Dynamic Spectrum Access and Overlay Systems

Friedrich K. Jondral,* Ulrich Berthold,† Dennis Burgkhardt,* and Timo A. Weiß‡

Abstract—Dynamic spectrum access (DSA) plays a key role for an efficient use of spectrum and therefore is currently a rapidly growing research field, impacting not only the engineering community but also political, regulatory and economical mechanisms. The growing market of wireless communications requires a paradigm shift in all of these areas. In this paper we first give a general overview over the current developments in DSA with a focus on technological approaches and advances. Especially in the transition phase to greater general dynamics in spectrum access, overlay systems are a promising approach, because they allow coexistence and fast deployment in the same frequency band as an already existing system, without a need for changing the infrastructure or regulations of the primary system. In overlay systems the reliable and fast detection of upcoming primary users is of paramount interest. Therefore, this topic is thoroughly discussed, first of all for centralized and afterwards for decentralized overlay systems.

I. INTRODUCTION

Ubiquitous wireless communications more and more is taken for granted in everyday life. Starting from two directions, namely with the second and third generation cellular telephone networks on the one and with the internet access over wireless local area networks (WLANs) on the other side, mobile communications is moving towards the all in one mobile internet where wireless broadband access will replace digital subscriber lines.

Standardization and regulation processes evolve quite slowly compared to the rapid technological advances in communications. But newly assigned spectrum and prices paid for it, e.g. in the auctions for the Universal Mobile Telecommunications System (UMTS) frequencies in Europe, indicate the growing demand as well as the economic potential for wireless digital communications.

Thanks to the rapid development in integrated circuits and the connected computing powers of power-efficient and low-cost hardware, radios showing increasing flexibility may be employed: Software defined radios (SDRs) will implement all signal processing in software modules that can, dependent on actual needs, be exchanged fast and flexibly. Consequently, cognitive radios based upon SDRs will be able to autonomously adapt to spectral, local, power and policy needs [33], [34].

Many national regulation bodies in the meantime support a paradigm shift towards opening spectrum bands for dynamic access, e.g. [18]. Moreover, the operation of secondary communication systems is allowed in frequency ranges allocated to other, primary, services. Usually, changing software or hardware of these primary systems turns out to be difficult. Hence, secondary overlay systems will be forced to take care of not interfering with subscribers of these primary networks. The paradigm shift is motivated on the one hand by measurement campaigns [36], [5] indicating that wide frequency ranges are used only sporadically. On the other hand, frequency bands allowing unlicensed access get overcrowded with the extension of WLAN applications in densely occupied areas.

The license exempt spectrum usage by WLANs is a first example of DSA based on a simple listen-before-talk etiquette. Dynamic spectrum sharing between primary and overlay systems is now on the agenda, for example to allow secondary systems to operate in temporally and locally unused television (TV) channels [13].

The remainder of this paper is organized as follows: In Section II we introduce important notions concerning efficiency in spectrum use and give an overview of dynamic spectrum access technologies. The state of the art in spectrum sharing research as well as current standardization efforts in this field are reviewed in Section III. After that in Section IV, we take a closer look onto the special case of orthogonal frequency multiplexing (OFDM) based overlay systems possessing an access point and on the idea of cognitive radio (CR) in connection with spectrum pooling. The challenge related to the distributed acquisition of knowledge about the current spectrum usage by primary systems is covered in Section V. Section VI examines the more complex problem of an overlay system working in ad hoc mode. Finally, conclusions are drawn in Section VII.

II. SPECTRUM ALLOCATION AND EFFICIENCY

We constrain our discussion here to the frequency range from 300 MHz to 6000 MHz, i.e. to a total bandwidth of 5700 MHz. If WiMAX is regarded as a mobile system, we find that the frequency allocation to mobile systems in Europe is in total 1818.5 MHz or about 32% of the range under discussion. Out of this, 663.5 MHz are accessible through license exempt systems. The remaining 3881.5 MHz of spectrum are assigned to other systems as specified in the national frequency allocation plans. Several of these systems (navigation, astronomical sensing, etc.) must be protected absolutely from any disturbances. The frequency ranges of other systems (e.g. broadcast, radio relay networks, satellite links, RADAR, military radio, remote sensing) are candidates for the support of overlay, i.e. secondary users’ (SU) systems since the results of recent measurement campaigns [36], [5] as well as the structures of these systems suggest that huge parts of the spectrum allocated to them are only sporadically engaged. Whether an overlay system may be implemented in a certain frequency range therefore depends on time and location as well as on the primary users’ (PU) system.

A. Efficiency Concepts

The goal of dynamic spectrum access is to enhance the efficiency in spectrum use. Therefore, it is necessary to give
an explanation of efficiency concepts:

- The spectrum efficiency of a point-to-point radio link is defined by the number of bits that it can transmit per second and per Hertz, i.e., bits/Hz. Given the bandwidth of the link, this efficiency may for example be augmented by using advanced modulation methods or by implementing multiple input multiple output (MIMO) technology.

- The spectrum efficiency of a wireless system is defined as the number of bits that it can transmit within a fixed bandwidth per second and per Hertz (by all users) within a certain area. Given the area in square meters, its dimension is bits/Hz/m². This efficiency may for example be augmented by installing additional base stations or access points.

Both efficiencies introduced so far are upper limits that are defined with respect to parameters of the wireless link or system and derived from theoretical considerations.

- The efficiency in spectrum use¹, which is of interest here, is measured for a system in operation as the average number of bits that are transmitted within a fixed bandwidth per second and per Hertz (by all users and all systems) within a certain area. Therefore, its dimension is also bits/Hz/m². One way to augment this efficiency is the implementation of overlay systems that make use of the resources left idle by a PU system.

The spectrum efficiency of a point-to-point radio link and the efficiency in spectrum use can also be distinguished by either asking the question how to access the spectrum or when and where to access it (regarding the frequency range). Since both approaches are related to different layers, they complement one another and thus can also be combined.

1) How to access: This question includes all aspects regarding a single point-to-point transmission, under the prerequisite that the “when and where” is given. Accordingly, these optimizations are mainly performed in the physical layer. There is a variety of different concepts in this area, contributing to a more efficient spectral use, including OFDM [38], MIMO [40], or adaptive modulation [23]. All these concepts follow the goal of transmitting as much information as possible per time and frequency. One example is the transition from analog to digital TV broadcast: Instead of transmitting a single TV program, it is now possible to transmit two or more program channels within the same bandwidth.

2) When and where to access: Another possibility to increase spectral efficiency is to optimize the coordination of spectrum access. This is equivalent to avoiding “blocked spectrum”, i.e., idle, but reserved parts of the spectrum, which are therefore blocked for others. The demand for spectrum is time variant and therefore dynamic. Blocked spectrum originates from a spectrum allocation strategy which also may be dynamic, but does not perfectly match the demand. In order to avoid blocked spectrum, the allocation must be as dynamic as the demand. Note, that all following considerations are only feasible as long as there is overall enough (blocked or unblocked) spectrum available, i.e., meeting the demand for spectrum is only an issue of allocation strategy. Optimizing the strategies for spectrum allocation and constantly adapting it to the current demand leads to the concept of dynamic spectrum access. This actually means a different view on regulation—licenses must become much more dynamic and have a finer granularity with respect to bandwidth, time, and location. But not only the changed view on regulation is necessary, this required flexibility can only be achieved by tightly coupling policy, technological, and economical aspects.

On the way to full flexibility regarding spectrum allocation it is not always possible to introduce newly developed dynamic access strategies, as they are exemplarily presented in Section III-A, right away. There are some scenarios, e.g., during the transition phase, where it is not possible to avoid blocked spectrum.

One direction of our work is to find technical solutions that profit from idle resources (i.e., find and exploit blocked spectrum), e.g., to use temporarily idle frequencies by overlay systems and herewith to enhance the efficiency in spectrum use. One of these solutions, OFDM based overlay systems employing CRs, is discussed in detail in Section IV.

B. Categorizing Spectrum Access

Using a very simplified view, the electromagnetic spectrum is accessed by a variety of different transmitters and receivers designed for miscellaneous purposes, e.g., communication systems or sensor networks. In order to enable a successful coexistence, the spectrum access has to be coordinated, i.e., each transmitter needs to know exactly when and in which frequency range it is allowed to transmit. In the same manner, the receiver or receivers paired to that transmitter have to be informed at which frequency and with which bandwidth and modulation mode they may expect the respective transmission. In general, several receivers and transmitters are related to each other, forming a network that is supporting a certain standard. Due to physical limitations and hardware constraints, nodes within a specific network should be operating in a similar frequency range.

Based on these basic observations, spectrum access can be structured in a matrix as shown in Fig. 1. There are two main levels for the coordination of spectrum access. On the network level, only a coarse coordination of spectrum regions is performed, e.g., by regulatory bodies assigning parts of the spectrum to services and operators. Within this given frequency block, spectrum access regarding the involved nodes is coordinated by the applied communications standard. In the process of determining frequency and time of a node’s transmission, regulation is responsible for the coarse coordination

¹For simplicity reasons in the following we refer to “efficiency in spectrum use” as “spectral efficiency”.

Fig. 1. Different approaches and structure for the coordination of spectrum access.
(network coordination) and the standard then takes care of the detailed access coordination between participating nodes (node level). On the other hand, independent of the level, spectrum access can be divided into centralized and decentralized approaches, using either a global or local view, respectively. Regarding the network level, regulation embodies the centralized approach. A central instance, here the regulatory body, is in total control of all spectrum assignments. In case of a decentralized approach, we have the opposite situation: Several independent standards coexist in the same frequency region without any prioritization, as e.g. in the industrial, scientific and medical (ISM) bands. This is referred to as spectrum commons [45]. Note, that it is only important whether there is a central coordination instance or not. In [45] the term “dynamic exclusive use” is coined to establish a third category, grouping concepts that propose local dynamic spectrum assignment, e.g. [44], and spectrum property rights [24]. Nevertheless, these concepts have in common with regulation the time and location dependency, and also especially the central coordination instance. In fact, regulation varies frequency assignments in a much larger time scale and for larger geographical regions, but nevertheless it is not static. Therefore, we do not make a distinction in our categorization. On the node level, infrastructure based networks represent the centralized approach, since they rely on central base stations acting as masters in their dedicated cells and frequency bands. Accordingly, ad hoc networks stand for the distributed and decentralized approach for the coordination of spectrum access. Note, that there also are hybrid spectrum access strategies on both, the network and node level: For example, looking at the ISM bands, a dedicated frequency band is defined by regulation (centralized component), which is then used as spectrum commons (decentralized component). On node level, a base station can be connected with nodes in the cell, which are also connected to nodes outside the cell, using ad hoc mode.

III. CURRENT APPROACHES TO SPECTRUM SHARING

DSA and spectrum sharing research and newly developed frameworks are driven by the recognition of a seeming scarcity of spectral resources. This scarcity is often caused by inefficient and especially inflexible modes of spectrum usage. So first, a look onto the status quo in regulation shows one main motivation of DSA. The cognitive radio vision [37] is the other main driver of the very different approaches taken to achieve the desired spectral efficiency in an autonomous and dynamic way. This idea of autonomous and self-configuring radios and networks of them has inspired a tremendous range of research. In this section we aim at giving a coarse overview over this broad field. We continue with approaches that discuss the flexibilization of spectrum access and currently inflexible regulation. Besides a generic look onto proposed frameworks that use spectral resources left idle by an existing PU system by SU systems, such an approach will also be dealt with in detail by Sections IV and V. Afterwards, we discuss some research on competitive and distributed cognitive radio networks. After a short overview of current standardization efforts, a comment about ongoing regulatory discussions concludes this section.

A. Improving Spectrum Usage in the Current Situation

By issuing licenses, a regulating authority seeks to guarantee exclusive and interference free access to spectral resources for a certain wireless communication, measurement, or observation system. The granted licenses have legal character and hence, in the case of wireless communication systems, the operators and users of the system can rely on a service level that is not disturbed by arbitrary operations of other, nonlicensed systems. For providers of mobile communications who need to offer their services to paying customers, these guarantees for exclusive use are an essential part of their economic existence.

However, compared to the fast improvements in transmission technologies and standardization, a drawback of fixed allocation of licenses seems to be slowness with respect to adaptation. Whereas systems currently under development incorporate spectral efficiency, e.g. measured in bit/s/Hz/m², as a main design goal, many existing systems leave spectrum holes [25] in significant parts of their licensed frequency bands and operational regions.

The example of ISM bands, especially the most prominent one at 2.4 GHz, shows that the absence of specific licensing constraints to certain services encourages a plethora of new services to be established. But taking only two widely spread services like WiFi and Bluetooth, commonly used spectrum will lead to mutual interferences which renders a certain channel overlappingly defined in each of the systems useless when trying to simultaneously operate both systems.

On the other hand, Mitola developed the vision of cognitive radios in his seminal paper [37]. A cognitive-enabled software radio would act autonomously and user-centric. It would observe its environment with respect to spectral occupancy and other cognitive radios available, and it would have knowledge about its users’ communication needs. By employing learning and reasoning algorithms, the radio would then automatically adapt to these circumstances by switching parameters like frequency band, modulation, protocols, etc. After these adaptations, the cognition cycle starts over and observations continue.

Keeping in mind that there will be a multitude of regulatory constraints, generic spectrally agile cognitive radios need technical aid in order to autonomously follow requirements imposed by regulation. When having knowledge of policies in effect for the present location of the cognitive radio, it can scan appropriate frequency ranges and use them compliantly. [2] and [16] present formal policy description languages that can be evaluated within the cognitive radios so that spectrum can be classified with respect to its usability for the CR. Also side constraints can be coded, e.g. maximum transmission powers. The DARPA XG program also estimates a formal description of locally effective policies from radio protocols as essential for a widely deployable CR network.

However, even describing current regulatory policies in a formal manner does not per se enable a reduction of time scales in which these policies might change. In the case of licenses for cellular access networks, it appears that especially service providers assign a monetary value to transmission licenses. As a consequence, transmission licenses for certain frequencies, regions and periods should be made available as a tradable good [11]. In this way, the regulating bodies would still set the framework, e.g. rules for a certain transmission
standard and maximum transmission power values. But instead of declaring the frequency ranges as spectrum commons, licenses can be sold to service providers. The first step of selling the licenses from the regulator to service providers is common practice. But to adjust allocation to current demand afterwards, further trading of licenses between service providers should be allowed [11] to form a secondary market for the licenses. In this way, owners of licenses can have monetary profit from selling or temporarily renting their licenses to others who may require the transmission resources to satisfy their customers’ communication needs. Nevertheless, such dynamic license distribution still needs adherence of all participants and a central instance must be present in order to control and enforce policies of flexible spectrum management [15].

The proposal of the DIMSUMnet framework [9] introduces a “central radio access network manager” for the negotiation and leasing of spectrum usage rights in a coordinated access band. The usage of these hands is controlled by a regional radio access network manager. From there potential service-providers get chunks of spectrum needed for their services. With the assumption of tradable spectrum licenses, [21] proposes inter-operator auctions for service-providers owning base stations in overlapping service areas. By renting idle channels from each other, these operators can optimize the services to their customers.

In the case that it is not desirable or even impossible to exchange widely deployed but inefficient communications systems, the employment of independent secondary overlay systems enables increasing efficiency in spectrum usage. A first approach to using spectrum holes left idle by existing PU systems was presented in [42]. The central objective of this system is that the PUs do not have to change any protocol or transmission behavior. The SUs should be invisible to the PU system. OFDM with its flexibility in arbitrarily switching on and off subcarriers is an appropriate candidate as a transmission scheme for overlay systems. Hence, Sections IV–VI discuss OFDM-based overlay systems in more detail. Main tasks for any overlay system are sensing and predicting spectrum holes as well as coordinating communication links between SUs. It can be seen that these are also parts of the cognition cycle introduced by Mitola.

The idea of [42] was taken up and is the basis for frameworks providing dynamic secondary spectrum access by overlay systems. The CORVUS framework [8], [10] proposes to use spectrum of different existing PU systems by declaring spectrum holes as virtually unlicensed spectrum. Different possibilities for the coordination between SU cognitive radios are discussed including a dedicated beacon channel in an exclusively allocated frequency range. The OFDM-based cognitive radio (OCRA) network described in [1] proposes a framework for the physical and medium access control layers (PHY, MAC) procedures for an overlay system operating in a heterogenous spectral environment. The IEEE standardization project 802.22 ([30], c.f. Section III-C) is developing an OFDM-based overlay system to work in idle TV frequencies.

B. Competition for Common Spectrum

When talking about overlay systems, the actions of a single secondary system with respect to the resources left idle by the primary system are studied. By moving to situations where multiple systems try to use given resources, game theory turns out to be an appropriate means to analyze the effectiveness of spectrum allocation strategies. Depending on the research fields published, “system” may refer to several access points with their associated users, or it assumes pairs of CRs which seek to establish a wireless link to each other, probably as a part of an ad hoc mesh network. Game theory provides the mathematical framework to analyze such situations for the existence of equilibria and the way in which these equilibria are achieved. Also, it is used to discuss optimality of such equilibria states.

To get a first idea of game theory in the field of communications systems, [35] gives a basic introduction to its application and the taxonomy used. To further distinguish situations, [25] identifies competitive and cooperative systems: In the first case, competing systems do not exchange any information about their needs for spectral resources. They only react indirectly to the actions taken by other systems by means of, e.g. measuring the spectrum occupancy or, in case of a code division multiple access (CDMA)-like access scheme, the noise level. Based on this information, systems decide whether and how to occupy the channel. Decisions are made in a selfish way only looking to the individual advantages of the respective system. For example, [19] discusses systems that try to maximize their own throughput by regularly updating their power allocation to given channels. Each channel can be used by more than one system and hence interference occurs.

But in general, such uncoordinated spectrum games can lead to inefficient allocation: This is the “price of anarchy” [32]. So cooperation between systems is introduced, if there is some form of channel etiquette or protocol to which all participating systems adhere. That protocol in turn is designed to lead to optimal spectrum occupancy. Here, approaches differ mainly in the definition of “optimal,” which can mean to maximize overall throughput or any kind of fairness measure, which are the main optimization goals in spectrum allocation.

In [28], systems interchange information about the negative impact, i.e., signal to interference-and-noise ratio (SINR) loss, of a channel occupation by other systems. Each system then tries to maximize its own throughput and to simultaneously minimize the negative effect upon other systems. In contrast to selfish systems, [43] uses the following assumption inspired by anthropology: The systems are averse to disparities in spectrum usage and seek to enforce a fair partitioning. To achieve a local adaptation of allocation to communication needs of systems, [12] introduces bargaining between neighboring systems, thereby defining a distributed protocol for spectrum access. When considering device-centric spectrum occupancy following as simple rules as possible, [46] compares the efficiency of several of such rules, in presence as well as in absence of possible PU systems. Introducing a central instance and an auction protocol between CDMA systems, in [29] systems are charged for their used transmit power or for the interference induced at a certain measurement point. Also using bargaining and pricing, [32] describes a distributed pricing model that avoids a powerful and omniscient central instance by using belief functions that describe other systems’ private information.
C. Standardization Efforts

The IEEE 802.22 working group [30] is developing an OFDM-based standard for overlay usage of locally unused TV channels. The standard aims at wireless broadband access for the coverage of rural and suburban areas [13]. License-free so-called customer premise equipments (CPE) in a radius of up to about 100 km around a base station (BS) connected to the backhaul network are served. With average distances of about 30 km the system is also referred to as wireless regional area network (WRAN).

Under coordination of the base stations, the CPEs and BSs perform regular measurements in common “quiet periods” to acquire currently idle TV channels. [27] proposes to introduce frequency hopping groups, so that occupancy measurements can be performed in the TV channels currently not used by the SU system. As the characteristics of existing PU systems are known—TV stations and wireless microphones licensed in the US operate in these bands—the complexity of measurements can be reduced and specialized methods using spectral correlations of the possible PU signals are proposed [22]. After the measurement, BSs coordinate the TV channels used for WRAN transmissions. Thereby, not only single channels may be used, but also parts of one channel or concatenated neighboring channels. Interference to wireless microphone PUs is mitigated by switching off specific subcarriers. The protocol is also prepared for coordination of common channel allocation between BSs of different service-providers even by using the air interface for BS to BS communication. Of course, well-known technologies like automatic power control and adaptive modulation are incorporated.

The IEEE Standards Coordinating Committee 41, Dynamic Spectrum Access Networks, (SCC 41) has amongst other duties the responsibility to vote on approval of proposed IEEE standards and to develop proposed standards in the area of dynamic spectrum access networks [31]. It was originally established by the IEEE Communications Society (ComSoc) and the Electromagnetic Compatibility (EMC) Society as the IEEE 1900 Standards Committee on Next Generation Radio and Spectrum Management in the first quarter of 2005, and was reorganized in the first half of 2007, thereby changing its name.

As far as world-wide regulation is concerned the World Radio Communication Conference recently held in Geneva (WRC-2007) resolved to invite the International Telecommunications Union (ITU) “to study whether there is a need for regulatory measures related to application of technologies of cognitive radio systems; to study whether there is a need for regulatory measures related to the application of software defined radio, [and] further resolves that WRC-2011 consider the results of these studies and take appropriate actions.”

IV. OFDM Overlay Systems

In this section we present the OFDM spectrum pooling [42] system which we will use as a basis for the considerations in the remainder of this contribution. In a spectrum pooling system spectrum owners bundle their resources left idle by the PUs and make them accessible to SUs who form the overlay system. The SU system uses OFDM [38], [23], [17] as transmission technology and it is assumed that the following two rules are accepted [34]:

- The PUs should not be disturbed by the SUs.
- The PUs’ equipment (infrastructure and terminals) remains unchanged when the SU system is introduced.

The second rule enforces that all signal processing that has to be done in order to avoid interferences from the SU system to the PU system has to be implemented in the SU system’s devices.

For the sake of simplicity we consider the PU system to use time/frequency division multiple access (TDMA/FDMA) as access mode (i.e. similar to the global system for mobile communications, GSM). Fig. 2 shows the allocation of a PU system in the time/frequency plane and the resulting spectrum holes. The PU system has a total bandwidth $B$ that is divided into subbands with equal bandwidth $b$, e.g. representing different channels. The occurring spectrum holes are the spectral resources available for the SU system. For exploiting them, the SU system must dynamically adapt its configuration, restricting transmission to the subbands not used by the PU system. For this reason, OFDM is particularly suitable for the SU system, because it is easy to switch on and off groups of subcarriers in this transmission technology and therefore providing a basis for dynamic and efficient use of the secondary spectrum resources.

For the coexistence of a SU system and a PU system in the same frequency band, reduction of mutual interference plays a vital role. In this context, PHY and MAC layers deal with different aspects of interference reduction. The PHY layer is responsible for interference mitigation in frequency direction, while the focus of the MAC layer lies on avoidance of interference in time direction. Thereby, detection is a central component that has impact on the interference reduction in both layers. Due to its fundamental role, detection will be discussed in detail in Sections V and VI.

A SU system has to periodically perform allocation measurements that must be initiated by the MAC layer. When taking the time/frequency plane in Fig. 2, looking at a specific subchannel of the PU system and performing a cut parallel to the time axis, the subfigure describing the periodical measurements is derived. The time reserved for an allocation measurement is called detection period and the interval between two detection periods is referred to as update interval. The update interval is bounded by the maximum amount of time for which a PU system can tolerate interference when allocating new parts of the spectrum. From the perspective of the SU system a long update interval reduces the time spent for detection (and thus more time is available for the actual transmissions), but also increases the probability for collisions with the PU system (there leading to more destroyed data blocks) [3]. Note, that during the detection periods all SUs must be silent so that only signals from the PU system are detected.

The focus of the physical layer lies on out-of-band radiation, i.e. interference between active subchannels of the PU system and adjacent subcarriers of the overlay system. For this case, we keep the time constant and perform a cut parallel to the frequency axis of the time/frequency plane in Fig. 2. The resulting subfigure explains how instantaneous spectrum holes in the PU system can be filled with OFDM subcarriers of the SU system that has the same total bandwidth $B$ as the PU system. We assume that one PU subband is resolved by a set of $a$ (in Fig. 2: $a = 4$) SU OFDM subcarriers, so that $b = a \cdot \Delta f$ where $\Delta f$ is the subcarrier spacing. Below the subcarriers it
is indicated which of them see an idle PU subband (“0”) and which of them see a busy subband (“1”). Read from left to right, these subband indicators form the allocation vector.

Especially subcarriers of the SU system that are close to active channels of the PU system contribute to out-of-band radiation impacting the PU system. The SU system’s physical layer must suspend the resulting interference to a considerable degree to not degrade the performance of the PU system. One approach is to simply deactivate the adjacent subcarriers [42], thus inserting a dynamic guard band. Another promising possibility is the employment of cancellation carriers [6], [7]. Here, the resulting sidelobes of an OFDM symbol are calculated and the modulation of some defined subcarriers is adapted to minimize the out-of-band radiation.

V. DISTRIBUTED DETECTION IN OFDM OVERLAY SYSTEMS

When inserting an overlay system, reliable detection of PU signals by the SU system is of paramount importance. Only if this reliability is satisfactory, the PUs will tolerate the SU system sharing their spectrum. Therefore, the SUs’ probability of detection $P_D$ for PU signals should be high (e.g. 99.9%). On the other hand, the SU’s false alarm probability $P_F$ should be small because the SU systems’ efficiency decreases with increasing the false alarm probability. As we know, $P_D$ and $P_F$ cannot be independently optimized [41]. We can only specify either $P_D$ or $P_F$ while optimizing the other probability. According to its outstanding importance, we are going to specify $P_D$ and then to minimize $P_F$.

In general, the detection of a primary system will get more reliable, as a priori knowledge and signal processing power of the detector increase. When knowing exactly the modulation and frequency parameters of the PU, the SU can directly scan for active transmission by demodulating them. For a detector not having this knowledge, there are many proposals to scan for signal features, most commonly cyclostationary properties of a modulated signal. But the major drawback of these methods is the tremendous calculation effort which is infeasible to perform in a low-cost and low-power device under real-time constraints. Therefore, we use the energy detector model described in detail in the following Section, as it can make use of the existing FFT modules on the SU’s OFDM hardware and as it offers low calculation complexity.

A. The Detector Model

Fig. 3 shows the block diagram of a SU’s detector for PU signals. $s(k)$ is the complex baseband signal originating from a PU as it is received by the SU. The sum of $s(k)$ and the disturbing white noise component $n(k)$ forms the signal

$$ r(k) = s(k) + n(k) \quad (1) $$

that is presented to the SU’s detector. A fast Fourier transform (FFT) maps $r(k)$ into the frequency domain. The output samples of the FFT representing a single PU system subband are stored in a memory. In time direction the FFT operates until $M$ complex samples are collected. The content of the memory is then represented by the $2M$-component vector

$$ z = (x_1, x_2, \ldots, x_M; y_1, y_2, \ldots, y_M)^T \quad (2) $$

where $x_m$ and $y_m$ denote the real and imaginary parts of $z_m = x_m + jy_m$, the $m^{th}$ component of $z$.

To derive the probability density function (pdf) of $z$, we have to make some assumptions. First, we suppose that there is no line of sight (NLOS) between the transmitting PU and the detecting SU. Moreover, many multipath signals superimpose at the SU’s antenna. Consequently, we may apply the central limit theorem and conclude that $s(k)$ may be interpreted as a zero-mean complex Gaussian random variable.

We further assume that the higher layers of the SU’s protocol stack initiate silent periods for the SU’s access point and all mobile stations to ensure that the only spectral energy that may be detected on the wireless medium stems from a PU and that the detection process is not interfered by SUs. The silent periods are periodically repeated and may be announced by the broadcast of the SU’s access point (just like the beacon signal in IEEE 802.11).

According to (1) it turns out that also $r(k)$ is a complex zero-mean Gaussian random variable. Since the FFT is a linear operation, also the samples in the detector’s memory (c.f. Fig. 3) are complex zero-mean Gaussian random variables represented by

$$ z_m = x_m + jy_m = R(m) + jI(m); m = 1, 2, \ldots, M. \quad (3) $$

In (3) $\{S(m) = S_x(m) + jS_y(m)\}_{m=1}^M$ and $\{N(m) = N_x(m) + jN_y(m)\}_{m=1}^M$ are the FFTs of $\{s(k)\}_{k=1}^M$ and
\{n(k)\}_{k=1}^M$. We denote the mean powers of $S(m)$ and $N(m)$ by $2\sigma_S^2$ and $2\sigma_N^2$, respectively. Since all random variables $S(m)$ and $N(m)$; $m = 1, 2, \ldots, M$; are independent, the conditional pdf of $z$ under the condition that a PU signal is present, is given by

$$f_{z|\text{PU}}(z|\text{PU}) = \left(2\pi\sigma_N^2\right)^{-M/2} \exp\left\{-\frac{1}{2} z^T \left(\mathbf{C}_{SS} + \sigma_N^2 \mathbf{I}\right)^{-1} z\right\}. \quad (4)$$

$\mathbf{C}_{SS}$ is the covariance matrix of $\mathbf{S} = (S(1), S(2), \ldots, S(m))^T$ and $\mathbf{I}$ is the identity matrix. Both matrices are of dimension $2M \times 2M$.

If no PU signal is present (indicated by $\overline{\text{PU}}$) in the subband the conditional pdf simplifies to

$$f_{z|\overline{\text{PU}}}(z|\overline{\text{PU}}) = \left(2\pi\sigma_N^2\right)^{-M} \exp\left\{-\frac{1}{2} \frac{z^T z}{\sigma_N^2}\right\}. \quad (5)$$

For the further discussion we put on record that both pdfs (4), (5) are unimodal, i.e. they both possess a unique local maximum.

Because we are interested in the receiver operating characteristics of the SU’s detector, we follow the Neyman-Pearson strategy [41] starting with the definition of the false alarm probability $P_F$ and the detection probability $P_D$:

$$P_F = \int_{\mathbf{A}_{\text{PU}}} f_{z|\text{PU}}(z|\text{PU}) dz,$$

$$P_D = \int_{\mathbf{A}_{\overline{\text{PU}}}} f_{z|\overline{\text{PU}}}(z|\overline{\text{PU}}) dz. \quad (6)$$

In (6) $\mathbf{A}_{\text{PU}} \subset \mathbb{R}^{2M}$ is the set of vectors $z$ for which a decision in favor of the presence of a PU signal is made.

To find a practical solution for the decision problem, we define the likelihood ratio (LR)

$$\hat{\Lambda}(z) := \frac{f_{z|\text{PU}}(z|\text{PU})}{f_{z|\overline{\text{PU}}}(z|\overline{\text{PU}})} \frac{\mathbf{A}_{\text{PU}}}{\mathbf{A}_{\overline{\text{PU}}}} \hat{\lambda}_0. \quad (7)$$

$\hat{\Lambda}(z)$ may be interpreted as a scalar random variable with pdf $f_{\hat{\Lambda}}(\hat{\lambda})$ and corresponding conditional pdfs $f_{\hat{\Lambda}|\text{PU}}(\hat{\lambda}|\text{PU})$ and $f_{\hat{\Lambda}|\overline{\text{PU}}}(\hat{\lambda}|\overline{\text{PU}})$. If $\hat{\Lambda}(z)$ is greater than a threshold $\hat{\lambda}_0$, a PU is present. If $\hat{\Lambda}(z)$ is smaller than $\hat{\lambda}_0$, no PU is present. With the notations just introduced, (6) translates to

$$P_F = \int_{\hat{\lambda}_0}^{\infty} f_{\hat{\Lambda}|\text{PU}}(\hat{\lambda}|\text{PU}) d\hat{\lambda},$$

$$P_D = \int_{\hat{\lambda}_0}^{\infty} f_{\hat{\Lambda}|\overline{\text{PU}}}(\hat{\lambda}|\overline{\text{PU}}) d\hat{\lambda}. \quad (8)$$

As we already mentioned, all random variables $S(m)$ and $N(m); m = 1, 2, \ldots, M$; are independent. I.e. also all the $2M$ components of $z$ in (2) are independent. Therefore, the covariance matrix $\mathbf{C}_{SS}$ in (4) turns out to be

$$\mathbf{C}_{SS} = \sigma_S^2 \mathbf{I}. \quad (9)$$

Now we keep (9) in mind, insert (4) and (5) into (7), take the natural logarithm, ignore constants and end up with the LR test:

$$\Lambda(z) = z^T z = \sum_{m=1}^{M} \sigma_m^2 + \sum_{m=1}^{M} y_m^2 \frac{0}{\lambda} \Lambda_0 \quad (10)$$

The problem is now to find the conditional pdfs $f_{\Lambda|\text{PU}}(\lambda|\text{PU})$ and $f_{\Lambda|\overline{\text{PU}}}(\lambda|\overline{\text{PU}})$ and to determine the optimal threshold $\lambda_0$ by either fixing $P_F$ or $P_D$. As already mentioned above, the task of inserting a SU system into a PU system, it seems to be reasonable to specify $P_D$ in order to guarantee a (high) detection probability for PU signals by the SU system.

From (3) we conclude that all the (real) zero-mean Gaussian random variables $x_m$ and $y_m; m = 1, 2, \ldots, M$; have the variance $\sigma_S^2 + \sigma_N^2$. I.e. $\Lambda(z)$ as a sum of $2M$ squared identically distributed Gaussian random variables (10) is centrally $\chi^2$-distributed [41] with $2M$ degrees of freedom and

$$f_{\lambda|\text{PU}}(\lambda|\text{PU}) = \frac{2^M (\sigma_S^2 + \sigma_N^2)^M (M-1)!}{\lambda^{M-1}} \lambda^{\lambda - 2(\sigma_S^2 + \sigma_N^2)} - \exp\left\{-\frac{\lambda}{2(\sigma_S^2 + \sigma_N^2)}\right\} \quad (11)$$

To determine $P_D$ we use the integral table [20] and get

$$P_D = \int_{\lambda_0}^{\infty} f_{\lambda|\overline{\text{PU}}}(\lambda|\overline{\text{PU}}) d\lambda \exp\left\{-\frac{\lambda_0}{2(\sigma_S^2 + \sigma_N^2)}\right\} \lambda - 2(\sigma_S^2 + \sigma_N^2) + \sum_{m=1}^{M-1} 2(\sigma_S^2 + \sigma_N^2)\right\} \quad (12)$$

The calculation of $P_F$ is carried out in the same way and leads to $\sigma_S^2 = 0$ in (12):

$$P_F = \frac{1}{(2\alpha_N^2)^M} \left(\frac{M-1}{\lambda_0} \right) \lambda^{M-1} \exp\left\{-\frac{\lambda_0}{2(\sigma_N^2)}\right\} \lambda - 2(\sigma_S^2 + \sigma_N^2) + \sum_{m=1}^{M-1} 2(\sigma_S^2 + \sigma_N^2)\right\} \quad (13)$$

The receiver operating characteristics (ROC) for different signal-to-noise ratios (SNR) $\sigma_S^2/\sigma_N^2$ and $M = 8$ are depicted in Fig. 4. From these curves we conclude that the detection strategy discussed so far is insufficient for the reliable detection of PUs by SUs. This is especially true when we take into account that the calculations leading to Fig. 4 represent something like a best-case scenario in which it is assumed that the PU detector in the SU receivers is matched to the actual PU signal statistic.
B. Distributed Detection

In order to make a SU overlay system attractive to PUs, the detection probability for PU signals by the SU system should be greater than 99.9%. As already mentioned this is not realizable with the detector presented in Section V-A. Adjusting $P_D$ to 99.9% or more would lead to a false alarm probability that is much too high and that cannot be tolerated because with this the capacity of the SU system would decrease severely. Also an extension of the detection length $M$ will not eliminate this drawback. Large numbers of FFT cycles cause a large temporal overhead and this degrades the throughput of the SU system. This disadvantage could be mitigated by using a finer FFT resolution with respect to the PU system’s subbands. On the other hand there is a much more powerful technique to decrease $P_F$ by several orders of magnitude while still maintaining the desired $P_D$ of 99.9% or more [26].

The detection performance is significantly improved by a diversity approach. I.e. not a single SU performs spectral detections but all SUs associated to a given access point as shown in Fig. 5a; SUs that do not detect the PU, for example because of fading or shadowing effects, are connected with the PU by dashed lines. In this case, it is sufficient that only one SU actually detects the appearance of a PU. This leads to a high detection probability at SU system level. With this approach the false alarm probability may be reduced. Moreover, this method is beneficial in overcoming the fading effects of the mobile radio channel.

At the end of every SU system’s detection period the mobile terminals transmit their results to the access point (c.f. Fig. 5b). Here, these Boolean variables are combined by a simple OR operation. Afterwards, the access point transmits the combined results back to the mobile terminals as indicated in Fig. 5c.

For the analysis of this distributed detection approach, the SU system’s detection probability $P_{DS}(L)$ is introduced

\[ P_{DS}(L) = 1 - (1 - P_F)^L \]  

where $L$ indicates the number of SUs. $P_{DS}(L)$ represents the probability that the appearance of a PU signal is correctly detected at system level. Similarly, the SU system’s false alarm probability $P_{DF}(L)$ is defined. It describes the probability that the appearance of a PU signal is erroneously detected by the SU system.

We assume that the detection results of all SU terminals are independent and that all SU terminals have approximately the same distance from the PU. Hence the SU system detection probability $P_{DS}(L)$ and false alarm probability $P_{DF}(L)$ can be expressed as

\[ P_{DS}(L) = 1 - (1 - P_F)^L \]  

\[ P_{DF}(L) = 1 - (1 - P_D)^L \]  

(14)

where $P_F$ and $P_D$ are the false alarm and detection probabilities of a single SU.

In order to meet the PUs’ requirements $P_{DS}(L)$ should be larger than 99.9%. Therefore, in our following considerations, we take $P_{DS}$ as fixed and independent of $L$.

Of course, the SU system false alarm probability $P_{DF}(L)$ will increase when the distributed detection approach is realized. But the gain in the lower $P_{DS}(L)$ value overcompensates this drawback because of the ROC’s convexity: If we solve the second equation in (14) for $P_D$ we find

\[ P_D = 1 - \sqrt[2L]{1 - P_{DS}(L)}. \]  

(15)

$P_D$ and $P_F$ are related over the ROC. From Fig. 4 we infer that $P_F$ as function of $P_D$ is strictly monotonically increasing.

Inserting (15) into the first equation in (14) yields

\[ P_{DS}(L) = 1 - \left[ 1 - P_F \left( 1 - \sqrt[2L]{1 - P_{DS}(L)} \right) \right]^L. \]  

(16)

The parentheses on the right hand side of (16) indicate the dependence of $P_F$ from $P_D$ (15). It can be shown by differentiation of (16) that $P_{DS}(L)$ is a strictly monotonically decreasing function of $L$ for a given $P_{DS}$. It follows that the more stations participate in the distributed detection the larger the diversity gain will be. As an example we look onto Table Ia for which $P_D$, $P_F$ and $P_{DS}(L)$ were calculated for $M = 8$, $SNR = 2 dB$, $P_{DF} = 0.999$. We see that distributed detection leads to a substantial improvement of the SU system’s detection performance. For $L = 4$ or more participating stations the SU system’s false alarm probability $P_{DF}(L)$ is below 0.13. If 10 stations are involved $P_{DF}(L)$ is as low as 0.009.

<table>
<thead>
<tr>
<th>L</th>
<th>$P_D$</th>
<th>$P_F$</th>
<th>$P_{DS}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.999</td>
<td>0.853</td>
<td>0.853</td>
</tr>
<tr>
<td>2</td>
<td>0.968</td>
<td>0.280</td>
<td>0.488</td>
</tr>
<tr>
<td>3</td>
<td>0.900</td>
<td>0.088</td>
<td>0.248</td>
</tr>
<tr>
<td>4</td>
<td>0.822</td>
<td>0.033</td>
<td>0.124</td>
</tr>
<tr>
<td>10</td>
<td>0.499</td>
<td>0.001</td>
<td>0.009</td>
</tr>
<tr>
<td>20</td>
<td>0.292</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>
VI. DISTRIBUTED DETECTION IN AD HOC OVERLAY SYSTEMS

Now a scenario is considered where the SU system operates as a mesh network in ad hoc mode. This results in additional challenges, since there is no access point that can provide central coordination and processing capabilities. As we pointed out in the previous section, distributed detection increases the SU system’s performance substantially. Based on these results, we are now interested in the performance of distributed detection applied in ad hoc networks. Therefore, we first give a short overview on some basics of graph theory, define the network detection probability, and analytically derive a theoretical bound for the necessary node density to achieve a predefined network detection probability. Then we propose a basic structure for a MAC protocol allowing for a coordination of the SU system’s detection period. Based on this basic structure we finally discuss an approach for distributed detection and signaling in SU ad hoc systems.

A. Network Detection Probability

The properties of ad hoc networks can be described with the help of geometric random graph theory [39], where a node is equivalent to a SU. The considered system consists of SUs placed in two-dimensional space \( \mathbb{R}^2 \) and is represented by an undirected graph \( G = G(X; r) \) with the node set \( X \subset \mathbb{R}^2 \) and undirected edges connecting all pairs of SUs that have a distance smaller than \( r \). This implies that all SUs in \( G \) have the same transmission range \( r \). Two SUs are neighbors if they are connected by an edge. The degree \( d(u) \) of a node \( u \in G \) is defined as the node’s number of neighbors.

\[ N = |X| \]

\[ A \]

\[ A \]

\[ A \]

Note, that (20) also gives the average detection probability of a single SU. It is equal to the network’s detection probability since we disregard border effects.

In a similar way, the SU system’s network false alarm probability \( P_F^N(\nu) \) can be expressed when using the first equation of (14):

\[ P_F^N(\nu) = \sum_{D=0}^{N-1} \frac{\nu^D}{D!} e^{-\nu} \cdot \left( 1 - (1 - P_D) D + 1 \right) \]  \( (21) \)

For an ad hoc overlay system Table Ib displays the resulting network false alarm probabilities for several combinations of \( \nu \) and \( P_D \), assuming the best case scenario.

VII. CONCLUSIONS

Dynamic spectrum access is an essential approach for increasing efficiency in spectrum use, and thus to counteract the observed spectrum scarcity. In this contribution we discussed current approaches to dynamic spectrum sharing first, including standardization efforts. One concept to utilize occurring spectrum holes in the time/frequency plane are overlay systems that are deployed in the same frequency band as a licensed system. To enable a successful coexistence of these two systems and to avoid collisions as well as mutual interference, the overlay system has to periodically perform measurements to detect the allocation of the licensed system and dynamically adapt its system parameters. OFDM is a flexible and suitable transmission technology for overlay systems, since the integrated FFT can also be used for performing PU signal detections and groups of subcarriers can be easily switched off to adapt to the current allocation. Due to the importance of detection for an overlay system, we investigated a detector model and analyzed the detection and false alarm probabilities. To achieve better receiver operating characteristics, we presented a distributed detection approach, allowing for less requirements regarding the precision of a single detection. Furthermore, these results were then extended to ad hoc networks and we developed an approach for distributed detection and signaling, where no central access point is required and all necessary actions are performed in a decentralized manner. The resulting detection and false alarm probabilities were also investigated for this ad hoc overlay scenario.
ACKNOWLEDGEMENT

The authors gratefully acknowledge that the work presented in this paper was partially funded by the Deutsche Forschungsgemeinschaft within the focus project TakeOFDM.

REFERENCES


