paper. Final conclusion is stated in Section V.

II. Decentralized Framework Description

We consider a decentralized framework (see Fig. 1) where each cell (MC or FC) constitutes an independent agent that performs autonomous spectrum assignment decisions with the objective of improving cell's SINR (signal to interference plus noise ratio) while guaranteeing cell users' quality of service (QoS). In the following, we focus on a single cell. Assume that the cell has U users

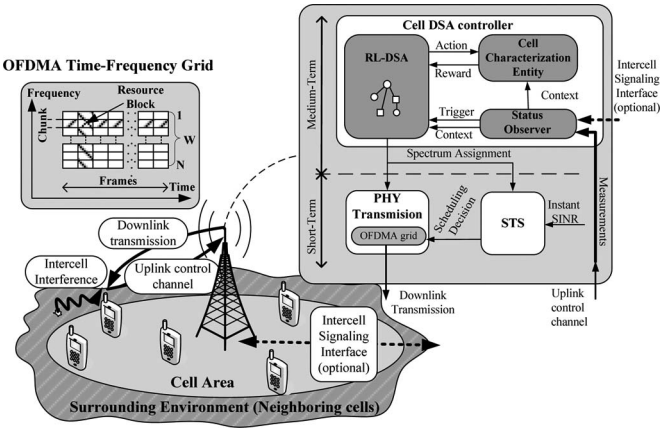

Fig. 1. Proposed decentralized framework for spectrum assignment.

$\{1, \ldots, u, \ldots, U\}$. A generalized OFDMA radio interface is considered in downlink for users' data transmission, where a common system bandwidth W for the service area is divided into N *chunks* $\{1, \ldots, n, \ldots, N\}$. Each chunk is a group of contiguous OFDM subcarriers with bandwidth $B = W/N$ Hz. Moreover, time is divided into *frames*. The minimum radio resource block assignable to users is one chunk per frame. On the other hand, there is an uplink control channel where users send instantaneous (frame-by-frame) *measurements report* messages from which the SINR in the different chunks can be obtained. These reports are necessary to perform the link adaptation for each established downlink communication in terms of an adequate modulation and coding rate scheme for a given SINR in a given frame.

The operation of a cell is as follows. In the short term (i.e., at the frame-time scale), the cell handles users' traffic and performs OFDMA fast link adaptation following a channel-aware strategy such as proportional fair (PF) [19]. On the other hand, spectrum assignment is done in the medium term (i.e., from tens of seconds to tens of minutes). More precisely, each cell tries to learn the best spectrum assignment by executing the RL-DSA algorithm (see the next section) in an *execution period* of L frames. It is important to remark that RL-DSA is run in each cell assuming that the spectrum assignment in adjacent cells is not varying during its execution to assure the self-organization of the MAS to a stable solution. Notice that if the periods were aligned between cells (i.e., all cells execute RL-DSA simultaneously), then adjacent cells would be simultaneously varying their respective spectrum assignments, and thus, convergence to a stable solution could be compromised. Trying to avoid this, we consider that, after switch-on, a cell randomly selects an initial time to execute for the first time, and then the execution period L is followed from the initial time. Hence, since large values of L are expected for a medium-term execution of RL-DSA, the probability that adjacent cells select the same initial time becomes negligible. For instance, we have taken in this paper $L = 60\,000$ frames, what leads to a probability of simultaneous execution of around 10^{-4} assuming a value of six adjacent cells. In the case of simultaneous execution, this could be easily detected, e.g., due to consecutive performance degradation in the cell, and then simply another initiation time for the cell would be randomly selected.

In order to perform a reliable autonomous spectrum assignment decision, the cell should estimate the spectrum usage of adjacent cells to calculate the potential intercell interference and then their own chunks' capacities. We consider two ways of obtaining such information: 1) a *communicative spectrum assignment* scheme, where cells explicitly

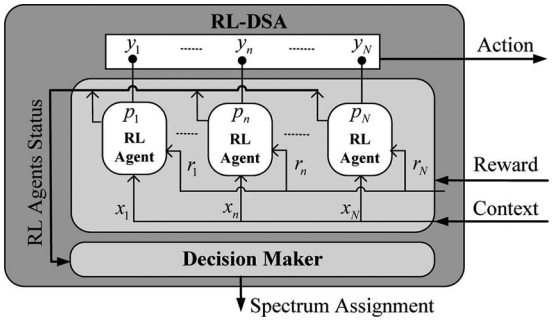

Fig. 2. RL-DSA functional scheme.

exchange the current spectrum assignment through an intercell interface; and 2) a *noncommunicative spectrum assignment* scheme, where the spectrum usage of adjacent cells is estimated from users' measurements reports. The communicative spectrum assignment scheme is only possible when a signaling interface between cells is available (e.g., the *X*2 interface in MC scenarios [20]). On the other hand, the noncommunicative spectrum assignment scheme is possible in most of the scenarios as long as there are measurement reports from users. This is especially useful for FC scenarios, where the random location of FCs makes challenging an intercell signaling interface.

III. Cell Dynamic Spectrum Assignment Controller

The aim of the *cell DSA controller* shown in Fig. 1 is to find a suitable spectrum assignment that improves the SINR of the cell while the QoS of the users in terms of a minimum throughput is assured. Its functional blocks *status observer*, *RL-DSA algorithm*, and *cell characterization entity* (CCE) are explained next.

A. Status Observer

The status observer entity in each cell is responsible for triggering the RL-DSA algorithm following periods of L frames and appropriately selecting the initial execution frame. In addition, the status observer entity collects and builds necessary inputs for the spectrum assignment decision task. Hence, local measurements to the cell are averaged over a *measurement averaging period* of l frames, where $l \ll L$ to favor that inputs at the time of execution reflect the latest up-to-date measurements. Then, the status observer provides to the RL-DSA algorithm and the CCE (see Section III-C) the execution context composed by 1) the average number of users in the cell, which are denoted $\bar{U}$, and 2) information regarding spectral usage in nearby cells. In the communicative scheme, this latter information is the spectrum assignment of adjacent cells explicitly retrieved from the intercell signaling interface. In turn, in the noncommunicative scheme this information is implicitly included in the probability density function (pdf) of the average SINR for each chunk $\bar{\gamma}_n$, which is denoted $s_{\bar{\gamma}_n}(\bar{\gamma}_n)$.

B. RL-DSA

The RL-DSA algorithm performs the decision task to select an appropriate spectrum assignment for the cell. The inherent optimization behavior of RL over a reward signal is exploited by RL-DSA to dynamically find proper spectrum assignments in the cell depending on current traffic load variations. Also, learning is retained to be exploited in subsequent spectrum assignment tasks. RL-DSA functional architecture is shown in Fig. 2. A feedforward network composed of N RL-agents is used to implement the RL-DSA algorithm in each cell, where the $n\text{th}$ RL-agent in the feedforward network is devoted to learn whether the $n\text{th}$ chunk should be assigned to the cell or not. The learning procedure given in [18] has been adopted, which is briefly described next.

RL-DSA interacts with the CCE on a step-by-step basis to obtain a reward. In each RL step t, the action chosen by the RL system to interact with the CCE is a binary assignment vector $\Upsilon(t) = (y_1(t), y_2(t), \ldots, y_N(t))$ that contains the chunk-to-cell assignment so that it is considered that the $n\text{th}$ chunk is assigned to the cell if the output $y_n(t)$ is 1 (and not assigned otherwise). It is worth to remark that outputs are Bernoulli random variables that depend on an internal probability $p_n(t)$ corresponding to the probability that the output $y_n(t)$ is equal to 1 at a given RL step. This random nature and the learning algorithm allow RL-DSA to explore the solution space in a direction where reward is globally maximized [18], [21]. Specifically, in [21], the general REINFORCE methods, being the algorithm proposed in this paper as a specific case, have been proven to converge to a global maximum of reward signal. Moreover, we set the inputs $x_n(t) = \bar{U}$ for all n and t. This input remains constant during an RL-DSA execution so that RL-DSA is able to associate solutions to different loads of the cell. Then, RL-DSA maximizes a reward signal which is appropriately defined in terms of the average SINR in the cell and the QoS of users' communications. It is considered that the transmitted power per chunk is constant for a given cell, so that an increment in the SINR for a given distribution of the users in the cell is equivalent to a reduction of the interference from other cells. Finally, *Decision Maker* stops RL-DSA after a sufficiently high number of steps MAX_STEPS. Then, it examines RL-DSA status, which is a vector with agents' internal action–selection probabilities $p_n(t)$ for $n = 1, 2, \ldots, N$, and assumes for the new spectrum assignment that a chunk has to be used in the cell if $p_n(t) > 0.5$ and not assigned otherwise. We showed in [18] that RL-DSA tends to converge to an optimal solution for a reduced number of steps compared with the size of the space solution.

For each action of RL-DSA, CCE returns a reward signal $r \in \mathbb{R}$ that is common for all RL-agents in a given cell so that all agents are involved in the same goal oriented problem [22]. Notice that in essence, RL-DSA involves learning in a MAS, and thus, having a common reward indirectly makes the actions of each RL-agent dependent on actions taken by other RL-agents in that cell. This fact, jointly with the random nature of the outputs of RL-agents, provides adequate coordination between them to converge to a stable solution. The aim of RL-DSA algorithm is to maximize a reward signal $r(t)$. To give a physical meaning to the learning procedure, the reward signal should be linked to a system performance metric. Since the Cell DSA Controller in this paper aims at maximizing cell's SINR assuring a given QoS in terms of a minimum average user throughput, $r(t)$ is defined as

$$r(t) = \begin{cases} 0, & \text{if } \widehat{th}(t) < th_{\text{target}} \\ \hat{\gamma}(t), & \text{otherwise} \end{cases} \tag{1}$$

where $\hat{\gamma}(t)$ is the estimated average SINR in the cell and $\widehat{th}(t)$ is the estimated average user throughput for the cell. Notice that the reward signal is zero for the cell if the average user throughput is below a given *user satisfaction throughput target* th_{target}, so that the reward signal retains QoS constraints. Otherwise, RL-DSA maximizes the SINR. This reward signal is built by CCE for each action of RL-DSA in a cell. Thus, coordination among different cells results from the fact that reward signal captures the SINR resulting from the actions of the other cells. Notice that this reward is nonstationary since different rewards can be obtained for the same action in different executions of RL-DSA depending on the current status of the cell and its environment (i.e., the spectrum assignment chosen by other cells in the surroundings). In the following, we describe how CCE can estimate $\hat{\gamma}(t)$ and $\widehat{th}(t)$.

C. Cell Characterization Entity

The CCE constitutes the environment for the RL-DSA and tries to mimic the response of the cell for a given spectrum assignment. It is worth noticing that the accuracy of the estimation could affect the accuracy of the proposed solution. However, as results in Section IV will reveal, the models given here are adequate since the proposed solution by RL-DSA certainly improves performance over the real network.

For each candidate spectrum assignment $\Upsilon(t)$ given by RL-DSA in step t, the CCE returns the reward value reflecting the suitability of each action as given by (1). Hence, the reward signal should be estimated for each action. To this end, CCE should compute estimations of the average SINR in the cell $\hat{\gamma}(t)$ and the average user throughput $\widehat{th}(t)$ for a certain spectrum assignment selected by RL-DSA in a given step. We propose two estimation methods depending on whether cooperation exists between cells or not.

1) Communicative Spectrum Assignment Scheme: In the communicative spectrum assignment scheme, cells exchange the spectrum assignment so that a given cell knows the spectrum assignment in adjacent cells. Thus, it can compute the set of cells $\Phi_n(t)$ that cause interference to each chunk n in a given RL step. In this paper, the communicative spectrum assignment is only considered for an MC scenario where an intercell signaling interface is feasible and cell's deployments are controlled by the operator. Then, let us assume an interference limited MC scenario with omnidirectional antennas (i.e., noise can be neglected) and uniformly distributed users per cell. The average signal to interference ratio (SIR) per chunk in the cell for a given candidate spectrum assignment in an RL step can be estimated as

$$\hat{\gamma}_n(t) = \iint_A \frac{1}{A}\mathrm{SIR}(\Phi_n(t), \rho, \theta)\rho d\,\rho d\,\theta \tag{2}$$

where $\mathrm{SIR}(\Phi_n(t), \rho, \theta)$ is the SIR at a given point (ρ, θ) of the cell in polar coordinates. Hence, (2) averages the SIR for all points in the actual area A covered by the cell. Considering that any interfering cell j is located, in polar coordinates, at a point (d_j, ϕ_j) with respect to the reference cell, $\mathrm{SIR}(\Phi_n(t), \rho, \theta)$ is written as

$$\mathrm{SIR}(\Phi_n, \rho, \theta) = \frac{P_n K_{PL}\rho^{-\chi}}{\sum\limits_{j\in\Phi_n} P_n K_{PL}\rho_j(\rho,\theta)^{-\chi}} \tag{3}$$

for $\rho_{\min} \le \rho \le \rho_{\max}$ and $0 \le \theta < 2\pi$. $\rho_{\min} > 0$ is the minimum distance between users and the base station due to base station antenna height, and $\rho_{\max}$ is the maximum distance to the base station in the cell's coverage area. Constant chunk power P_n is assumed for all chunks and pathloss is modeled as $K_{\mathrm{PL}}\rho^{-\chi}$, being K_{PL} a pathloss constant and χ the pathloss exponent dependent on the scenario. Since we are interested in average results in the medium term, slow and fast varying fading has not been considered in (3). The distance between the interfering cell and the point of interest in the reference cell can be written as $\rho_j(\rho,\theta) = \sqrt{d_j^2 + \rho^2 - 2\rho d_j \cos(\theta - \phi_j)}$. Then, (3) can be reduced to

$$\mathrm{SIR}(\Phi_n, \rho, \theta)^{-1} = \sum_{j\in\Phi_n}\left(1 + d_j^2/\rho^2 - 2d_j\cos(\theta-\phi_j)/\rho\right)^{-0.5\chi}. \tag{4}$$

Finally, we average $\hat{\gamma}_n(t)$ for all assigned chunks to the cell in a given RL step:

$$\hat{\gamma}(t) = \frac{1}{|C(t)|}\sum_{n\in C(t)}\hat{\gamma}_n(t) \tag{5}$$

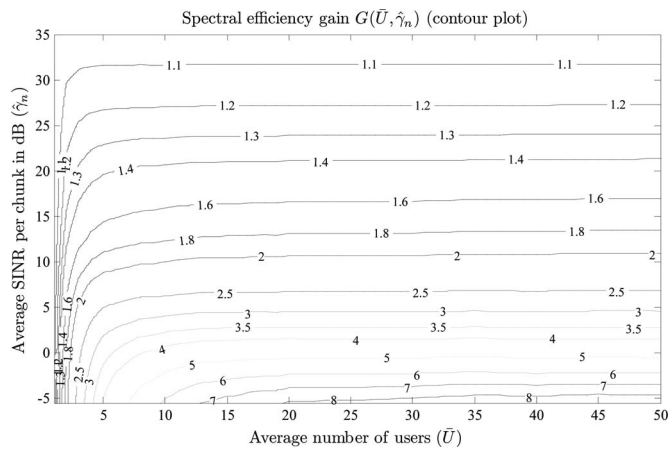

Fig. 3. Spectral efficiency gain versus average SINR and number of users.

where $C(t)$ is the set of chunks assigned to the cell by the RL-DSA for a given action (i.e., $n \in C(t)$ if $y_n(t) = 1$), and consequently, $|C(t)|$ stands for the number of assigned chunks.

On the other hand, the average user throughput in the cell can be obtained as

$$\widehat{th}(t) = \frac{B\,|C(t)|\,\hat{\eta}(t)}{\bar{U}} \tag{6}$$

where B denotes the chunk bandwidth, $\bar{U}$ is the average number of users in the cell, and $\hat{\eta}(t)$ is an estimation of the average cell spectral efficiency for a given spectrum assignment. Similar to $\hat{\gamma}(t)$, $\hat{\eta}(t)$ can be obtained as

$$\hat{\eta}(t) = \frac{1}{|C(t)|}\sum_{n\in C(t)} G(\bar{U}, \hat{\gamma}_n(t)) \times \iint_A \frac{1}{A} q(\mathrm{SIR}(\Phi_n(t), \rho, \theta))\rho d\rho d\theta \tag{7}$$

where $q(\mathrm{SIR}(\Phi_n(t), \rho, \theta))$ is the spectral efficiency in bits/(s Hz) for a given value of the SIR. For instance, the function q can be the mapping table given in [1, Tab. 8.1]. $G(\bar{U}, \hat{\gamma}_n(t))$ is a gain factor that captures the multiuser diversity features [1] of the short-term scheduling strategy used in the cell as a function of the average number of users $\bar{U}$ and the average estimated SINR per chunk $\hat{\gamma}_n(t)$. In particular, Fig. 3 shows the gain factor obtained for a PF scheduler from simulations (simulation details are in Table I). Notice that spectral efficiency gain tends to be constant for numbers of users above a certain value because multiuser diversity cannot be further exploited (i.e., after reaching a certain number of users, it is always very probable that a user with a good channel will be found).

2) Noncommunicative Spectrum Assignment Scheme: In the case of noncommunicative spectrum assignment scheme, the average SINR and user throughput are estimated from measurements, because the spectrum usage of adjacent cells is unknown. Let $s_{\bar{\gamma}_n}(\bar{\gamma}_n)$ be the average SINR pdf for the chunk n computed by status observer from users' measurement reports. Then, the average cell SINR per chunk is obtained as

$$\hat{\gamma}_n(t) = \int_{-\infty}^{\infty}\bar{\gamma}_n s_{\bar{\gamma}_n}(\bar{\gamma}_n)d\bar{\gamma}_n \tag{8}$$

and hence, the average cell SINR is computed as

$$\hat{\gamma}(t) = \frac{1}{|C(t)|}\sum_{n\in C(t)}\hat{\gamma}_n(t). \tag{9}$$

TABLE I
SIMULATION PARAMETERS

Default parameters	
Frame time	2 ms
Chunk bandwidth [B]	375 kHz
Number of chunks [N]	24 chunks
UE thermal noise	-174 dBm/Hz
UE noise factor	9 dB
Short Term Scheduling method	Proportional Fair [19]
PF Averaging window	50 frames
MC parameters	
Cell Radius	500 m
Minimum distance to BS	35m
Antenna Pattern	Omnidirectional
Power per chunk	30 dBm
Path Loss at d Km in dB	$128.1+37.6\log_{10}(d)$
Shadowing standard deviation	8 dB
Shadowing decorrelation distance	5 m
Small Scale Fading Model	ITU Ped. A
FC parameters	
Cell Radius	20 m
Minimum distance to BS	1m
Antenna Pattern	Omnidirectional
Power per chunk	-3.8 dBm
Path Loss at d m in dB	$37+30\log_{10}(d)$
Shadowing standard deviation	8 dB
Small Scale Fading Model	ITU Ped. A
External wall penetration loss	15 dB
DSA parameters	
Measurements averaging period [l]	2500 frames
RL-DSA execution period [L]	60000 frames
RL-DSA [18] parameters [α , β , σ , Δ]	[100, 0.01, 0.05, 10^{-6}]
RL-DSA exploratory probability [p_{explore}]	0.1%
RL-DSA steps [MAX_STEPS]	1000000
Margin factor [24]	3.0

On the other hand, an estimation of the average spectral efficiency can be obtained as

$$\hat{\eta}(t) = \frac{1}{|C(t)|} \sum_{n \in C(t)} G(\bar{U}, \hat{\gamma}_n(t)) \int_{-\infty}^{\infty} q(\bar{\gamma}_n) s_{\bar{\gamma}_n}(\bar{\gamma}_n) d\bar{\gamma}_n \quad (10)$$

where, as in the communicative spectrum assignment scheme, $G(\bar{U}, \hat{\gamma}_n(t))$ is the spectral efficiency gain factor. Finally, similar to the communicative spectrum assignment scheme, the average user throughput in the cell can be estimated by following (6) and considering (10).

IV. SIMULATION RESULTS

Performance results for the proposed distributed framework have been obtained in three different case studies for an MC, an FC, and a combined MC and FC scenario, respectively. In all case studies, the number of available chunks is $N = 24$. The modulation and coding rate SINR thresholds in [1, Tab. 8.1] have been considered to perform the short-term link adaptation. Other simulation parameters for these case studies can be found in Table I.

A. Case Study 1: MC Scenario

We consider a downlink OFDMA-based MC scenario composed of 19 hexagonal cells where the central cell is rounded by two rings of six and 12 cells, respectively. Users are distributed homogeneously within a cell, and they move at the speed of 3 km/h following a random walk model [23]. Handovers are not considered, so users always remain within their cell. Users always have data ready to be sent (i.e., full-buffer traffic model), so that each user tries to obtain as much capacity as possible. We compare the proposed decentralized framework based on RL-DSA with FRF strategies FRF1, FRF3, and PR [4], which deploy a static spectrum assignment over the cellular network. Also, comparison with the centralized counterpart strategy [18] and with another decentralized spectrum assignment strategy proposed in [24] has been performed. This latter strategy (denoted in the following as "Bernardo *et al.* [24]") uses a heuristic algorithm to determine the spectrum assignment and transmission power for each cell taking into account throughput QoS constraints as in this paper. For fair comparison purposes, the transmission power optimization in [24] has not been simulated.

The performance of the system is evaluated during 1 h to capture changes in the spatial distribution of the load (users) and to focus on the dynamic response of the algorithms. In that respect, three types of cells can be distinguished in the scenario. At the beginning, all cells are equally loaded with 15 users. After 25 min, the central cell increases the number of active users in two users per minute. The six cells in the first ring increase the number of users in one user per minute whereas the 12 cells in the second ring decrease the number of users in one user per minute. These variations take place only during a 10-min period between 25 and 35 min so that, after that period of time, traffic load has been grouped in the central cell.

Fig. 4 shows a spectral efficiency comparison between considered schemes. It is clear that RL-DSA strategies overcome the performance attained by the rest of strategies. As is expected, the centralized approach achieves the best performance due to handling global information. However, decentralized approaches presented here demonstrate a very close spectral efficiency performance although each cell only has a partial observation of the assignment problem. Furthermore, RL-DSA strategies show a very satisfactory behavior in terms of users' QoS compliance. Concretely, Fig. 4 shows the average dissatisfaction probability defined as the probability that the user throughput is below the target throughput th_{target}. Here, th_{target} per user is set to 256 kb/s (the impact on results of changing this target can be seen in case study 2). As can be seen in the figure, static reuse schemes are not adapted to heterogeneous distribution of the load (from 35 min), and thus, they obtain poor performance in dissatisfaction probability. In contrast, RL-DSA and Bernardo *et al.* [24] maintain a reduced dissatisfaction thanks to a dynamic adaptation of the spectrum assignment in both homogeneous and heterogeneous spatial distributions of the traffic load. Nevertheless, RL-DSA performance in terms of spectral efficiency is better than the obtained by Bernardo *et al.* [24].

Comparing the communicative and the noncommunicative schemes, the communicative spectrum assignment scheme shows slightly better performance than the noncommunicative scheme because a given cell handles the precise spectrum assignment in adjacent cells at the moment of RL-DSA execution. On the other hand, the noncommunicative scheme relies on measurements to estimate the spectrum usage of adjacent cells. These measurements need to be averaged during a certain period of time (i.e., the measurements averaging period). In case that the spectrum assignment in adjacent cells changes during the measurements averaging period, the estimation of the spectrum usage in adjacent cells could be less accurate. However, results show that the noncommunicative scheme attains similar performance to the communicative scheme. It happens thanks to the random selection of the initial execution time, the long RL-DSA execution period L, and a sufficiently short measurement averaging period (i.e., $l \ll L$). Moreover, the noncommunicative spectrum assignment scheme avoids the need of signaling between cells. Hence, it raises a good choice in deployments

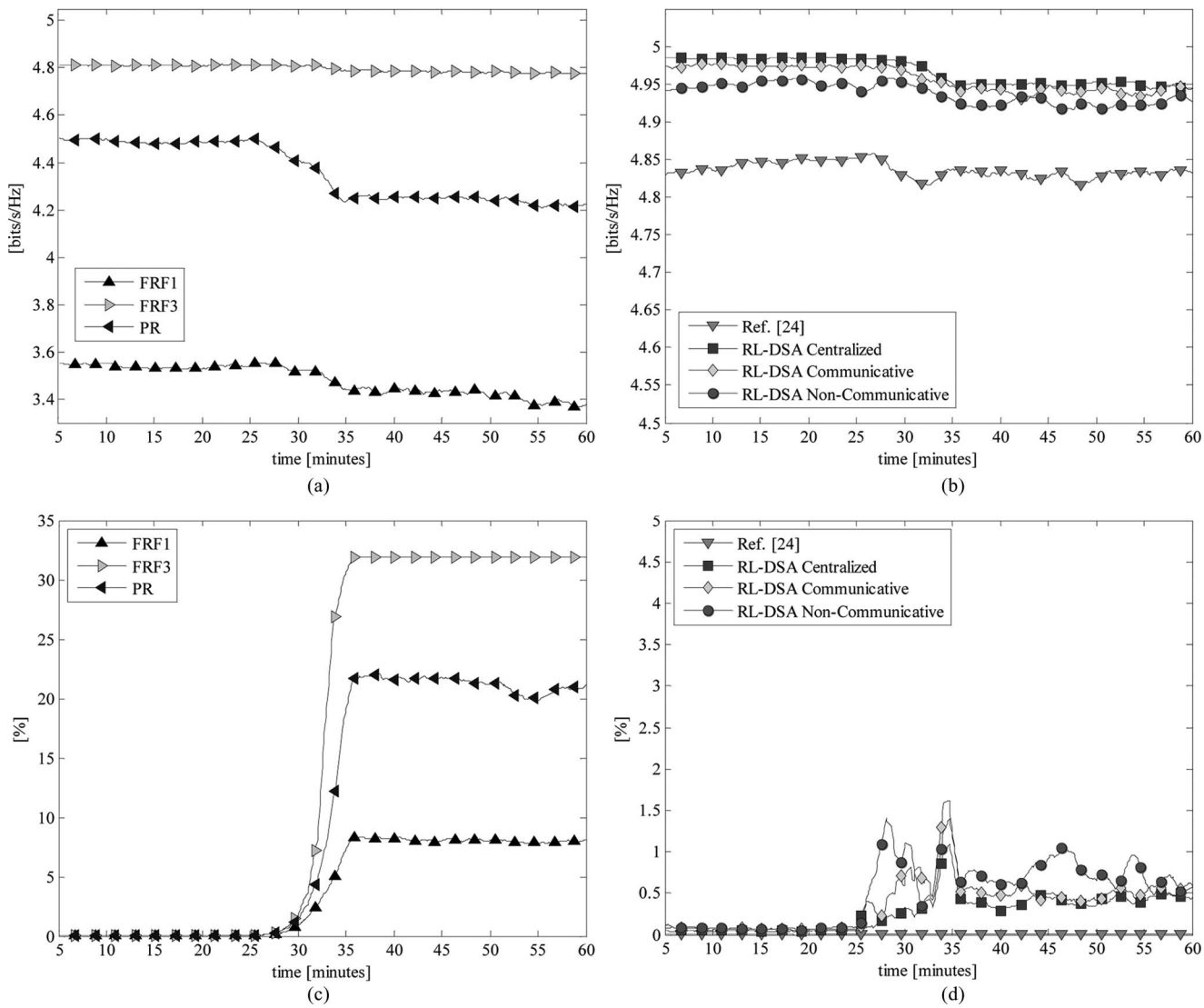

Fig. 4. Performance comparison for case study 1. Spectral efficiency for (a) static schemes and (b) dynamic schemes. Dissatisfaction probability for (c) static schemes and (d) dynamic schemes.

where the intercell signaling interface is not available, as in FC deployments.

It is worth noticing the close performance shown by the decentralized strategies in relation to the centralized strategy. This certainly favors the usage of the distributed approaches because they also add robustness against failure of either a given cell or the centralized controller in charge of executing the centralized algorithm. Also, decentralized strategies show better scalability. For instance for each cell that is added, the solution space for the centralized RL-DSA strategy increases in 2^N solutions and the signaling increases in the order of N/L bits/s [18]. However, the solution space does not increase in the decentralized approach. In turn, signaling increases in the same amount as in the centralized approach in the communicative scheme, and remains constant for the noncommunicative approach. Moreover, decentralized approaches are more flexible since they can be used in FC scenarios where a centralized architecture is not practical due to the high number and random positions of FCs' access points as the scenario tested in the following.

Finally, we have studied the convergence behavior of the decentralized RL-DSA methods in terms of the root mean squared error (RMSE) between the reward eventually achieved by RL-DSA after a certain number of RL steps and the optimal reward in a given scenario. In order to make feasible the computation of the optimal reward, we have set a particular scenario with 19 cells, 19 chunks, and five users per cell so that any spectrum assignment providing one different chunk per cell (i.e., no intercell interference) was considered to be optimum, i.e., attained the best reward defined in (1). Considering the decentralized RL-DSA configuration values given in Table I ($\alpha = 100$, $\sigma = 0.05$), an RMSE up to 1% is reached in 10^5 RL steps for both the communicative and noncommunicative schemes, what denotes a good convergence behavior since the solution space for the tested scenario involves $2^{19\times 19}$ ($=4.69 \times 10^{108}$) different assignments. It is worth to remark that this result is similar to the one obtained by the centralized RL-DSA [18].

B. Case Study 2: FC Scenario

In this case study, an FC scenario without an MC layer is studied. The performance of the noncommunicative spectrum assignment scheme based on RL-DSA is compared with Bernardo *et al.* [24] and with the FRS_x strategy as defined in [9], which deploys a randomly selected

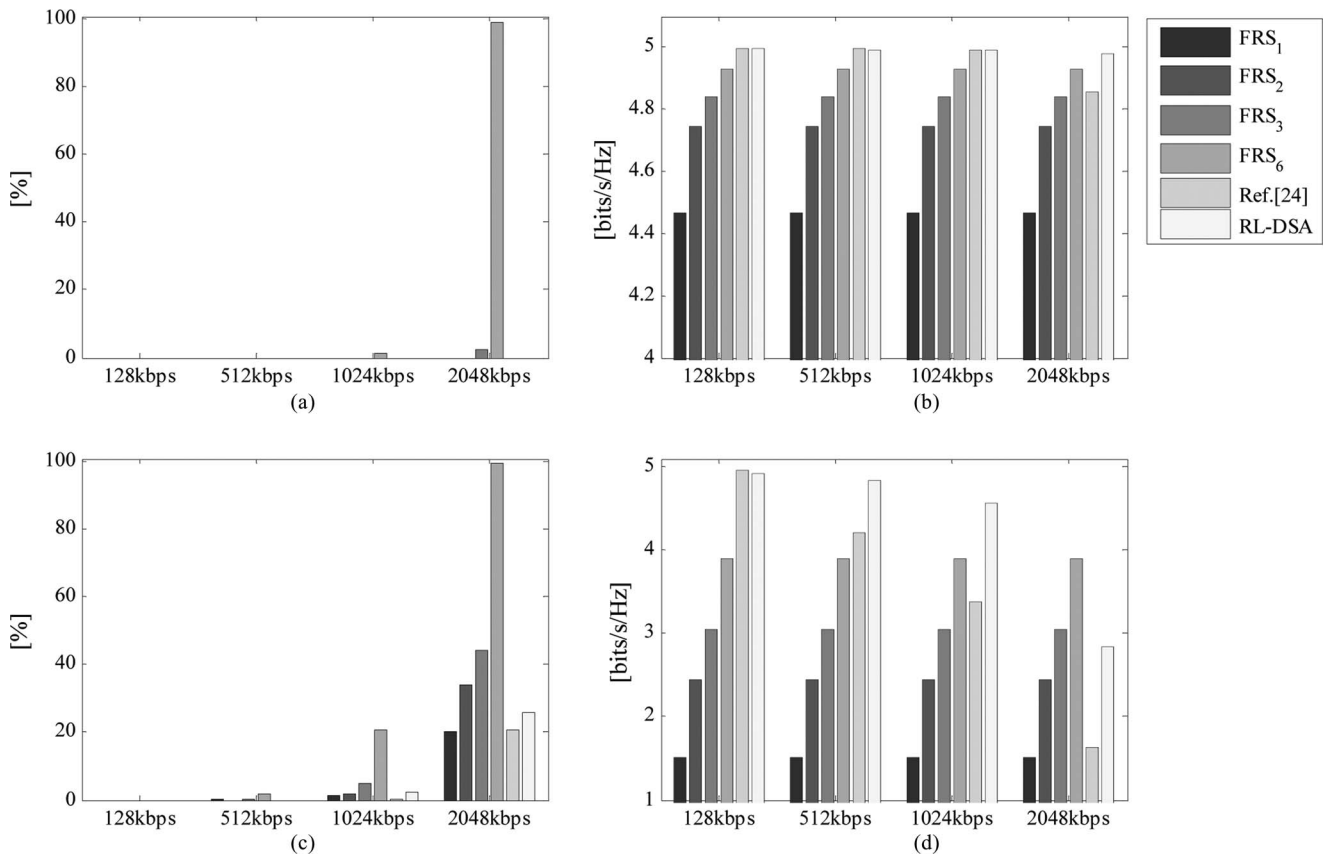

Fig. 5. Performance comparison case study 2. (a) Average dissatisfaction (10 FCs). (b) Average spectral efficiency (10 FCs). (c) Average dissatisfaction (100 FCs). (d) Average spectral efficiency (100 FCs).

subband of N/x chunks per FC. Concretely, we compare RL-DSA with FRS_1, FRS_2, FRS_3, and FRS_6. Ten and 100 circular FCs are randomly deployed with a uniform distribution in a square area of $500 \times 500\,\mathrm{m}^2$ to obtain the performance comparison under two different densities of FCs. The MC pathloss model is used between FCs for interference considerations plus an additional wall penetration loss detailed in Table I. Users in an FC remain static and the target throughput th_{target} per user is varied from 128 to 2048 kb/s. The number of users per FC is randomly selected, where 4 is the maximum number of users in any case. Results are averaged for 100 simulations with different FCs positions and users' distributions.

Fig. 5 shows the average performance statistics for the case study in terms of dissatisfaction probability and spectral efficiency for the two densities of FCs. In both cases, RL-DSA maintains the dissatisfaction probability at the lowest level even for high QoS throughput targets. To this end, RL-DSA adapts the number of chunks per cell to cope with the demanded traffic. Logically, an increment in the assigned bandwidth per cell increases the possibility of having intercell interference. Thus, in the high density scenario with strong intercell interference due to cells' proximity, RL-DSA experiments a reduction of the spectral efficiency. On the other hand, the greater the value of x in the FRS_x, the lower the probability that two adjacent FCs use the same portion and interfere each other, which turns into an increment of the spectral efficiency. However, a high x reduces the available bandwidth in each FC, which leads to an increase of dissatisfaction probability due to the lack of capacity (this is especially remarkable for FRS_6 with a target throughput of 2048 kb/s). In all, RL-DSA is the strategy that obtains the best tradeoff between dissatisfaction probability and spectral efficiency.

C. Case Study 3: MC and FC Scenario

In this case study, we focus on a coexistence MC and FC scenario, with cochannel spectrum assignment. The layout of 19 MCs of case study 1 is combined with ten or 100 FCs randomly positioned in the coverage area of the central MC. A total of 15 users per MC and four users per FC are uniformly deployed, requiring 256 and 512 kb/s per user, respectively. Closed access is assumed [25], that is, MC users cannot connect to FCs. This produces the worst-case MC-to-FC interference patterns.

We compare the performance of using different spectrum assignment strategies in the MC layer and the FC layer, respectively. In particular, FRF1, FRF3 or RL-DSA are tested in MC deployment at the same time that FRS_1, FRS_3, or RL-DSA are used in the FC layer, having a total of nine possible combinations. In the case of RL-DSA strategy, the communicative spectrum assignment scheme was used at the MC layer whereas the noncommunicative scheme was employed at the FC layer.

Tables II and III show the average spectral efficiency obtained in the central MC and the FC layers for the deployment of 10 FCs and 100 FCs, respectively. The employment of RL-DSA strategies in both the MC and FC layers brings important spectral efficiency improvements, being the improvement particularly sensitive at the FC layer and when RL-DSA is used at both the MC and FC layers. Hence, the inclusion of self-organization at both layers is clearly beneficial. Finally, Fig. 6

TABLE II
SPECTRAL EFFICIENCY IN MCS AND FCS (10 FCS)

FC Spectrum Assignment	MC Spectrum assignment (bits/s/Hz)					
	FRF1		FRF3		RL-DSA	
	Macro	Femto	Macro	Femto	Macro	Femto
FRS_1	3.34	2.48	5.00	3.95	5.00	4.39
FRS_3	3.35	2.60	5.00	4.12	5.00	4.63
RL-DSA	3.35	2.59	5.00	4.64	5.00	4.95

TABLE III
SPECTRAL EFFICIENCY IN MCS AND FCS (100 FCS)

FC Spectrum Assignment	MC Spectrum assignment (bits/s/Hz)					
	FRF1		FRF3		RL-DSA	
	Macro	Femto	Macro	Femto	Macro	Femto
FRS_1	2.94	1.56	4.28	2.19	4.28	2.42
FRS_3	3.15	2.10	4.85	3.23	4.85	3.73
RL-DSA	3.15	2.09	4.95	3.84	4.99	4.31

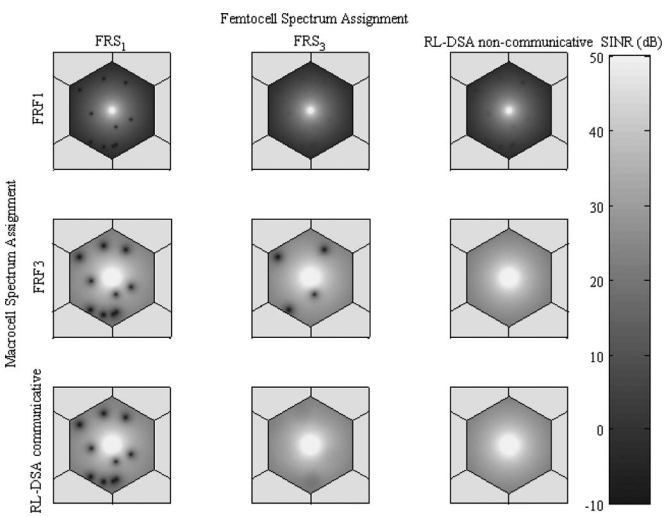

Fig. 6. SINR comparison in the central MC of an MC–FC scenario.

illustrates the average SINR improvement in the central MC when RL-DSA strategy is used. In the figure, each row represents a spectrum assignment for the MC layer whereas each column represents a spectrum assignment for the FC layer. Notice that FCs produce SINR dead zones, where the SINR for MC users highly decays in the proximities of an FC (dark spots). However, for other spectrum assignment schemes, these zones can be avoided. In fact, RL-DSA strategy in both the MC and FC layers is the best approach, demonstrating a better average SINR pattern in the central MC.

V. CONCLUSION

The decentralized approach for spectrum assignment proposed in this paper based on learning in MAS can help on the task of implementing autonomic procedures in the network nodes and thus reduce operational costs. Concretely, the proposed framework has shown its effectiveness in several cellular deployments in terms of spectral efficiency and SINR improvement, and QoS fulfillment. Also, the solution envisages communication between cells if an intercell signaling interface is available between cells. Since this is not the case for FC deployments, a complete independent (noncommunicative) spectrum assignment scheme has been also considered, where knowledge of the environment is obtained from measurements taken in the own cell. This approach could be of interest of mobile network operators that currently face the challenge of managing a network composed of thousands or even millions of nodes with the exploitation of FC technologies, where clearly, centralized approaches are unpractical. Future work could investigate global learning techniques in the MAS, where the learning rules in each cell (agent) explicitly consider the actions taken by other cells, making each cell orient its decisions, taking into account, for instance, the predicted behavior of other cells in the surroundings. Also, the possibility of having different goals per cell can be studied, constituting, in this case, a heterogeneous MAS.

ACKNOWLEDGMENT

The authors would like to thank thier colleagues from E^3 consortium for their contributions. 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.

REFERENCES

[1] B. Furht and S. A. Ahson, Eds., *Long Term Evolution 3GPP LTE Radio and Cellular Technology*. Boca Raton, FL: Auerbach, 2009.
[2] Z. Wang and R. A. Stirling-Gallacher, "Frequency reuse scheme for cellular OFDM systems," *Electron. Lett.*, vol. 38, no. 8, pp. 387–388, 2002.
[3] M. Mustonen, K. Hooli, J. Ylitalo, and A. Tölli, "Application of intra-ran flexible spectrum use to networks with multi-rate services," in *Proc. 17th IEEE Int. Symp. Pers., Indoor, Mobile Radio Commun.*, pp. 1–5.
[4] M. Sternad, T. Ottosson, A. Ahlen, and A. Svensson, "Attaining both coverage and high spectral efficiency with adaptive ofdm downlinks," in *Proc. IEEE 58th Veh. Technol. Conf.*, Oct. 2003, vol. 4, pp. 2486–2490.
[5] D. López-Pérez, A. Jüttner, and J. Zhang, "Dynamic frequency planning versus frequency re-use schemes in ofdma networks," in *Proc. IEEE Veh. Technol. Conf.*, Apr. 26–29, 2009, pp. 1–5.
[6] A. L. Stolyar and H. Viswanathan, "Self-organizing dynamic fractional frequency reuse in OFDMA systems," in *Proc. 27th IEEE Conf. Comput. Commun.*, Apr. 2008, pp. 691–699.
[7] B. W.-K. Ling, L. Benmesbah, V. Chandrasekhar, X. Chu, and M. Dohler, "Decentralized spectral resource allocation for OFDMA downlink of co-existing macro/femto networks using filled function method," in *Proc. 7th IEEE, IET Int. Symp. Commun. Syst., Netw. Dig. Signal Process.*, Jul. 21–23, 2010, pp. 881–885.
[8] V. Chandrasekhar, J. Andrews, and A. Gatherer, "Femtocell networks: A survey," *IEEE Commun. Mag.*, vol. 46, no. 9, pp. 59–67, Sep. 2008.
[9] D. López-Pérez, A. Valcarce, G. de la Roche, and J. Zhang, "OFDMA femtocells: A roadmap on interference avoidance," *IEEE Comm. Mag.*, vol. 47, no. 9, pp. 41–48, Sep. 2009.
[10] Y. Bai, J. Zhou, L. Liu, L. Chen, H. Otsuka, "Resource coordination and interference mitigation between macrocell and femtocell," presented at the 20th Pers., Indoor, Mobile Radio Commun. Symp., Tokyo, Japan, Sep. 13–16, 2009.
[11] E. Bogenfeld, I. Gaspard, Eds., (Dec. 2008), "Self-x in radio access networks," [Online]. Available: https://www.ict-e3.eu/project/dissemination/whitepapers/whitepapers.html..
[12] G. Tesauro, "Reinforcement learning in autonomic computing: a manifesto and case studies," *IEEE Internet Comput.*, vol. 11, no. 1, pp. 22–30, Jan.–Feb. 2007.
[13] P. Stone and M. Veloso, "Multiagent systems: A survey from a machine learning perspective," *Auton. Robots*, vol. 8, no. 3, pp. 345–383, Jul. 2000.
[14] E. Alonso, M. D'Inverno, D. Kudenko, M. Luck, and J. Noble, "Learning in multi-agent systems," *Knowl. Eng. Rev.*, vol. 16, no. 3, pp. 277–284, 2001.
[15] L. Busoniu, R. Babuska, and B. De Schutter, "A comprehensive survey of multiagent reinforcement learning," *IEEE Trans. Syst., Man, Cybern., C, Appl. Rev.*, vol. 38, no. 2, pp. 156–172, Mar. 2008.
[16] K. M. Dresner and P. Stone, "A multiagent approach to autonomous intersection management," *J. Artif. Intell. Res.*, vol. 31, pp. 591–656, 2008.
[17] F. Bernardo, R. Agustí, J. Pérez-Romero, and O. Sallent, "Distributed spectrum management based on reinforcement learning," in *4th Int. Conf.*

Cognitive Radio Oriented Wireless Netw. Commun., Jun. 22–24, 2009, pp. 1–6.

[18] F. Bernardo, R. Agustí, J. Pérez-Romero, and O. Sallent, "An application of reinforcement learning for efficient spectrum usage in next-generation mobile cellular networks," *IEEE Trans. Syst., Man, Cybern., C, Appl. Rev.*, vol. 40, no. 4, pp. 477–484, Jul. 2010.

[19] C. Wengerter, J. Ohlhorst, and A.G.E. von Elbwart, "Fairness and throughput analysis for generalized proportional fair frequency scheduling in OFDMA," in *Proc. IEEE 61st Veh. Technol. Conf.*, Jun. 2005, vol. 3, pp. 1903–1907.

[20] 3GPP TS 36.300 v8.7.0, "3GPP E-UTRA and E-UTRAN," Overall description, Stage 2 (Rel. 8), 2009.

[21] V. V. Phansalkar and M. A. L. Thathachar, "Local and global optimization algorithms for generalized learning automata," *Neural Comput.*, vol. 7, no. 5, pp. 950–973, Sep. 1995.

[22] M. A. L. Thathachar and P. S. Sastry, "Varieties of learning automata: an overview," *IEEE Trans. Man, Cybern., B*, vol. 32, no. 6, pp. 711–722, Dec. 2002.

[23] 3GPP, TR 25.814 v7.1.0, "Physical layer aspects for evolved Universal Terrestrial Radio Access (UTRA)," Rel. 7, Sep. 2006.

[24] F. Bernardo, R. Agustí, J. Cordero, and C. Crespo, "Self-optimization of spectrum assignment and transmission power in ofdma femtocells," in *Proc. 6th Adv. Int. Conf. Telecommun.*, May 9–15, 2010.

[25] V. Chandrasekhar and J. G. Andrews, "Spectrum allocation in tiered cellular networks," *IEEE Trans. Commun.*, vol. 57, no. 10, pp. 3059–3068, Oct. 2009.