Energy-aware Dynamic Selection of Overlay and Underlay Spectrum Sharing for Cognitive Small Cells

Pavel Mach, Zdenek Becvar, *member, IEEE*

***Abstract*—The small cell base stations (SCs) with cognitive capabilities are seen as an efficient way to cope with interference between the SCs and macrocells (MBSs). The cognitive SCs may access the spectrum by means of overlay or underlay mode. An efficiency of the overlay mode fully depends on the activity of macrocell users (MUEs), since insufficient resources remain for the small cell users (SUEs). Contrary, the main weakness of the underlay mode is that it can result in low transmission efficiency due to restricted transmission power. Apart from the transmission efficiency of both modes, an energy consumption of the SCs should not be disregarded. Thus, we propose a centralized scheme selecting the spectrum sharing mode in downlink according to both the SUEs' throughput and the energy consumption of the SCs. The objective is to maximize the overall performance of the SCs while their energy consumption is taken into account. Then, we also propose a distributed algorithm in order to decrease complexity and signaling overhead of the centralized scheme. The results show that the proposed dynamic selection significantly outperforms all competitive schemes in terms of the SUEs' throughput while the throughput of MUEs is intact and only negligible signaling overhead is generated in case of the proposed distributed algorithm. Moreover, the proposed algorithm is able to notably decrease the energy consumption of the SCs.**

***Index Terms*—cognitive small cells, overlay/underlay spectrum sharing, hybrid spectrum sharing, 5G, energy consumption.**

I. Introduction

As forecasted in [1], a simple incremental enhancement of 4G-based mobile networks will not be able to satisfy users demands in near future. In this regard, new generation of mobile networks, known as 5G, emerges. Examples of major features enabling the enormous data rates by means of 5G networks are [2]: 1) extreme densification of base stations, 2) utilization of millimetres waves, or 3) exploitation of massive MIMO techniques [3]. Dense deployment of small cells (SCs) can be seen as a convenient approach for cellular operators to cheaply increase the capacity of their contemporary mobile networks towards 5G requirements. Nevertheless, high density of the SCs inevitably results in severe interference to the macrocell base stations (MBSs) and their users. Moreover, the mutual interference among the SCs cannot be discarded as well. The interference problem is emphasized if the SCs use closed access [4] (typical especially for femtocells) and if a co-channel deployment is exploited [5].

A promising option for interference avoidance in the co-channel deployment is to implement SCs with cognitive capabilities, also known as cognitive SCs [6]. The users connected to the cognitive SCs are considered to be the secondary users (SUs) with a lower priority while the users attached to the MBSs are the primary users (PUs) having a higher priority. The cognitive SCs may access the spectrum of the MBS either in an underlay or an overlay mode [7]-[9].

In case of the underlay spectrum sharing (USS), both the MBSs (PUs) and the cognitive SCs (SUs) use the same radio frequencies (in OFDMA systems, the same radio resource blocks (RBs))[10]-[14]. In general, the main drawback of the USS approaches is that the SC users (SUEs) may not be able to attach to the SC as its coverage is limited by a low transmission power. Moreover, the restricted transmission power may result in utilization of less effective modulation and coding scheme (MCS). In case of the overlay spectrum sharing (OSS), the cognitive SCs access only those RBs that are not currently occupied by the MBS [15]-[20]. The main disadvantage of the OSS is that the performance of the SCs is strongly dependent on the activity of the MBSs. In the worst case scenario, there could be no RBs available to the SCs if the MBS is loaded heavily.

A feasible way how to address the problem of both above mentioned spectrum sharing approaches is to use a hybrid spectrum sharing (HSS). In [21], we have proposed a scheme that dynamically selects the downlink spectrum sharing mode for the SCs that is currently more efficient for the SUEs in terms of throughput. The selection itself is based on the amount of resources available in each mode and on data transmission efficiency experienced by the SUEs. Several HSS schemes for conventional cognitive systems without the SCs have been recently proposed in the literature (e.g., [22]-[28]). The general idea of these schemes is to use the OSS if the PU is idle. Contrary, the USS is used whenever the PU is active. This maximizes throughput of the SUs. Nevertheless, [22]-[28] cannot be efficiently adopted by cognitive SCs in up-to-date cellular networks because of following reasons. In case of cognitive SCs using the OSS, the amount of RBs available for the SC depends strongly on the load of MBS and the number of SCs in the OSS mode. Thus, even if the SUEs can

Manuscript received xxx; revised August xxx; accepted xxx. Date of publication xxx; date of current version. This work has been supported by Grant No. 13-24931P funded by the Czech Science Foundation and by Grant No. 8G15008 funded by MEYS in frame of Czech-Israel project in call MSMT-10795/2015-1.

The authors are with the Department of Telecommunication Engineering, Faculty of Electrical Engineering, Czech Technical University in Prague, Prague, 166 27 Czech Republic (email: machp2@fel.cvut.cz; zdenek.becvar@fel.cvut.cz).

enjoy highly efficient transmission, the amount of available RBs may not be sufficient for all the SUEs. Although the USS offers more available RBs when compared to the OSS, its problem consists in low efficiency of transmission, which is a result of restricted SCs' transmission power to minimize interference to the MBS users (MUEs). With respect to both above-mentioned limitations, the HSS scheme suitable for the cognitive SCs must be able to estimate how much RBs would be available to the SC in each allocation mode and select the one, which is more beneficial in terms of SUEs' throughput.

Moreover, the HSS schemes proposed in [22]-[28] solve the problem only from capacity point of view. As pointed out in [29], the introduction of huge amount of SCs into contemporary wireless networks, as supposed in 5G networks, may lead to many-fold increase in energy consumption. In this regard, many recent studies focus on energy savings if the SCs are incorporated into the network by means of either sleep-mode strategies (e.g., [30]-[32]) or various power allocation techniques (e.g., [33][34]) at the side of the SCs. None of these, however, takes into account the selection of spectrum sharing mode (either the USS or the OSS) with the energy consumption of the SCs in mind. The energy consumed by the SC's downlink communication for each mode depends on current amount of RBs used for transmission (usually, the SC in the USS transmits over more RBs than in case of the OSS) and actual transmission power of the SCs (usually, transmission power of the SC in the OSS is significantly higher than in the USS). As a consequence, the energy consumption of each spectrum sharing mode can fluctuate rapidly depending on above-mentioned parameters and should not be disregarded in proper selection of the spectrum sharing mode.

In this paper, we extend our former work presented in [21] towards Energy-aware Dynamic Spectrum Sharing (EDSS) algorithm, selecting the most appropriate mode in downlink for all individual SCs in the system. The contribution of this paper, considering also [21], is as follows:

- First, we propose a centralized hybrid mode selection scheme maximizing the performance of the system in terms of transmission efficiency and downlink energy consumption of the SCs for selected weight of these two metrics. The MBS exhaustively searches among all possible combinations of the SCs' allocation modes and dynamically selects the one maximizing the performance of the system. Although this scheme is characterized by high complexity and huge signaling overhead, it shows theoretical upper bound performance of the whole concept.
- Second, we design a distributed scheme lowering complexity and the amount of signaling overhead of the centralized approach in order to incorporate the proposed hybrid spectrum spectrum sharing into real networks. Moreover, with respect to the HSS in [21], we address also a situation when the MBS changes its radio resource allocation pattern. This protects the MUEs and the SUEs against potential high interference caused by the change of MBS's allocation pattern. We also elaborate signaling mechanism enabling the implementation of distributed algorithm into mobile networks.
- Third, we evaluate the performance of the proposed scheme not only in terms of served data ratio (as in [21]) but we analyze also energy consumed by downlink transmission of the SCs. Furthermore, we discuss a trade-off between the transmission efficiency and the energy consumption and we investigate suitable weighing of both metrics. Moreover, we analyze how the frequency of MBS' allocation pattern impacts the performance of individual schemes in terms of the MUEs and SUEs throughput.

The rest of this paper is organized as follows. The next section describes related work done in scope of cognitive SCs and hybrid cognitive radio. Then, in Section 3, we introduce a system model used further for the proposal description. Section 4 explains the basic principle of the proposed mode selection and describes both the centralized and distributed approach. The evaluation methodology and simulation results are addressed in Section 5 and Section 6, respectively. The last section gives our conclusion and outlines potential future research work.

II. Related Works

This section describes in more detail work related to the cognitive spectrum sharing. Basically, the related work can be divided into research addressing either individual USS or OSS for the SCs.

The schemes utilizing various power control techniques at the side of the SCs in order to mitigate interference to the MBS can be considered as a specific case of the USS [10]-[12]. In [10][11], the power allocation setting to the SCs exploiting Q-learning is proposed to mitigate the interference to the MBS. In both works, the SCs represent the agents, which set their actions to maximize rewards. In [12], two centralized power control schemes benefiting from a context information of the SUEs and the MUEs is proposed. The power control techniques utilizing the cognitive principle (especially the sensing techniques) adopting the USS have been introduced, for example, in [13][14]. The authors in [13] propose downlink transmit power allocation on each channel exploiting information on downlink radio resource usage obtained from the sensing. Similarly, the optimal power allocation in order to minimize interference to the PUs is proposed in [14].

Like in case of the USS, a lot of work has been done regarding the cognitive SCs utilizing the OSS. In [15], the SCs autonomously sense the radio frequencies used by the MBS and schedule their transmission to unoccupied radio resources. A dynamic spectrum reuse is proposed in [16]. If the MUE is close to the SC and suffers from the SC's interference, the SC does not occupy the same radio resources as the MUE. In [17], the Gale-Shapley Spectrum Sharing (GSOIA) scheme is suggested. The GSOIA is based on multichannel opportunistic sensing resulting in no collisions among the MBS and the SCs. The authors in [18] propose Cognitive Hybrid Division Duplex (CHDD) where the MBSs use frequency division duplex (FDD) while the SCs access the available band by means of time division duplex (TDD). Another OSS scheme is presented in [19], where the SCs perform sensing and then

access only those channels that are assumed to be unoccupied (i.e., not utilized by the MBS and its users). A specific case of the OSS is considered in [20], where the MBSs and the SCs use mutually exclusive sub-channels by means of fractional spectrum reuse. This approach, however, decreases the spectral efficiency.

There are several hybrid spectrum sharing schemes for coexistence between the MBS and SCs. However, these focus on combination of co-channel deployment and dedicated channel deployment [35]-[38], but not on combination of the OSS and the USS (note that we focus solely on co-channel deployment). Consequently, in this paper we propose EDSS, which addresses the disadvantages of both the OSS and the USS and it is suitable for OFDMA-based cognitive SCs.

III. System model

We define system model with seven MBSs in hexagonal grid (note that in Fig. 1, only central MBS is depicted for better clarity). Inside the coverage of each MBS, N SCs are deployed. Out of these SCs, O SCs are in the OSS and U SCs are in the USS (i.e., $N = O + U$). Note that the amount of SCs in individual modes is supposed to be changing over time.

As illustrated in Fig. 1, each SC may have one or more direct neighbors (e.g., SC2 has SC1 and SC3 as the direct neighbors). Consequently, the i-th SC has O_i^d and U_i^d direct neighbors exploiting the OSS and the USS, respectively. Similarly, the i-th SC has O_i^n (U_i^n) non-direct neighbors utilizing the OSS (the USS), i.e., the SCs that are not the direct neighbors of the i-th SC. In our model, the k-th SC is considered to be the direct neighbor of the i-th SC if:

$$RSS_{ki} > NI_i - \sigma_n \tag{1}$$

where RSS_{ki} is the received signal strength from the k-th SC at the i-th SC if the k-th SC transmits with the maximum power P_{max}, NI_i stands for thermal noise and interference level from all adjacent MBSs and their underlying SCs at the location of the i-th SC, and σ_n represents a direct neighbor interference threshold. If the σ_n is low (or even negative), the SCs have a low number of direct neighbors but a strong interference can be generated among them if the OSS is applied. Contrary, higher value of σ_n protects the SCs against

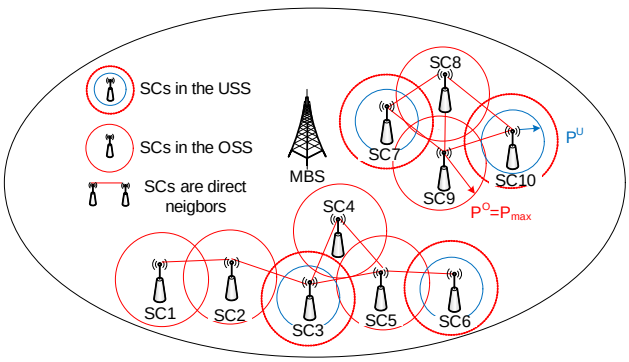


Fig. 1: System model.

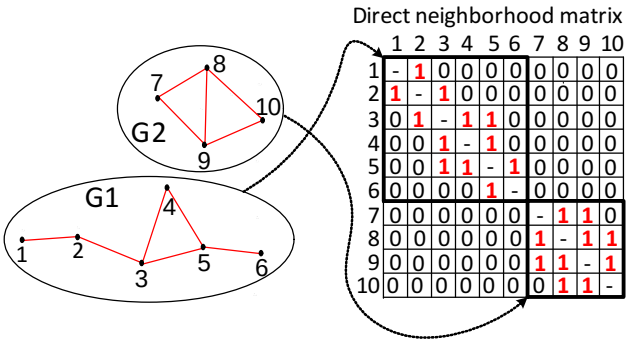


Fig. 2: Determination of direct neighbors and creation of direct neighborhood matrix.

interference from other SCs while less amount of resources is available for the SCs.

The direct neighbors are determined by means of graph theory. To this end, we denote each graph as $G_x = (V_x, E_x)$, where V_x is the amount of SCs in the x-th graph and E_x represents the connection between SCs corresponding to potential interference from the direct neighbors (i.e., fulfilling (1)). The connection between direct neighbors is expressed by a direct neighborhood matrix, which is shown in Fig. 2 (this matrix is derived from deployment depicted in Fig. 1). The direct neighbors of the SCs are represented by value "1" in the matrix. Note that the direct neighborhood matrix is changed only if a new SC is added, existing SC is removed, or if the SC is switched on/off. In Fig. 2, two graphs are created (G_1 and G_2), representing two clusters of the SCs. Notice that the SCs do not interfere to the SCs in different clusters each other even in the OSS due to the mutual distance.

In the model, we distinguish three types of users: 1) users connected to the SC in the OSS, 2) users connected to the SC in the USS, and 3) users connected to the MBS. The SINR of the f-th SUE attached to the i-th SC in the OSS (γ_{fi}^O) or the USS (γ_{fi}^U) and SINR of the MUEs (γ_m) are calculated as:

$$\gamma_{fi}^O = \frac{|h_{fi}|^2 p_i^O}{NI + \sum_{o=1}^{O_i^n} |h_{fo}|^2 p_o^O + \sum_{u=1}^{U_i^n} |h_{fu}|^2 p_u^U} \tag{2}$$

$$\gamma_{fi}^U = \frac{|h_{fi}|^2 p_i^U}{NI + \sum_{o=1}^{O_i^n} |h_{fo}|^2 p_o^O + \sum_{u=1}^{U_i^n + U_i^d} |h_{fu}|^2 p_u^U + |h_{fb}|^2 P} \tag{3}$$

$$\gamma_m = \frac{|h_{mb}|^2 P}{NI + \sum_{u=1}^{U} |h_{mu}|^2 p_u^U}, \tag{4}$$

where $|h_{fi}|^2$, $|h_{fo}|^2$, $|h_{fu}|^2$, $|h_{fb}|^2$, $|h_{mb}|^2$, and $|h_{mu}|^2$ are the channel gains between the SUE and its serving SC, the SUE and individual interfering SCs in the OSS and the USS, the SUE and the MBS, the MUE and the MBS, and the MUE and the SCs in the USS, respectively. The p^O, p^U, and P are the transmission powers of the SCs in the OSS, the SCs in the USS, and of the MBS, respectively. In case of the OSS,

the p^O is set to the maximal power P_{max} as the SCs occupy only RBs not used by the MBS so interference to the MUEs is eliminated. In case of the USS, the setting of p^U is more complicated. In conventional cognitive systems, the power is set to such value so that the interference temperature imposed to the PU is below a specified threshold [40] or to guarantee target SINR of the PU [13]. In our case we set p^U according to the latter approach, thus, p^U is set to guarantee that:

$$\gamma_m \geq \gamma_{m(noSC)} - \delta \tag{5}$$

where $\gamma_{m(noSC)}$ is SINR of the MUEs if no SCs are considered, and δ stands for allowed decrease in γ_m. This way, however, we cannot guarantee that all active SUEs (by active SUE we consider the SUEs receiving data from the SC at the moment or the SUE to which the SC has data to be sent) can still attach to the SC since p^U could be decreased significantly. Hence, some of the active SUEs may experience insufficient channel quality resulting in connection drops. In this case, the proposed scheme prefers to use the OSS as will be explained later.

In case of γ^O_{fi}, interference to the f-th SUE is generated only by the SCs, which are not the direct neighbors of the i-th SC (see upper limits in sums in (2) and (3)). The direct neighboring SCs cause no interference to the SUEs attached to the SC in the OSS since these SCs use orthogonal RBs.In case of γ^U_{fi}, interference to the f-th SUE originates from the SCs in the l-th SC's direct neighborhood using the USS and also from all the SCs not in the direct neighborhood of the i-th SC since the same RBs may be occupied in the USS.

Without loss of generality, the MBS and the SCs adopt a system where data at physical layer is sent in frames (e.g., like in LTE-A, where the duration of one frame is 10 ms). Some of the frames are dedicated solely to sensing as suggested in [15]. During the sensing frames, the SCs determine received interference power ι_r at individual RBs. If the ι_r at the RB is above the sensing interference threshold σ_s (i.e., if $\iota_r > \sigma_s$), the RB is supposed to be occupied by either the MBS or the SCs. The sensing frames are scheduled periodically with period n_s (in our paper $n_s = 0.2s$ according to [15]).

According to [41] the smallest unit for data transmission in downlink is a resource element (RE) spanning over one sub-carrier and one OFDM symbol. Each RB is composed of n_{RE} REs (e.g., in LTE-A [41], each RBs occupies 12 consecutive sub-carriers in a frequency domain and 7 OFDM symbols in a time domain, i.e., $n_{RE} = 84$). The number of RBs available for the SC in one frame (n_{RB}) depends on selected channel bandwidth [42]. The number of RBs per frame required for the transmission of data to a active SUEs at the l-th SC is derived as:

$$n_{RB,r,i} = \sum_{z=0}^{a} ceil\left(\frac{\xi_z}{\Gamma_z\left(n_{RE} - n_{RE,OH}\right)}\right), \tag{6}$$

where ξ_z is the amount of data (bits) sent in downlink to the z-th SUE, $n_{RE,OH}$ stands for the amount of REs dedicated to signaling in each RB, and Γ_z represents the amount of bits sent in one RE that is derived from MCS of z-th SUE. The MCS is assigned with respect to γ^O_{fi} or γ^U_{fi}. Since $\gamma^O_{fi} >> \gamma^U_{fi}$, significantly less amount of RBs is required for data transmission in the OSS than in the USS (i.e., $n^O_{RB,r,i} << n^U_{RB,r,i}$). On the other hand, considerably less amount of RBs are available for the SC in OSS when compared to the USS.

In our model, the number of RBs available to the SC in the OSS depends on two factors. The first factor is the amount of RBs assigned to the MUEs by the MBS ($n_{RB,M}$). The reason is that the SC in the OSS cannot exploit these RBs to avoid interference to the MUEs. The second factor is the number of neighboring SCs that are in an active state, that is the SCs having data to be transmitted for their users ($n_{A,i}$) in a buffer. As a consequence, the SCs in the OSS use non-overlapping RBs with respect to the neighboring SCs (both in the OSS and the USS) to avoid interference to the SUEs. In our model, the SC in the OSS shares RBs not occupied by the MBS equally with its active neighboring SCs disregarding their requirements. Allocation of $n^O_{RB,i}$ depending on the SCs' load is left for future research. From the above, the number of RBs available for the i-th SC in the OSS ($n^O_{RB,i}$) can be formulated as:

$$n^O_{RB,i} = \frac{n_{RB} - n_{RB,M}}{n_{A,i} + 1}, \tag{7}$$

The number of available RBs for the SCs in the USS depends solely on the number of active neighboring SCs in the OSS ($n^O_{A,i}$). This is because of the fact that the SCs in the USS are not able to exploit the RBs occupied by the neighboring SCs in the OSS since interference from these SCs would be too high. On the other hand, the SCs in the USS are allowed to use the same RBs as their power is restricted by (5). Then, the number of RBs in the USS ($n^U_{RB,i}$) is expressed as:

$$n^U_{RB,i} = n_{RB} - n^O_{A,i} n^O_{RB,i}, \tag{8}$$

The transmission efficiency of both the OSS and the USS modes of the l-th SC can be defined as:

$$\eta^O_i = \frac{n^O_{RB,i}}{n^O_{RB,r,i}}, \tag{9}$$

$$\eta^U_i = \frac{n^U_{RB,i}}{n^U_{RB,r,i}}. \tag{10}$$

Both (9) and (10) express the ratio between the number of RBs available to the SC and the number of RBs required by the SC. Thus, the value can theoretically vary from 0 (no RBs are available to the SCs) to ∞ (the SUEs have no requirements). In case that $\eta_i < 1$, the SC is not able to transmit all generated data to its SUEs since more RBs are required to transmit all data. Contrary, if $\eta_i \geq 1$, the SC is able to serve all its users because required amount of RBs is not higher than the number of available RBs. Moreover, from (9) (and (10)), it can be observed that the higher η^O_i (η^U_i) is, the more efficient transmission mode is. This is due to the fact that our goal is to keep $n^O_{RB,r,i}$ or $n^U_{RB,r,i}$ low while $n^O_{RB,i}$ or $n^U_{RB,i}$ should be high. The η^O_i is influenced by $n^O_{RB,i}$ according to (7). Contrary, the η^U_i is influenced mainly by $n^U_{RB,i}$ according to (8) and by p^U_i according to (1) since lower transmission power

results in less efficient MCS and more RBs are required for data transmission.

Besides transmission efficiency, the mode selection in our proposal considers also downlink energy consumption of the SCs in both modes. We define downlink energy consumption of individual modes ($\epsilon_i^O, \epsilon_i^U$) as:

$$\epsilon_i^O = \begin{cases} n_{RB,r,i}^O * p_i^O & , n_{RB,r,i}^O \leq n_{RB,i}^O \\ n_{RB,i}^O * p_i^O & , n_{RB,r,i}^O > n_{RB,i}^O \end{cases} \quad (11)$$

$$\epsilon_i^U = \begin{cases} n_{RB,r,i}^U * p_i^U & , n_{RB,r,i}^U \leq n_{RB,i}^U \\ n_{RB,i}^U * p_i^U & , n_{RB,r,i}^U > n_{RB,i}^U \end{cases} \quad (12)$$

where p_i^O is set to P_{max} and p_i^U is set according to (1) as described above.

Our objective is to design a flexible scheme, which enables to select the mode solely with respect to transmission efficiency (most profitable for the SUEs), according to energy consumption (profitable for the network), or taking both into account. Thus, we define mode selection function as:

$$M = \begin{cases} \text{OSS: } \alpha\eta_i^O + (1-\alpha)\frac{1}{\epsilon_i^O} > \alpha\eta_i^U + (1-\alpha)\frac{1}{\epsilon_i^U} \\ \text{USS: } \alpha\eta_i^U + (1-\alpha)\frac{1}{\epsilon_i^U} > \alpha\eta_i^O + (1-\alpha)\frac{1}{\epsilon_i^O} \end{cases} \quad (13)$$

where $\alpha = \langle 0, 1 \rangle$ gives preferences to transmission efficiency (if $\alpha = 1$, energy consumption is not considered) or energy consumption of the SCs (if $\alpha = 0$, transmission efficiency is not considered).

IV. Proposed Selection of Allocation Mode for the SCs

This section firstly describes the fundamental principle of the proposed principle. Then, both centralized and distributed allocation modes are introduced in detail.

A. Fundamental principle and and basic assumptions

The fundamental principle of the proposal is to change the allocation mode of the SC dynamically between the USS and the OSS in order to maximize overall performance of the SUEs. The selection of allocation mode and its change is demonstrated by an example in Fig. 3, where SC1 and SC2 are direct neighbors and, without loss of generality, α is set to 1 (i.e., preference is given solely to the transmission efficiency). The example shows three states of the system for different loads of the MBS. In *State 1*, the MBS load is heavy (in the example, $n_{RB,M} = 14$ out of 16 RBs) and the USS is more appropriate for both SCs, since $n_{RB,i}^U = 16 >> n_{RB,i}^O = 1$ for both SCs (the SCs in the OSS can use only RBs not occupied by the MBS or direct neighboring SCs in the OSS). In *State 2*, the load of the MBS is decreased ($n_{RB,M} = 8$). Consequently, $n_{RB,i}^O$ is increased to 4 RBs (see (7)) while $n_{RB,i}^U$ is decreased depending on the number of SCs in direct neighborhood using the OSS (see (7)). Therefore, the SC1 changes its mode to the OSS since $\eta_1^O > \eta_1^U$. On the other hand, the SC2 stays in the USS because $\eta_2^U > \eta_2^O$ since the SC2 is able to transmit more data in the USS ($n_{RB,2}^O = 3$ while $n_{RB,2}^U = 10$). As Fig. 3 illustrates, the SC2 cannot use the RBs occupied by the SC1 in the OSS due to interference. In *State 3*, the MBS load is further lowered ($n_{RB,M} = 4$) resulting in a change of SC2 mode to the OSS. The reason is that $\eta_2^O > \eta_2^U$ due to decreased $n_{RB,2}^U$ to 10 RBs and increases $n_{RB,2}^O$ to 6 RBs. If the whole situation is reversed and load of the MBS increases, the SCs change their mode back to the USS (follow reverse direction from the *State 3* to the *State 1* in Fig. 3).

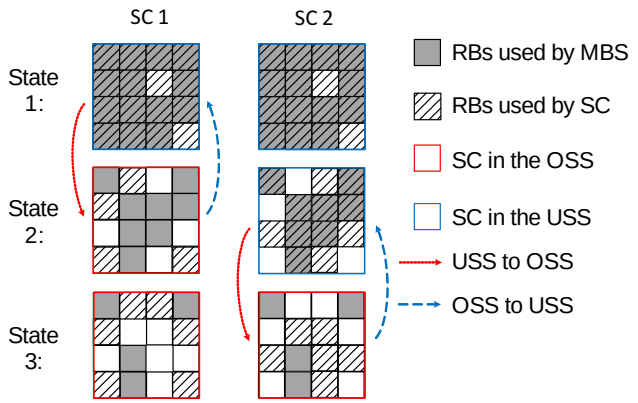


Fig. 3: The principle of change of allocation mode according to the MBS load ($\alpha = 1$ is assumed).

Before detail description of both proposed mode selection algorithms, we give several basic assumptions considered in designing of the proposal. First, we assume that the SUEs attached to the SCs are either static or slowly moving (i.e., pedestrian users) while users moving with high speed (e.g., vehicular users) are not allowed to connect. This is common assumption for the SCs because of their low coverage, where fast users would generate significant amount of handovers. Consequently, the channel is varying slowly and coherence time is significantly longer than the computation time of both centralized and distributed algorithms needed to make a decision. Second, we assume that in case of the centralized algorithm, the MBS is aware of all SCs requirements and that it has perfect CSI knowledge in order to calculate (6), (7), and (8). Third, we assume that in case of the distributed algorithm, the SC determines the RBs occupied by the MBS through the eavesdropping technique consisting in listening broadcast information of the MBS (see [17] for more details). The eavesdropping technique allows to determine the amount of RBs used by the MBS ($n_{RB,M}$) and this information is necessary to calculate $n_{RB,i}^O$ (see (7)).

In the next subsections, we describe the proposed centralized and distributed algorithms for a dynamic selection of the allocation mode.

B. Centralized selection of allocation mode for the SCs

To determine upper bound maximum performance of the SCs in terms of the transmission efficiency and their energy consumption, we define a centralized algorithm, where the MBS itself selects the most appropriate mode for all the SCs within its coverage. In this regard, the MBS evaluates all possible combinations of allocation modes for all underlying

SCs and selects the most efficient allocation for each SCs taking into account both transmission efficiency and energy consumption of the SCs.

The selection process is initiated whenever the MBS changes allocation of RBs to its own MUEs (e.g., the requirements of the MUEs are increased/decreased). In such a situation, the amount of available RBs for the SCs according to (7) and (8) is changed affecting η and ϵ. Moreover, the MBS initiates the selection process whenever the SCs' requirements are changed. This happens mostly if the SUEs demands are altered or if the channel quality between the SUEs and the SCs varies. As a result, η and ϵ of both allocation modes are influenced since the amount of RBs needed by them is different and the transmission power in case of the USS may be adapted as well.

The centralized selection of allocation mode is presented in Algorithm 1. In case that the MBS changes its allocation pattern, all the SCs are influenced. Since the SCs in each clusters are not mutually interfered by the SCs in other clusters (see Fig. 2, where two independent clusters are created), the MBS performs allocation for each cluster separately (see Algorithm 1, line 2). The selection of allocation mode is then performed as follows. First, the MBS sets allocation mode for each SC in the cluster (see line 4). Note that in the first step, all the SCs are considered to be in the USS mode and, subsequently, all possible combinations of all USS/OSS allocations are exhaustively checked out in the next steps. After the MBS sets new allocation pattern to each SC, the MBS determines n^O_{RB} for all SCs in the OSS according to (7) and n^U_{RB} for each SC in the USS according to (8) (see line 6). This can be easily done by the MBS since the MBS knows $n_{RB,M}$ and also the number of neighboring SCs in the USS and in the OSS. Then, the MBS also derives the amount of RBs required by individual SCs (i.e., $n^O_{RB,r}$ and $n^U_{RB,r}$) as indicated in line 7 (note that this is feasible because of second assumption in Section IV.A). If the MBS finds out that there is the SC in the USS that is not able to serve all active users due to low channel quality, this allocation option is discarded by the MBS algorithm goes back to line (line 8-10). Contrary, if all active SUEs are able to attach to the SC in selected allocation pattern, the MBS calculates mode transmission efficiency η for each SC within the cluster according to (9) or (10) and energy consumption ϵ according to (11) or (12), respectively (lines 11-12). Finally, the MBS calculates overall η and ϵ of this particular allocation (see line 14). After the set of steps on lines 4-14 are done for each possible combination of the USS and OSS allocations, the MBS selects the one resulting in maximum performance for selected α (line 16). In the last step, the MBS sends information on selected mode and available RBs to each SC within the cluster (both for the SCs in the OSS and the USS modes). The whole process is then repeated for each cluster.

Algorithm 1 : Centralized algorithm for selection of allocation mode

1: **if** change of MBS alloc. pattern *or* SC requirements **then**
2: **for** 1:cl (cl = no. of clusters in the MBS) **do**
3: **for** 1:c (c = no. of possible allocations in cl) **do**
4: set new allocation pattern for SCs
5: **for** 1:m_{cl} (m_{cl} = no. of SCs in cluster) **do**
6: determine n^O_{RB} acc. (7) or n^U_{RB} acc. (8)
7: determine $n^O_{RB,r}$ or $n^U_{RB,r}$ acc. (6)
8: **if** $\exists$ SUEs with low channel quality **then**
9: disregard all. pattern, go back to line 4
10: **end if**
11: calculate η acc. (9) or (10)
12: calculate ϵ acc. (11) or (12)
13: **end for**
14: calculate $\sum \eta$ and $\sum \epsilon$
15: **end for**
16: select allocation with $\max(\alpha \sum \eta + (1-\alpha)\frac{1}{\sum \epsilon})$
17: sent to all SCs in a cluster assigned all. mode
18: **end for**
19: **end if**

The centralized allocation mode selection leads to maximum performance for selected α. But, at the same time, the centralized approach is of high complexity and not feasible for implementation in real system. In the worst case, if all the SCs within the MBS coverage form one cluster, the complexity of Algorithm 1 is $O(2^m)$, where m is the number of the SCs within the MBS's coverage. In other words, the MBS has to check 2^m possible combinations for mode allocation to identify maximum performance. In case of high value of m, the computation time of the algorithm could be longer than channel coherence time, especially for ultra-dense deployment of the SCs in the future. In addition, the centralized algorithm requires perfect CSI knowledge of all involved paths (i.e., the paths between all the SUEs and their serving SCs) resulting in significant signaling overhead via backhaul. Because of above-mentioned drawbacks, the centralized algorithm is not feasible for real network and it is considered here exclusively for benchmarking purposes to show the upper bound performance.

C. Distributed selection of allocation mode for the SCs

When compared to the centralized mode selection algorithm, each SC selects the allocation mode that is more beneficial for it in the distributed algorithm. When compared to the centralized case, the decision itself is done only after appointed sensing frames scheduled periodically with a period equal to n_s as indicated in the system model. This way, the distributed approach minimizes signaling overhead and the amount of information exchanged among the SCs via backhaul.

The important aspect of the distributed algorithm is decision on frames, which should be used for sensing purposes (called as sensing frames). The sensing frames of the SC should not overlap with the sensing frames of its direct neighbors as shown in Fig. 4. This way, the SC is able to determine RBs, which are currently exploited by its direct neighbors in the OSS. Consequently, the SCs use only orthogonal RBs and the interference can be avoided. The SC can negotiate sensing frame's allocation with its neighbors when firstly deployed.

A decision whether the mode should be changed or not is done every sensing frame. Nevertheless, if the MBS changes

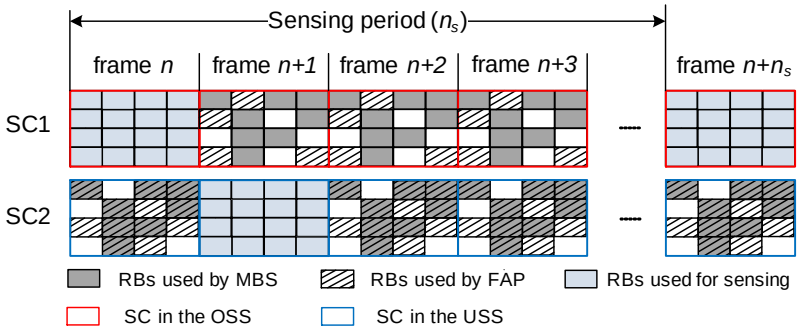


Fig. 4: The principle of the proposed sensing mechanism.

its allocation pattern (determined by eavesdropping technique implemented at the side of the SCs), the mutual interference among the SCs in the OSS and the MUEs may occur. To avoid interference to the MUEs, the distributed algorithm forces the SCs in the OSS to switch to the USS until the next sensing frame. After that, the SCs may change back to the OSS mode if it is beneficial to the SUEs. It is obvious that if the MBS changes its allocation pattern more often than the decision interval itself, the proposed scheme tends to use the USS more often even if the OSS would be more efficient. We study the impact of frequency of MBS' allocation pattern changes on the performance later in the paper.

In order to select the most beneficial allocation mode, each SC needs to estimate γ_{fi}^{O*} and γ_{fi}^{U*} (note that in our paper the estimated values are distinguished by "*"). We set σ_n to 10 dB in order to reduce interference from the SCs that are not in the direct neighborhood (i.e., $\sum_{o=1}^{O_i^n} |h_{fo}|^2 p_o^O + \sum_{u=1}^{U_i^n} |h_{fu}|^2 p_u^U << NI$ in (2) and (3). This allows to simplify estimation of γ_{fi}^{O*} and γ_{fi}^{U*} so that the SC estimates γ_{fi}^{O*} and γ_{fi}^{U*} as:

$$\gamma_{fi}^{O*} \approx \frac{|h_{fi}|^2 p_i^O}{NI}, \tag{14}$$

$$\gamma_{fi}^{U*} \approx \frac{|h_{fi}|^2 p_i^U}{NI + \sum_{u=1}^{U_i^d} |h_{fu}|^2 p_u^U + |h_{fb}|^2 P} \tag{15}$$

where interference from the SCs that are not neighbors is neglected when compared to (2) and (3), respectively. The above-mentioned simplification of SINR estimation, can result in an inaccurate estimation of SINR and wrong selection of the MCS. If this situation occurs, the SCs corrects wrongly estimated SINR value as explained in the next paragraph describing principle of Algorithm 2.

The distributed algorithm for selection of allocation mode to individual SCs is described in Algorithm 2. In general, the SCs have to be able to estimate when the change of allocation mode is beneficial for them. After the SC performs the sensing (line 1 in Algorithm 2) and if the USS is used, the SC needs to evaluate whether the OSS would be more efficient at the moment. As the first step, the SC determines $n_{RB,i}^{O*}$ and $n_{RB,r,i}^{O*}$. The $n_{RB,i}^{O*}$ is calculated according to (7). To derive $n_{RB,r,i}^{O*}$ for all active SUEs, we need to estimate p_i^{O*}, NI^* (line 3) and calculate γ_{fi}^{O*} using (14) for each active SUE (lines 4-7). The p_i^{O*} is simply derived as $p_i^{O*} = P_{max}$ since the SC in the OSS always transmits with P_{max} as already explained in Section III. The NI^* can be taken from the time when the SC has been in the OSS last time. The reason is that NI^* depends on noise and transmission power of the MBSs that both are almost constant values (the SCs in the area of other MBSs are below noise even if those are in the OSS due to their low coverage). After that, the SC estimates γ_{fi}^{O*} for each SUE (line 5). Since the estimation of γ_{fi}^{O*} could be incorrect because of the neglecting the non-direct neighbors interference in (14), the SC corrects the estimated values of γ_{fi}^{O*} for each SUE by means of correction factor ξ_{fi}^{O} obtained from previous estimations (lines 6). The correction factor is defined as a difference between the real observed γ_{fi}^{O} according (2) after the SC switches to the OSS and the estimated value according to (14). Various learning and prediction techniques can be applied to further minimize the estimation errors, but this is left for future research. Then, the SC calculates η_i^{O*} according to (9) and ϵ_i^{O*} according to (11) (line 9). If the OSS is estimated to be more beneficial when both transmission and energy consumption are considered, the SC switches to the OSS (line 11) and advertise its direct neighbors. After moving to the OSS, the SC can use RBs currently occupied by neither the MBS nor neighboring SCs. Since γ_{fi}^{O*} is estimated value and channels change over time, an error may occur even if ξ_{fi}^{O} is considered. This could lead to selection of an incorrect MCS resulting in overestimating efficiency of the OSS. In such a case, the SC updates ξ_{fi}^{O} for each active SUE (lines 13-15), switches back to the USS and advertise its direct neighbors (line 16).

If the SC is in the OSS, it evaluates if the USS would be more appropriate (see Algorithm 2, lines 20-42). The SC determines $n_{RB,i}^{U*}$ and $n_{RB,r,i}^{U*}$. The $n_{RB,i}^{U*}$ is calculated according to (8). To derive $n_{RB,r,i}^{U*}$ we need to determine Γ_z for all active SUEs (see (6)). As explained before, Γ_z could be calculated if we know γ_{fi}^{U*} for each active SUE. To estimate γ_{fi}^{U*} according to (15) we exploit the fact that p_i^{U*} is set in line with (5). Then, estimated γ_{fi}^{U*} is corrected by the correction factor ξ_{fi}^{U} similarly as in previous case (line 24). In case that γ_{fi}^{U*} for some active SUE(s) is lower than minimal required SINR (γ_{min}) the algorithm is stopped and the SC remains in the OSS for the time being to protect the SUEs from any connection drops (line 25-29). Otherwise, we calculate $n_{RB,r,i}^{U*}$

Algorithm 2 : Distributed algorithm for selection of allocation mode

1: SC performs sensing (every n_s interval)
2: **if** (SC is in the USS) **then**
3: estimate $n^{O*}_{RB,i}$ acc. (7), p_i^{O*} and NI^*
4: **for** z=1:a (a is the number of all active SUEs) **do**
5: estimate γ_{fi}^{O*} acc. (14)
6: $\gamma_{fi}^{O*} = \gamma_{fi}^{O*} + \xi_{fi}^{O}$
7: **end for**
8: calculate $n^{O*}_{RB,r,i}$ acc. (6)
9: determine η_i^{O*} acc. (9) and ϵ_i^{O*} acc. (11)
10: **if** $\left(\alpha\eta_i^{O*} + (1-\alpha)\frac{1}{\epsilon_i^{O*}} > \alpha\eta_i^{U} + (1-\alpha)\frac{1}{\epsilon_i^{U}}\right)$ **then**
11: change to the OSS, change advertise to neigh.
12: **if** $\left(\alpha\eta_i^{O} + (1-\alpha)\frac{1}{\epsilon_i^{O}} < \alpha\eta_i^{U} + (1-\alpha)\frac{1}{\epsilon_i^{U}}\right)$ **then**
13: **for** z=1:a **do**
14: $\xi_{fi}^{O*} = \gamma_{fi}^{O} - \gamma_{fi}^{O*}$
15: **end for**
16: change to the USS, change advertise to neigh.
17: **end if**
18: **end if**
19: **end if**
20: **if** (SC is in the OSS) **then**
21: estimate $n^{U*}_{RB,i}$ acc. (8), p_i^{U*}, NI^*, $(\sum_{u=1}^{U_i^d} |h_{fu}|^2 p_u^U)^*$
22: **for** z=1:a (a is the number of all active SUEs) **do**
23: estimate γ_{fi}^{U*} acc. (15)
24: $\gamma_{fi}^{U*} = \gamma_{fi}^{U*} + \xi_{fi}^{U}$
25: **if** $\gamma_{fi}^{U*} < \gamma_{min}$ **then**
26: skip the algorithm and stay in the OSS
27: **else**
28: calculate Γ_z
29: **end if**
30: **end for**
31: calculate $n^{U*}_{RB,r,i}$ acc. (6)
32: determine η_i^{U*} acc. (10) and ϵ_i^{U*} acc. (12)
33: **if** $\left(\alpha\eta_i^{U*} + (1-\alpha)\frac{1}{\epsilon_i^{U*}} > \alpha\eta_l^{O} + (1-\alpha)\frac{1}{\epsilon_i^{O}}\right)$ **then**
34: change to the USS, change advertise to neigh.
35: **if** $\left(\alpha\eta_i^{U} + (1-\alpha)\frac{1}{\epsilon_i^{U}} < \alpha\eta_i^{O} + (1-\alpha)\frac{1}{\epsilon_i^{O}}\right)$ **then**
36: **for** z=1:a **do**
37: $\xi_{fi}^{U*} = \gamma_{fi}^{U} - \gamma_{fi}^{U*}$
38: **end for**
39: change to the OSS, change advertise to neigh.
40: **end if**
41: **end if**
42: **end if**

for all active SUEs in line with (6) (line 31). As the next step, η_i^{U*} is determined according to (10) and ϵ_i^{U*} according to (12). If the USS is estimated to be more beneficial, the SC changes its mode to the USS (line 33) and advertisement is sent to all direct neighbors. The estimation whether mode should be changed to the USS or not can be erroneous since γ_{fi}^{U*} is estimated according to (15) while real value corresponds to (3). Hence, if the OSS proves to be more efficient than the USS, the SC updates correction factor ξ_{fi}^{U} for all SUEs (line 36-38) and returns back to the OSS (line 39).

The distributed algorithm significantly reduces the complexity of the centralized approach described in previous subsection. To be more precise, the complexity of Algorithm 2 is $O(a^2)$, where a is the number of active SUEs attached to the SC. Since the number of SUEs attached to the SC is limited, the proposed distributed algorithm is feasible for real networks (note that the complexity is the same as for HSS according to [21]). Still, to implement the algorithm into real networks, a signaling mechanism needs to be introduced. In this regard, each SC needs to know the activity status and mode status of its direct neighbors. This information facilitates to determine $n_{A,i}$ and $n^{O}_{A,i}$, which is further exploited in estimation of $n^{O}_{RB,i}$ and $n^{U}_{RB,i}$, respectively (see (7) and (8)). Hence, each SC sends notification to all its direct neighbors whenever it changes its activity status (i.e., if it becomes active/inactive) by means of Activity_Status_Change message. Analogously, the SC informs its neighbors whenever it changes its mode (i.e., if the SC switches from the OSS to the USS or vice versa) through Mode_Status_Change message. Both messages contain only address of recipient SCs and information on status change. Hence, the size of each signaling message with consideration of overhead from all layers of protocol stack is approximately 800 bits [43].

V. Simulation methodology

This section describes the methodology used in the subsequent simulations. The simulations are performed in MATLAB. We assume FDD LTE-A release 12 with parameters

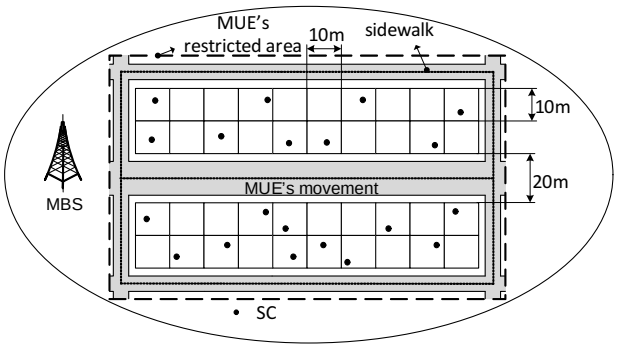


Fig. 5: Simulation model and deployment (position of the SCs is illustrative for one random drop).

TABLE I: Parameters and setting for simulations

Parameter	**Value**
Carrier frequency f [GHz]	2.0
MBS/SC channel bandwidth BW [MHz]	20/20
Max./min. SC transmission power P_{max} / P_{min} [dBm]	10/-20
MBS transmission power [dBm]	43
Noise [dBm/Hz]	BW4pW/GHz
Number of SCs/SUEs/MUEs [-]	20/40/20
Loss of internal wall, external wall, window [dB]	5, 10, 3
γ_{min} [dB]	-9.478 [46]
Sensing period [s]	0.2
Indoor path loss model	ITU-RP.1238
Outdoor path loss model	COST 231
Simulation time[s]/Number of drops	10 000/10

set-up aligned with Small cell forum as presented in Table I. We consider dual strip model [44] with 20 SCs randomly placed in 40 apartments (see example in Fig. 6, where the neighboring MBSs are not depicted for the sake of simplicity). The simulations are done for 10 drops (the SCs location is randomly generated for each drop) and final results are averaged out. Note that 10 drops are sufficient in our case since the averaged results for 5 and more drops are the same. In addition, we consider that the locations of dual strip block is at the edge of the MBS cell since the performance of the cell edge users is the most critical (the MUEs receive strong interference signal from neighboring MBSs equal roughly to -70 dBm). We assume 20 MUEs moving along the sidewalk (see Fig. 6). Initial position and movement direction of each MUE is selected randomly. After that, the MUEs are moving along straight trajectories with speed of 1 m/s. If the MUE leaves the MUE's restricted area, a new MUE enters the scenario at the opposite side. The SUEs are moving according to the mobility model specified in [45], where the movement is limited by apartment boundaries. Consequently, none of the SUE leaves the apartment within the simulation time.

We consider FTP traffic model for the SUEs [44] resulting in mean SCs load varying from 1 Mbit/s to 15 Mbits/s and maximum SC loads (i.e., load when all SUEs are simultaneously active) varying from 2.5 Mbit/s to 37.5 Mbit/s. The activity of the MUEs is modeled also as FTP traffic . For the purpose of our simulation, the mean load of the MBS varies from 0% to 100%. While the former case represents one extreme scenario when no MUEs activity occurs, the latter corresponds to opposite extreme scenario when the MBS is fully loaded all the time and the SCs have to access all RBs in the conventional USS mode.

In the simulation, the energy consumption of the SCs is measured in terms of transmission power of the SCs required to serve SUEs over the whole simulation cycle. The energy consumption of individual modes is derived from (11) and (12), and it is proportional to the amount of RBs used for transmission together with transmission power allocated to these RBs. Note that in figures presented in the next section, the energy consumption is related to the highest energy consumption generated by the SCs if the OSS is used and the MBS load is 0%.

In the simulation, we distinguish three hybrid schemes: 1) distributed HSS scheme taken without changes from [21] (labeled as "HSS [21]"), 2) distributed EDSS scheme improved as described in Section 4.C. (labeled as "EDSS-dis"), and 3) centralized EDSS as described in Section 4.B. acting as a benchmark for the performance (labeled as "EDSS-cen"). In addition, we consider the USS when the SCs restrict their transmission power on all the RBs indiscriminately whether this are used by the MBs or not (based on [10]-[14]). Moreover, the OSS enabling the SCs to utilize only the RBs currently not used by the MBS is assumed (based on [15]-[20]). In the simulation, we consider that sensing is not always perfect and, hence, false alarms (SC determines that RB is occupied by the MBS or SC despite the fact that this RB is actually free) or missed detections of primary signal (SC determines that RB is not occupied by the MBS or SC while the RB is not free) can occur. In line with [47], we consider 10% probability of both missed detection and false alarms.

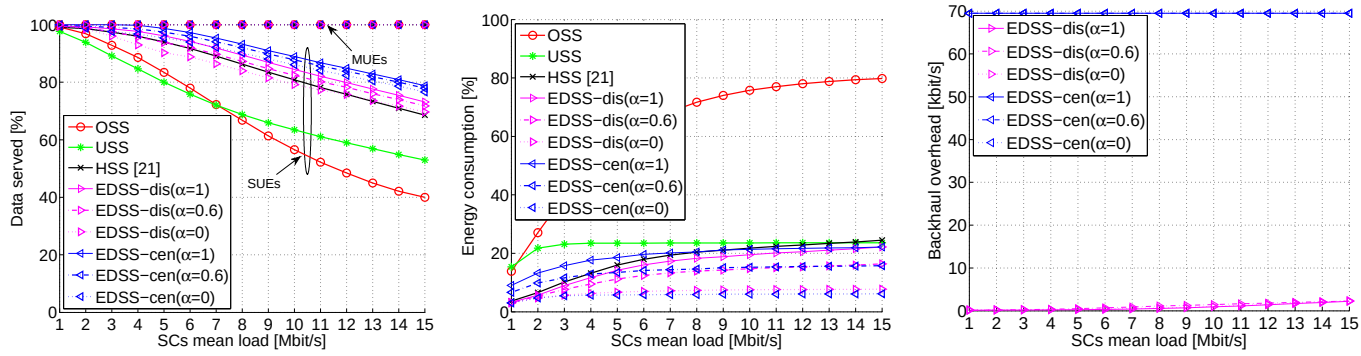
VI. Simulation results

Simulation results are divided into three sub-sections. The first two sections show the SUEs and MUEs data served, energy consumption of the SCs and generated backhaul signaling for various SCs loads and MBS loads, respectively. The third section analyzes how the ratio of served data of the SUEs and the MUEs is affected by various frequency of MBS allocation pattern change.

A. Performance evaluation for varying SCs load

This subsection shows the performance of individual schemes depending on the SCs load while mean MBS load is equal to 50%. The amount of data served for both the SUEs and the MUEs is depicted in Fig. 6a. The served data of the MUEs is intact for all compared schemes. The reason is that the MBS is able to serve all MUEs data at mean 50% load despite occasionally higher interference to the MUEs. Regarding the amount of data served for the SUEs, the results are influenced by the selection of α. If the preferences are given solely to the transmission efficiency (i.e., $\alpha = 1$), the EDSS-cen outperforms all schemes in terms of SUEs served data up to 38.8% (OSS), up to 26% (USS) and up to 10.3% (HSS [21]). At the same time, the performance gap between the EDSS-cen and EDSS-dis is 5.7% at most. Note that the performance gain of EDSS-dis over HSS [21] (up to 4.6%) is enabled by the SCs switching to the USS whenever the MBS changes its allocation pattern. Hence, the interference at the RBs is suppressed resulting in a higher served data. In case that the SCs select mode only according to energy consumption (i.e., $\alpha = 0$), the data served for the SUEs is decreased by up to 3.3% for the EDSS-cen and up to 5.9% for the EDSS-dis comparing to case with $\alpha = 1$. As a consequence, the EDSS-dis performs slightly worse than HSS [21] if SCs load is below 11 Mbit/s. Still, at higher SCs load, the served data by the EDSS-dis and HSS [21] equals and both schemes are able to significantly outperform either the OSS or the USS in terms of SUEs' throughput.

Fig. 6b depicts the energy consumption of the SCs. The highest energy consumption is observed for the OSS since the SCs transmit with the maximal transmission power. For this case, the energy consumption varies (depending on the load of SC) between 13.8% and 79.8% of the maximum energy consumption reached by the SC in OSS with 0% load of MBS. In case of the USS, the energy consumption is lowered roughly up to one third and varies between 15.4% and 23.6%. With respect to the USS, energy consumption is further significantly reduced by the HSS [21] if the SCs load is low (below 5 Mbit/s). The EDSS-dis slightly outperforms the HSS [21] by up to 2.3% even if mode selection is done only with respect to transmission efficiency (it varies between 3.2% and 22.2%). The reason is that in case of the EDSS-dis, the SCs are more often in the USS if the MBS changes its allocation patter (see explanation in Section 4C) and the USS is less energy demanding, in general. Contrary, the energy



(a) Amount of served data for the SUEs and MUEs

(b) Energy consumption of the SCs

(c) Amount of backhaul signaling overhead

Fig. 6: Performance of individual schemes depending on varying SCs load (MBS load=50%).

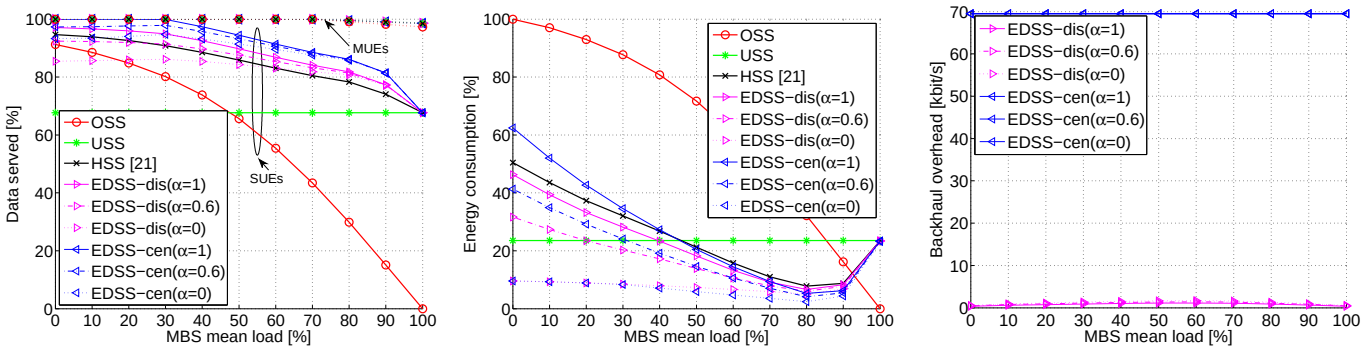
consumption of the EDSS-cen is higher when compared to EDSS-dis (between 9% and 22.2%) if energy consumption has priority in mode selection ($\alpha = 1$). If the preference is switched solely to the energy consumption of the SCs $\alpha = 0$, the energy consumption of the SCs for both the EDSS-cen and EDSS-dis is significantly reduced comparing to $\alpha = 1$ (see Fig. 6b). In case of the EDSS-cen, the energy consumption is lowered by up to 16% for the highest SCs load. This corresponds to approximately up to 13 times less energy consumption of the SCs if compared to the OSS and nearly 4 times less energy consumption with respect to the USS and the HSS [21]. Similarly, the EDSS-dis is able to minimize energy consumption by up to 14.5% resulting in roughly 10 times less consumption (compared to the OSS) and 3 times less consumption (compared to the USS and HSS [21]).

Fig. 6a and Fig. 6b also illustrate suitable trade-off between data served to the SUEs and energy consumption of the SCs for $\alpha = 0.6$. In this case, the reduction of energy consumption of the SCs is by up to 6.4% (EDSS-cen) and 5.8% (EDSS-dis). This corresponds to reduction of energy consumption by approximately 5 times (compared to the OSS) and 1.5 times (compared to the USS) for both the EDSS-cen and the EDSS-dis. At the same time, a decrease in the data served of the SUEs is only marginal (by up to 1.3% for EDSS-cen and 2.3% for the EDSS-dis) when compared to scenario with $\alpha = 1$ (see Fig. 6b). Hence, both proposed scheme are still able to outperform the HSS [21], the OSS, and the USS.

Fig. 6c illustrates the signaling overhead generated over backhaul that enable to implement both the EDSS-cen and EDSS-dis. While the EDSS-cen generates roughly 70 kbit/s over the backhaul, the EDSS-dis significantly reduces backhaul overhead disregarding α (up to 2.5 kbit/s backhaul overhead is generated). The main reason for high backhaul overhead of the EDSS-cen with respect to the EDSS-dis is that in case of the EDSS-cen, the MBS has to sent new allocation pattern to all its SCs whenever it is beneficial to the SUEs. Contrary, in case of the EDSS-dis, the SCs exchange only limited amount of information among its direct neighbors quite infrequently.

B. Performance evaluation for varying MBS load

The detail impact of varying MBS load on the performance of all individual schemes is depicted in Fig. 7. Note that the SCs load remains the same for all MBS loads and equals to 8 Mbit/s for each SC. Fig. 7a shows that the data served to the MUEs is affected just marginally and only at heavy MBS loads. To be more precise, at fully loaded MBS, the performance of MUEs is degraded by 2.7% (OSS), by 1.3% (USS), by 1.6% (HSS [21]), and by 1.4% (EDSS-cen, EDSS-dis). Regarding the data served to the SUEs, we can observe that if the MBS is not loaded at all, the performance of all schemes except the USS reaches its maximum. In the opposite extreme (i.e., if the MBS load is always 100%), the results are the same for all schemes but the OSS, which has no resources available left (note that the USS is always used by the HSS [21], EDSS-cen and EDSS-dis at full MBS load). If the preferences are given solely to the SUEs ($\alpha = 1$), the EDSS-cen outperforms the USS by up to 32.4%, the OSS by up to 67.7% and the HSS [21] by up to 9.1% (see Fig. 7a). At the same time, the performance gap between the EDSS-cen and EDSS-dis is always less than 5%. If the preferences are changed to the energy consumption ($\alpha = 0$), the data served by the EDSS-cen and the EDSS-dis is decreased by up to 6.7% and 11.6%, respectively. Nevertheless, the data served to the SUEs is notable decreased only at very low MBS load (less than 20%). With higher MBS load, both the EDSS-cen and the EDSS-dis results are getting closer to the scenario with $\alpha = 1$. Fig. 7b shows the energy consumption of the SCs. Similarly as in Fig. 6b the highest energy consumption is observed by the OSS (except the cases where the mean MBS load is 90% and higher as the SCs have only small amount of RBs at their disposition). In case of the USS, the energy consumption is notably lowered comparing to the OSS and remains constant for all MBS loads (23.5%). The energy consumption of the HSS [21] varies between 50.4% and 7.8% depending on the MBS load. To be more precise, the highest energy consumption is observed at low MBS loads (40% and lower). This is due to the fact that the SCs in the HSS [21] at low MBS loads utilizes the OSS more frequently due to its high transmission efficiency. Contrary, at MBS loads varying between 50% and 90%, the HSS [21]



(a) Amount of served data for the SUEs and MUEs

(b) Energy consumption of the SCs

(c) Amount of backhaul signaling overhead

Fig. 7: Performance of individual schemes depending on varying MBS load (SCs load = 8 Mbit/s).

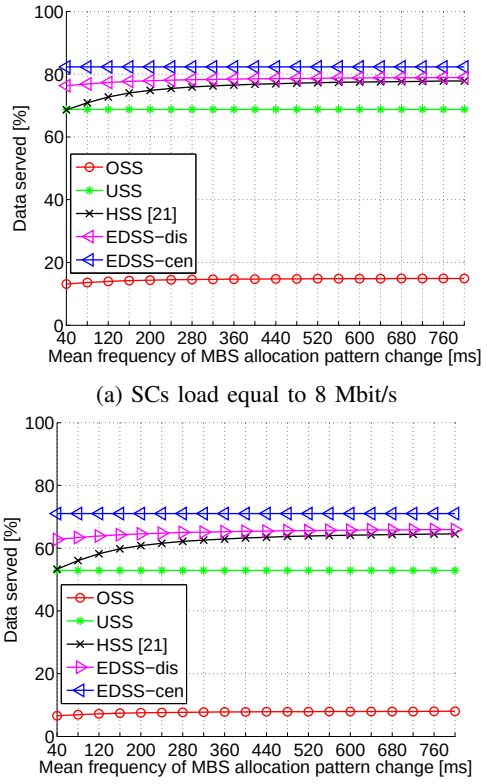
is more energy efficient than either the OSS or the USS. Analogously, energy consumption of both the EDSS-cen and EDSS-dis scheme are similar to the HSS [21] if $\alpha = 1$ (EDSS-dis slightly outperforms the EDSS-cen and HSS [21]). On the other hand, if we give preferences solely to the energy consumption ($\alpha = 0$), both the EDSS-cen and EDSS-dis are able to significantly lower energy consumption of the SCs, especially at lower MBS loads. Consequently, energy consumption of the SCs is decreased by these schemes by 7 times (compared to the OSS), by 2.6 times (USS), and by 2.8 times (HSS [21]) on average.

Fig. 7a and Fig. 7b further show that a suitable trade-off between data served to the SUEs and energy consumption of the SCs can be found for $\alpha = 0.6$. In such case, the energy consumption is notably lowered even for low MBS loads. Simultaneously, the decrease in amount of data served to the SUEs is only marginal for low MBS loads. As a consequence, the data served to the SUEs is decreased only by up to 2.6% (EDSS-cen) and by up to 4.6% (EDSS-dis) when compared to scenario with $\alpha = 1$ (see Fig. 6b).

Fig. 7c shows the amount of overhead generated over the backhaul. Analogously to Fig. 6c, the use of EDSS-cen results in huge signaling overhead when compared to the EDSS-dis. As already explained earlier the reason is that the EDSS-cen requires exchanging of significant amount of information over the backhaul in order to select the allocation mode for each SC.

C. Performance evaluation of proposal for various frequency of MBS allocation pattern change

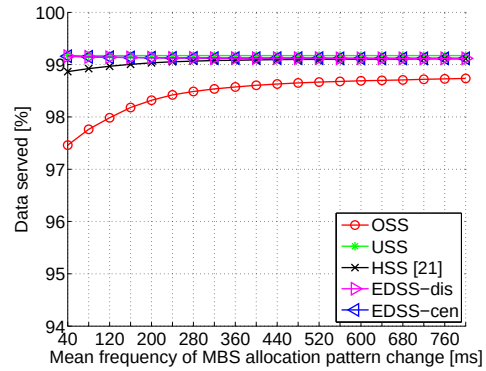
The important aspect of the EDSS-dis is that its performance is dependent on the frequency with which the MBS changes allocation pattern. In previous figures we consider that the MBS changes allocation pattern each 200 ms on average (as explained in Section IV this corresponds to the sensing interval). How the different mean frequency of change of allocation pattern influences the data served to the SUEs is shown in Fig. 8 for mean MBS load equal to 90% and SCs load equal to 8 Mbit/s (Fig. 8a) and 15 Mbit/s (Fig. 8b). The change of allocation pattern does not influence the USS and the EDSS-cen. Contrary, if the change of MBS allocation pattern occurs frequently the HSS [21] is the worst affected (decrease in SUEs served data by up to 9.2% and 11.2% for SCs load of 8 Mbit/s and 15 Mbit/s, respectively). In these cases the data served by the HSS [21] is similar to the USS. The reason for degradation of HSS [21] performance is that it is not able to sufficiently react to changes of MBS allocation pattern and high interference from the MBS may occur (i.e., if the MBS change allocation pattern, the SCs wait until sensing interval to select the more beneficial mode). The performance of the EDSS-dis is also degraded by fast changes of MBS allocation pattern. However, the degradation in SUEs served data is only



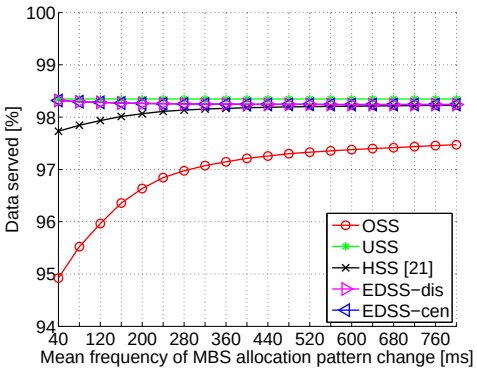
(a) SCs load equal to 8 Mbit/s

(b) SCs load equal to 15 Mbit/s

Fig. 8: The amount of SUE's data served depending on mean frequency of MBS allocation pattern.



(a) SCs load equal to 8 Mbit/s



(b) SCs load equal to 15 Mbit/s

Fig. 9: The amount of MUE's data served depending on mean frequency of MBS allocation pattern.

by up to 2.6% (SCs load of 8 Mbit/s) and 3.2% (SCs load of 15 Mbit/s) if the MBS allocation pattern is shortened from 800 ms to 40 ms. Fig. 8 further illustrates that the amount of data served of the SUEs is marginally decreased since the higher interference is observed at the RBs if the MBS suddenly changes its allocation pattern.

Fig. 9 depicts the data served of the MUEs for different mean frequency of allocation pattern change. Similarly as in case of the SUEs, the performance of the MUEs is not negatively influenced for the USS and the EDSS-cen. Contrary, the highest degradation of MUEs served data is observed by the OSS (up to 2.5%). The reason is high interference when the MUEs and the SCs occupy the same RBs. Fig. 9 also illustrates that the HSS [21], and EDSS-dis affect the MUEs negligibly as only up to 0.5% (HSS [21]), and 0.1% (EDSS-dis) decrease in their performance is demonstrated even if the change of allocation pattern occurs frequently.

From Fig. 8 and Fig. 9 can be concluded that we are able to significantly boost the performance of the EDSS-dis when compared to the HSS [21] by enhancement described in Section IVC. Even if the change of MBS allocation pattern is quite often, the SUEs' data served by the EDSS-dis is decreased only marginally and only for fast changes of allocation pattern (up to 200 ms) and the EDSS-dis always outperforms the HSS [21]. If the change of allocation pattern is not that often, the performance of the EDSS-dis converges to its maximum.

VII. Conclusion

In this paper, we have proposed dynamic selection of overlay/underlay mode for cognitive OFDMA-based small cells that can be easily implemented in emerging 5G mobile networks. The main idea is to dynamically select the allocation mode, which is at the moment more beneficial solely in terms of small cell users throughput, only with respect to the energy consumption of the SCs, or if combination of both is assumed. To accomplish that, we have proposed the centralized algorithm for dynamic mode selection. However, this solution is of high complexity and generates significant amount of signaling over backhaul. Hence, we have also proposed distributed algorithm lowering the both complexity and signaling overhead due to its distributive nature. The results indicate that the dynamic switching between overlay and underlay spectrum sharing modes can significantly improve performance of small cell users while the performance of the macrocell users is not or only marginally degraded. In addition, energy consumed by transmission of the SCs can be significantly reduced by our proposal when compared to other competitive schemes if energy consumption of the SCs in spectrum sharing modes is considered.

As the future work, the proposed mechanism can incorporate various learning techniques to minimize the probability of wrongly estimated parameters during the selection of more beneficial mode. Furthermore, a power control for protection of UEs against interference in USS can be optimized.

References

[1] Cisco, Visual Networking Index, white paper at Cisco.com, 2014.
[2] J. G. Andrews, et. al, "What Will 5G Be?" *IEEE J. Sel. Areas Commun.*, 32 (6), 1065-1082, 2014.
[3] A. L. Swindlehurst, E. Ayanoglu, P. Heydari, and F. Capolino, "Millimeter-Wave Massive MIMO: The Next Wireless Revolution?", *IEEE Commun. Mag*, 52 (9), 56-62, 2014.
[4] A. Golaup, M. Mustapha. M., and L.B. Patanapongpibul, "Femtocell Access Control Strategy in UMTS and LTE", *IEEE Commun. Mag*, 47 (9), 117-123, 2009.
[5] V. Chandrasekhar, J.G, Andrews, "Spectrum Allocation in Shared Cellular Network", *IEEE Trans. Commun.*, 57 (10), 30593068, 2009.
[6] S.A. Rubaye, A.A. Dulaimi, and J. Cosmas, "Cognitive Femtocells", *IEEE Vehicular Technology Magazine*, 6 (1), 44-51, 2011.
[7] A. U. Ahmed, M. T. Islam, and M. Ismail, "A review on Femtocell and its Diverse Mitigation Techniques in Heterogeneous Network", *Wireless Personal Communications*, 78, 85-106, 2014.
[8] W. Wang, G. Yu, and A. Huang, "Cognitive Radio Enhanced Interference Coordination for Femtocell Networks", *IEEE Communications Magazine*, 51 (6), pp. 37-43, 2013.
[9] H. O. Kpojime, G. A. Safdar, "Interference Mitigation in Cognitive Radio based Femtocells", to be published in *IEEE Communications Surveys & Tutorials*, 2014.
[10] H. Saad, A. Mohamed, and T. ElBatt, "A Cooperative Q-learning Approach for Online Power Allocation in Femtocell Networks", *IEEE Vehicular Technology Conference (VTC fall)*, pp. 1-6, 2013.
[11] H. Wang, R. Song, "Distributed Q-Learning for Interference Mitigation in Self-Organised Femtocell Networks: Synchronous or Asynchronous?", *Wireless Personal Communications*, 71 (4), pp. 2491-2506, 2013.
[12] R. Kurda, L. Boukhatem, T.A. Yahiya, and M. Kaneko, "Power adjustment mechanism using context information for interference mitigation in two-tier heterogeneous networks", *IEEE Symposium on Computers and Communications (ISCC)*, pp. 1-6, 2014.
[13] D. Sun, X. Zhu, Z. Zeng, and S. Wan, "Downlink Power Control in Cognitive Femtocell Networks", In *WCSP*, 1-5, 2011.
[14] X. Tao, Z. Zhao, R. Li, J. Palicot, and H. Zhang, "Downlink interference minimization in cooperative cognitive LTE-femtocell networks", *EURASIP Journal on Wireless Communications and Networking*, 2013.

[15] L.S. Yu, T.Ch. Cheng, Ch.K. Cheng, S.Ch. Wei, "Cognitive Radio Resource Management for QoS Guarantees in Autonomous Femtocell Networks" in Proc. *IEEE ICC*, pp.1-6, 2010.

[16] I. Demirdogen, I. Guvenc, H. Arslan, "Capacity of Closed-Access Femtocells Networks with Dynamic Spectrum Reuse", in Proc. *IEEE PIMRC*, pp. 1315-1320, 2010.

[17] L. Huang, G. Zhu, and X. Du, "Cognitive femtocell networks: an opportunistic spectrum access for future indoor wireless coverage", *Wireless Commun.*, 20 (2), 44-51, 2013.

[18] Y. S. Soh, T. Q. S. Quek, M. Kountouris, G. Caire, "Cognitive Hybrid Division Duplex for Two-Tier Femtocell Networks", *IEEE Trans. Wireless Commun.*, 12 (10), 4852-4865, 2013.

[19] J. Lee, et al. , "Traffic Pattern-based Opportunistic Spectrum Access of Cognitive Femto Base Stations for Decentralized Cross-tier Interference Management", In *ICUFN*, 352-356, 2014.

[20] F. Tariq, L. S. Dooley, A. S. Poulton, "An interference-aware virtual clustering paradigm for resource management in cognitive femtocell networks", *Computers and Electrical Engineering*, 40 (2014), pp. 587-598, 2014.

[21] P. Mach, Z. Becvar, "Distributed Hybrid Spectrum Sharing for OFDMA-based Cognitive Femtocells in 5G Networks", *European Wireless*, 1-6, 2015.

[22] J. Oh, W. Choi, "A Hybrid Cognitive Radio System: A Combination of Underlay and Overlay Approaches", In *IEEE VTC-fall*, 1-5, 2010.

[23] M. Elalem, L. Zhao, "Optimal Access Strategy for Capacity Optimization in Cognitive Radio System" In *IEEE WCNC*, 3382-3386, 2013.

[24] A. K. Karmokar, S. Semthuran, A. Anpalagan, "Physical layer-optimal and cross-layer channel access policies for hybrid overlay-underlay cognitive radio networks", *IET Communications*, 8 (15), 2666-2675, 2014

[25] X. Tan, H. Zhang, and J. Hu, "Achievable Transmission Rate of the Secondary User in Cognitive Radio Networks with Hybrid Spectrum Access Strategy", *IEEE Commun. Lett.*, 17 (11), 2088-2091, 2013.

[26] Bansal, G.; Hossain, M.J.; Bhargava, V.K.; Tho Le-Ngoc, "Subcarrier and Power Allocation for OFDMA-Based Cognitive Radio Systems With Joint Overlay and Underlay Spectrum Access Mechanism," *IEEE Transactions on Vehicular Technology*, 62 (3), pp. 1111-1122, March 2013.

[27] Senthuran, S.; Anpalagan, A.; Das, O., "Throughput Analysis of Opportunistic Access Strategies in Hybrid Underlay-Overlay Cognitive Radio Networks," *IEEE Transactions on Wireless Communications*, 11 (6), pp. 2024-2035, June 2012.

[28] Y. Chen, Q. Lei, and X. Yuan, "Resource allocation based on dynamic hybrid overlay/underlay for heterogeneous services of cognitive radio networks", *Wireless Personal Communication*, 2014 (79), 1647-1664.

[29] X. Huang, T. Han, and N. Ansari, "On Green-Energy-Powered Cognitive Radio Networks", *IEEE Communications Surveys & Tutorials*, 17 (2), 2015.

[30] M. Wildemeersch, T. Q. S. Quek, C. H. Slump, and A. Rabbachin, "Cognitive Small Cell Networks: Energy Efficiency and Trade-Offs, *IEEE Transactions on Communications*, 61 (9), pp. 4016 - 4029, 2013.

[31] Z. Luo, M. Ding, and H. Luo, "Dynamic Small Cell ON/OFF Scheduling Using Stackelberg Game", *IEEE Communications Letters*, 18 (9), pp. 1615-1618, 2014.

[32] Y. Qu, Y. Chang, Y. Sun, and D. Yang, "Equilibrated Activating Strategy with Small Cell for Energy Saving in Heterogeneous Network", *IEEE Vehicular Technology Conference (VTC-fall)*, pp. 1-6, 2014.

[33] C. Yang, J. Li, X. Jiang, and A. Anpalagan, "Interference-Aware Spectral-and-Energy Efficiency Tradeoff in Heterogeneous Networks", *IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 819 - 824, 2015.

[34] J. Tang, D. K. C. So, E. Alsusa, K. A. Hamdi, and A. Shojaeifard, "Resource Allocation for Energy Efficiency Optimization in Heterogeneous Networks", *IEEE Journal on Selected Areas in Communications*, 33 (10), pp. 2104 - 2117, 2015.

[35] Y. Bai, J. Zhou, L. Chen, "Hybrid Spectrum Usage for Overlaying LTE Macrocell and Femtocell", *GLOBECOM*, pp. 1-6, 2009.

[36] L. Yang, L. Zhang, T. Yang, W. Fang, "Location-Based Hybrid Spectrum Allocation and Reuse For Tiered LTE-A Networks", *VTC-spring*, pp. 1-5, 2011.

[37] W. Liu, X. Chen, W. Wang, "Price-Based Uplink Resource Allocation for Hybrid-Spectrum Femtocell Networks", *WCNC*, 2665-2670, 2012.

[38] Y. Bai, L. Chen, "Hybrid Spectrum Arrangement and Interference Mitigation for Coexistence between LTE Macrocellular and Femtocell Networks", *EURASIP Journal on Wireless Communication and Networking*, 2013.

[39] P. Mach, Z. Becvar, "Enhancement of Hybrid Cognitive Approach for Femtocells", *IEEE Vehicular Technology Conference (IEEE VTC-Spring 2015)*, 2015.

[40] W. Wang, T. Peng, and W. Wang, "Optimal Power Control Under Interference Temperature Constraints in Cognitive Radio Network", In *IEEE WCNC*, 116-120, 2007.

[41] 3GPP TS 36.211: Physical channels and modulation (Release 13), V13.1.0, 2016.

[42] 3GPP TS 36.300: Overall description; Stage 2 (Release 12), 2014.

[43] T. Vanek, M. Rohlik, "Alternative Protocols for Femtocell Backbone security", *Wireless and Mobile Networking Conference*, pp. 1-4, 2011.

[44] 3GPP TR 36.814: Further advancements for E-UTRA physical layer aspects (Release 9), 2010.

[45] Becvar, Z, et. al, "D21: Scenarios and requirements", Project TROPIC deliverable, 2013.

[46] J. Fan, et. al., "MCS Selection for Throughput Improvement in Downlink LTE Systems", *ICCCN*, 2011.

[47] S. Shellhammer, V. Tawil, G. Chouinard, M. Muterspaugh, and M. Ghosh, "Spectrum sensing simulation model", *IEEE 802.22*, 2006.

Pavel Mach received his MSc and PhD degree from the Czech Technical University in Prague in 2006 and 2010, respectively. At the present, he is a researcher at the Department of Telecommunication Engineering within the same university. He has been actively involved in several national and international projects founded by European Commission in FP6 and FP7 framework such as ROCKET, FREEDOM or TROPIC. In 2015, he has joined 5G mobile research lab funded at Czech Technical University focusing on key aspects and challenges related to future mobile networks and emerging wireless technologies. He has co-authored more than 50 papers in international journals and conferences. His research interest is mainly related to radio resource management in emerging wireless technologies.

Zdenek Becvar received M.Sc. and Ph.D. in Telecommunication Engineering from the Czech Technical University in Prague, Czech Republic in 2005 and 2010, respectively. Now, he is Associate Professor at the Department of Telecommunication Engineering, Czech Technical University in Prague, Czech Republic. From 2006 to 2007, he joined Sitronics R&D center in Prague focusing on speech quality in VoIP. Furthermore, he was involved in research activities of Vodafone R&D center at Czech Technical University in Prague in 2009. He was on internships at Budapest Politechnic, Hungary (2007), CEA-Leti, France (2013), and EURECOM, France (2016). In 2013, he becomes representative of the Czech Technical University in Prague in ETSI and 3GPP standardization organizations. In 2015, he founded 5Gmobile research lab at CTU in Prague focusing on research towards 5G mobile networks and beyond. He is a member of more than 15 program committees at international conferences or workshops and he published 3 book chapters and more than 60 conference or journal papers. He works on development of solutions for future mobile networks (5G and beyond) with special focus on optimization of radio resource management, mobility support, device-to-device communication, self-optimization, power control, architecture of radio access network, and small cells.