Cognitive Radio Systems

Cognitive Radio Systems

Edited by
Wei Wang

In-Tech
intechweb.org

Preface

Cognitive radio is a hot research area for future wireless communications in the recent years. In order to increase the spectrum utilization, cognitive radio makes it possible for unlicensed users to access the spectrum unoccupied by licensed users. Cognitive radio let the equipments more intelligent to communicate with each other in a spectrum-aware manner and provide a new approach for the co-existence of multiple wireless systems.

The goal of this book is to provide highlights of the current research topics in the field of cognitive radio systems. The book consists of 17 chapters, addressing various problems in cognitive radio systems.

Chapter 1 ~ Chapter 3 introduce the concepts and the state-of-the-art strategies for cognitive radio systems.

Chapter 4 ~ Chapter 6 focus on the spectrum sensing techniques for cognitive radio systems, including cooperative spectrum sensing, cyclostationary-based signal detection and sensing-based traffic pattern prediction.

Chapter 7 ~ Chapter 10 provide some recent research achievements on the transmission technologies. Two of them describe the interference related transmission, while the other two chapters study the multi-antenna transmission for cognitive radio systems.

Chapter 11 ~ Chapter 15 present the efficient solutions for various resource allocation problems, including MAC protocol, subband selection, power allocation, resource management and the transmission for multimedia services.

Chapter 16 ~ Chapter 17 discuss some practical issues in cognitive radio systems. The platform for inter-radio system switching and the power amplification are investigated in the two chapters respectively.

The editor appreciates the authors of the chapters for their contribution. We hope that this book will provide the readers an efficient way for understanding both the basic ideas of cognitive radio and the cutting-edge research results.

Wei Wang
Institute of Information and Communication Engineering
Department of Information Science and Electronic Engineering
Zhejiang University
P.R. China

Contents

A Brief Survey on Cognitive Radio

Wei Wang
Institute of Information and Communication Engineering
Zhejiang University
P.R. China

1. Introduction

With the increasing demand of wireless application, the insufficiency of spectrum is more and more serious; on the contrary, the utilization of some licensed spectrum is always low [FCC, 2003]. In order to increase the spectrum utilization, cognitive radio makes it possible for unlicensed users to access the spectrum unoccupied by licensed users.

The concept of cognitive radio is proposed first by Mitola (Mitola & Maguire, 1999), and the language for cognitive function is investigated in (Mitola, 2000). In (Haykin, 2005), the detailed expositions of signal processing and adaptive procedures are presented. In (Akyildiz et al, 2006), the major characteristics of cognitive radio networks are presented from physics layer to transport layer, as well as cross-layer design.

The spectrum agility of cognitive radio brings new challenges. The chapters in the rest of the book illustrate the wide variety of new problems for cognitive radio. The state-of-the-art strategies are presented in this chapter. To understand the general idea of cognitive radio better, the reader is encouraged to complete the brief survey before studying the chapters on specific techniques.

2. Cognitive Radio Models

Cognitive radio is a hot research topic in recent years. The wireless communication systems with cognitive radio are modelled as different models. Until now, there have been many research works on cognitive radio. Most of the works can be concluded as one of the following four kinds of cognitive radio models.

2.1 Initial Cognitive Cycle
When cognitive radio is proposed, an intelligent communication technology is expected, including observe, orient, plan, learn, decide and act (Mitola & Maguire, 1999; Haykin, 2005). The basic idea of the initial cognitive cycle is concluded as Fig. 1. The receivers obtain the channel quality information and the interference information from the surrounding radio environment by observing. After the transmitters receive the necessary feedback

information from their corresponding receivers, they determine the strategies, which react to the radio environment. For more intelligent function, machine learning is adopted for estimating the utilities of possible strategies to improve system performance.
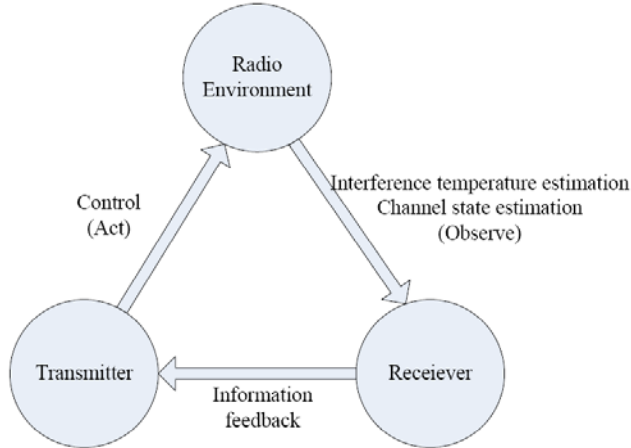
Fig. 1. Basic cognitive cycle

2.2 Dynamic spectrum model

Based on the initial cognitive cycle model, cognitive radio is studied to be utilized further for spectrum sharing between licensed/primary users and unlicensed/secondary users in licensed spectrum. In that case, the secondary users are not allowed to cause too large interference that may interrupt the communication or decrease the service quality of primary users.

In the dynamic spectrum model (Peng et al, 2006), it is assumed that the primary users may not always use the spectrum. Hence, the secondary users can opportunistically utilize the spectrum when it is not being occupied by the primary users, as shown in Fig. 2. According to the primary users' spectrum usage pattern, based on the experimental results in (Motamedi and Bahai, 2007) and (Geirhofer et al, 2007), the spectrum usage can be modelled as an ON-OFF process: ON (OFF) state represents when the spectrum is occupied (unoccupied) by primary users. The spectrum dynamics can be modelled as a semi-Markov process as in (Kim and Shin, 2008).

In this model, with perfect spectrum sensing, which means that the secondary users detect the spectrum status error-freely and justify the status in time if some primary user comes back, the secondary users and primary users do not interfere with each other. The research challenge focuses on how to discover and utilize the spectrum opportunities more efficiently. Considering the error of spectrum sensing, the possible interrupt to primary users should be investigated. The schemes need to achieve a balance of the tradeoff between the utility of secondary users and the influence to primary users.
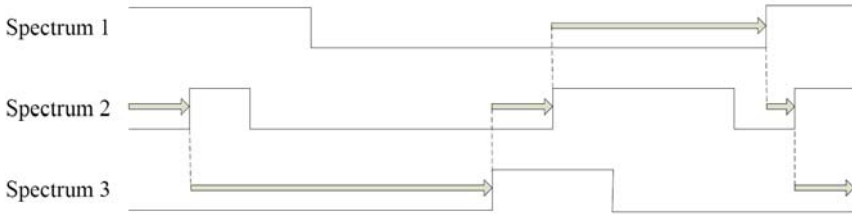
Fig. 2. Dynamic spectrum model

2.3 Interference temperature model

In the interference temperature model (Xing, 2007a), both primary and secondary users can co-exist on the same spectrum. The secondary users' interference to the primary receivers should not exceed a threshold. Interference temperature is introduced into cognitive radio by Federal Communications Commission (FCC) as a metric for the measurement of interference in a radio environment. In order to prevent the negative impact to the primary users, the interference temperature limit is used to indicate the allowed worst RF environment. In order to protect the primary users' communications, the interference caused by secondary users must be kept below the interference temperature limit at the primary receivers. That is, the primary users' Quality-of-Service (QoS) is considered acceptable if the secondary users' interference is kept below a given interference temperature limit. The maximum interference tolerance can be calculate as

$$Q_{\max} = \xi T_{\max} \tag{1}$$

where ξ is Boltzmann's constant, $T_{\max}$ is the interference temperature limit.
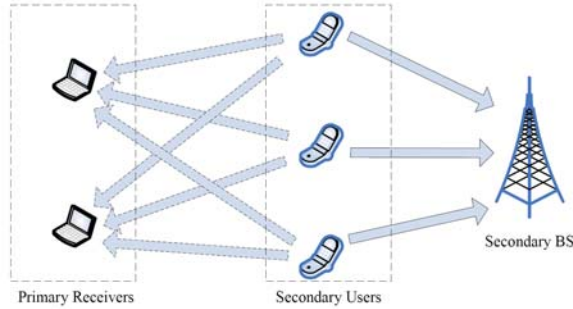
Fig. 3. Interference temperature model

(Ghasemi & Sousa, 2007) analyzes the capacity of cognitive user with the assumption that the cognitive user estimates the statistic results of its interference to the primary user through various fading channels. The average and peak interference constraints are considered respectively in (Ghasemi & Sousa, 2007). With this model, an extra interference temperature constraint is added into the problems compared with conventional wireless communication systems, as shown in Fig. 3.

2.4 Cognitive Cooperation

In (Devroye et al, 2005) and (Devroye et al, 2006), it is assumed that the cognitive user can obtain and transmit the messages that the primary user will send. The capacities of both primary users and secondary users are obtained. Based on these, (Jovicic & Viswanath, 2006) analyzes the capacity of a cognitive user who transmits simultaneously with a primary user, in the condition that the primary user can achieve the data rate just as it would in the absence of the cognitive radio user. (Cheng et al, 2007) extends the results of (Jovicic & Viswanath, 2006) to multiple access channels (MAC) and gives a heuristic scheme to achieve the maximum sum-rate.
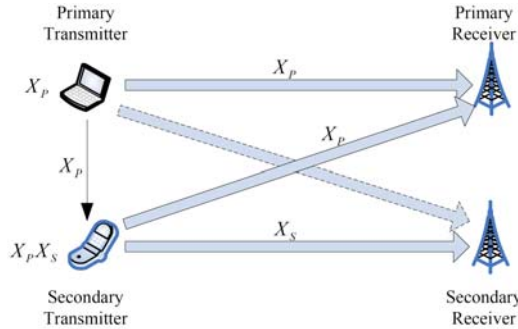
Fig. 4. Cognitive cooperation model

In this model, there exists a tradeoff that the secondary transmitter sends primary data or secondary data, as shown in Fig. 4. Transmitting primary data can increase the primary throughput and improve the capability of interference tolerance of primary users. On the other hand, transmitting secondary data can increase the secondary throughput and decrease the interference to primary users.

3. Research on Cognitive Radio Systems

3.1 PHY-layer Spectrum Sensing

Spectrum sensing is a neccesary technology of cognitive radio. With efficient spectrum sensing, the spectrum opportunities could be discovered. From PHY-layer view, the spectrum sensing can be divided into three categories, non-coherent detection, coherent detection and feature detection (Sahai et al, 2004).

The most usual non-coherent detection method is energy detection. The advantages of energy detection are short sensing time and low complexity. In addtion, it does not need any aprior information. However, because of the uncertainty of noise, there exists a SNR wall. The signal can not be detected if its SNR is lower than the SNR wall. As the signal is detected according to the signal strength, it can not distinguish different kinds of signal.

When the signal has the corresponding pilot, the coherent detection can be adopted. The matched filter is one of the coherent detection methods, but the performance is affected by

high complexity, unstable time clock and the length of pilot. Because of these factors, the implement is limited in practice.

Feature detection utilizes the properties of signal to detect whether there is any primary user nearby. As the signal has periodic features because of frame structures but the noise does not have any period, cyclostationary detection can be used to distinguish the signal and the noise. Using pattern recognition, different kinds of signal can be distinguished by comparing the cyclostationary properties of the detected signal with aprior known signal properties. Although the performance is better than energy detection, the SNR wall still exists. If the signal strength is not too low, the signal can be recognized from the unstable noise.

3.2 MAC-layer Spectrum Sensing

On spectrum sensing, there exists a tradeoff between sensing time and sensing veracity. The sensing methods which have high veracity always need long sensing time. A two-level spectrum structure is proposed in (IEEE 802.22, 2006) to balance the tradeoff between these two aspects. Energy detection is adopted to discover primary users cursorily. Then, if it is possible that there exists any primary user, more elaborate spectrum sensing is deployed.

Because of the fading effect, the spectrum sensing results of one user is not always accurate. Therefore, the cooperation between secondary users is necessary (Mishra et al, 2006). There are two kinds of cooperation, centralized cooperative spectrum sensing and distributed cooperative spectrum sensing.

In centralized cooperative spectrum sensing, a centralized controller collects the sensing results from different users, and fuses the collected data altogether to obtain a table for available spectra. The results obtained by centralized cooperative spectrum sensing are accurate relatively, but it needs long sensing time, large computational capability and heave overhead.

Distributed cooperative spectrum sensing lets each user detect the signal and obtain the table of available spectrum respectively. By communicating with the neighbour users, the chosen spectrum is determined. How to sense the spectrum accurately by exchanging limited information is still an open problem.

3.3 Radio Resource Allocation

Dynamic spectrum management is an efficient method to avoid the interference between primary users and secondary users. When some spectrum is idle, the cognitive radio systems choose the spectra which have low interference. If the primary users come back to use the spectrum occupied by secondary users, the cognitive radio systems should obtain the information in time. Based on the information, the secondary users choose another spectrum from the candidate spectrum set, or decrease the transmit power to avoid too large interference to primary users if there is no other candidate spectra.

In cognitive radio networks, the power control schemes need to consider not only their own utilities, but also the influence to primary users. Game theory is an efficient method for

distributed power control (Zhu & Liu, 2007; Wang et al, 2007a). Spectrum allocation and power control affect each other, so joint spectrum allocation and power control are investigated (Wang et al, 2009a). For multi-hop networks, routing is also an important issue. The performance of cognitive radio networks can be optimized by designing appropriate routing, spectrum allocation and power control schemes.

3.4 Spectrum Marketing

On spectrum pricing, in (Buddhikot et al, 2005), a framework for coordinating dynamic spectrum is proposed. In (Xing et al, 2007b), the dynamic pricing strategy is proposed for competitive agile spectrum access markets. Sharply value in cooperative game is used to evaluate the contribution of each system in spectrum marketing (Wang et al, 2007b). The investigation on spectrum pricing is also introduced into IEEE 802.22 standardization (IEEE 802.22, 2007).

On the contrary of spectrum pricing, spectrum auction (Gandi et al, 2007; Zhou et al, 2008) is also a practical way for spectrum marketing. Each system announces a price to other systems according to the utilities and costs if it can win the auction and get the spectrum. Based the economic theory, the systems can approach the optimal performance by maximizing their own profits.

3.5 Application and Standardization

Cognitive radio is used widely in several areas of wireless communication. In (Wang et al, 2007c), the application of cognitive radio in wireless emergence networks is investigated combined with relaying to enhance the coverage performance in the disasters. In (Hinman, 2006), cognitive radio is employed for military application.

IEEE 802.22 is the first wireless standard applying cognitive radio. The secondary users use TV spectrum to improve the spectrum utilization, when it is unoccupied by nearby TV transmitters. Besides IEEE 802.22, other wireless standardization, such as IEEE 802.11n and IEEE 802.16h, also adopt cognitive radio for interference coordination among users in the same system, rather than between two systems. Many researchers are trying to use the idea of cognitive radio in LTE networks.

4. A Perspective of Future Research on Cognitive Radio

Cognitive radio is one of the research frontiers in wireless communication field. Both academic and industry researchers have large interest to cognitive radio and gained many achievements. However, there are still some research challenges as follows (Wang, 2009b).

1) Cooperative Sensing: Distributed cooperative spectrum sensing needs further research to balance the tradeoff between accurateness and overhead better.

2) Cognitive Relaying: Using additional user to relay the data can increase the throughput for either primary links or secondary links. In addition, relaying for primary links can increase the data transmission for more spectrum opportunities, and relaying for secondary links can decrease the interference to primary users.

3) Cognitive MIMO: MIMO can decrease the interference by adjusting the signal orthogonal to the interference channel to primary users (Zhang & Liang, 2008). Therefore, using multiple antennas is helpful in cognitive radio networks to increase the throughput of secondary users and decrease the interference to primary users.

4) Femtocell: As the characteristics of femtocells, the interference decreases a lot because the signal usually penetrates walls, which is very favourable for cognitive radio to avoid interference.

5) Robust Cognitive Radio: In most of the exist research works, the radio resource allocation is investigated based on perfect spectrum sensing results. Considering the error of spectrum sensing, the resource allocation schemes should restrict the outage probability that secondary users interrupt the communication of primary users.

5. References

Akyildiz, I.F.; Lee, W.Y.; Vuran, M.C. & Mohanty S. (2006). NeXt generation/dynamic spectrum access/cognitive radio wireless networks: A survey. *Computer Networks (Elsevier)*, Vol. 50, No. 13, pp. 2127 – 2159

Buddhikot, M.M.; Kolodzy, P.; Miller, S.; Ryan, K. & Evans, J. (2005). DIMSUMnet: new directions in wireless networking using coordinated dynamic spectrum. *Proc. of IEEE WoWMoM 2005*, pp. 78 - 85

Cheng, P.; Yu, G.; Zhang, Z.; Chen, H. & Qiu, P. (2007). On the achievable rate region of gaussian cognitive multiple access channel. *IEEE Commun. Letters*, Vol. 11, No. 5, pp. 384 - 386

Devroye, N.; Mitran, P. & Tarokh, V. (2005). Cognitive multiple access networks, *Proc. of IEEE ISIT 2005*, pp. 57 - 61

Devroye, N.; Mitran, P. & Tarokh, V. (2006). Achievable rates in cognitive radio channels. *IEEE Trans. Inf. Theory*, Vol. 52, No. 5, pp. 1813 - 1827

FCC (2003). ET Docket No 03-222. *Notice of proposed rule making and order*.

Geirhofer, S.; Tong, L. & Sadler, B.M. (2007). Dynamic spectrum access in the time domain: Modeling and exploiting white space. *IEEE Commun. Mag.*, Vol. 45, No. 5, pp. 66–72

Ghasemi, A. & Sousa, E. (2007). Fundamental limits of spectrum-sharing in fading environments. *IEEE Trans. Wireless Commun.*, Vol. 6, No. 2, pp. 649 - 658

Haykin, S. (2005). Cognitive radio: Brain-empowered wireless communications. *IEEE J. on Sel. Areas in Commun.*, Vol. 23, No.2, pp. 201-220

Hinman, R. (2006). Application of Cognitive Radio Technology to Legacy Military Waveforms in a JTRS (Joint Tactical Radio System) Radio. Proc. of IEEE MILCOM 2006, pp.1-5

IEEE 802.22-06/0003r3 (2006). *A PHY/MAC Proposal for IEEE 802.22 WRAN Systems. Part 2: The Cognitive MAC.*

IEEE 802.22-07/0113r0 (2007). *Draft Standard for Wireless Regional Area Networks Part 22: Cognitive Wireless RAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: Policies and procedures for operation in the TV Bands*

Jovicic, A. & Viswanath, P. (2006). Cognitive radio: An information-theoretic perspective. *Proc. of IEEE ISIT 2006*, pp. 2413 – 2417

Kim, H. & Shin, K.G. (2008). Efficient discovery of spectrum opportunities with MAC-layer sensing in cognitive radio networks. *IEEE Trans. Mobile Computing*, Vol. 7, No. 5, pp. 533–545

Mishra, S.M.; Sahai, A. & Broderson, R.W. (2006). Cooperative sensing among cognitive radios. *Proc. of IEEE ICC 2006*, pp. 1658 - 1663

Mitola, J. & Maguire, G. (1999). Cognitive radio: Making software radios more personal. *IEEE Personal Communications*, Vol. 6, No. 4, pp. 13-18

Mitola, J. (2000). *Cognitive radio: An integrated agent architecture for software defined radio*. Doctor of Technology, Royal Institute of Technology, Stockholm, Sweden

Motamedi, A. & Bahai, A. (2007). MAC protocol design for spectrum agile wireless networks: Stochastic control approach. *Proc. of IEEE DySPAN 2007*, pp. 448–451

Neel, J.; Buehrer, R.M.; Reed, J.H. & Gilles, R.P. (2002). Game theoretic analysis of a network of cognitive radios. Proc. of 45th MWSCAS 2002, pp. 409 – 412

Peng, C.; Zheng, H. & Zhao, B.Y. (2006). Utilization and fairness in spectrum assignment for opportunistic spectrum access. *ACM Mobile Networks and Applications (MONET)*, Vol. 11, No. 4, pp. 555 - 576

Sahai, A.; Hoven, N. & Tandra, R. (2004). Some fundament limits on cognitive radio. *Proc. of 42nd Allerton coference on communication, control and computing*

Xing, Y.; Mathur, C.N.; Haleem, M.A.; Chandramouli, R. & Subbalakshmi K.P. (2007a). Dynamic Spectrum Access with QoS and Interference Temperature Constraints. *IEEE Trans. Mobile Computing*, Vol. 6, No. 4, pp. 423–433

Xing, Y.; Chandramouli, R. & Cordeiro, C. (2007b). Price dynamics in competitive agile spectrum access markets. *IEEE J. on Sel. Areas in Commun.*, Vol. 25, No. 3, pp. 613 – 621

Gandi, S.; Buragohain, C.; Cao, L.; Zheng, H. & Suri, S. (2007). A General Framework for Wireless Spectrum Auctions. *Proc. of IEEE DySPAN 2007*, pp. 22 - 33

Wang, W.; Cui, Y.; Peng, T. & Wang, W. (2007a). "Noncooperative Power Control Game with Exponential Pricing for Cognitive Radio Network", *IEEE VTC2007-Spring*

Wang, W.; Liu, W.; Peng, T. & Wang, W. (2007b). Pricing Negotiation for Cooperative Spectrum Marketing Based on Cognitive Radio. *Proc. of FTC 2007*

Wang, W.; Gao, W.; Bai, X.; Peng, T.; Chuai, G. & Wang, W. (2007c). A Framework of Wireless Emergency Communications Based on Relaying and Cognitive Radio. *Proc. of IEEE PIMRC 2007*

Wang, W.; Wang, W.; Lu, Q. & Peng T. (2009a). An Uplink Resource Allocation Scheme for OFDMA-Based Cognitive Radio Networks. *Wiley Int. J. of Commun. Sys.*, Vol.22, No. 5, pp. 603 – 623

Wang, W. (2009b). Investigation of Radio Resource Management Investigation on Radio Resource Management Algorithms for Cognitive Radio Networks. Ph.D. Dissertation, Beijing University of Posts and Telecommunications.

Zhang, R. & Liang, Y.C. (2008). Exploiting multi-antennas for opportunistic spectrum sharing in cognitive radio networks. *IEEE J. Sel. Topics in Signal Processing*, Vol. 2, No. 1, pp. 88-102

Zhou, X.; Gandi, S.; Suri, S. & Zheng, H. (2008). eBay in the Sky: Strategy-Proof Wireless Spectrum Auctions. *Proc. of ACM MobiCom 2008*

Zhu, J. & Liu, K.J.R. (2007). Dynamic spectrum sharing: A game theoretical overview. *IEEE Commun. Mag.*, Vol. 45, No. 5, pp. 88 – 94, May 2007

Information Theory of Cognitive Radio System

F. G. Awan[1], N. M. Sheikh[1] and M. F. Hanif *
[1]University of Engineering and Technology
Lahore, 54890 Pakistan
*University of the Punjab
Quaid-e-Azam Campus
54590, Lahore
Pakistan

1. Introduction

Cognitive radio (CR) carries bright prospects for very efficient utilization of spectrum in future. Since cognitive radio is still in its early days, many of its theoretical limits are yet to be explored. In particular, determination of its maximum information rates for the most general case is still an open problem. Till today, many cognitive channel models have been presented. Either achievable or maximum bit rates have been evaluated for each of these. This chapter will summarize all the existing results, makes a comparison between different channel models and draw useful conclusions.

The scarcity of the radio frequency (RF) spectrum along with its severe under utilization, as suggested by various government bodies like the Federal Communications Commission (FCC) in USA and Ofcom in UK like Rafsh [1] and [2] has triggered immense research activity on the concept of CR all over the world. Of the many facets that need to be dealt with, the information theoretic modeling of CR is of core importance, as it helps predict the fundamental limit of its maximum reliable data transmission. The information theoretic model proposed in [3] represents the real world scenario that the CR will have to encounter in the presence of primary user (PU) devices. Authors in [3] characterize the CR system as an interference channel with degraded message sets (IC-DMS), since the spectrum sensing nature of the CR may enable its transmitter (TX) to know PU's message provided the PU is in close proximity of the CR. Elegantly using a combination of rate-splitting [4] and Gel'fand Pinsker (GP) [5] coding, [3] has given an achievable rate region of the so called CR-channel or IC-DMS. Further, in [3] time sharing is performed between the two extreme cases when either the CR dedicates zero power ("highly polite") or full power ("highly rude") to its message. A complete review of information theoretic studies can be found in [6] and [7].

This chapter then discusses outer-bounds to the individual rates and the conditions under which these bounds become tight for the symmetric Gaussian CR channel in the low interference gain regime. The CR transmitter is assumed to use dirty paper coding while deriving the outer-bounds. The capacity of the CR channel in the low interference scenario is

known when the CR employs "polite" approach by devoting some portion of its power to transmit PU's message that will help calculating quality of service for the CR users. Then, we will focus on the scenario when the CR goes for the "rude" approach i.e., does not relay PU's message and tries to maximize its own rates only. It will be derived that when both CR and the PU operate in low interference gain regime, then treating interference as additive noise at the PU receiver and doing dirty paper coding at the CR is nearly optimal.

2. Cognitive Radio Network Paradigms

Since its introduction in [8], the definition of cognitive radio has evolved over the years. Consequently, different interpretations of cognitive radio and different visions for its future exist today. In this section we describe a few communication models that have been proposed for cognitive radio. We broadly classify them into *overlay* or *known interference* models, *underlay* or *interference avoidance* models.

Underlay Paradigm:
The underlay paradigm encompasses techniques that allow communication by the cognitive radio assuming it has knowledge of the interference caused by its transmitter to the receivers of all noncognitive users [7]. In this setting the cognitive radio is often called a *secondary user* which cannot significantly interfere with the communication of existing (typically licensed) users, who are referred to as *primary users*. Specifically, the underlay paradigm mandates that concurrent noncognitive and cognitive transmissions may occur only if the interference generated by the cognitive devices at the noncognitive receivers is below some acceptable threshold. The interference constraint for the noncognitive users may be met by using multiple antennas to guide the cognitive signals away from the noncognitive receivers, or by using a wide bandwidth over which the cognitive signal can be spread below the noise floor, then despread at the cognitive receiver. The latter technique is the basis of both spread spectrum and ultrawideband (UWB) communications.
The interference caused by a cognitive transmitter to a noncognitive receiver can be approximated via reciprocity if the cognitive transmitter can overhear a transmission from the cognitive receiver's location. Alternatively, the cognitive transmitter can be very conservative in its output power to ensure that its signal remains below the prescribed interference threshold. In this case, since the interference constraints in underlay systems are typically quite restrictive, this limits the cognitive users to short range communications.
While the underlay paradigm is most common in the licensed spectrum (e.g. UWB underlays many licensed spectral bands), it can also be used in unlicensed bands to provide different classes of service to different users.

Overlay Paradigm:
The enabling premise for overlay systems is that the cognitive transmitter has knowledge of the noncognitive users' codebooks and its messages as well [7]. The codebook information could be obtained, for example, if the noncognitive users follow a uniform standard for communication based on a publicized codebook. Alternatively, they could broadcast their codebooks periodically. A noncognitive user message might be obtained by decoding the message at the cognitive receiver. However, the overlay model assumes the noncognitive message is known at the cognitive transmitter when the noncognitive user begins its

transmission. While this is impractical for an initial transmission, the assumption holds for a message retransmission where the cognitive user hears the first transmission and decodes it, while the intended receiver cannot decode the initial transmission due to fading or interference. Alternatively, the noncognitive user may send its message to the cognitive user (assumed to be close by) prior to its transmission. Knowledge of a noncognitive user's message and/or codebook can be exploited in a variety of ways to either cancel or mitigate the interference seen at the cognitive and noncognitive receivers. On the one hand, this information can be used to completely cancel the interference due to the noncognitive signals at the cognitive receiver by sophisticated techniques like dirty paper coding [9]. On the other hand, the cognitive users can utilize this knowledge and assign part of their power for their own communication and the remainder of the power to assist (relay) the noncognitive transmissions. By careful choice of the power split, the increase in the noncognitive user's signal-to-noise power ratio (SNR) due to the assistance from cognitive relaying can be exactly offset by the decrease in the noncognitive user's SNR due to the interference caused by the remainder of the cognitive user's transmit power used for its own communication. This guarantees that the noncognitive user's rate remains unchanged while the cognitive user allocates part of its power for its own transmissions. Note that the overlay paradigm can be applied to either licensed or unlicensed band communications. In licensed bands, cognitive users would be allowed to share the band with the licensed users since they would not interfere with, and might even improve, their communications. In unlicensed bands cognitive users would enable a higher spectral efficiency by exploiting message and codebook knowledge to reduce interference.

Interweave Paradigm:

The 'interweave' paradigm is based on the idea of *opportunistic communication*, and was the original motivation for cognitive radio [10]. The idea came about after studies conducted by the FCC [8] and industry [2] showed that a major part of the spectrum is not utilized most of the time. In other words, there exist temporary space-time frequency voids, referred to as *spectrum holes*, that are not in constant use in both the licensed and unlicensed bands.

These gaps change with time and geographic location, and can be exploited by cognitive users for their communication. Thus, the utilization of spectrum is improved by opportunistic frequency reuse over the spectrum holes. The interweave technique requires knowledge of the activity information of the noncognitive (licensed or unlicensed) users in the spectrum. One could also consider that all the users in a given band are cognitive, but existing users become primary users, and new users become secondary users that cannot interfere with communications already taking place between existing users.

To summarize, an interweave cognitive radio is an intelligent wireless communication system that periodically monitors the radio spectrum, intelligently detects occupancy in the different parts of the spectrum and then opportunistically communicates over spectrum holes with minimal interference to the active users. For a fascinating motivation and discussion of the signal processing challenges faced in interweave cognitive radio, we refer the reader to [11].

Table 1 [12] summarizes the differences between the underlay, overlay and interweave cognitive radio approaches. While underlay and overlay techniques permit concurrent cognitive and noncognitive communication, avoiding simultaneous transmissions with

noncognitive or existing users is the main goal in the interweave technique. We also point out that the cognitive radio approaches require different *amounts* of side information: underlay systems require knowledge of the interference caused by the cognitive transmitter to the noncognitive receiver(s), interweave systems require considerable side information about the noncognitive or existing user activity (which can be obtained from robust primary user sensing) and overlay systems require a large amount of side information (non-causal knowledge of the noncognitive user's codebook and possibly its message). Apart from device level power limits, the cognitive user's transmit power in the underlay and interweave approaches is decided by the interference constraint and range of sensing, respectively. While underlay, overlay and interweave are three distinct approaches to cognitive radio, hybrid schemes can also be constructed that combine the advantages of different approaches. For example, the overlay and interweave approaches are combined in [7].

Before launching into capacity results for these three cognitive radio networks, we will first review capacity results for the interference channel. Since cognitive radio networks are based on the notion of minimal interference, the interference channel provides a fundamental building block to the capacity as well as encoding and decoding strategies for these networks.

Underlay	Overlay	Interweave
Channel Side Information: Cognitive (secondary) transmitter knows the channel strengths to noncognitive (primary) receiver(s).	Codebook Side Information: Cognitive nodes know channel gains, codebooks and the messages of the noncognitive users	Activity Side Information: Cognitive user knows the spectral holes in space, time, or frequency when the noncognitive user is not using these holes.
Cognitive user can transmit simultaneously with noncognitive user as long as interference caused is below an acceptable limit.	Cognitive user can transmit simultaneously with noncognitive user; the interference to noncognitive user can be offset by using part of the cognitive user's power to relay the noncognitive user's message.	Cognitive user transmits simultaneously with a noncognitive user only in the event of a false spectral hole detection.
Cognitive user's transmit power is limited by the interference constraint.	Cognitive user can transmit at any power, the interference to noncognitive users can be offset by relaying the noncognitive user's message.	Cognitive user's transmit power is limited by the range of its spectral hole sensing.

Table 1. Comparison of underlay, overlay and interweave cognitive radio techniques.

3. Interference-Mitigating Cognitive Behavior: The Congnitive Radio Channel

This discussion is has been taken from Natasha's paper. This consideration is simplest possible scenario in which a cognitive radio could be employed. Assume there exists a primary transmitter and receiver pair (S1 — R1), as well as the cognitive secondary transmitter and receiver pair (S2 — R2). As shown in Fig. 1.(a), there are three possibilities for transmitter cooperation in these two point-to-point channels. We have chosen to focus on transmitter cooperation because such cooperation is often more insightful and general than receiver-side cooperation [12, 13]. Thus assume that each receiver decodes independently. Transmitter cooperation in this figure is denoted by a directed double line. These three channels are simple examples of the cognitive decomposition of wireless networks seen in [14]. The three possible types of transmitter cooperation in this simplified scenario are:

1. **Competitive behavior:** The two transmitters transmit independent messages. There is no cooperation in sending the messages, and thus the two users compete for the channel. This is the same channel as the 2 sender, 2 receiver interference channel [14, 15].

2. **Cognitive behavior:** Asymmetric cooperation is possible between the transmitters. This asymmetric cooperation is a result of S2 knowing S1's message, but not vice-versa. As a first step, we idealize the concept of message knowledge: whenever the cognitive node S2 is able to hear and decode the message of the primary node S1, we assume it has full a priori knowledge. This is called the genie assumption, as these messages could have been given to the appropriate transmitters by a genie. The one way double arrow indicates that S2 knows S1's message but not vice versa.

This is the simplest form of asymmetric non-causal cooperation at the transmitters. Usage of the term cognitive behavior is to emphasize the need for S2 to be a "smart" device capable of altering its transmission strategy according to the message of the primary user. We can motivate considering asymmetric side information in practice in three ways:

• Depending on the device capabilities, as well as the geometry and channel gains between the various nodes, certain cognitive nodes may be able to hear and/or obtain the messages to be transmitted by other nodes. These messages would need to be obtained in real time, and could exploit the geometric gains between cooperating transmitters relative to receivers in, for example, a 2 phase protocol [3].

• In an Automatic Repeat reQuest (ARQ) system, a cognitive transmitter, under suitable channel conditions (if it has a better channel to the primary transmitting node than the primary receiver), could decode the primary user's transmitted message during an initial transmission attempt. In the event that the primary receiver was not able to correctly decode the message, and it must be retransmitted, the cognitive user would already have the to-be-transmitted message, or asymmetric side information, at no extra cost (in terms of overhead in obtaining the message).

• The authors in [16] consider a network of wireless sensors in which a sensor S2 has a better sensing capability than another sensor S1 and thus is able to sense two events, while S1 is only able to sense one. Thus, when they wish to transmit, they must do so under an asymmetric side-information assumption: sensor S2 has two messages, and the other has just one.

3. **Cooperative behavior:** The two transmitters know each others' messages (two way double arrows) and can thus fully and symmetrically cooperate in their transmission. The channel pictured in Fig. 1. (c) may be thought of as a two antenna sender, two single antenna receivers broadcast channel [17].

Many of the classical, well known information theoretic channels fall into the categories of competitive and cooperative behavior. For more details, we refer the interested reader to the cognitive network decomposition theorem of [13] and [18]. We now turn to the much less studied behavior which spans and in a sense interpolates between the symmetric cooperative and competitive behaviors. We call this behavior asymmetric cognitive

behavior. In this section we will consider one example of cognitive behavior: a two sender, two receiver (with two independent messages) interference channel with asymmetric and a priori message knowledge at one of the transmitters, as shown in Fig. 1. (b).

Certain asymmetric (in transmitter cooperation) channels have been considered in the literature: for example in [19], the capacity region of a multiple access channel with asymmetric cooperation between the two transmitters is computed. The authors in [20] consider a channel which could involve asymmetric transmitter cooperation, and explore the conditions under which the capacity of this channel coincides with the capacity of the channel in which both messages are decoded at both receivers. In [21, 18] the authors introduced the cognitive radio channel, which captures the most basic form of asymmetric transmitter cooperation for the interference channel. We now study the information theoretic limits of interference channels with asymmetric transmitter cooperation, or cognitive radio channels.
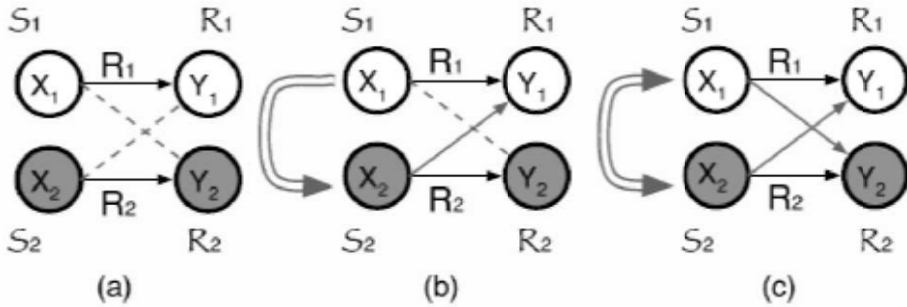
Fig. 1. (a) Competitive behavior, the interference channel. The transmitters may not cooperate. (b) Cognitive behavior, the cognitive radio channel. Asymmetric transmitter cooperation. (c) Cooperative behavior, the two antenna broadcast channel. The transmitters, but not the receivers, may fully and symmetrically cooperate.

The channel is thus expressed via following pair of equations:

$$Y_p = X_p + aX_s + Z_p \tag{1}$$

$$Y_s = bX_p + X_s + Z_s \tag{2}$$

While deriving the channel capacity, an assumption of low interference- gain has been made. Low interference regime corresponds to the scenario where the cognitive user is assumed to be near its own base station rather than that of primary user, which normally is the case. When applied to the interference channel in standard form, this situation corresponds to $a \leq 1$.

At the same time the two devices are assumed to work in an environment where the *co-existence conditions* exist, ensuring that cognitive radio generates no interference with the primary user in its vicinity and the primary receiver is a single user decoder. With all these assumptions, the channel is named as the *cognitive* (1,a,b,1) *channel*. The capacity *R s* (in bps) of the cognitive radio under the conditions existing as above is expressed in a closed form relation as:

$$R_s^* = \frac{1}{2}\log(1+(1-\alpha^*)P_s) \quad \text{provided } a \leq 1 \tag{3}$$

where $\alpha^* \in [0,1]$ and its value is determined using the following arguments:

To ensure that the primary user remains unconscious of the presence of the cognitive device and communicates at a rate of $\frac{1}{2}\log(1+P_p)$, the maximum achievable rate R_p^* of the primary system is found to be:

$$\frac{1}{2}\log\left(1+\frac{(\sqrt{P_p}+a\sqrt{\alpha P_s})^2}{1+a^2(1-\alpha)P_s}\right) = \frac{1}{2}\log(1+P_p) = R_p^* \tag{4}$$

Now for $0 < a < 1$, using Intermediate Value Theorem, this quadratic equation in a always has a unique root in [0, 1]:

$$\alpha^* = \left(\frac{\sqrt{P_p}(\sqrt{1+a^2P_S(1+P_p)}-1)}{a\sqrt{P_S}(1+P_p)}\right)^{\frac{1}{2}} \tag{5}$$

It is to be noted that the above capacity expressions hold for any $b \in R$. For detailed proofs the reader is referred to [8].

A few important points are worth mentioning here. Since the cognitive radio knows both *mp* (the message to be transmitted by primary user) and *ms* (the message to be transmitted by the cognitive device), it generates its codeword X_{ns} such that it also incorporates the codeword X_{np} to be generated by the primary user. By doing so, the cognitive device can implement the concept of dirty paper coding that helps it mitigating the interference caused by the primary user at the cognitive receiver. Thus the cognitive device performs superposition coding as follows:

$$X_s^n = \widehat{X}_s^n + \sqrt{\frac{\alpha P_S}{P_p}}X_p^n \tag{6}$$

Where $\widehat{X}_s^n$ encodes m_s and is generated by performing dirty paper coding, treating $(b+\sqrt{\frac{\alpha P_S}{P_p}})X_p^n$ as a known interference that will affect the secondary receiver. It is evident from (6) that the secondary device uses part of its power αPs to relay the primary user's message to the primary receiver. This relaying of message from the secondary user results in an elevated value of SNR at the primary receiver. At the same time, the secondary user's message with power (1 - α) P_s, transmitted towards the primary receiver balances the increase in SNR and as a result the primary device remains completely oblivious to the presence of the cognitive user. This approach has been named as *selfless approach* in [23].

In [8], results corresponding to high interference regime $a > 1$ have also been presented as ancillary conclusions. But such results are of not much significance as they present the

scenario of high interference caused by the secondary user which is an event with low probability of occurrence.

Similar results have also been obtained in [3], [16] and [15]. The results in [15] correspond to the *selfish approach* of [23] and thus represent an upper bound on information rates of cognitive radio but with interference, because in this case the cognitive user does not spend any of its power in relaying the message to the primary receiver. Similarly authors in [16] have shown that, in the Gaussian noise case, their capacity region is explicitly equal to that of [8] and, numerically, to that of [23]. Very recently [24] has also extended the results of [8] to the case of Gaussian Multiple Access channel (MAC) with *n* cognitive users. [24] has simply used the results for a general MAC channel i.e., the achievable rate for n-users is the sum of the achievable rates of individual users. Using this, together with the result of [8], it has determined the achievable rate region of a Gaussian MAC cognitive channel.
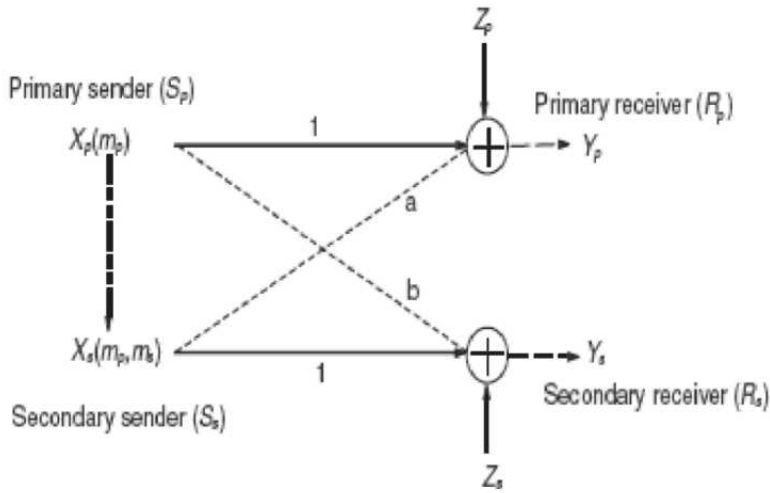
Fig. 2. The Gaussian interference channel in its standard form

4. Interference Avoiding Channel

This channel model, as devised in [22] and [6], works on the principle of *opportunistic communication* i.e., the secondary communication takes place only when the licensed user is found to be idle and a *spectrum hole* is detected. Thus this model conforms to the basic requirement of the cognitive device not interfering with the licensed users. The secondary sender S_S and receiver R_S are assumed to have a circular sensing region with them being in the center. The secondary transmitter and receiver sensing regions are circular with each having radius R_r. The distance between them is d. They are further supposed to be communicating in the presence of primary users *A, B* and *C* as shown in Fig. 3.
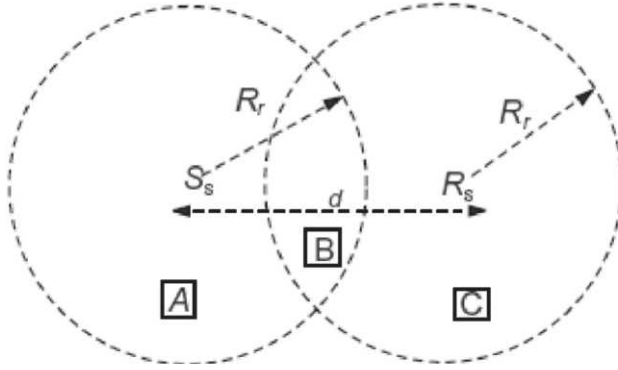
Fig. 3. The scenario for interference avoiding channel

The cognitive sender *SS* can detect a spectral hole when both *A* and *B* are inactive whereas the secondary receiver *RS* determines this when it finds both *B* and *C* to be not involved in a communication scenario. Since secondary transmitter and receiver do not have complete knowledge of primary users activity in each other's sensing regions, the spectral activity in their respective regions corresponds to the notion of being *distributed.* Similarly, the primary user activity sensed by the secondary transmitter-receiver pair continues to change with time i.e., different primary users become active and inactive at different time. Thus, the spectral activity is also assumed to be *dynamic.* To incorporate both these features, the conceptual model of Fig. 3 is reduced to the two switch mathematical model as shown in Fig. 4.
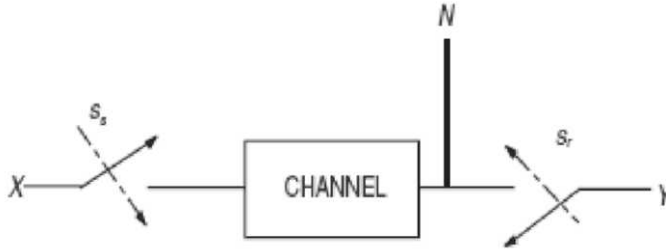
Fig. 4. The two switch model

The two switches s_s, s_r are treated as binary random variables with s_s, $s_r \in \{0, 1\}$. The value of $s_s = 1$ or $s_r = 1$ means that there are no primary users in the sensing region of the secondary transmitter or receiver and vice versa holds if either of these two values is zero. The input *X* is related to the output 7 via following equation:

$$Y = (X_{s_s} + N)s_r \tag{7}$$

where *N* is additive white Gaussian noise (AWGN) at the secondary receiver. s_r is either 1 or 0 as mentioned above. So when it is multiplied as done in (7), it simply shows whether the secondary receiver has detected the primary device or not.

If s_s is only known to the transmitter and s_r is only available to the receiver, the situation corresponds to communication with *partial side information.* Determination of capacity with partial side information involves input distribution maximization that is difficult to solve. This is not done and instead a tight capacity upper and lower bound is obtained for this communication channel. A capacity upperbound is determined by assuming that the receiver knows both *ss* and *sr* whereas, the transmitter only knows s_s. The expression of capacity Cs_s* of secondary user for this case is [23]:

$$C_{s_s,*}(P) = Prob[s_s = s_r = 1] \log \left(1 + \frac{P}{Prob[s_s = 1]}\right) \tag{8}$$

where P is the secondary transmitter power constraint. For the capacity lowerbound, [23] uses the results of [25]. For this a genie g argument is used. It should be noted that utilization of genie concept represents the notion that either the sender or receiver is provided with some information noncausally. To determine the capacity lower bound the genie is supposed to provide some side information to the receiver.

So if the genie provides some information it must have an associated entropy rate. The results in [25] suggest that the improvement in capacity due to this genie information cannot exceed the entropy rate of the genie information itself. Using this argument and that the genie provides information to the receiver every T channel uses, it is easy to establish that the capacity lower bound approaches the upper bound even for very highly dynamic environments.

It is assumed that the location of primary users in the system follows Poisson point process with a density of X nodes per unit area. And that the primary user detection at the secondary transmitter and receiver is perfect. The capacity expression in (8), as given in [23], is evaluated to be:

$$C = e^{-\lambda \left(2R_r^2(\pi - \arccos(\frac{d}{2R_r})) + dR_r\sqrt{1 - \frac{d^2}{4R_r^2}}\right)} \log\left(1 + Pe^{\lambda \pi R_r^2}\right) \tag{9}$$

where, again, P is the secondary power constraint.

5. Colloborative Cognitive Channel

In this channel, the cognitive user is modeled as a relay, with no information of its own, working between the primary transmitter and receiver. Capacity limits for *collaborative communications* have been recently explored [26] that suggest sufficient conditions on the geometry and the signal path loss of the transmitting entities for which performance close to the genie bound can be achieved [Natasha's tutorial].

Consider three nodes, source (s), relay (r) and destination *(d)* as illustrated in Fig. 5.
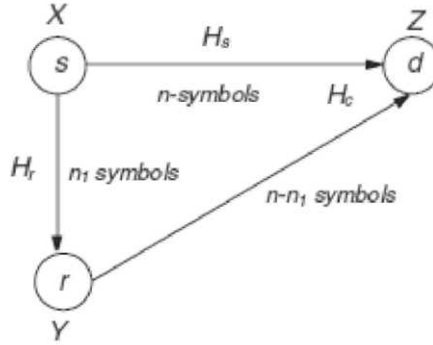
Fig. 5. The collaborative communication channel incorporating the source s, relay r and the destination d node

The relay node is assumed to work in *half duplex* mode, meaning that, it cannot receive and transmit data simultaneously. Thus the system works in two phases i.e., *the listening phase* and *the collaborative phase.* During the first phase the relay node receives data from the source node for the first $n1$ transmissions while for the collaborative phase the relay transmits to the destination for the remaining $n - n_1$ transmissions, where n are the number of channel uses, in which the source node wishes to transmit the 2^{nR} messages. Taking the channels as AWGN, considering X and U as column vectors representing the transmission from the source and relay node respectively and denoting by Y and Z the received messages at the relay and destination respectively, the listening phase is described via following equations:

$$Z = H_s X + N_z \tag{10}$$

$$Y = H_r X + N_y \tag{11}$$

where N_z and N_y represent complex AWGN with variance $1/2$, H_s is the fading matrix between the source and destination nodes and H_r is the fading matrix between the source and relay nodes. In the collaborating phase:

$$Z = H_c \left[X^T, U^T \right]^T + N_z \tag{12}$$

where *Hc* is the channel matrix that contains *Hs* as a submatrix. It is well known that a Multiple Input Multiple Output (MIMO) system with Gaussian codebook and rate R bits/channel can reliably communicate over any channel with transfer matrix H such that

$R \leq \log_2 \det(I + HH^{\dagger}) := C(H)$ where I denotes the identity matrix and $H^{\dagger}$ represents the conjugate transpose of H. Before providing an explicit formula for rates, a little explanation is in order here.

During the first phase, relay listens for an amount of time $n1$ and since it knows *Hr,* it results in $nR \leq n_1 C(Hr)$. On the other hand, the destination node receives information at the rate of *C(Hs)* bits/channel during the first phase and at the rate of C(H$_c$) bits/channel during the second phase. Thus it may reliably decode the message provided that $nR \leq n_1 C(H_s)+(n-$

$n_1)C(H_c)$. Taking limit n^><x> the ratio n_1/n tends to a fraction f such that the code of rate R for the set of channels (H_r, H_c) satisfies:

$$R \le fC(H_s) + 1(1 - f)C(H_c) \tag{13}$$

$$R \le fC(H_r) \tag{14}$$

where $f \in [0,1]$

Similarly [23] presents a corollary which suggests that if the cognitive user has no message of its own, it can aid the primary system because it knows the primary's message, resulting in an improvement of primary user's data rates.

New outer bounds to the individual rates and the conditions under which these bounds become tight for the symmetric Gaussian cognitive radio (CR) channel in the low interference gain regime are presented in [27]. The CR transmitter is assumed to use dirty paper coding while deriving the outer bounds. The capacity of the CR channel in the low interference scenario is known when the CR employs "polite" approach by devoting some portion of its power to transmit primary user's (PU's) message. However, this approach does not guarantee any quality of service for the CR users. Hence, focus will be on the scenario when the CR goes for the "rude" approach, does not relay PU's message and tries to maximize its own rates only. It is shown that when both CR and the PU operate in low interference gain regime, then treating interference as additive noise at the PU receiver and doing dirty paper coding at the CR is nearly optimal.

6. Comparsions

The channel model presented in the first section uses complex coding techniques to mitigate channel interference that naturally results in higher throughput than that of the channel model of the second section. But there is one serious constraint in the first channel model, the information throughput of the cognitive user is highly dependent upon the distance between the primary transmitter and the cognitive transmitter. If this distance is large, secondary transmitter spends considerable time in obtaining the primary user's message. After obtaining and decoding the message, the cognitive device dirty paper codes it and sends it to its receiver. This message transfer from the primary to the cognitive transmitter results in lower number of bits transmitted per second by the cognitive radio and hence results in reduced data rates.

The capacity of the two switch model is independent of the distance between the two transmitters as the secondary transmitter refrains from sending data when it finds the primary user busy. Thus the benefit of the channel interference knowledge in the first channel model quickly disappears as the distance between the primary and secondary transmitter tends to increase. Accurate estimation of the primary system's message and the interference channel coefficient needed in the first channel model for dirty paper coding purposes is itself a problem. Inaccurate estimation of these parameters will result in a decrease in the rates of the cognitive radio because the dirty paper code, based on the knowledge of primary user's message and channel interference coefficient, will not be able to completely mitigate the primary user's interference. At the same time determination of

channel interference coefficient requires a handshaking protocol to be devised which is a serious overhead and may result in poor performance.

On the other hand, the interference avoiding channel cannot overcome the *hidden terminal problem*. This problem naturally arises as the cognitive user would not be able to detect the presence of distant primary devices. The degraded signals from the primary users due to multipath fading and shadowing effects would further aggravate this problem.

This requires the secondary systems to be equipped with extremely sensitive detectors. But very sensitive detectors have prohibited long sensing times. Thus a protocol needs to be devised by virtue of which the sensed information is shared between the cognitive devices.

Finally, the role of CR as a relay in the third channel model restricts the real usefulness of the concept of dynamic spectrum utilization. Although limited, yet significant gains in data rates of the existing licensed user system can be obtained by restricting a CR device to relays of PU's message only.

7. References

1. Federal Communications Commission Spectrum Policy Task Force, "Report of the Spectrum Efficiency Working Group," *Technical Report 02-135*, no. November, 2002.
2. Shared Spectrum Company, "Comprehensive Spectrum occupancy measurements over six different locations,"August 2005.
3. N. Devroye, P. Mitran, and V. Tarokh, "Achievable rates in cognitive radio channels," *IEEE Transactions on Information Theory*, vol. 52, no. 5, pp. 1813–1827, May 2006.
4. T. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Transactions on Information Theory*, vol. 27, no. 1, pp. 49–60, 1981.
5. S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Problems of Control and Information Theory*, vol. 9, no. 1, pp. 19–31, 1980.
6. F. G. Awan and M. F. Hanif, "A Unified View of Information-Theoretic Aspects of Cognitive Radio," in *Proc. International Conference on Information Technology: New Generations*, pp. 327–331, April 2008.
7. S. Srinivasa and S. A. Jafar, "The throughput potential of cognitive radio - a theoretical perspective," *IEEE Communications Magazine*, vol. 45, no. 5, pp. 73–79, 2007.
8. Jovicic and P. Viswanath, "Cognitive radio: An information theoretic perspective," *2006 IEEE International Symposium on Information Theory*, July 2006.
9. M. H. M. Costa, "Writing on dirty paper," *IEEE Transactions on Information Theory*, vol. 29, no. 3, pp. 439–441, May 1983.
10. Joseph Mitola, "Cognitive Radio: An Integrated Agent Architecture for Software Defined Radio," *PhD Dissertation, KTH, Stockholm, Sweden*, December 2000.
11. Paul J. Kolodzy, "Cognitive Radio Fundamentals," *SDR Forum, Singapore*, April 2005.
12. C.T.K.Ng and A. Goldsmith, "Capacity gain from transmitter and receiver cooperation," in Proc. IEEE Int. Symp. Inf. Theory, Sept. 2005.
13. N. Devroye, P. Mitran, and V. Tarokh, "Cognitive decomposition of wireless networks," in Proceedings of CROWNCOM, Mar. 2006.
14. Carleial, "Interference channels," IEEE Trans. Inf. Theory, vol. IT-24, no. 1, pp. 60–70, Jan. 1978.
15. N. Devroye, P. Mitran, and V. Tarokh, "Achievable rates in cognitive radio channels," IEEE Trans. Inf. Theory, vol. 52, no. 5, pp. 1813–1827, May 2006.

16. W. Wu, S. Vishwanath, and A. Arapostathis, "On the capacity of the interference channel with degraded message sets," IEEE Trans. Inf. Theory, June 2006.
17. H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian MIMO broadcast channel," IEEE Trans. Inf. Theory, vol. 52, no. 9, pp. 3936–3964, Sept. 2006.
18. N. Devroye, P. Mitran, and V. Tarokh, "Achievable rates in cognitive networks," in 2005 IEEE International Symposium on Information Theory, Sept. 2005.
19. E. C. van der Meulen, "Three-terminal communication channels," Adv. Appl. Prob., vol. 3, pp. 120–154, 1971.
20. I.Maric, R. Yates, and G. Kramer, "The strong interference channel with unidirectional cooperation," in Information Theory and Applications ITA Inaugural Workshop, Feb. 2006.
21. N. Devroye, P. Mitran, and V. Tarokh, "Achievable rates in cognitive radio channels," in 39th Annual Conf. on Information Sciences and Systems (CISS), Mar. 2005.
22. S. Jafar and S. Srinivasa, "Capacity limits of cognitive radio with distributed dynamic spectral activity," in *Proc. of ICC*, June 2006.
23. S. Srinivasa and S. A. Jafar, "The throughput potential of cognitive radio: A theoretical perspective," in *Fortieth Asilomar Conference on Signals, Systems and Computers, 2006.*, Oct. 2006.
24. P. Cheng, G. Yu, Z. Zhang, H.-H. Chen, and P. Qiu, "On the achievable rate region of gaussian cognitive multiple access channel," *IEEE Communications Letters*, vol. 11, no.5, pp. 384–386, May. 2007.
25. S. A. Jafar, "Capacity with causal and non-causal side information-a unified view," *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5468–5474, Dec. 2006.
26. P. Mitran, H. Ochiai, and V. Tarokh, "Space-time diversity enhancements using collaborative communication*," IEEE Transactions on Information Theory*, vol. 51, no.6, pp. 2041–2057, June 2005.
27. F. G. Awan, N. M. Sheikh and F. H. Muhammad, "Outer Bounds for the Symmetric Gaussian Cognitive Radio Channel with DPC Encoded Cognitive Transmitter," in *Proc. The World Congress on Engineering 2009 (WCE 2009) by the International Association of Engineers (IAENG)*, London, UK, July 2009.

Cooperative Cognitive Systems

Lorenza Giupponi and Christian Ibars

1. Introduction

Cognitive Radio is a new paradigm in wireless communications to enhance utilization of limited spectrum resources. It is defined as a radio able to utilize available side information, in a decentralized fashion, in order to efficiently use the radio spectrum left unused by licensed systems. The basic idea is that a secondary user (SU)(a cognitive unlicensed user) is able to properly sense the spectrum conditions and, to increase efficiency in spectrum utilization, it seeks to underlay, overlay or interweave its signals with those of the primary (licensed) users (PUs), without impacting their transmission. In this sense, the cognitive radio paradigm defines a set of rules for the coexistence of two or more radio systems in a given spectrum allocation. These systems are given different usage rights and a set of rules to abide. While coordination between different systems is not a requirement, it may improve the performance of both primary and secondary users, as it is argued in this text. In particular, the book chapter is organized in two parts. In the first part (section 2), we provide several mechanisms which require cooperation and improve different functions of the cognitive radio mechanism. Secondary spectrum usage, requires more sophisticated spectrum management and coexistence techniques than licensed or unlicensed spectrum operation. Different degrees of cooperation are possible: from simply following the spectrum regulation and keeping transmission power below the specified mask, to accurate sensing and tracking of the primary licensee, or contribution of the SUs to the detection of the primary signal. A higher degree of cooperation entails a higher degree of complexity, both in terms of hardware design and in terms of network coordination. This results in a tradeoff between performance gain and complexity increase, which may result in the adoption of a particular cooperative solution. Then, we focus on mechanisms requiring cooperation at the physical layer level, in terms of signal design and we address the most significant relaying techniques. In the second part of the chapter 3, we focus on the decisions the cooperative cognitive radios have to make. These decisions strongly depend on those made by the other radios, since the licensed users performances are limited by the aggregated interference generated by all the cognitive radios simultaneously transmitting in their band. This is why the performance is analyzed using game-theoretic tools, already proven good at modeling interactions in decision processes. As a result, the focus of the second part of the chapter is to take advantage of the cooperative schemes introduced in the first part in order to model and control the interference generated at the licensed users by the cooperative and cognitive system. We propose a particular kind of games, characterized by favorable convergence characteristics, i.e. potential games, and we describe how to design and identify these games, in both cases of complete and incomplete information. Finally, we summarize the chapter conclusions in chapter 4.

2. Cooperation Mechanisms for Cognitive Radio

2.1 Cognitive Radio Techniques: Interweave, Underlay, Overlay

Different strategies for cognitive radio environments have been defined, depending on how secondary usage of the spectrum is carried out. A first strategy, which is in line with the original idea of secondary spectrum usage, consists in utilizing the *spectrum holes* left by primary systems to establish communications between SUs. This approach has been referred to as *interweave* Srinivasa (2007). A second approach consists in ensuring that secondary transmissions are always below the maximum allowable interference temperature at the primary receivers[1]. This approach is referred to as *underlay*. Finally, a third approach consists in the SUs cooperating with the primary transmission while transmitting their own signal. This scheme is known as *overlay*. These schemes are schematically represented in Figure 1.
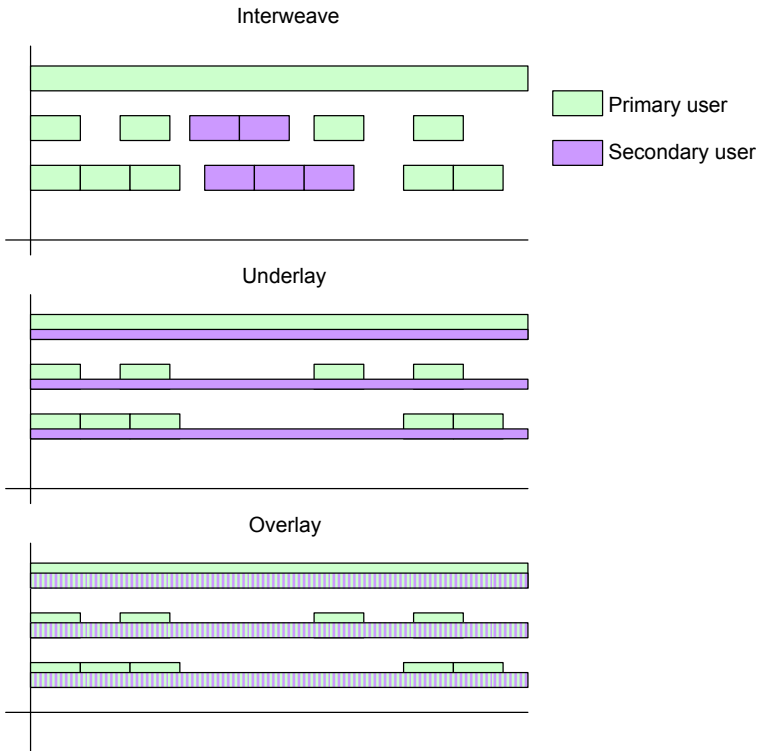

Fig. 1. Different types of cognitive radio schemes: interweaved transmission of primary and secondary signals, underlay of secondary signal, and cooperative overlay of secondary signal.

---

[1] Interference temperature is a measure of the noise and interference power level at the receiver.

2.1.1 Interweave Cognitive Radio

Interweave cognitive radio follows an interference avoidance strategy. SUs are only allowed to use the spectrum when primary users are inactive. Such method is opportunistic in the sense that SUs take advantage of spectrum that otherwise goes unused, and implements the original idea that cognitive radio should exploit the so-called spectrum holes left by the primary licensee. The main difficulty in the interweave scheme is that of sensing and predicting the activity of the primary user in several radio channels, i.e. detecting the spectrum holes. This task becomes more difficult if PUs are highly dynamic, i.e. their spectral activity changes fast, and it requires secondary transmission equipment to be very agile in switching on and off, and in switching frequency channels. It also becomes more difficult when the range of secondary transmissions increases, as the primary activity sensed by the secondary transmitter and secondary receiver may vary due to different signal strengths of the primary signal. These factors decrease the correlation between the spectrum sensed at the transmitter and at the receiver, and therefore reduce the effectiveness of secondary spectrum utilization.

2.1.2 Underlay Cognitive Radio

The underlay approach is a more conservative choice. Rather than tracking the primary user activity and adapting to it, it consists in transmitting at very low power to ensure that the interference temperature of the primary user does not exceed a predefined limit. This method has very different requirements on the secondary transceiver equipment than the interweave scheme. Rather than time and frequency agile radios, the secondary transceiver must be able to operate at very low SNR (Signal to Noise Ratio). This typically restricts underlay cognitive radio to low data rate applications or very short range applications.

Underlay cognitive radio has been adopted by regulatory bodies worldwide, and is allowed when SUs transmit using the Ultra-wideband (UWB) signal format. The UWB signal format is limited by a very strict spectral mask, allowing very low power transmission in a large bandwidth which overlaps with other licensed services (the actual bandwidth varies in different countries and is defined by their corresponding regulatory bodies). Due to its large bandwidth, the spectrum used by UWB spans several primary services. The spectral mask, along with additional restrictions on its operation (passive beacons, for example, are not allowed, and outdoor usage is restricted to handheld devices) ensures that secondary (UWB) transmission does not interfere with primary bandwidth usage, as the received UWB power at any primary receiver is typically well below the noise floor. However, such restrictive regulation limits the applicability of UWB to very short range applications (below 10m), such as personal area networks or cable replacement applications. Longer range operation, up to 300m, is possible but at very low data rates.

2.1.3 Overlay Cognitive Radio

In the overlay approach, SUs devote part of its transmit power to enhance the primary signal and facilitate its detection at the primary receiver. In exchange, they may be allowed to increase the interference temperature level further than the underlay approach. The fact that they contribute to improve the detection of the PU may also contribute to a higher acceptance of this technique by primary licensees. Thus, the overlay approach can be seen as an evolutionary step from the underlay technique, where a tighter degree of integration between primary and secondary is necessary, and higher performance is achieved. However, the overlay approach has not been yet implemented. The basic principle is that SUs that are close to the primary transmitter have access to a high quality primary signal which they are able to

successfully decode. Then, they use knowledge of the primary message to produce a signal that complements the primary signal and improves the detection probability of the primary receiver. At the same time, the secondary transmitter communicates with the secondary receiver with a low-power signal. It may be assumed that either the secondary transmitter or secondary receiver have knowledge of the primary signal, or both. Depending on the assumption made, several techniques, such as dirty paper coding, or successive interference cancelation, may be applied.

The main advantage of this approach is the higher interference temperature that can be tolerated, and better detection of the primary signal. On the other hand, this technique requires a higher degree of complexity in the secondary transceivers, and assumes that these are able to produce a compatible signal format when aiding the primary transceiver. Furthermore, this scheme may require knowledge of the channel state in order to guarantee that the primary signal is successfully detected. Finally, a power control mechanism must be put in place which determines the power that secondary transmitters devote to the primary and secondary signals, respectively.

2.2 Cooperation in Interweave Cognitive Radio Systems

In the initially proposed interweave cognitive radio scheme, where the SUs track the activity and spectrum usage of the primary licensee and adapt to it, several mechanisms have been proposed to improve their spectrum utilization.

In the event of lack of coordination, SUs must use sophisticated algorithms to track and predict the activity of PUs. Such mechanisms, may be effective for static, predictable primary licensees, such as television broadcasters. However, several factors decrease the efficiency of an uncoordinated approach, as argued earlier.

If one allows for a certain degree of cooperation between secondary users, these can exchange the outcome of their spectrum sensing and collectively provide a better notion of the primary spectrum occupancy Ghasemi & Sousa (8-11 Nov. 2005, Baltimore, USA). For example, a shadowed user may learn from the presence of a primary through the exchange of spectrum occupancy information with other SUs.

A further step is to allow coordination between primary and secondary systems. An effective approach is that the PU signals with beacons the spectrum occupancy in its channels. SUs may then monitor the beacon channel and learn about spectrum holes. In addition, such beacons can be designed to be easily detected with strong modulation and channel coding formats. Several options are possible in a beacon system: to use grant beacons to signal spectrum availability; to use denial beacons to warn about the activity of the primary system, so that SUs do not transmit; or to use a dual beacon approach, which is better in the case of multiple primary transmitters Mangold (2006). In this case, users must wait until they hear a grant beacon and, at the same time, no denial beacon is detected. The advantage of this approach is that a SU that is "hidden" from the primary transmitter (and does not receive the denial beacon) may not receive a grant beacon either, thus refrain from transmitting and interfering the primary communication.

In the following section we study a second type of cooperation where SUs actively contribute to enhance the quality of the primary signal at the primary receiver.

2.3 Cooperative Transmission for Overlay Cognitive Radio

In the previous section, several techniques were described in order to facilitate the coexistence of primary and secondary signals in a cognitive radio spectrum. Of the techniques described,

the overlay of secondary signals requires SUs to relay the primary signal in order to compensate for the increased interference temperature. Such requirement is implemented using so-called physical layer cooperative transmission. The most important aspect of this technique is that SUs shall be able to relay the primary signal in such a way that its detectability at the primary receiver is increased. In this section we shall review the concept of physical layer cooperative transmission, and outline the most effective techniques.

The wireless channel is a shared medium, where a transmission intended to a particular user is overheard by many others. While this often creates unwanted interference, it also provides the transmitted signal to neighouring nodes *for free*. The concept of physical layer cooperation takes advantage of this property by reinforcing the direct transmission through additional transmissions from other nodes in the network which have overheard the original signal. In a multiuser network, such transmissions can take place along with each user's own transmission.

The foundations of such scheme lay in the information-theoretic relay channel model, as well as in the models described in Sendonaris et al. (2003), among others. In qualitative terms, the following benefits can be derived from physical layer cooperation:

- Spatial diversity: cooperating nodes provide antenna diversity in a similar fashion to multiple antenna terminals. Spatial diversity can be exploited to make the received signal more robust to channel impairments and decrease the outage probability

- Increased range: relaying has been typically used to extend the range of transmissions. In a cognitive radio environment, relaying can be used to extend the range of the primary signal.

- Increased availability: service availability is typically limited by shadowing of obstacles in the coverage area, such as buildings or mountains. Physical layer cooperation can be used to go around obstacles and therefore reduce the number and size of shadowed areas.

The system model for overlay cooperation is shown in Figure 2; the primary transmitter signal is detected by the SUs. These transmit two signals, one intended for the primary receiver and a second one intended to the secondary receiver.

2.3.1 Relaying Techniques

Relaying techniques have been thoroughly studied. Classically, relaying has been performed at the network layer, where packets are forwarded from one hop to the next according to the information in the routing tables. At the physical layer, relaying should be seen as a transmission technique focused on improving the end-to-end reliability of a wireless transmission involving multiple hops. Therefore, rather than being the transition from one hop to the next, relaying assumes that the receiver will decode the packet using both the information received directly from the source and the relayed signal. We shall focus on the simpler case of the parallel relay channel, where one or more relays communicate with source and destination, but not among them. The more complex case of serial relaying deserves a more elaborate treatment.

A first broad classification of relaying techniques is regenerative versus non-regenerative relaying. Regenerative techniques assume that the relay is able to decode the source signal, and re-process it in order to increase the effectiveness of relaying. Non-regenerative techniques assume that the relay is not able to retrieve any information about the received signal, and therefore the relaying operation is not able to distinguish between desired signal, noise,
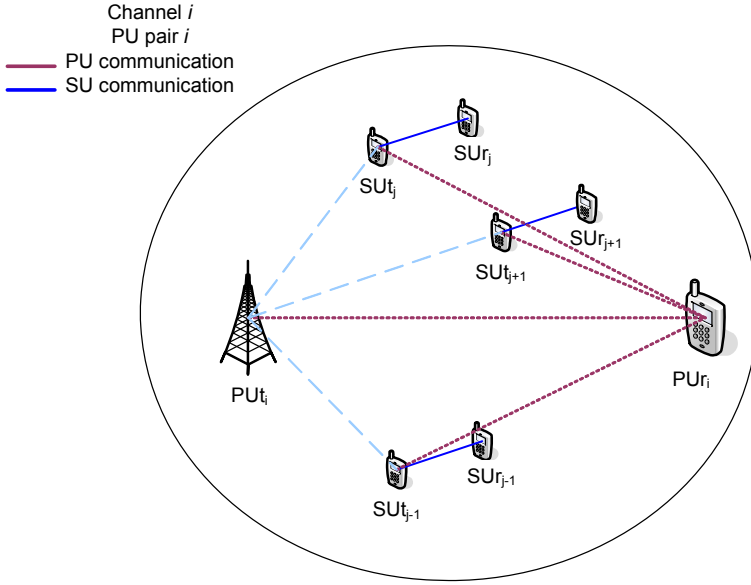
Fig. 2. Implementation of overlay cognitive radio through SU cooperation. SUs relay the primary signal and contribute to improve its reception at the primary receiver. In exchange, they are allowed to increase the interference temperature by communicating in that same band.

or interference. While non-regenerative relaying shall, in general, underperform regenerative techniques, it has the advantage of allowing more users to participate in the cooperative transmission, since they do not need to be able to decode the transmitted signal.

Another important aspect is the implementation of the cooperative scheme. We first shall distinguish between full-duplex relays, able to transmit and receive simultaneously in the same frequency band, and half-duplex relays. The former is difficult to implement in practice and more of an academic interest, therefore the latter shall be assumed hereafter. If half-duplex relaying is implemented, relays first listen to the source transmission and then occupy the channel to communicate with the destination. They may do so simultaneously (in a space-time coded signal or using beamforming), taking turns in a time division multiplexing scheme, or upon request, in an ARQ-like style. These implementations are shown in Figure 3.

In the following we address relaying techniques in more detail.

Decode and Forward: With this approach, the relay decodes the source signal and retrieves the transmitted data bits. These are then reencoded and transmitted to the destination. One of the main advantages of this technique is its flexibility. Relays may use different strategies when reencoding the signal, from regenerating the source signal to transmitting incremental redundancy. Moreover, they may choose different encoding schemes depending on the relay-to-destination channel. On the other hand, decode and forward requires the relay to
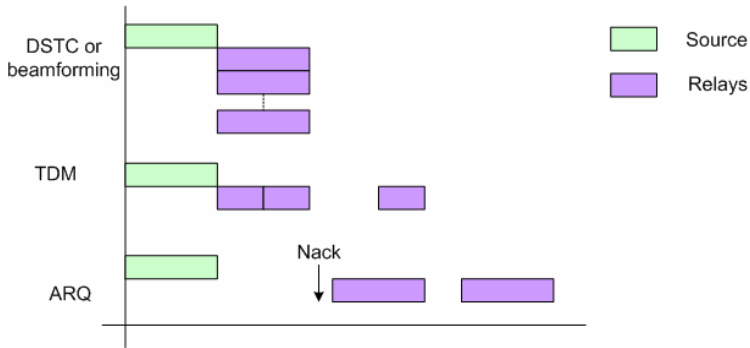
Fig. 3. Implementation schemes of parallel relaying: distributed space-time coding (DSTC) or beamforming, where relays transmit simultaneously; time-division multiplexing (TDM), where relays are organized in consecutive time slots; Automatic Repeat Request (ARQ), where relays transmit upon receiving a negative acknowledgment from the destination.

decode the source signal in order to cooperate with the transmission. In many instances, this is not possible and severely limits the number of potential cooperating users. As a result, the diversity gain of this technique is typically smaller than that of non-regenerative techniques.

Linear Relaying: In non-regenerative relaying, the relay takes the signal at its input and reproduces it at its output, after some processing. The main difference with decode and forward is that the relay is not able to recognize the desired signal from noise and interference. In linear relaying, the relay processing is limited to a linear operation. Amplify and forward constitutes the simplest form of linear relaying schemes, where the output is an amplificated version of the input. In more general case of linear relaying, the output is a linear function of past inputs. The linear relaying relay channel can be characterized as a multipath channel. However, unlike traditional multipath channels, the relays amplify noise and interference. With this technique, the beamforming or TDM schemes can be used (this type of relay would not understand a NACK from the transmitter, therefore the ARQ-like approach is not feasible). For the beamforming scheme, the relay should be able to retrieve phase information for the incoming signal, and then transmit coherently to the destination.

Compress and Forward: Compress and forward constitutes a non-linear approach to non-regenerative relaying. In this technique, relays quantize the signal received from the source, compress it, and transmit it to the destination with the appropriate channel coding. Either lossy or loss-less compression can be used. In this technique, relays transmit different data, therefore the TDM or ARQ schemes should be used. While this technique shares the advantage of linear relaying that all relays can cooperate, it is limited by the capacity of the relay-destination channel to transmit all the quantized data. This limitation becomes more severe as the number of relays increases.

Figure 4 del Coso & Ibars (2009) shows the performance of different relaying schemes as a function of the relative distance between source and relay.
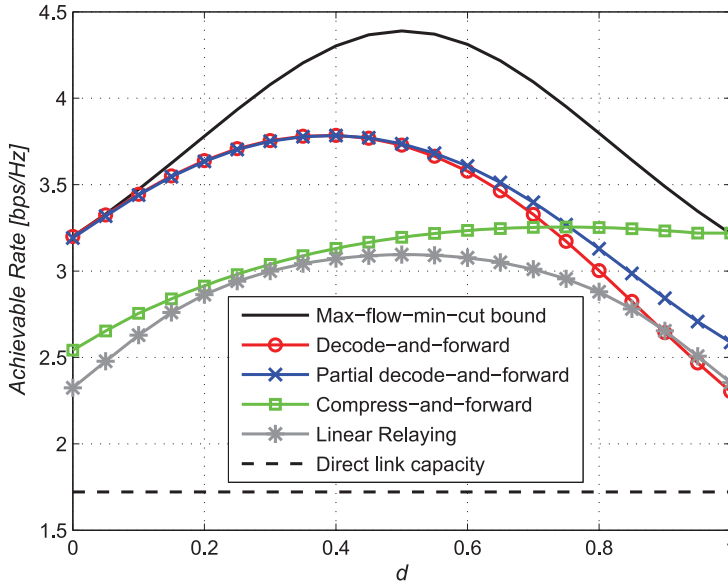
Fig. 4. Performance comparison of parallel relaying schemes, in a two-relay scenario where the source-relay distance varies from 0 to 1 and source-destination distance is fixed at 1: regenerative techniques are nearly optimal when the relays are close to the source, and non-regenerative techniques have an advantage when they are close to the destination. All techniques outperform the direct link capacity. A beamforming / DSTC scheme is assumed in all cases.

3. Game-theoretic Interference Management

The performance of cognitive radio networks are ultimately limited by interference, so that smart algorithms for power and channel control become key element in the network design. An additional difficulty arises because cognitive radios, independently make decisions about transmission power level and frequency channel, to maximize their own benefit. Such actions affect not only their performance, but also that of the entire network.

The cognitive radio network can be naturally modeled using a game theoretic framework, in which the players of the game are the radio terminals, their actions are their choices of transmission parameters (e.g. transmission powers, frequency channels, access probability, relay nodes), and their utilities are their defined performance measure that each radio tries to maximize. The players choose their actions independently, but their choice impacts on all the users in the network. The players are also assumed to be rational, i.e. they act in their best interest, to maximize their own utility. The underlay paradigm for cognitive radio, already introduced in the previous subsections, can be modeled by means of game theoretic approaches Nie & Comaniciu (8-11 Nov. 2005, Baltimore, USA). The main drawback of this approach is that the maximization of the game utility function represents an incentive to reduce the interference at the PU's receiver, but not a guarantee that the aggregated interference generated by the SUs is

maintained below a certain threshold, especially in scenarios where the spatial reuse is most challenging, e.g. where PU's receivers are passive or where SU's transmitters are very close to PU's receivers. In this context, cooperation of SUs and PUs (overlay approach) can significantly reduce the interference at the PU's receivers. As a result, in this section, after briefly introducing the basic concepts of game theory (3.1), we consider this discipline to model cooperation among SUs and PUs (3.2). Specifically, we present a game to model channel and power allocation for cooperative cognitive radios. We will consider both the hypothesis where complete and incomplete information about the other players' channel conditions and actions is available across the SUs. To solve the problem of incomplete information a Bayesian game has to be considered.

3.1 Game Theory

Game theory is a discipline to model interactive decision making processes. A significant amount of work in wireless communications is related with the use of game theory. A game consists of a finite set of N players. Each player $i \in N$ selects a strategy $s_i \in S_i$ with the objective of maximizing a utility u_i. The strategy profile s is the vector containing the strategies of all players: $\mathsf{s} = (s_i)_{i \in N} = (s_1, s_2, ..., s_n)$. It is customary to denote s_{-i} the collective strategies of all players except player i. The space of strategy profiles is defined as the Cartesian product of the individual strategy spaces: $\mathsf{S} = \times_{i \in N} S_i$. Finally, the utility function characterizes each player's sensitivity to everyone's actions. It is therefore a scalar valued function of the strategy profile: $u_i(s_i, s_{-i}) : \mathsf{S} \longrightarrow \Re$.

The most well known equilibrium concept in game theory is the Nash equilibrium. The Nash equilibrium is a joint strategy where no player can increase its utility by unilaterally deviating. That is:

Definition 1: A strategy profile $\mathsf{s} \in \mathsf{S}$ is a Nash equilibrium if $u_i(s_i, s_{-i}) \geq u_i(s_i', s_{-i}), \forall s_i' \in \mathsf{S}_i, \forall i \in N$.

An alternative interpretation of the definition of Nash equilibrium is that it is a mutual best response from each player to other players' strategies.

Definition 2: The best response $br(s_{-i})$ of player i to the profile of strategies s_{-i} is a strategy s_i such that: $br_i(s_{-i}) = argmax_{s_i \in S_i} u_i(s_i, s_{-i})$.

The first step towards solving a game is to investigate the *existence* of a Nash equilibrium. Once we have verified that a Nash equilibrium exists, we have to determine whether it is a *unique* equilibrium point. If the players have identified various Nash equilibria, it still might be difficult for them to coordinate on which one to choose. In case that there exists only one Nash equilibrium, we would be guaranteed that the player would play it. In order to have this favorable convergence characteristics, some mathematical properties have to be imposed on the utility functions. In particular, certain classes of games have shown to always converge to a pure Nash Equilibrium when a best response adaptive strategy is applied. An example of them is the class of Exact Potential Games. This kind of games is characterized by complete information about the utility functions across the multiple players. However, to model a broader class of real life situations, incomplete information should be considered by means of the so called bayesian potential games.

Finally, it is useful to identify a method to assess the efficiency of the reached equilibrium point. This method is based on the comparison of the strategy profiles using the concept of Pareto-optimality. To introduce this concept, we first define Pareto-superiority.

Definition 3: The strategy profile s is Pareto-superior to the strategy profile s' if for any player i: $u_i(s_i, s_{-i}) \geq u_i(s_i', s_{-i}')$, with strict inequality for at least one player. In other words, the

strategy profile s is *Pareto-superior* to the strategy profile s$'$, if the utility of a player i can be increased by changing from s to s$'$, without decreasing the utility of other players. Based on the concept of Pareto-superiority, we can identify the most efficient strategy profile.

Definition 4: The strategy profile s^p is Pareto-optimal if there exists no other strategy profile s$'$ that is Pareto-superior to s^p.

In a Pareto-optimal strategy profile, one cannot increase the utility of player i without decreasing the utility of at least one other player. Therefore, using the concept of Pareto-optimality, we can eliminate poor Nash equilibria by selecting those with a Pareto-superior strategy profile.

3.1.1 Potential games

Potential games were defined and discussed together with their properties in D. & B. (1996). A game $\Gamma = \{N, \{S_i\}_{i \in N}, \{u_i\}_{i \in N}\}$ is an exact potential game if there exists a function $Pot :$ S $\longrightarrow \Re$ such that, for all $i \in N, s_i, s_i' \in S_i$,

$$Pot(s_i, s_{-i}) - Pot(s_i', s_{-i}) = u(s_i, s_{-i}) - u(s_i', s_{-i}) \tag{1}$$

The function Pot is called *Exact Potential Function* of the game Γ. The potential function reflects the change in utility for any unilaterally deviating player. As a result, if Pot is an exact potential function of the game Γ, and s* $\in \{argmax_{s \in S} Pot(s)\}$ is a maximizer of the potential function, then s* is a Nash equilibrium of the game. In particular, the best reply dynamic converges to a Nash Equilibrium in a finite number of steps, regardless of the order of play and the initial condition of the game, as long as only one player acts at each time step, and the acting player maximizes its utility function, given the most recent actions of the other players. Notice that, for a finite exact potential game, i.e. characterized by a finite strategy space and a finite player set, the game has at least one pure-strategy Nash equilibrium. In addition to this, another significant property is that if the players follow a better response dynamic (whenever a player has an opportunity to revise its strategy, it will chose one characterized by a higher utility than the current one), all the repeated games where each stage is the same finite exact potential game and all players are myopic, converge to the Nash equilibria of the stage game. Considering the interesting properties of potential games, it would be useful to know how to recognize or design one of them. There exist some properties that might be helpful in recognizing these games.

Definition 5: A *coordination game* is a game in which, for any choice of pure strategies, all the players receive the same utility. That is, $u_i(s) = u_j(s), \forall i, j \in N, \forall s \in S$. As a result, there exists some function $V : S \longrightarrow \Re$ such that $u_i(s) = V(s) \forall i \in N, \forall s \in S$. All the maximizers of V are Nash equilibria, and at least one of them must be Pareto efficient.

Definition 6: A *dummy game* is a game in which each player's utility is a function of only the actions of other players, so that unilateral deviations do not produce any change in the utility of the deviating player: $u_i(s) = D_i(s_{-i}), \forall i \in N$.

Considering these two definitions, the result is that any exact potential game can be written as the sum of a coordination and a dummy game. That is, there exist functions $V : S \longrightarrow \Re$ and $D_i : S_{-i} \longrightarrow \Re$, such that $u_i(s) = V(s) + D_i(s_{-i}), \forall i \in N, \forall s \in S$.

As a result of that, one way of identifying an exact potential game is to try to separate the game into a coordination and a dummy game.

Finally, it is worth noting that for continuous and twice differentiable utility functions, a game is potential if and only if:

$$\frac{\partial P}{\partial s_i} = \frac{\partial u_i}{\partial s_i}, \forall i \in N \tag{2}$$

and

$$\frac{\partial^2 P}{\partial s_i \partial s_j} = \frac{\partial^2 u_i}{\partial s_i \partial s_j} = \frac{\partial^2 u_j}{\partial s_i \partial s_j}, \forall i, j \in N \tag{3}$$

3.1.2 Bayesian Potential games

When some players do not know the utility of the others, the game is said to have incomplete information. The case of perfect knowledge of utilities is a simplifying assumption that may be a good approximation in some cases. A game of incomplete information is defined as: $\Gamma = \{N, \{S_i\}_{i \in N}, \{\eta_i\}_{i \in N^+}, \{f_{H_i}(\eta_i)\}_{i \in N}, \{u_i\}_{i \in N}\}$, where, besides the players, the strategy space and the player's utility, we also have to define the player's type and its distribution probability. The player's type embodies any information that is not common knowledge to all players and is relevant to the players' decision making. This may include the player's utility function, his belief about other player's utility functions, etc. Each player is assumed to observe perfectly its type, but is unable to observe the types of its neighbors. As for the game with complete information, we need to find an equilibrium point from which no player has anything to gain by unilaterally deviating. In a Bayesian game, this point is a Bayesian Nash equilibrium, that is, a Bayesian Nash equilibrium is a Nash equilibrium of a Bayesian game. In particular, a strategy profile $\mathsf{s}^* = (s_1^*, ..., s_N^*)$ is a Bayesian Nash equilibrium if $s_i^*(\eta_i)$ solves (4), assuming that types of different players are independent.

$$s_i^*(\eta_i) \in \arg \max_{s_i \in S} \sum_{\eta_{-i}} f_H(\eta_{-i}) u_i(s_i, s_{-i}; \eta_i, \eta_{-i}) \tag{4}$$

As it is proven in Fuderberg & Tirole (1991), the existence of a Bayesian Nash equilibrium is an immediate consequence of the Nash existence theorem. As a result, considering that the potential games have shown to always converge to a Nash Equilibrium when a best response adaptive strategy is applied, it can be derived that for the Bayesian Potential game Γ there exists a Bayesian Nash equilibrium, which maximizes the expected utility function Facchini et al. (1997).

3.2 Game Theoretic Modeling of Cooperative Cognitive Radios

In this section we model joint channel and transmission power selection in a cognitive radio scenario as the output of a game where the players are the N SUs, the strategies are the choice of the transmission power and of the frequency channel, and the utility is a function of: (1) the interference each SU causes to the surrounding PUs and SUs simultaneously operating in the same frequency channel, (2) the interference each SU receives from the surrounding SUs simultaneously operating in the same frequency channel, (3) the satisfaction of each SU. The SUs are aware of the interference they receive, but to evaluate the interference they cause to the surrounding PUs and SUs, they need information about the wireless channel gains of their neighbors.

To retrieve this information, we consider two cases. In the first case, we foresee the existence of a CCC where all the users in the scenario share their transmission information, so that the decisions of the SUs are made with complete information. Much attention has recently been paid to this kind of channels, some examples are the Cognitive Pilot Channel (CPC) Perez-Romero et al. (17-20 April, Dublin, Ireland) proposed by the E2R2/E3 consortium, or

the radio enabler proposed by the P1900.4 Working Group. In the second case, taking into account that the hypothesis of the existence of a CCC has often been rejected in the cognitive radio literature, we provide a more realistic and feasible proposal by avoiding the need of the CCC and assuming that the decisions of the SUs are made with incomplete information.

3.2.1 System Model

The cognitive radio network we consider consists of M transmitting-receiving PUs pairs, and N transmitting-receiving SUs pairs. We will indicate the transmission power levels of the PUs' transmitters as $p_i^P, i = 1, ..., M$, and the transmission power levels of the SUs' transmitters as $p_j^S, j = 1, ..., N$. PUs and SUs, both transmitters and receivers, are randomly and uniformly distributed in a circular coverage region of a primary network with radius R_{max}. The SUs are in charge of sensing the channel conditions and of choosing a transmission scheme which does not disrupt the communication of the PUs. A SU distributively selects the frequency channel and the transmission power level to maximize its throughput while at the same time not causing harmful interference to the PUs. On the other hand, according to a cooperative paradigm, besides selecting the transmission power and the frequency channel, the SUs devote part of their transmission power to relaying the primary transmission. As a result, the SU's transmission power level is split in two parts, 1) a power level $p_j^{S'}, j = 1, ..., N$ for its own transmission, and 2) a cooperation power level $p_j^{S''}, j = 1, ..., N$ for relaying the PU's message on the selected band, where $p_j^S = p_j^{S'} + p_j^{S''}$. The cooperative scheme used by the SUs is shown in Fig. 5, where half-duplex operation is assumed. The PU transmission is divided into frames, and each frame further into slots. Relays are assumed to operate in half-duplex mode. Therefore, each relay listens to the primary transmission during one slot and transmits during the next. The relay will choose, as part of its strategy, whether to listen during even or odd slots. Decode & Forward relay mode is assumed.
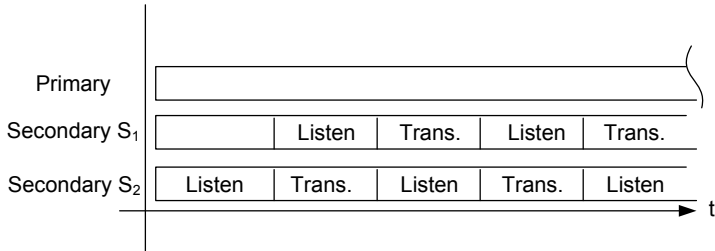
Fig. 5. Half-duplex relaying scheme for SUs. Each user chooses one slot to listen to the primary and retransmits in the following slot. Secondary users choose in which slot to transmit as a part of their strategy.

3.2.2 Game Model with complete information

Rather than relying on a network operator to decide on the power and channel allocation of the SUs, suppose that each user i is free to choose its own power allocation with the goal to

minimize the interference at the PUs and at the other SUs, and to maximize its own through-out. The resulting interaction among SUs and PUs leads to a non-cooperative game, which is defined as follows:

i) N is the finite set of players, i.e. the SUs.

ii) The strategies for player $i \in N$ are:

- a power level p_i^S in the set of power levels $P^S = (p_1^S, \cdots, p_m^S)$;

- the power level $p_i^{S'}$ that the player devotes to its own transmissions, in the set of power levels $P^{S'} = (p_1^{S'}, \cdots, p_q^{S'})$, where q is the order of set $P^{S'}$;

- the cooperative power level $p_i^{S''}$ that the player devotes to relaying a PU transmission and which is computed as $p_i^{S''} = p_i^S - p_i^{S'}$. The set of these power levels, $P^{S''}$, is the same as $P^{S'}$;

- a channel c_i in the set of channels $C = (c_1, \cdots, c_l)$.

- a slot subset sl_i from the two possible subsets $\mathcal{S}_1$ (even) and $\mathcal{S}_2$ (odd)

These strategies can be combined into a composite strategy $s_i = (p_i^S, p_i^{S'}, p_i^{S''}, c_i, sl_i) \in S_i$. We define $S = \times S_i, i \in N$ as the strategy space.

iii) The utility of each player i is defined as follows:

$$
\begin{aligned}
u(s_i, s_{-i}) &= -\sum_{j=1}^{M} p_i^{S'} h_{ij}^{SP} f(c_i, c_j) \\
&\quad -\sum_{j=1, j\neq i}^{N} p_j^{S'} h_{ji}^{SS} f(c_j, c_i) f'(sl_j, sl_i) - \sum_{j=1, j\neq i}^{N} p_i^{S'} h_{ij}^{SS} f(c_i, c_j) f'(sl_i, sl_j) \\
&\quad + b\log(1 + p_i^{S'} h_{ii}^{SS}) + \sum_{j=1}^{M} p_i^{S''} h_{ij}^{SP} f(c_i, c_j) f''(\gamma_i^{PS} > \rho)
\end{aligned}
\tag{5}
$$

where:

$$
f\left(c_i, c_j\right) \doteq \begin{cases} 1 & \text{if } c_i = c_j \\ 0 & \text{if } c_i \neq c_j \end{cases}
\tag{6}
$$

and

$$
f'\left(sl_i, sl_j\right) \doteq \begin{cases} 1 & \text{if } sl_i = sl_j \\ 0 & \text{if } sl_i \neq sl_j \end{cases}
\tag{7}
$$

In addition, in the Decode & Forward approach, the SU must be able to correctly decode the primary signal to relay it. In order to do that, the Signal to Noise and Interference Ratio (SINR) of the primary signal at SU i, which is given by

$$
\gamma_i^{PS} = \frac{p_j^P h_{ji}^{PS}}{\sigma^2 + \sum_{k\neq i}^{N} p_k^S h_{ki}^{SS} f(c_k, c_i) f'(sl_k, sl_i)}, \quad i = 1, \cdots, N
\tag{8}
$$

must be above the sensitivity threshold, ρ. We define the function

$$
f''\left(\gamma_i^{PS} > \rho\right) \doteq \begin{cases} 1 & \text{if } \gamma_i^{PS} > \rho \\ 0 & \text{otherwise} \end{cases}
\tag{9}
$$

It can be demonstrated that for a game with the utility function in 5, it can be found a potential function with the above described properties, so that the game is characterized by a pure Nash equilibrium.

3.2.3 Game Model with incomplete information

In a more realistic and feasible scenario, we should not rely on the existence of a CCC where SUs share their transmission information. As a result, we consider a situation where incomplete knowledge is available at the decision making agents. As a result, to model joint channel and transmission power selection for cognitive radios with incomplete information, we rely on the theory of Bayesian Potential game. The resulting game is defined as the one with complete information, but also the player's type have to be defined. For every $i \in N^+$, H_i is the finite set of possible types of player i, $\eta_i = (h_{1i}^{SS}, \cdots, h_{i-1i}^{SS}, h_{i+1i}^{SS}, \cdots, h_{Ni}^{SS}) \in H_i$, which includes the wireless channel gains of player i. Finally, also $f_{H_i}(\eta_i)$ have to be defined, being a probability distribution on H $= \times H_i, i = 1, \cdots, N$, with the a priori probability density function (PDF) on H defining the wireless channel gain PDF.

4. Conclusions

In this chapter, we have described how cooperation can benefit the different phases of the so called cognitive radio cycle. In particular we have focused on physical layer cooperation, showing that benefits can be obtained for both the primary and the secondary system in terms of spatial diversity, increased range and increased availability. In addition, we have modeled the critical interference management problem in a cooperative and cognitive system through a game theoretical approach, as well as providing design guidelines for games with good convergence characteristics, in both cases of complete and incomplete information.

5. References

D., M. & B., W. S. (1996). Potential games, *Games Econ. Behav.* **14**(1): 124–143.

del Coso, A. & Ibars, C. (2009). Achievable rates for the AWGN channel with multiple parallel relays, *IEEE Transactions on Wireless Communications, to appear* .

Facchini, G., Van Megen, F., Borm, P. & Tijs, S. (1997). Congestion Models and Weighted Bayesian Potential Games, *Theory and Decision (Kluwer Academic Publ.)* **42**: 193–206.

Fuderberg, D. & Tirole, J. (1991). *Game Theory*, MIT Press.

Ghasemi, A. & Sousa, E. (8-11 Nov. 2005, Baltimore, USA). Collaborative spectrum sensing for opportunistic access in fading environments, *Proceedings of IEEE Dyspan 2005*, pp. 269–278.

Mangold, S.; Jarosch, A. M. C. (2006). Operator assisted cognitive radio and dynamic spectrum assignment with dual beacons - detailed evaluation, *Proceedings First International Conference on Communication System Software and Middleware, (Comsware)*.

Nie, N. & Comaniciu, C. (8-11 Nov. 2005, Baltimore, USA). Adaptive channel allocation spectrum etiquette for cognitive radio networks, *Proceedings of IEEE Dyspan 2005*, pp. 269–278.

Perez-Romero, J., Sallent, O., Agusti, R. & Giupponi, L. (17-20 April, Dublin, Ireland). A Novel On-Demand Cognitive Pilot Channel enabling Dynamic Spectrum Allocation, *Proceedings of IEEE Dyspan 2007*.

Sendonaris, A., Erkip, E. & Aazhang, B. (2003). User cooperation diversity. part i. system description, *IEEE Transactions on Communications* **51**(11): 1927–1938.

Srinivasa, S.; Jafar, S. (2007). The throughput potential of cognitive radio: A theoretical perspective, *IEEE Communications Magazine* **45**(5): 73–79.

Cooperative Spectrum Sensing

Jinlong Wang, Qihui Wu, Xueqiang Zheng
and Juan Chen
Institute of Communications Engineering
PLA University of Science and Technology
China

1. Introduction

With the rapid growth of wireless applications and services in the recent decade, spectrum resources are facing huge demands. Traditionally, the frequency spectrum is licensed to users by government agencies in a rigid manner where the licensee has the exclusive right to access the allocated band. Recent measurements by Spectrum Policy Task Force (SPTF) within the FCC indicate that, the allocated spectrums are vastly underutilized sporadically and geographically with a high variance in time (McHenry, 2005). There are increasing interests in increasing the spectrum efficiency. The exciting findings shed light on the problem of spectrum scarcity and motivate a new direction to solve the conflicts between spectrum scarcity and spectrum under-utilization.

Cognitive radio (CR) enables much higher spectrum efficiency by dynamic spectrum access (Mitola & Maguire, 1999). Therefore, it is a potential technique for future wireless communications to mitigate the spectrum scarcity issue. CR implementations face many technical challenges, including spectrum sensing, dynamic frequency selection, adaptive modulation, and wideband frequency-agile RF front-end circuitry. These challenges are compounded by the inherent transmission impairments of wireless links, user mobility, non uniform legacy radio resource allocation policies, and user dependent economic considerations.

Obviously, spectrum sensing is a critical functionality of CR networks, it allows cognitive users to detect spectral holes and to opportunistically use under-utilized frequency bands without causing harmful interference to legacy systems.

In cognitive radio system, when cognitive users are sensing the channel, the sampled received signal of cognitive users has two hypotheses. Hypothesis H_1 denotes the primary user is active, and hypothesis H_0 denotes the primary user is inactive.

$$
\begin{aligned}
H_1 &: y(k) = h * s(k) + n(k) \\
H_0 &: y(k) = n(k)
\end{aligned}
\tag{1}
$$

where $y(k)$ is the signal received by cognitive user, $s(k)$ is the transmitted signal of the primary user. The signal $s(k)$ is distorted by the channel gain h, which is assumed to be constant during the detection interval, and is further corrupted by the zero-mean additive white Gaussian noise $n(k)$ with the variance σ_n^2. Without loss of generality, $s(k)$ and $n(k)$ are assumed to be independent of each other. $r = |h|^2 E_s / \sigma_n^2$ denotes the signal-to-noise ratio, where E_s is the signal energy of the primary user.

Given a single frequency band, the first challenge for CR is to reliably detect the existence of primary user to minimize the interference to existing communications. A number of different methods are proposed for identifying the presence of signal transmission, such as matched filter detection, energy detection, feature detection techniques and wavelet approach (Arslan & Yucek, 2007).

However, the hidden terminal problem occurs when cognitive user is shadowed, in severe multipath fading or inside buildings with high penetration loss while primary user is operating in the vicinity. Due to the hidden terminal problem, the sensing performance for one cognitive user will be degraded.

To prevent the hidden terminal problem, the CR network could fuse the sensing results of multiple cognitive users and exploit spatial diversity among distributed cognitive users to enhance the sensing reliability.

In this way, a network of spatially distributed cognitive users, which experience different channel conditions from the target, would have a better chance of detecting the primary user by exchanging sensing information. Therefore, collaborative spectrum sensing can alleviate the problem of corrupted detection by exploiting the built-in spatial diversity to reduce the probability of interfering with primary users.

Since collaborative sensing is generally coordinated over a reporting channel, efficient cooperation schemes should be investigated to reduce bandwidth and power requirements while optimizing the sensing reliability. Important design considerations include the overhead reduction associated with sensing information exchange and the feasibility issue of reporting channels. In general, operating characteristics (such as false alarm versus detection probabilities) of the detector should be selected by considering the achievable opportunistic throughput of cognitive users and the probability of no colliding with primary user.

2. Classification of Cooperative Spectrum Sensing

Cooperative sensing can be implemented in two fashions: centralized or distributed (Akyildiz et al, 2006). These two fashions will be explained in the following sections.

2.1 Centralized Sensing

In centralized sensing, a fusion centre collects sensing information from cognitive users. The fusion centre identifies the available spectrum, and broadcasts the fused result to other cognitive users or directly controls cognitive user's traffic. According to whether cognitive users exchange sensing information themselves, the centralized sensing fashion can also be divided to two categories.

A Partially Cooperative Network
Each of CR users detects the spectrum independently and directly transmits its sensing information to the fusion centre.
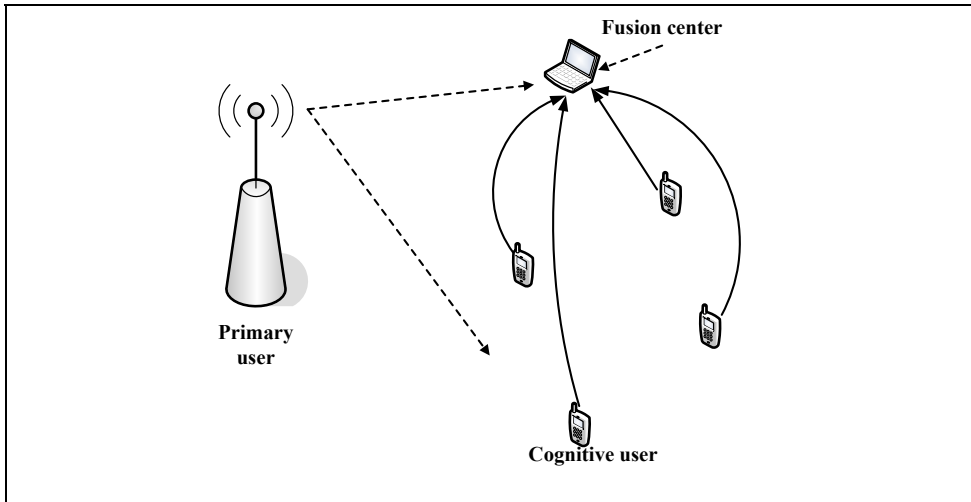

Fig. 1. Scheme of the partially cooperative network

B Totally Cooperative Network
Cognitive users cooperatively transmit each others sensing information, and then send sensing information to the fusion centre.
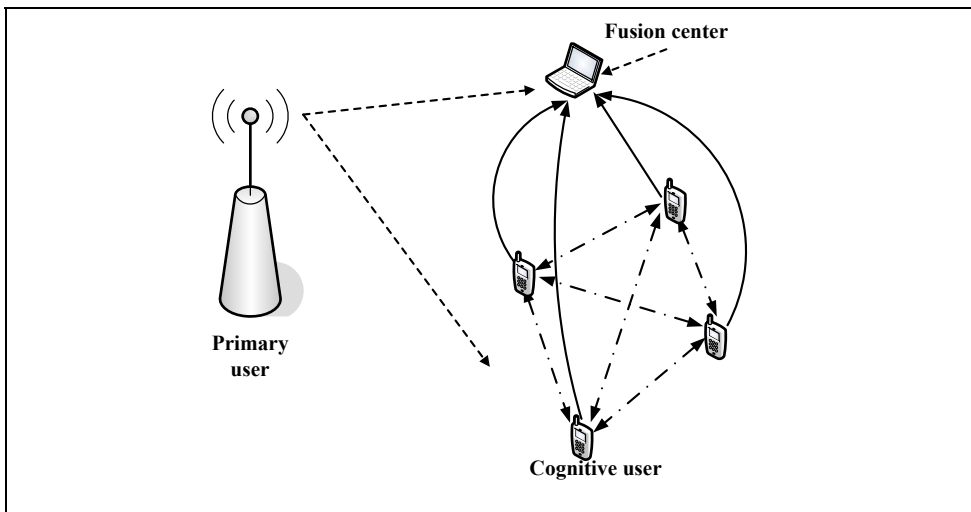

Fig. 2. Scheme of the totally cooperative network

2.2 Distributed Sensing

In the case of distributed sensing, cognitive users share information among each other but they make their own decisions as to which part of the spectrum they can use. Distributed sensing is more advantageous in the sense that there is no need for a backbone infrastructure.
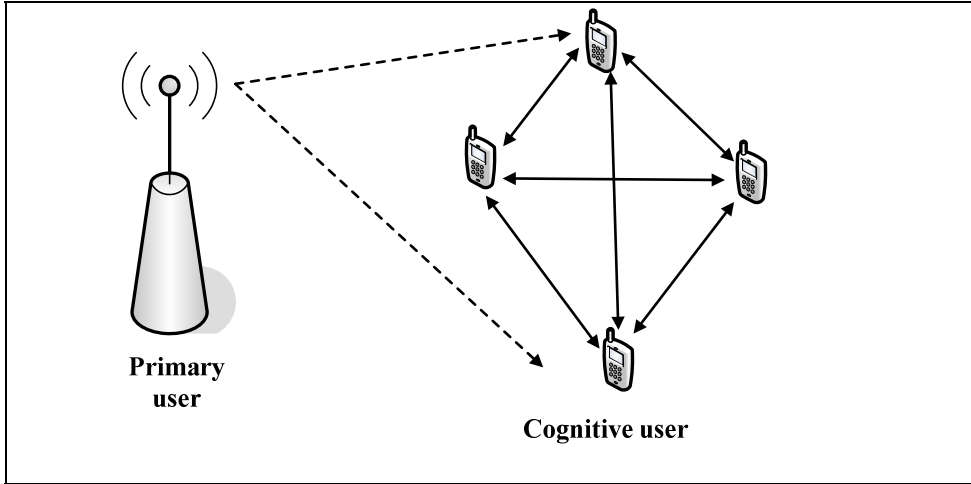
Fig. 3. Scheme of the distributed sensing

We focus on centralized sensing in the following sections.

3. The Fusion Rules

The cognitive users collaborate by reporting their local observations to the fusion centre, which makes the final decision on whether H_1 or H_0 is true. Due to bandwidth and other practical constraints, the cognitive users may report only quantized observations. Here, two classes of fusion algorithms, hard information-combining and soft information-combining algorithm, are considered. In the former, each cognitive user i performs a local hypothesis test and reports a binary variable B_i=1 if it believes H_1 is true, and B_i=0 otherwise. In the latter, each user reports the full observation to the fusion centre. The sensing performance of cooperative spectrum sensing is related to the local observations and the fusion rules.

3.1 Hard-decision Fusion Rules

One of the simplest suboptimal solutions to the data fusion problem is the Counting Rule (Aalo & Viswanathan, 1992) (also referred to as the Voting Rule), which counts the number of cognitive users that vote for the presence of the signal and compares it against a given threshold. Equivalently, the decision is based solely on the type of the received vector of bits. Based on voting rules, in (Ghasemi & Sousa, 2005), a fusion rule known as the OR logic operation was used to combine decisions from several cognitive users, the sensing performance is given as follows:

$$\begin{cases} P_{OR\_D} = 1 - \prod_{i=1}^{N}(1 - P_{Di}) \\ \\ P_{OR\_F} = 1 - \prod_{i=1}^{N}(1 - P_{Fi}) \end{cases} \tag{2}$$

where N is the number of cognitive users, P_Di and P_Fi are the probability of the detection and false alarm.

In (Visotsky et al, 2005), hard decision with the AND logic operation was proposed, the sensing performance is given as follows,

$$\begin{cases} P_{AND\_D} = \prod_{i=1}^{N} P_{Di} \\ \\ P_{AND\_F} = \prod_{i=1}^{N} P_{Fi} \end{cases} \tag{3}$$

A general suboptimal solution to the fusion problem was given in (Unnikrishnan & Veeravalli, 2008) when the signal received by cognitive users are correlated. This solution uses partial statistical knowledge and gives better performance than the one obtained by ignoring the correlation information completely.

An optimum data fusion structure has been derived in (Chair & Varshney, 1986) which combines the decisions from the individual detectors while minimizing the overall probability of error. Individual decisions are weighted according to their reliability, that is, the weights are a function of the probability of detection and false alarm of the individual user. The fusion and decision rule is as follow.

$$\alpha_0 + \sum_{i=1}^{N} \alpha_i u_i \underset{\substack{< \\ H_0}}{\overset{\substack{H_1 \\ >}}{}} 0 \tag{4}$$

where u_i is the local decision of the i th cognitive user, the weighting factor α_i is defined as follows.

$$\alpha_i = \begin{cases} \log \dfrac{P_1}{P_0}, & i = 0 \\ \\ \log \dfrac{P_{Di}}{P_{Fi}}, & u_i = +1 \\ \\ \log \dfrac{1 - P_{Fi}}{1 - P_{Di}}, & u_i = -1 \end{cases} \tag{5}$$

3.2 Soft-decision Fusion Rules

If the channel which cognitive users report the sensing results to the fusion centre is not limited by bandwidth, each cognitive user can report sensing information without any compression. Then the performance of collaborative spectrum sensing will be greatly enhanced. In (Visotsky et al., 2005), both soft information-combining strategy and hard information-combining strategy are introduced to detect the primary user. It was shown that the soft information-combining strategy has better detecting performance.

When cognitive users did not know any prior knowledge of the primary user, energy fusion method is the optimal fusion of the detection statistics. The test statistic for energy detector is $\theta = \sum_{k=1}^{2u} |y(k)|^2$, where u is the time-bandwidth product. Given an instantaneous r , θ follows the distribution (Visotsky et al., 2007).

$$f(\theta|r) \sim \begin{cases} x_{2u}^2 & H_0 \\ x_{2u}^2(2r) & H_1 \end{cases} \tag{6}$$

where r is exponentially distributed with the mean value $\bar{r}$, u is the time-bandwidth product of the energy detector, x_{2u}^2 represents a central chi-square distribution with $2u$ degrees of freedom, and $x_{2u}^2(2r)$ represents a non-central chi-square distribution with $2u$ degrees of freedom and a non-centrality parameter $2r$.

If cognitive users know the prior knowledge of the primary users, soft decision using the likelihood ratio test (LRT) is optimal. The LRT-based optimal fusion rule (Lim et al., 2008) involves a quadratic form as a result of high computational complexity. The closed expressions of detection probability and false alarm probability can not be derived.

The optimal linear framework for cooperative spectrum sensing in cognitive radio networks is proposed in (Quan et al., 2008). The proposed methods optimize the detection performance by operating over a linear combination of local test statistics from individual cognitive users, which combats the destructive channel effects between the target primary user and the opportunistic cognitive users.
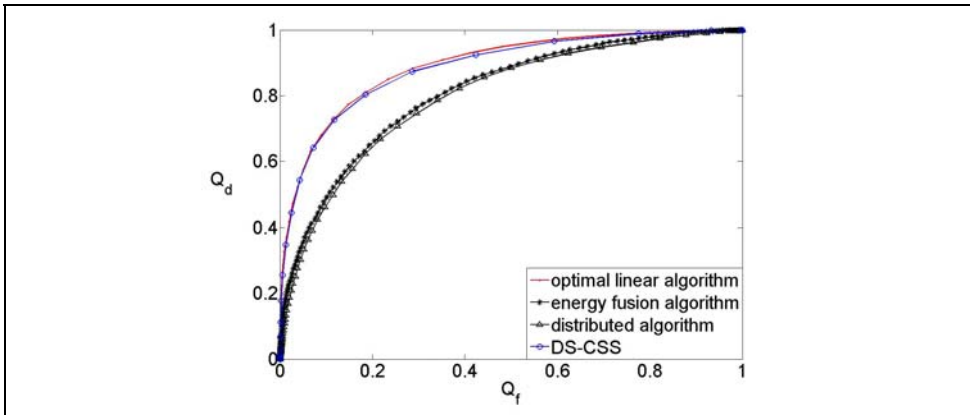

Fig. 4. ROC of DS-CSS

The optimal linear detection provides performance comparable to that achievable by the optimal LRT-based fusion in many situations. But the fusion centre needs to know the prior knowledge of the primary user and all cognitive users.

In order to efficiently detect the licensed users under prior knowledge constraints in the fusion centre, a novel cooperative spectrum sensing algorithm based on Dempster-Shafer theory (DS-CSS) is proposed in (Zheng et al, 2008). In this novel algorithm, the credibility and uncertainty are calculated based on the local sensing result of cognitive users, then fused at the fusion centre using Dempster-Shafer theory. Fig.4 shows that the performance of the novel algorithm is better than energy fusion method.

4. Network Overhead Reduction

Except the sensing performance, network overhead for cooperation is also important for the cooperative spectrum sensing. At present, there are mainly two ways to reduce the network overhead. One is to reduce the sensing information which cognitive users send to the fusion centre. Another is to reduce the number of reporting cognitive radios.

4.1 Compression of Sensing information

Usually, the first way to reduce the network overhead is to reduce the amount of information which each cognitive user transmits through the reporting channel, that is, the results of the local perception compression to achieve the purpose of network overhead reduction.

At present most of the hard-information combination algorithms fall into this category, such as the AND criteria, OR criteria K/M criteria and so on.
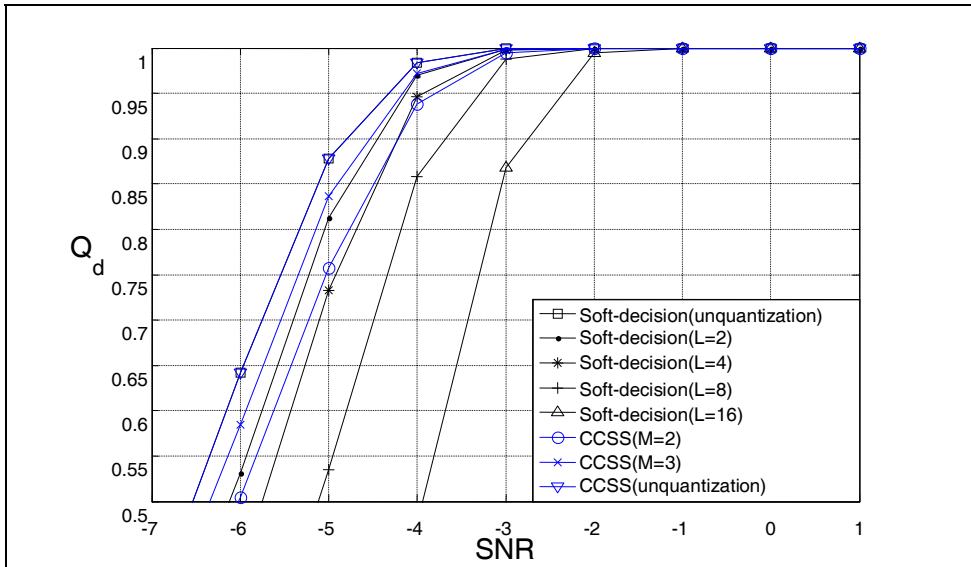

Fig. 5. The sensing performance of CCSS algorithm

Quantization of the local sensing information is another way to reduce the amount of the reporting information. The quantization in signal detection problems were studied in (Hashlamoun & Varshney, 1996) (Helstrom, 1988) (Chamberland & Veeravalli, 2004)(Blum, 1999). The influence of the number of quantization bits were studied in (Kaligineedi et al., 2008) (Taherpour et al., 2007). A new softened hard combination scheme with two-bit overhead for each user is proposed in (Ma et al., 2008), and this scheme can achieve a good tradeoff between sensing performance and complexity.

To reduce the network overhead of soft information-combining strategy with little expense of performance loss, a novel credible cooperative spectrum sensing (CCSS) algorithm was given in (Zheng et al., 2009). We considered the potential of using the reliability of cognitive users for cooperative spectrum sensing in cognitive radio systems. Analysis showed that it was possible to achieve high probabilities of detection as soft information-combining strategy by proper fusion of the cognitive users' decisions. In particular, the close-form expressions for probabilities of detection and false-alarm were derived for novel algorithm, and expression for average overhead used for cooperation was given. The threshold designing method for the algorithm was also discussed. Fig.5 and Fig.6 show that the novel algorithm can perform as well as soft information-combining strategy and its overhead was much less than soft information-combining strategy (W in Fig. 6).
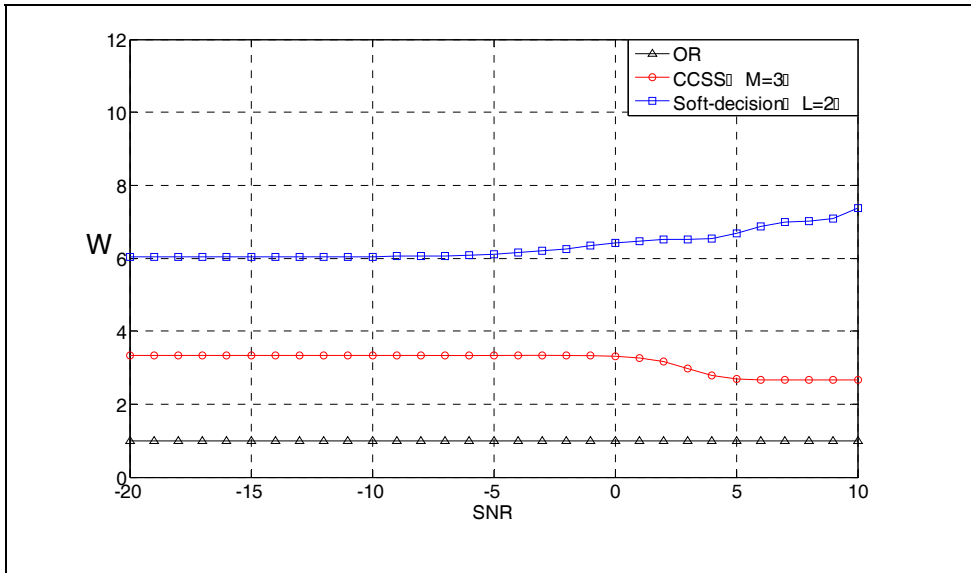

Fig. 6. The overhead of CCSS algorithm

4.2 sensor selection

Except compressing the sensing information of cognitive users, another method of network overhead reduction for cooperative spectrum sensing is to reduce the number of reporting cognitive users. For hard and soft information-combining algorithms, the sensor selection methods are different.

A Hard information-combining algorithms

Firstly, for hard information-combining algorithm, there are some topics about it. In (Edward & Liang, 2007), when cooperative spectrum sensing algorithms use AND and OR fusion schemes with energy detector, the optimum probabilities of detection and false alarm can be achieved by cooperating a certain number of users with the highest primary user's signal to noise ratio rather than cooperating all the users within the network. In (Zhang et al., 2008), the performance of cooperative spectrum sensing in cognitive radio networks have been studied. It has been found that the optimal decision fusion rule to minimize the total error probability is the half-voting rule. Moreover, the optimal detection threshold of energy detection is determined numerically. In particular, an efficient spectrum sensing algorithm has been proposed which requires fewer than the total number of cognitive radios in cooperative spectrum sensing while satisfying a given error bound.

In (Sun et al., 2008), a novel sensing user selection scheme is proposed. The optimal data fusion rule is used to simulate the sensing performance before and after the sensing nodes selection. Analysis and simulation results suggest that this method can reduce the number of sensing users remarkably and effectively without degrading the performance of sensing system obviously. Compared to the censoring method with quantization, this scheme can also reduce the performance deterioration when there are many cognitive users experience the same shadowing effects.

B Soft information-combining algorithms

Generally, when the fusion rule is soft information-combining rule, the sensing performance is best when all cognitive user in cognitive radio system participate cooperative spectrum sensing. However, the network overhead is the most in that situation. Considering the sensing performance and network overhead, the optimum number of cognitive users in collaborative spectrum sensing is derived in (Chen, 2008) for lognormal shadowing channels, static additive white Gaussian noise channels and Rayleigh fading channels, when the efficiency of resources usage is considered in the system design.

When there are N cognitive users, the network efficiency is

$$\eta(n) = \frac{N-n}{N} \tag{7}$$

The optimal number of cognitive radios is defined by weighting the sensing performance and the network efficiency in a target function

$$J(n) = (1-\alpha) \cdot P_d(n) + \alpha \cdot \eta(n) \tag{8}$$

where $J(n)$ is the target function, α is the impact factor, $P_d(n)$ is the probability of detection, where n is the number of cognitive users participate in cooperative sensing.

$$P_d(n) = P(H_0)(1-Q_f) + P(H_1)(1-Q_m) \tag{9}$$

Q_f and Q_m are the probability of false alarm and miss detection respectively.

$$\begin{cases} Q_f = P(H_1 \mid H_0) \\ Q_m = P(H_0 \mid H_1) \end{cases} \tag{10}$$

5. Sensing Information Transmission

Most present researches on cooperative spectrum sensing are based on the following assumptions: the reporting channel between the fusion centre and cognitive users is an ideal error-free transmission channel. But sometimes it is very difficult in practice to achieve the above-mentioned transmission conditions, the cognitive user's distance from the fusion centre may be far away, or there are obstacles in the transmission block, then the sensing information transmission are likely to be wrong.

When there exist reporting errors, the upper and lower limits of the false alarm probability was given in (Sun et al., 2007), but the solutions were not given. When the fusion rules are AND and OR logical operation, the probability of false alarm and miss detection are

$$\begin{cases} Q_f^{OR} = 1 - \prod_{i=1}^{N} \left[\left(1 - P_{f,i}\right)\left(1 - P_{e,i}\right) + P_{f,i} P_{e,i} \right] \\ Q_m^{OR} = \prod_{i=1}^{N} \left[P_{m,i}\left(1 - P_{e,i}\right) + \left(1 - P_{m,i}\right) P_{e,i} \right] \end{cases} \tag{11}$$

$$\begin{cases} Q_f^{AND} = \prod_{i=1}^{N} \left[\left(1 - P_{f,i}\right) P_{e,i} + P_{f,i}\left(1 - P_{e,i}\right) \right] \\ Q_m^{AND} = 1 - \prod_{i=1}^{N} \left[\left(1 - P_{m,i}\right)\left(1 - P_{e,i}\right) + P_{m,i} P_{e,i} \right] \end{cases} \tag{12}$$

Suppose that the local spectrum sensing conducted by cognitive user i results in $P_{f,i} = P_f$ and $P_{m,i} = P_m$, for all $i = 1, 2, \dots, K$, and that the probabilities of reporting errors are identical for all CRs $P_{e,i} = P_e$, then the probability of false alarm and miss detection can be bounded by

$$Q_f^{OR} \geq \overline{Q}_f \triangleq \lim_{P_f \to 0} Q_f = 1 - \left(1 - P_e\right)^N \approx N P_e \tag{13}$$

$$Q_m^{AND} \geq \overline{Q}_m \triangleq \lim_{P_m \to 0} Q_m = 1 - \left(1 - P_e\right)^N \approx N P_e \tag{14}$$

It has been shown that using multiple cognitive users may improve the detection probability over realistic sensing and reporting channels. But the performance is limited by the probability of reporting error P_e which is due to imperfect reporting channels.

In (Ganesan & Li, 2008), two schemes have been analyzed considering sensing and reporting at the same time. The schemes employ the Amplify-and-forward (AF) cooperation protocol to reduce the detection time. It was shown that cooperation between cognitive users increases the overall agility of the network.

Cooperative spectrum sensing taking into account the Decode-and-forward (DF) cooperation protocol on the decision reporting is studied in (Zhang & Letaief, 2008). A transmit diversity method is given which applies some existing space-time (ST) coding and space-frequency (SF) coding for multiple antennas systems to the CR network by viewing cognitive users as distributed antenna arrays. It was found that cooperative spectrum sensing with DF cooperation protocol could increase the detection probability with reporting errors.

Space-time coding can be used for cooperative spectrum sensing but the performance is constrained because the information exchange between cognitive users may not be correctly performed. Network coding (NC) admits a larger rate region than routing at polynomial complexity. We propose a novel cooperative spectrum sensing algorithm based on network coding (NC-CSS) by considering the use of network coding for such information exchanges. The simulation results show that the performance of the NC-CSS algorithm is better than the above cooperative spectrum sensing algorithm based on space-time coding (ST-CSS), and the novel algorithm can meet the situation when there exist destroyed reporting paths. When there exists one destroyed reporting path, cooperative spectrum sensing with network coding can achieve better sensing performance, as in Fig.7.
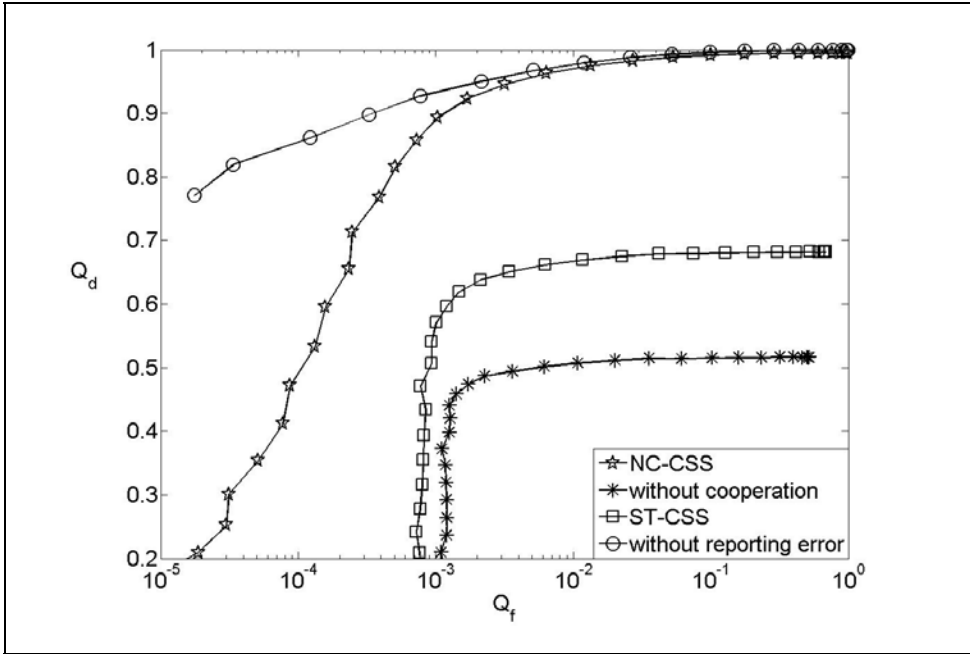

Fig. 7. ROC of NC-CSS algorithm

At present, there are still many issues unresolved in non-ideal reporting channel conditions, how to consider the synergy of the design spectrum sensing algorithm is particularly important.

6. Conclusion

With the increasing demand for radio spectrum on one hand and inefficient usage of the licensed bands on the other, cognitive radio can efficiently utilize the spectrum holes of the licensed channels in different locations and times. To detect the spectrum holes accurately and quickly, spectrum sensing is a critical component in cognitive radio systems. By allowing a number of cognitive users to perform local spectrum sensing independently and fusing their local decision results together at the fusion centre, the spectrum sensing performance can be greatly enhanced.

In this chapter we investigate the main issues associated with the design of cooperative spectrum sensing functionality for cognitive radio system. Sensing performance is decided by the local decisions and fusion rules in the fusion centre. The network overhead for cooperation is also important for cooperative spectrum sensing, a good cooperative spectrum sensing algorithm should have better sensing performance with less network overhead. Cooperative spectrum sensing algorithms using space-time coding and network coding are also discussed when the reporting channel is imperfect.

An important venue for further research is the design of cooperative spectrum sensing considering the positions and mobility of cognitive users. Further research on cooperative spectrum sensing can be envisioned on wideband sensing.

7. References

Aalo V & Viswanathan R.(1992) Asymptotic performance of a distributed detection system in correlated Gaussian noise, *IEEE TRANSACTIONS ON SIGNAL PROCESSING*, Vol. 40, pp. 211–213, 1992.

Akyildiz I F, Lee W Y, Vuran M C & Mohanty S.(2006) Next generation/dynamic spectrum access/cognitive radio wireless networks:a survey, *ElSEVIER COMPUTER NETWORKS JOURNAL*, Vol. 50, No. 13, pp. 2127–2159.

Arslan H & Yucek T.(2007) Spectrum Sensing for Cognitive Radio Applications, *Cognitive Radio, Software Defined Radio, and Adaptive Wireless Systems*, pp.263–289, Springer.

Blum R. (1999) Distributed Detection for Diversity Reception of Fading Signals in Noise, *IEEE TRANSACTIONS ON INFORMATION THEORY*, Vol. 45, No. 1, pp.158-164.

Chair Z & Varshney P K.(1986) Optimal Data Fusion in Multiple Sensor Detection Systems. *IEEE TRANSACTIONS ON AEROSPACE AND ELECTRONIC SYSTEMS*, Vol. AES-22, No. 98,pp. 98-101.

Chamberland J & Veeravalli V. (2004) The impact of fading on decentralized detection in power constrained wireless sensor networks, *ICASSP 2004*,pp.837-840.

Chen Y.(2008) Optimum Number of Secondary Users in Collaborative Spectrum Sensing Considering Resources Usage Efficiency. *IEEE COMMUNICATIONS LETTERS*, Vol. 12, No. 12, pp.877-879.

Digham F.F, Alouini M.S, and Simon M.K.(2007) On the Energy Detection of Unknown Signals Over Fading Channels, *IEEE TRANSACTIONS ON COMMUNICATIONS*, Vol.50, No.1, pp. 21-24.

Edward P & Liang Y C.(2007) Optimization for Cooperative Sensing in Cognitive Radio Networks, *WCNC 2007*, pp.27-32.

Ganesan G & Li Y.(2007) Cooperative Spectrum Sensing in Cognitive Radio, Part I: Two User Networks[J]. *IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS*, Vol.6, No.6, pp. 2204-2213.

Taherpour A, Nasiri-Kenari M, Jamshidi A.(2007) Efficient Cooperative Spectrum Sensing in Cognitive Radio Networks, *PIMRC2007.*

Ghasemi A & Sousa E.(2005) Collaborative Spectrum Sensing for Opportunistic Access in Fading Environments[C]. IEEE DySPAN 2005, pages: 131 – 136.(Ghasemi & Sousa, 2005)

Hashlamoun W. A & Varshney P K.(1999) Near-Optimum Quantization for Signal detection[J]. *IEEE TRANSACTIONS ON COMMUNICATIONS*, Vol. 44, No. 3, pp.294-297.

Helstrom C. (1988) Improved MuItilevel Quantization for Detection of Narrowband Signals[J]. *IEEE TRANSACTIONS ON AEROSFACE AND ELECTRONIC SYSTEMS* , Vol. 24, No. 2, Pp.141-147.

Kaligineedi P, Khabbazian M, Bhargava V.(2008) Secure Cooperative Sensing Techniques for Cognitive Radio Systems. *ICC 2008.*

Lim T J, Zhang R, Liang Y C & Zeng Y. (2008) GLRT-Based Spectrum Sensing for Cognitive Radio.*GLOBECOM 2008*.

Ma J, Zhao G & Li Y.(2008) Soft Combination and Detection for Cooperative Spectrum Sensing in Cognitive Radio Networks. *IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS*, Vol. 7, No. 11, pp.4502-4507.

McHenry M A.(2005) NSF Spectrum Occupancy Measurements Project Summary, shared spectrum co. report.

Mitola J & Maguire Q.(1999) Cognitive radio: Making software radios more personal, *IEEE PERSONAL COMMUNICATIONS*, Vol. 6, No. 4, pp. 13–18.

Quan Z, Cui S & Sayed A H. (2008) Optimal Linear Cooperation for Spectrum Sensing in Cognitive radio Networks. *IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING*, Vol. 2, No.1, pp.28-40.

Selen Y, Tullberg H, Kronander J.(2008) Sensor Selection for Cooperative Spectrum Sensing. *DYSPAN 2008*, pages: 1-11.

Sun C, Zhang W & Letaief K B.(2007) Cluster-Based Cooperative Spectrum Sensing in Cognitive Radio Systems. *ICC 2007*, pp.2511-2515.

Sun Y, Hu H, Liu F et al.(2008) Selection of Sensing Nodes in Cognitive Radio System Based on Correlation of Sensing Information. *CROWNCOM 2008*.

Unnikrishnan J & Veeravalli V.(2008) Cooperative Sensing for Primary Detection in Cognitive Radio. *IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING*, Vol. 2, No. 1, pp.18-27.

Visotsky E, Kuffner S & Peterson R.(2005) On collaborative detection of TV transmissions in support of dynamic spectrum sensing, *Proc. IEEE DYSPAN*, USA, pp. 338-345.

Zhang W & Letaief K B.(2008) Cooperative Spectrum Sensing with Transmit and Relay Diversity in Cognitive Radio Networks[J]. *IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS*, Vol.7, No.12, pp. 4761-4766.

Zhang W, Mallik R K & Letaief K B.(2008) Cooperative Spectrum Sensing Optimization in Cognitive Radio Networks. *ICC 2008*.

Zheng X, Wang J, Wu Q & Chen J.(2008) Cooperative Spectrum Sensing Algorithm based on Dempster-Shafer Theory, *ICCS 2008*, pp.218-221.

Zheng X, Wang J, Wu Q & Shen L. (2009) A Novel Cooperative Spectrum Sensing
     Algorithm in Cognitive Radio Systems, *JOURNAL OF COMMUNICATIONS AND
     NETWORKS*, Vol.11,No.2,Apirl,2009,pp:115-121. (Zheng et al., 2009)

Cyclostationary Approach to Signal Detection and Classification in CognitiveRadio Systems

Hao Hu

Beijing University of Posts and Telecommunications
P. R. China

1. Introduction

Cognitive radio (CR) is a newly emerging technology (Mitola & Maguire, 1999, Mitola, 2001), which has been recently proposed to implement some kind of intelligence to automatically sense, recognize, and make wise use of any available radio frequency spectrum. With the increasing demand for wireless application, access to available spectrum is becoming increasingly difficult. On the other hand, most licensed spectrum go unused most of the time according to the FCC's Spectrum Policy Task Force Report (FCC, 2003). In order to solve these problems, cognitive radio is proposed for sharing the licensed spectrum to unlicensed users without harmful interference to licensed system.

Spectrum sensing is a key element in cognitive radio system which enables the cognitive radio to share the spectrum in licensed bands by detecting temporarily unused spectral resources (Haykin, 2005, Ghasemi & Sousa, 2005). Recently, several spectrum sensing techniques have been explored for cognitive radios, such as matched filter detection, energy detection and cyclostationary feature detection (Akyildiz at el., 2006). Many signals used in communication systems exhibit periodicities of their second order statistical parameters due to the operations such as sampling, modulating, multiplexing and coding. These cyclostationary properties, which are named as spectral correlation features, can be used for spectrum sensing. Moreover, spectrum sensing can not be restricted to simply monitor the power in some frequency bands of interest but must include detection and identification in order to avoid interference (Fehske at el., 2005). Therefore, cyclostationary feature detection is undoubtedly a good solution for primary user signal detection and recognition. In this chapter, several well-known spectrum sensing techniques are reviewed first. A survey of signal detection and classification for cognitive radios combining the spectral correlation analysis and support vector machine (SVM) is given in Section 3. Several spectral coherence characteristic parameters which are sensitive with modulation types and insensitive with SNR variation are chosen via spectral correlation analysis. In order to give better performance of the SVM, an alignment based kernel selection method is proposed in Section 4, which is used to choose the best kernel function for the SVM with spectral coherence characteristic training samples. A simple cross-validation method is also introduced to

choose the most appropriate kernel parameters and penalty parameters for the SVM. The performance analysis of the proposed approach is given over both Gaussian channel and IEEE 802.22 WRAN channel in Section 5. Compared to the existing methods including the classifiers based on binary decision tree (BDT) and multilayer linear perceptron network (MLPN), the proposed approach is more effective in the case of low SNR and limited training numbers.

2. Overview of Spectrum Sensing in Cognitive Radio System

Interference due to a cognitive radio network is deemed harmful if it causes the signal-to-interference ratio (SIR) at any primary receiver to fall below a certain threshold, which supplied by the regulatory bodies. This threshold depends on the receiver's robustness toward interference and varies from one primary band or service to another. In order to achieve spectrum sharing, the sensitivity of cognitive radio should be higher than that of primary receiver. For example, in order to share the radio spectrum between cognitive users and TV users in IEEE 802.22 standard, the sensitivity of cognitive receiver should exceed the TV receiver without hidden terminal problem. To improve the sensing accuracy, an additional margin of 30dB - 40dB should be added to the detection threshold. Moreover, due to the dynamic characteristic of the radio environment and difference between primary users as well as unknown influence of interference, spectrum sensing has become a challenging problem in cognitive radio. Generally, the spectrum sensing techniques can be classified as transmitter detection, cooperative detection, and interference-based detection, as shown in Fig. 1. In this chapter, we focus on the transmitter detection which is commonly used in the practical system. Transmitter detection approach is based on the detection of the weak signal from a primary transmitter. To achieve dynamic spectrum sharing, the cognitive radio transmitter should have capability to determine if a signal from primary user is locally present in a certain spectrum. Basic hypothesis model for transmitter detection can be defined as follows

$$x(t) = \begin{cases} n(t) & H_0 \\ hs(t) + n(t) & H_1 \end{cases} \tag{1}$$

where $x(t)$ is the signal received by the cognitive user, $s(t)$ is the transmitted signal of the primary user, $n(t)$ is the additive white Gaussian noise and h is the amplitude gain of the channel. H_Q is a null hypothesis, which states that there is no licensed user signal in a certain spectrum band. On the other hand, H_a is an alternative hypothesis, which indicates that there exists some licensed user signals.

Three approaches exist for transmitter detection, based on the sensing users' knowledge on the transmitted signals, which are illustrated in Fig. 1.
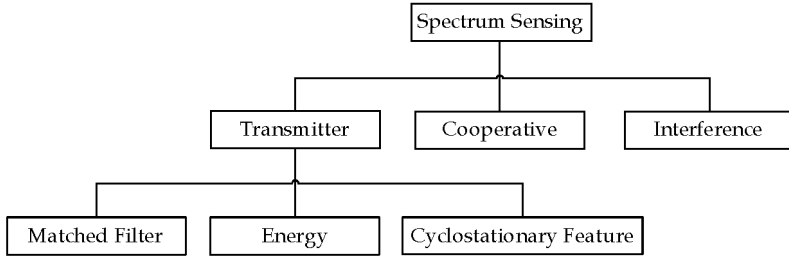
Fig. 1. Spectrum Sensing Techniques.

2.1 Matched Filter Detection

When the information of the primary user signal is well-known to the cognitive radio, the matched filter is the optimal linear filter for maximizing the signal to noise ratio (SNR) in the presence of additive Gaussian noise (Sahai at el., 2004, Akyildiz at el., 2006). In cognitive radio system, the matched filter is obtained by correlating a known signal, or primary user signal template, with an unknown signal to detect the presence of primary user signal in the unknown signal. This is equivalent to convolving the unknown signal with a time-reversed version of the template. According to the hypothesis model for transmitter detection, the received signal x(n) can be expressed as

$$x(n) = \begin{cases} w(n) & H_0 \\ \sqrt{1-\eta}s(n) + \sqrt{\eta}s_c(n) + w(n) & H_1 \end{cases} \tag{2}$$

where s(n) and $s_c(n)$ are original signal and pilot signal transmitted by primary user, which are orthogonal. η is the ratio of power allocated to pilot signal. The correlation function of received signal x(n) and unit vector of pilot signal $\hat{s}_c(n)$ is given by

$$T(x) = \frac{1}{N}\sum_{n=1}^{N} x(n)\hat{s}_c(n) \tag{3}$$

The Equation (3) can also be expressed as

$$T(x) = \begin{cases} \dfrac{\sum_{n=1}^{N} w(n)\hat{s}_c(n)}{N} & \sim N(0, \sigma^2/N) \quad H_0 \\[20pt] \dfrac{\sum_{n=1}^{N}\left\{\sqrt{1-\eta}s(n) + \sqrt{\eta}s_c(n) + w(n)\right\}\hat{s}_c(n)}{N} & \sim N(\sqrt{\theta P}, \sigma^2/N) \quad H_1 \end{cases} \tag{4}$$

where P is the power of the pilot signal, and a^2 is noise variance. Then the decision function can be given by

$$T(x) \underset{H_0}{\overset{H_1}{\underset{<}{>}}} \gamma \tag{5}$$

where the threshold y is determined by the probability of false alarm P_{FA}. The probability of false alarm P_{FA} and detection P_D are given as follows

$$P_{FA} = P\left(T(x) > \gamma \,|\, H_0\right) = Q\left(\gamma \middle/ \sqrt{\frac{\sigma^2}{N}}\right) \tag{6}$$

$$P_D = P\left(T(x) > \gamma \,|\, H_1\right) = Q\left(\gamma - \sqrt{\theta P} \middle/ \sqrt{\frac{\sigma^2}{N}}\right) \tag{7}$$

where $Q(\cdot)$ is the generalized Marcum Q-function, which can be defined by

$$Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \tag{8}$$

Combining Equation (6) and Equation (7), we can obtain N by eliminating y .

$$N = \left[Q^{-1}\left(P_D\right) - Q^{-1}\left(P_{FA}\right)\right]^2 \theta^{-1} SNR^{-1} \tag{9}$$

According to the Equation (9), O (1/ SNR) samples are required to meet a probability of error constraint, thus the main advantage of the matched filter is that it requires less time to achieve high processing gain. However, as a coherent detection method, matched filter detection requires a priori knowledge of the primary user signal such as the modulation type and order, the pulse shape, and the packet format. Although most of the priori knowledge can be obtained from pilot, preambles, synchronization word or spreading codes of the primary user network systems, an obvious shortcoming is that the cognitive radio user requires specific receivers for different types of primary user signal.

2.2 Energy Detection

If the receiver cannot gather sufficient information about the primary user signal, for example, if the power of the random Gaussian noise is only known to the receiver, the optimal detector is an energy detector (Akyildiz at el., 2006, Digham at el., 2003). In order to measure the energy of the received signal, the output signal of bandpass filter with bandwidth W is squared and integrated over the observation interval T. Finally, the output of the integrator, x(t), is compared with a threshold, y , to decide whether a licensed user is present or not.

$$s(t) \rightarrow \boxed{\text{Bandpass Filter}} \rightarrow \boxed{(\bullet)^2} \rightarrow \boxed{\frac{1}{T}\int_t^{t+T}(\bullet)\,dx} \rightarrow x(t)$$
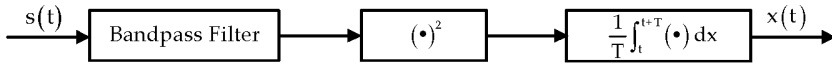
Fig. 2. The principal of energy detection.

According to the basic hypothesis model for transmitter detection, the probability of false alarm P_{FA} and detection P_D are given as follows

$$P_{FA} = P\{X > \gamma | H_1\} = Q\left(\sqrt{2\delta}, \sqrt{\gamma}\right) \tag{10}$$

$$P_D = P\{X > \gamma | H_0\} = \frac{\Gamma\left(m, 2/\gamma\right)}{\Gamma\left(m\right)} \tag{11}$$

Since it is easy to implement, the recent work on detection of the primary user has generally adopted the energy detection. Although this method can be implemented without any prior knowledge of the primar$_2$y user signal, it also has some drawbacks. Since energy detection is non-coherent, $O(1/SNR)$ samples are required to meet a probability of error constraint. Moreover, the threshold selection for energy detection is highly susceptible to uncertainty in background noise and interference, and it can only determine the presence of the signal without differentiating signal types. However, the largest advantage of energy detection is simple and low complexity.

2.3 Cyclostationary Feature Detection
The cyclostationary feature is intentionally embedded in the physical properties of a communication signal, which may be easily generated, manipulated, detected and analyzed using low complexity transceiver architectures. This feature is present in all transmitted signals, requires little signalling overhead and may be detected using short signal observation times, and thus it can be used for primary user signal detection and recognition. Recent research efforts exploit the cyclostationary features of signals via spectral correlation analysis as a method for spectrum sensing (Digham at el., 2003, Sahai at el., 2004, Ghasemi & Sousa, 2005), which has been found to be superior to simple energy detection and matched filtering. Energy detection can only detect whether or not a signal in present and utilizing a matched filter system requires extensive knowledge about the channel and signals that are to be identified. The method, which is not susceptible to in-band interference, can be used to detect and classify different types of signal. Conventional signal classification approaches are mainly based on decision theory (Polydoros & Kim, 1990, Sapiano & Martin, 1996, Sills, 1999, Wei & Mendel, 2000, Hang at el., 2001) and statistical pattern recognition (Nandi & Azzouz, 1995, Azzouz & Nandi, 1995, Azzouz & Nandi, 1996, Nandi & Azzouz, 1998). In the work by Hang (Hang at el., 2001), a binary decision tree is designed to signal classification. However, it's difficult to obtain the decision thresholds and rules, which needs a large amount of calculation. For more efficient and reliable performance, a novel approach based on multilayer linear perceptron network for signal classification in cognitive radio is studied by Fehske (Fehske at el., 2005). Support vector machine (SVM) is a new statistical pattern recognition approach, which is based on structural risk minimization principle (Vapnik,

1995). Compared with the conversional methods based on empirical risk minimization like artificial neural network (ANN), it has been found to give better generalization and better performance for small training examples. In the next section, a novel approach of signal classification for cognitive radios combining the cyclostationary features and SVM is proposed.

3. Spectrum Sensing based on spectral correlation analysis and SVM

3.1 System Framework

As our signal classification scheme combining spectral correlation analysis and SVM is based on statistical pattern recognition, which mainly consists of three modules (Han, 2003): feature extraction, classifier design and classification decision. Feature extraction is typically the first stage in any classification system in general, and in our spectrum sensing systems in particular. Given signal set to be classified, the feature parameters of different classes of signal and rules for classifier should be determined first. In order to achieve better classification performance, selected feature parameter should be insensitive with the SNR variation, and then a proper classifier is designed for specific classification problem using training data with known signal types. When the error probability of the classifier achieves a specific threshold, the classifier can be used for signal classification and recognition. In our scheme, there are three procedures adopted for primary user signal recognition:

1) Pro-processing procedure:

Several feature parameters are extracted via spectral correlation analysis first. The feature parameters insensitive with the SNR variation are selected as feature vector $x_1, x_2, \cdots, x_M$.

After that, the feature vector and signal type y_k are used to form the training set $(x_1, x_2, \cdots, x_M, y_k)$ for classifier.

2) Training and learning procedure:

The SVM classifier is trained using selected feature parameters in the training set. By utilizing a nonlinear SVM, an amount of calculation for training is performed offline, thus the computational complexity is reduced. The optimal classification plane for SVM is obtained in this procedure via training and learning.

3) Test procedure

Selected feature parameters extracted for received signals are inputted well-trained SVM classifier for primary user signal detection and recognition.

The framework of our scheme combining spectral correlation analysis and SVM is shown in Fig. 3.
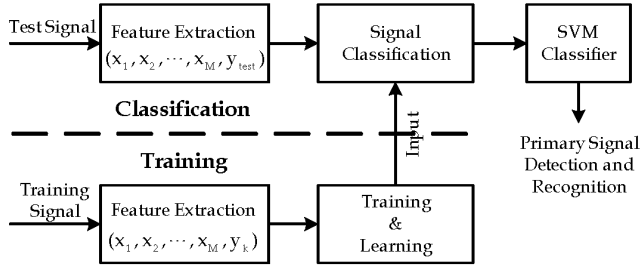
Fig. 3. System framework.

Although SVM is a better choice for the classifier, the selection of feature parameters has direct impact on the performance of the classification algorithm. In the next section, we will discuss the first step, how to choose spectral coherence characteristic parameters for our scheme.

3.2 Spectral Correlation Analysis

Many signals used in communication systems exhibit periodicities of their second order statistical parameters due to the operations such as sampling, modulating, multiplexing and coding. These cyclostationary properties, which are named as spectral correlation features, can be used for signal detection and recognition (Gardner, 1987).

In order to analyze the cyclostationary features of the signal x(t), two key functions are typically utilized. The cyclic autocorrelation function (CAF) is used for time domain analysis, which can be expressed as

$$
\begin{aligned}
R_x^\alpha(\tau) &= \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t+\frac{\tau}{2})\, x^*(t-\frac{\tau}{2})\, e^{-j2\pi\alpha t} dt \\
&= \left\langle x(t+\frac{\tau}{2})\, x^*(t-\frac{\tau}{2})\, e^{-j2\pi\alpha t} \right\rangle
\end{aligned}
\tag{12}
$$

The spectral correlation function (SCF), which exhibits the spectral correlation of the signal x(t), is obtained from the Fourier transform of the cyclic autocorrelation in Equation (12) (Gardner & Franks, 1975).

$$
\begin{aligned}
S_x^\alpha(f) &= \int_{-\infty}^{+\infty} R_x^\alpha(\tau)\, e^{-j2\pi f\tau}\, d\tau \\
&= \int_{-\infty}^{+\infty} \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t+\frac{\tau}{2})\, x^*(t-\frac{\tau}{2})\, e^{-j2\pi\alpha t} e^{-j2\pi f\tau} dt d\tau \\
&= \lim_{T\to+\infty} S_{xT}^\alpha(t,f)
\end{aligned}
\tag{13}
$$

Where α is the cycle frequency, f is the spectral frequency and $S_{xT}^{\alpha}(t,f)$ is the cyclic periodogram of $S_x^{\alpha}(f)$

$$S_{xT}^{\alpha}(t,f) = \left[X_T(t, f + \alpha/2) \cdot X_T^*(t, f - \alpha/2) \right] \Big/ T \qquad (14)$$

where the Fourier transform of the function x(u) on the bounded time interval $[t-T/2, t + T/2]$ is definedas

$$X_T(t,f) = \int_{t-T/2}^{t+T/2} x(u)\, e^{-j2\pi fu}\, du \qquad (15)$$

The correlation coefficient for the SCF between frequency components $f \pm \alpha/2$, which is known as spectral coherence coefficient (SCC), can be calculated by

$$C_x^{\alpha}(f) = \frac{S_x^{\alpha}(f)}{\sqrt{S_x^0(f + \alpha/2) \cdot S_x^0(f - \alpha/2)}} \qquad (16)$$

The magnitude of the SCC ranges from 0 to 1 with $a = 0$ for all f. Different signal classes (*i.e.* AM, ASK, FSK, PSK, MSK, QPSK) can be distinguished based on several characteristic parameters of SCF and SCC.

In practical situations, however, the number of observation samples at the sensor is limited. Therefore, the spectral correlation function needs to be estimated from a finite set of samples. In general, two methods are used for spectral correlation estimation including time-domain averaging and frequency-domain smoothing (Gardner & Spooner, 1988). In this section, the frequency-smoothing method is used for spectral correlation estimation, which can be expressed as follows.

$$\begin{cases} S_x^{\alpha}(f)_{\Delta f} = \dfrac{1}{M} \displaystyle\sum_{n=(M-1)/2}^{(M+1)/2} \dfrac{1}{\Delta t} X_{\Delta t}\left(t, f + \dfrac{\alpha}{2} + nF_S\right) \cdot X_{\Delta t}^*\left(t, f - \dfrac{\alpha}{2} + nF_S\right) \\[4mm] X_{\Delta t}(t,f) = \displaystyle\sum_{k=0}^{N-1} W_{\Delta t}(kT_S) \cdot X(t - kT_S) e^{-j2\pi f(t - kT_S)} \end{cases} \qquad (17)$$

where $F_S = 1/(N-1)T_S$ T_S is the frequency increment, T_S is the cyclic sampling interval and the length of the sample is $N = \Delta t / T_S$. Thus, the data-tapering window $W_{\Delta t}(kT_S)$ is of the width $\Delta f = M \cdot F_S$. The Spectral correlation functions of some typical signals can be obtained by frequency-smoothing method, which are shown in Fig. 4 (Gurman at el., 2008).
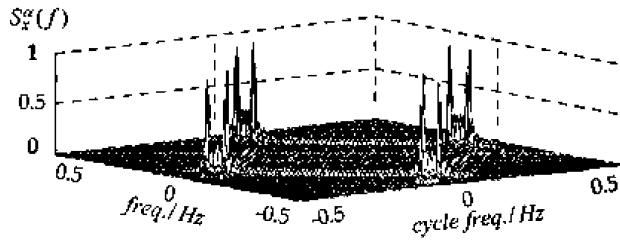
(a) FSK
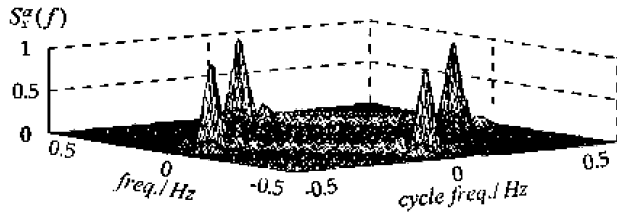
(b) PSK
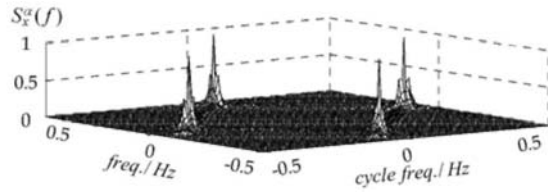
(c) FM
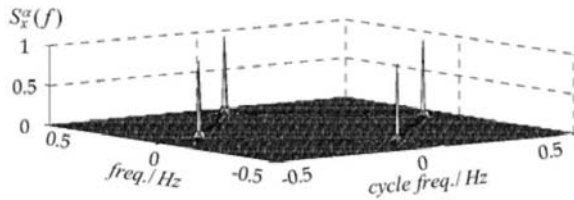
(d) AM

Fig. 4. Spectral correlation functions of some typical signal classes.

Even visually in the above figures, the spectral correlation functions of the different modulation types possess distinct characteristics. It is this fact that allows the successful application of the pattern recognition techniques to achieve primary user signal detection and recognition. In order to obtain better robustness of the proposed algorithm, some features less sensitive with SNR should be chosen for the classifier. Assumed that the

received signal s = s + n , where s , **n** are the transmitted signal and additional white Gaussian noise. The feature parameter of the received signal has a better classification where performance which is insensitive with SNR variation, if it satisfies

$$\left\| \hat{\mathbf{x}} - \mathbf{x} \right\| \leq \left\| \mathbf{n} \right\|_2 \tag{18}$$

X, x, n is the feature vectors of s, s, n .

Based on the calculation of the spectral correlation function, we can obtain the spectral correlation magnitude surface of different types of signal. According to above analysis, several spectral correlation features can be extracted for distinction of different modulation types. Typically, four key features x_1, x_2, x_3, x_4 which are sensitive with modulation types and insensitive with SNR variation are listed in Table. 1.

Feature	Signal Type					
	AM	ASK	FSK	PSK	MSK	QPSK
x_1	2	2	4	0	0	0
x_2	1	≥3	2	≥3	≥2	≥2
x_3	1	1	/	1	0	0
x_4	/	/	/	≥6	<6	<6

Table 1. Typical value of spectral correlation features. Four key features x_1, x_2, x_3, x_4 are described as follows.

- x_1: Number of 5 pulse on f - domain of SCF
  Let a = 0 in Equation (17), the SCF is transformed into $S''(f)$, thus x_1 can be obtained from the ichnography of $S^0_x(f)$.
- x_2: Number of cyclic spectral line on a -domain of SCF
  Let f = 0 in Equation (17), the SCF is transformed into $S^a (0)$, and x_2 can be obtained from the ichnography of $S^a (0)$.
- x_3: Average energy of cyclic spectral line on a - domain of SCF
  The Average energy of cyclic spectral line on a - domain of SCF can be computed by the equation as follows:

$$x_3 = \int_0^{+\infty} \left| S_x^\alpha (f) \right|^2 d\alpha \tag{19}$$

- x_4: Maximum value of SCC
  The spectral coherence coefficient can be obtained via Equation (16) and Equation (17). Then, the maximum value of SCC is computed as key feature x_4 .

In the simulation experiments, 4000 features are extracted from the signal for every trial. In order to prevent numerical computational errors, the features need to be normalized by

subtracting mean of each feature from the original feature and dividing the result by the standard deviation of the same feature.

$$x'_i = (x_i - \bar{x}_i)\big/\delta_{x_i} \qquad i = 1,2,3,4 \tag{20}$$

After normalization, the feature vector $x = (x'_1, x'_2, x'_3, x'_4)$ can be generated as the input of the signal classifier.

3.3 Support Vector Machine

The traditional statistical theory is primarily based on the asymptotic principle, which provides conclusion only for the situation where the sample size is tending to infinity. However, in most practical applications, the samples are usually limited so that it is difficult to achieve the desired results via existing methods. Statistical Learning Theory is a novel statistical theory based on small sample statistics by Vapnik (Vapnik, 1995). Compared to the conventional statistical theory, statistical learning theory mainly concerns the statistic principles when samples are limited, especially the properties of learning procedure in such cases. Statistical learning theory provides us a new framework for the general learning problem, which not only considers the asymptotic performance but obtains the optimal results under the condition of limited information. In order to study the generalization performance and the speed of uniform convergence, a series of indicators used to evaluate the learning performance of function sets are defined in statistical learning theory. One of the most important concepts is Vapnik-Chervonenkis(VC) dimension which was originally defined by Vladimir Vapnik and Alexey Chervonenkis in 1971. VC-dimension is a measure of learning machine complexity or the capacity of a statistical learning algorithm, which is the cardinality of the largest set of points that the algorithm can shatter. The learning machine is more complex with a greater VC-dimension. Statistical learning theory provides a novel strategy that balances the empirical risk and confidence interval. A nested subset sequence is chosen from the given set of functions according to the size of the VC dimension. For a given subset, the minimal value of the empirical risk can be obtained as the minimal true risk, which is illustrated in Fig. 5. This method is named as Structural risk minimization (SRM) which was also coined by Vapnik and Chervonekis in 1974. The principle of SRM is to provide a method to reach the trade-off between hypothesis space complexity (the VC dimension of approximating functions) and the quality of fitting the training data (Wang, 2007). The procedure is described in detail as follows.

Assumed a function set $Q(x, \alpha), \alpha \in \Lambda$ has a set S, which has a structure shown in Fig. 5. The structure is defined by a nested subset sequence $S_1, S_2, \cdots, S_n$, which satisfies

$$S_1 \subset S_2 \subset \cdots S_n \subset \cdots \tag{21}$$

The elements of the above structure have two properties as follows.

(1) The VC dimension of each subset h_k is limited and satisfies

$$h_1 \le h_2 \le \cdots h_n \le \cdots \tag{22}$$

(2)   Any element in the structure S_k contains a set of totally bounded function

$$0 \le Q(z, \alpha) \le B_k , \quad \alpha \subset \Lambda_k \tag{23}$$

or contains a function set which satisfies the following inequality for some (p, τ_K)

$$\sup_{\alpha \subset \Lambda_k} \frac{\left(\int Q^P(z, \alpha) dF(z)\right)^{1/p}}{\int Q(z, \alpha) dF(z)} \le \tau_k , \quad p > 2 \tag{24}$$

For a given set of observation set $z_1, z_2, \cdots, z_l$, SRM aims to choose the proper function $Q(z, \alpha_l^k)$ who makes the empirical risk minimal in subset S_k with smallest guaranteed risk. After performing empirical risk minimization on each subset, the subset whose sum of empirical risk and VC confidence is selected. The optimal function $Q(z, \alpha_l^k)$ that makes the empirical risk minimal in selected subset is the optimal solution.
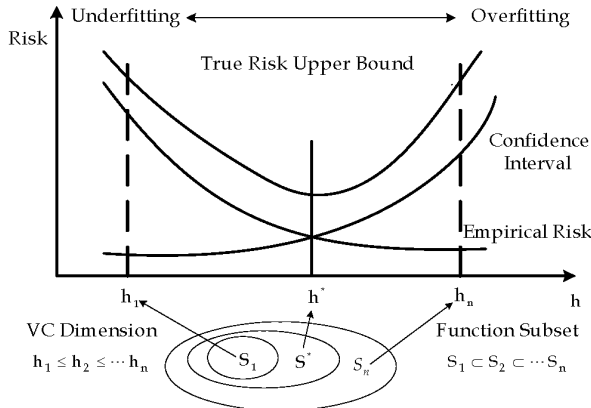
Fig. 5. The principle of Structural Risk Minimization.

Statistical learning theory also gives the required conditions for reasonable structure of function subset and the convergence property of actual risk in SRM principle. The actual risk is the sum of empirical risk and confidence interval. As the index of the elements in the structure increases, the empirical risk will be reduced with extended confidence interval. The smallest upper bound of the actual risk can be derived from a certain element in the structure. Support vector machine is a novel universal learning machine, which is widely used in the fields of pattern recognition, regression estimation and probability density. It is based on VC-dimension theory and SRM principle, which has a better generalization

performance by reaching a trade-off between model complexity with limited sample data and capacity of the learning algorithm. The support vector machine was coined by Vapnik in the late 1960s on the foundation of statistical learning theory. It was originally developed for binary classification problem. The optimal solution of SVM for a linearly separable case was introduced by Vapnik. Later this was extended to non-separable cases. In the previous research, a common solution to classification problem of communication signals is artificial neural network (ANN), such as multilayer linear perceptron network. After the first preliminary studies, SVM have shown a remarkable efficiency, especially when compared with traditional artificial neural networks. The main advantage of SVM, with respect to ANN, consists in the structure of the learning algorithm, characterized by the resolution of a constrained quadratic programming problem (CQP), where the drawback of local minima is completely avoided (Boser at el., 1992, Cortes & Vapnik, 1995, Scholkopf, 1995). Since the classification of communication signals is obvious to be a linearly non-separable problem, we will only discuss the computation of this optimization problem in this chapter. Given linearly non-separable classification problem, we suppose a training set is $\{(x_i, y_i)\}$, where $x \in R^N$, $y \in \{+1, -1\}$, $i = 1, ..., 1$. Assumed that the signal group can be classified by a hyperplane which is defined by

$$\mathbf{w}^T \Phi(\mathbf{x}) + b = 0 \qquad (25)$$

where the vector $\mathbf{w}$ defines the boundary of different classes of data, b is a scalar threshold and $\Phi(\mathbf{x}) = \|\mathbf{x}\|^2 / 2$.

In order to solve this non-separable problem, the non-minus slack variables ξ_i are introduced, and then the training vectors must satisfy

$$y_i(\mathbf{w}^T \mathbf{x}_i + b_i) - 1 \geq 1 - \xi_i, \qquad i = 1, 2, \cdots, 1 \qquad (26)$$

The hyperplane, which makes O(w) = ||w|| /2 to be minimum, is named as the optimal hyperplane. All of the training vectors are correctly classified by it and the vectors of each class are separated with a maximum margin (Burges, 1997).

Usually, structuring a hyperplane is solved as a quadratic optimization problem that can be formulated as

$$\min_{w,b,\xi_i} \quad \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^{1} \xi_i$$

$$\text{s. t.} \quad y_i(\mathbf{w}^T \Phi(\mathbf{x}_i) + b) \geq 1 - \xi_i \qquad (27)$$

$$\xi_i \geq 0, \quad i = 1, 2, \cdots, 1$$

where C is the penalty parameter, which is used to control the training error rate by different values.

Using a Lagrange multiplier technique, the optimization problem can be converted into

$$F = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{l}\xi_i - \sum_{i=1}^{l}\alpha_i[y_i(\mathbf{w}^T\Phi(\mathbf{x}_i) + b) - 1 + \xi_i] - \sum_{i=1}^{l}\beta_i\xi_i \qquad (28)$$

where α_i, $\beta_i > 0$ are Lagrange multiplier factors.

Given linearly non-separable classification problem, we can map the input data into a high dimensional feature space through some non-linear transformation which makes the data linearly separable (Devroye, 1996). Noted the above solution to linearly non-separable classification problem, only the inner product operation of the training samples is involved in the decision function. While structuring high dimensional feature space, the algorithm only use the inner product O $(\mathbf{x}_i)$«O $(\mathbf{x})$ in the space without separated O $(\mathbf{x})$ or O $(\mathbf{x}_i)$. If we can find a function K satisfying Mercer condition (Daniel & James, 2000), which can be denoted as

$$K(\mathbf{x}_i, \mathbf{x}) = \Phi(\mathbf{x}_i)\bullet\Phi(\mathbf{x}) \qquad (29)$$

where $K(\mathbf{x}_i, \mathbf{x})$ is the kernel function, which is utilized for mapping the input data to higher dimensional space in order to reduce the computational load. There are different kernel functions like polynomial, sigmoid and radial basis function (RBF) used in SVM, which are defined as follows.

1. Polynomial Kernel

$$K(\mathbf{x}, \mathbf{x}_i) = \left[(x \cdot x_i) + 1\right]^k \qquad (30)$$

A k-order polynomial classifier can be defined by Equation (30).

2. Radial Basis Function (RBF) Kernel

$$K(\mathbf{x}, \mathbf{x}_i) = \exp\{-\|x - x_i\|^2 / \sigma^2\} \qquad (31)$$

The width of the RBF kernel parameter o can be determined in general by an iterative process selecting an optimum value based on the full feature set. The main difference between RBF classifier and traditional RBF method is that each basis function in the RBF classifier corresponds to a support vector, which is automatically identified by the algorithm where the drawback of local minima is completely avoided.

3. Sigmoid Kernel

$$K(\mathbf{x}, \mathbf{x}_i) = \tanh\left(v(x \cdot x_i) + c\right) \qquad (32)$$

This kernel uses sigmoid function as inner product, which is equivalent to a multilayer perceptron with only one hidden layer. The number of node in hidden layer is automatically determined by algorithm.

Till now, the choice of the kernel functions was often used empirically, and this also became a theoretical drawback of SVM. A proper kernel function for a specific problem is dependent on the specific training sample data. In the practical applications, how to choose the proper model according to training sample set with better generalization ability is currently a research direction in the field of SVM. For a signal classification problem using cyclostationary features, we use an improved method of model selection based on kernel alignment, which will be described in Section 4.1 in detail. The choice of the kernel functions is studied via computer simulations and optimal results are achieved using radial-basis

function (RBF) kernel function. A typical classification experiment using RBF kernel function based SVM is illustrated as follows
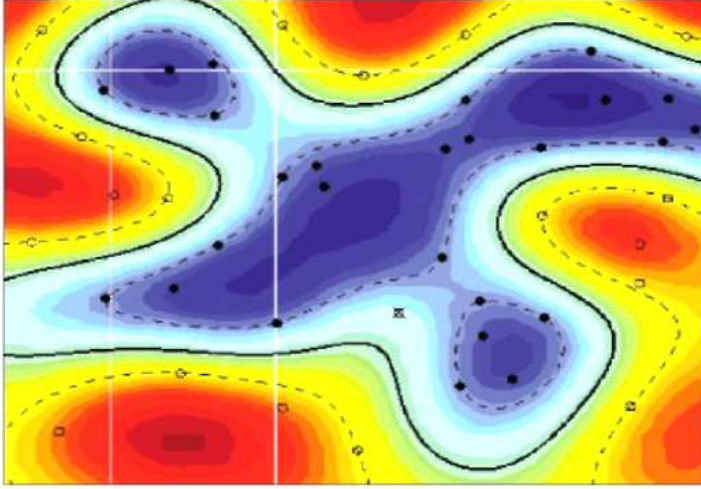
Fig. 6. A typical classification experiment using RBF kernel function (Sherrod, 2008).

After choosing the best kernel function, the dual representation of the optimization problem can be obtained by computing the derivatives with respect to w, b, ξ_i, which is described as

$$\max_{\alpha} \quad \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j)$$

$$\text{s. t.} \quad 0 \le \alpha_i \le C, i = 1, 2, \cdots, l \tag{33}$$

$$y^T \alpha = 0$$

The resulting decision function is obtained as follows:

$$f(x) = \text{sgn}\left\{ \sum_{i=1}^{l} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b \right\} \tag{34}$$

The architecture of the SVM classifier combining spectral correlation analysis is shown in Fig. 7.

Fig. 7. Architecture of the SVM classifier combining spectral correlation analysis.

4. Performance Evaluation and Analysis 4.1 Kernel Function and Parameters Selection

According to the definition of kernel function in the previous section, the kernel matrix can be defined as follows (Lanckriet at el., 2002)

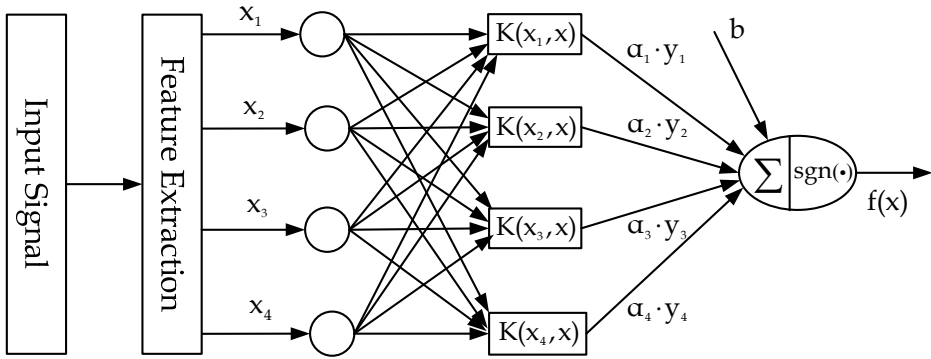$$\mathbf{K} = \left(k\left(x_i, x_j \right) \right)_{i,j=1}^{n} = \begin{bmatrix} K_{1,1} & K_{1,2} & \cdots & K_{1,l-1} & K_{1,l} \\ K_{2,1} & K_{2,2} & \cdots & K_{2,l-1} & K_{2,l} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ K_{l-1,1} & K_{l-1,2} & \cdots & K_{l-1,l-1} & K_{l-1,l} \\ K_{l,1} & K_{l,2} & \cdots & K_{l,l-1} & K_{l,l} \end{bmatrix} \tag{35}$$

where n is the number of the samples. It is a symmetric positive definite matrix, and since it specifies the inner products between all pairs of input elements, it completely determines the relative positions between those points in the embedding space.

In this section, we use an improved method of kernel selection based on kernel alignment to choose proper kernel function for our scheme (Cristianini at el., 2002). Assumed that $\mathbf{K_1}$ and $\mathbf{K_2}$ are kernel matrix of the kernel function k_1 and k_2, respectively. The (empirical) alignment of a kernel k_1 with a kernel k_2 with respect to the sample S is the quantity, which can be defined by

$$\hat{A}\left(S, k_1, k_2 \right) = \frac{\left\langle K_1, K_2 \right\rangle}{\sqrt{\left\langle K_1, K_1 \right\rangle \times \left\langle K_2, K_2 \right\rangle}} \tag{36}$$

Given a sample set $S = \left\{ x_i \,\middle|\, x_i \in S, i = 1, 2, \cdots, l \right\}$ for a specific signal classification problem, the sample can be divided into two types which are identified by +1 and -1. The signal type

set is denoted as $Y = \{y_i \,|\, y_i \in \pm 1 \,, i = 1, 2, \cdots, 1\}$ with the vector form $Y = \{y_1, y_1, \cdots, y_1\}^T$, and then alignment matrix can be defined by

$$\mathbf{K}_{ad} = YY^T = \begin{bmatrix} y_1 y_1 & y_1 y_2 & \cdots & y_1 y_{1\text{-}1} & y_1 y_1 \\ y_2 y_1 & y_2 y_2 & \cdots & y_2 y_{1\text{-}1} & y_2 y_1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ y_{1\text{-}1} y_1 & y_{1\text{-}1} y_2 & \cdots & y_{1\text{-}1} y_{1\text{-}1} & y_{1\text{-}1} y_1 \\ y_1 y_1 & y_1 y_2 & \cdots & y_1 y_{1\text{-}1} & y_1 y_1 \end{bmatrix} \tag{37}$$

This kernel selection method using kernel alignment is based on an important assumption that the kernel function has better performance if the kernel alignment of the kernel matrix and the alignment matrix is higher. Thus, if we consider $\mathbf{K} = \mathbf{K_1}$, $\mathbf{K_{ad}} = YY = \mathbf{K_2}$, then

$$\hat{A}\left(S, \mathbf{K}, YY^T\right) = \frac{\left\langle \mathbf{K}, YY^T \right\rangle}{\sqrt{\left\langle \mathbf{K}, \mathbf{K} \right\rangle \times \left\langle YY^T, YY^T \right\rangle}} = \frac{\left\langle \mathbf{K}, YY^T \right\rangle}{1\sqrt{\left\langle \mathbf{K}, \mathbf{K} \right\rangle}} \tag{38}$$

According to the above derivation, the optimal kernel function problem can be transformed into kernel alignment maximizing problem. In this section, the kernel alignment values of different kernel functions are compared via computer simulation by MATLAB 7.0. For the simulations, we define a signal set as {AM, ASK, FSK, PSK, MSK, QPSK}. To obtain the kernel alignment at different SNR, simulations are carried out with 1024 samples at SNR ranging from 0 dB to 20 dB. Simulation results show that the kernel alignment of the RBF kernel is the greater than that of other kernels, which is shown in Fig. 8. According to the simulation results, we choose RBF kernel as the kernel function of the SVM in our scheme. After the kernel function is selected, two key parameters of the SVW should be considered next. The first parameter, penalty parameter C of the SVM, is used for adjusting the range of the confidence interval to control the training error rate by different values. The second one, the width of the RBF kernel parameter o , can control the classification error by changing the largest VC dimension of linear classification plane. Therefore, these two parameters have a great impact on the classification performance (Chapelle at el., 2002). In this section, we use a simple cross-validation method to search the best parameters (C, o) .

In n-fold cross-validation, we first divide the training set into n subsets of equal size. Sequentially one subset is tested using the classifier trained on the remaining n-1 subsets. Thus, each instance of the whole training set is predicted once so the cross-validation accuracy is the percentage of data that are correctly classified. The process of cross-validation is described as follows.
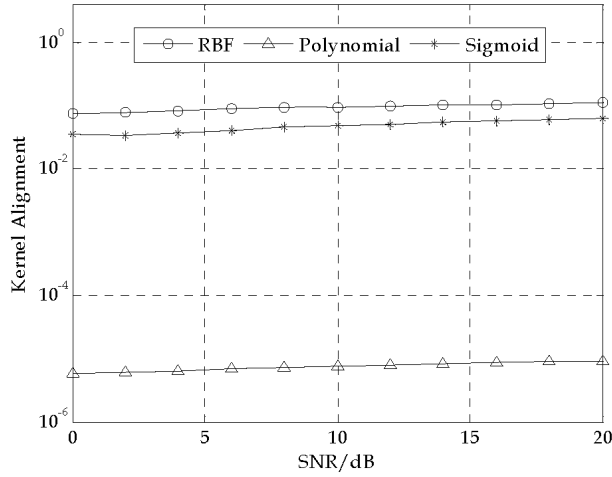
Fig. 8. Performance comparison of different kernel functions.

1. The training set is divided into n subsets (Γ_0, Λ_0) of equal size, and the elements of the subset are $\Gamma_0 = \{C_{0,1}, C_{0,2}, \cdots, C_{0,n}\}$, $\Lambda_0 = \{\sigma_{0,1}, \sigma_{0,2}, \cdots, \sigma_{0,n}\}$ n is typically from 4 to 10.

2. After training each $(C_{i,j}, \sigma_{i,j})$ using the classifier, the classification performance is tested.

3. The $(C_{i,j}, \sigma_{i,j})$ with best classification performance is used to form the parameter set

$$\Gamma_1 = \{C_{1,1}, C_{1,2}, \cdots, C_{1,n}\} \text{ and } \Lambda_1 = \{\sigma_{1,1}, \sigma_{1,2}, \cdots, \sigma_{1,n}\}, \sigma_{1,n} - \sigma_{1,1} \ll \sigma_{0,n} - \sigma_{0,1}$$

4. Repeat step 3 until the number of Γ_i and Λ_i reach to a preconcerted threshold. The $(C_{i,j}, \sigma_{i,j})$ will be the optimal choice for the specific SVM after iterations.

After performing cross-validation method for our scheme by MATLAB, we can obtain the best kernel parameters $\sigma^2 = 0,476$ and C = 0.212. After the best (C, σ) is found, the whole training set is trained again to generate the final SVM classifier.

4.2 Classifier Design
In order to compare the performance of different classifiers, two approaches based on existing methods, such as decision theory and artificial neural network, are introduced with spectral correlation features as training data.

4.2.1 Binary Decision Tree
After decades of research, decision theory has been widely studied in mathematics, statistics and communication concerned with identifying the values, uncertainties and other issues

relevant in a given decision and the resulting optimal decision. In the conventional decision theory, the binary decision tree (BDT) is a decision support tool that uses a graph or model of decisions mapping from observations to target value. Since it is simple and easy to understand, binary decision tree is widely used in signal detection and recognition. After observation of the value range the different features in Table. 1, it's easy to find that feature x can be used to classify the signals into three groups, which are {PSK, MSK, QPSK}, {FSK} and {AM, ASK}. Furthermore, feature x_2 and x_4 can be used to distinguish {ASK, AM} and {MSK, QPSK}, respectively. Thus, a binary decision tree is designed based on spectral correlation features for primary user signal recognition. In the decision algorithm given in Fig. 9, we make use of feature x to recognize the FSK signal in the first layer, and then feature x_2 and x_3 are used for the classification of ASK and AM signals and recognition of PSK signal in the second layer. In the third level, feature x_4 is utilized to distinguish MSK and QPSK signal.
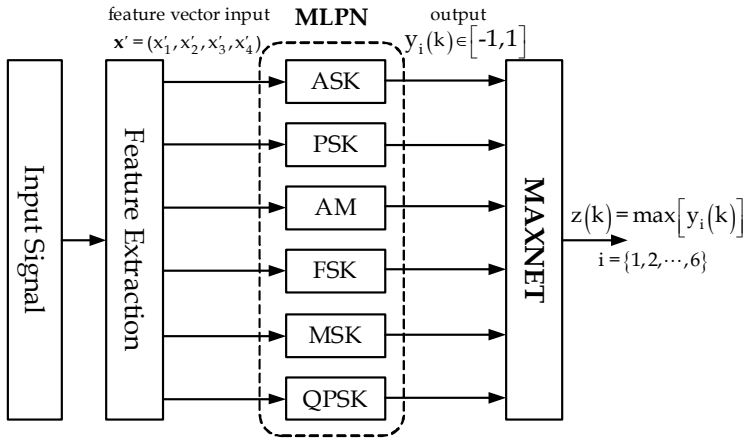

Fig. 9. Architecture of the BDT classifier based on spectral correlation analysis.

4.2.2 Multilayer Linear Perceptron Network

Artificial Neural networks have long been considered for pattern recognition and modulation classification and have proven to be robust to a variety of conditions such as interfering signals and noise. In order to compare the classifier performance of artificial neural network and support vector machine, a signal classification approach using spectral correlation and neural networks, which was proposed by A. Fehske (Fehske at el., 2005), is introduced below. Due to its simplicity, a multilayer linear perceptron network (MLPN) with 4 neurons in the hidden layer was used for each signal class, and each input layer uses the normalized spectral correlation feature vector $\mathbf{x'} = (x'_1, x_2, x'_3, x'_4)$ as input. Each MLPN was trained with a back propagation algorithm (Gupta, 2003) with an initial learning rate $\eta = 0.05$ decreasing with each epoch, a momentum constant $\alpha = 0.7$, and an activation function $\tanh(x)$. The output of each MLPN is a continuous value in the range (-1, 1). The MAXNET structure shown in Fig. 10 simply chooses the signal whose MLPN outputs the largest value. A typical gradient descent algorithm can be used to solve the linearly non-separable signal

classification problem, which can achieve minimal mean square error of expected output and actual output. The training results of all the MLPN are inputted into a simple MAXNET for final decision. The decision function of the MAXNET is defined by

$$z(k) = \text{argmax}\left[y_i(k)\right] \tag{39}$$

The signal classification approach using spectral correlation and MLPN is shown in Fig. 10.
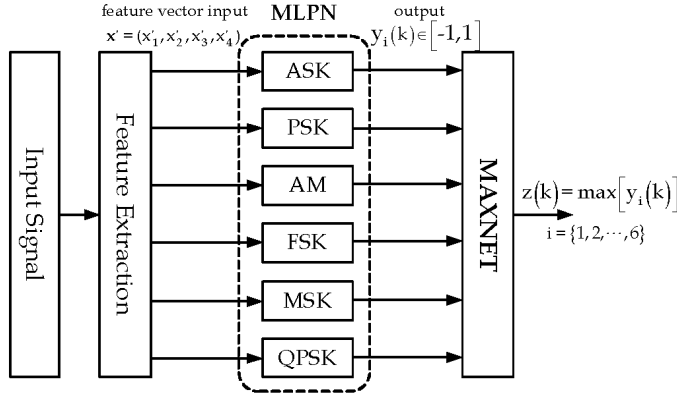

Fig. 10. Architecture of the MLPN classifier based on spectral correlation analysis.

5. Simulation Results

In this section, a variety of Monte Carlo simulations are presented to illustrate the performance of the algorithm. In the simulations, we define a signal set as {AM, ASK, FSK, PSK, MSK, QPSK}. For each type of signal, N signal samples constitute one frame which is used as an observation window to compare the performance of the algorithm with different data samples. To distinguish 6 modulation classes, simulation are carried out with 1100 frames at SNR ranging from 0 dB to 20 dB using three classifiers developed. 100 frames are used for training samples, and the remaining 1000 frames are used to calculate the probability of correct classification of different classifiers. The probability of correct classification (Pcc) can be defined by

$$P_{cc} = \frac{1}{N} \sum_{i=1}^{N} P_r\left[H_i \big| H_i\right] \tag{40}$$

where N is the number of simulations, $P_r\left[H_i \big| H_i\right]$ is the probability that the algorithm determine the signal class i correctly.

The radio channel models considered in the simulations include Gaussian channel and cognitive radio channel. In order to simulate the wireless environment of cognitive radios, WRAN channel model B recommended in IEEE 802.22 standard is used as cognitive radio channel (Sofer, E. & Chouinard, 2005). The WRAN channel reference model B determined

by the IEEE 802.22 standard group has a multi-path (6-path) delay profile, which is summarized in Table. 2.

In the IEEE 802.22 WRAN system, cognitive radio technology is considered to share the licensed spectrum of Digital TV, the typical service coverage is from 33 kilometres to 100 kilometres. The reference channel model B for the IEEE 802.22 WRAN are derived from a scenario that transmits the signals between the fixed BS and CPEs in wireless broadband environments. In such a dynamic channel environment, the delay extension is high with lower Doppler frequency. An Example channel responses for the nominal WRAN channel B is illustrated in Fig. 11.

Path	Delay(μ s)	Power(dB)	Doppler Freq.(Hz)
1	0	0	0
2	3	-7	0.10
3	8	-15	2.50
4	11	-22	0.13
5	13	-24	0.17
6	21	-19	0.37

Table 2. Multipath profiles of the IEEE 802.22 WRAN reference channel model B.
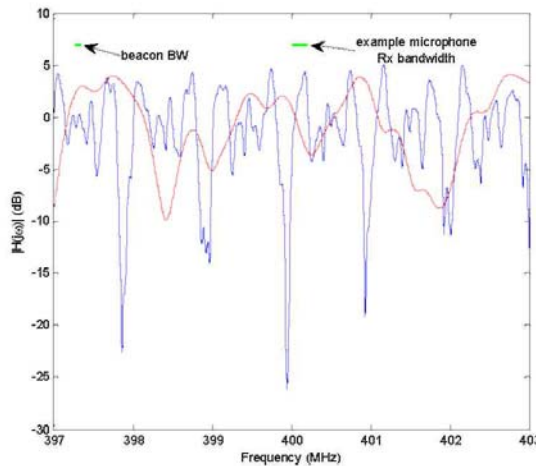
Fig. 11. An Example channel responses for the nominal WRAN channel B.

Table.3 and Table.4 indicate the probability of correct classification (P_{cc}) for each modulation type with the training data length of 1000 over Gauss channel and cognitive radio channel, respectively. Results show the overall correct rate is above 92.83% for a SNR of 4dB, and 97.32% for a SNR of 8dB. These good results for signals with low SNR in the cognitive radio environment show the proposed approach is insensitive with SNR variation, which come from the effects of the robustness of SVM classifier. According to Table.3, the proposed approach has better performance in both channel conditions.

Signal Type	SNR/dB					
	0dB	4dB	8dB	12dB	16dB	20dB
AM	96.92	100	100	100	100	100
ASK	95.23	100	100	100	100	100
FSK	92.07	97.23	100	100	100	100
PSK	86.34	93.02	98.24	99.91	100	100
MSK	88.58	95.54	99.29	100	100	100
QPSK	85.14	92.83	97.32	99.72	99.93	100

Table 3. Pec of SVM classifier over Gaussian channel.

Signal Type	SNR/dB					
	0dB	4dB	8dB	12dB	16dB	20dB
AM	94.32	96.61	99.71	99.92	100	100
ASK	91.33	96.14	99.88	100	100	100
FSK	89.87	95.22	99.91	100	100	100
PSK	87.29	92.36	98.29	99.91	100	100
MSK	85.61	94.19	97.90	99.89	100	100
QPSK	82.20	90.13	96.19	98.43	99.92	100

Table 4. P_{cc} of SVM classifier over WRAN channel B.

Fig. 12 and Fig. 13 show the performance of SVM classifier with data length as parameter over different channel models. When the data length is 100 and for a SNR of 4dB, the P_{cc} is up to 80.62% and with data length 200 and for a SNR of 6dB, the P_{cc} increases to 90%. When the data length is 1000 and for a SNR of l0dB, the Pcc is close to 100%. Above results show that the performance of the SVM classifier is high for small training data in both channel models. Fig. 14 and Fig. 15 are the performance comparison between BDT classifier, MLPN classifier and SVM classifier with the spectral correlation features over Gaussian channel and WRAN channel, which are calculated for different signals using data length of 200 and 1000, respectively. It is shown that when the SNR is lower, the MLPN classifier shows poor performance. While the SNR is higher, the probability of correct classification is increased. In lower SNR, the variation of the spectral correlation features (SCFs) is drastically due to the effect of the noise. Thus, the construction of neural network is not complete with small training data, which results in the performance degradation. The decision tree based classifier only use the partial information of the spectral correlation features (SCFs), therefore, the correct probability is lower than SVM classifier in the whole SNR range. All the results show the high performance of SVM classifier based on spectral correlation features (SCFs).
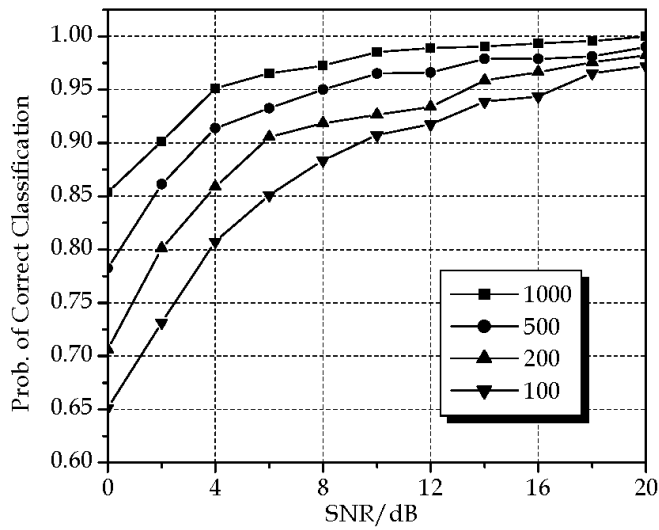
Fig. 12. The performance of SVM classifier with different data length over Gaussian channel.
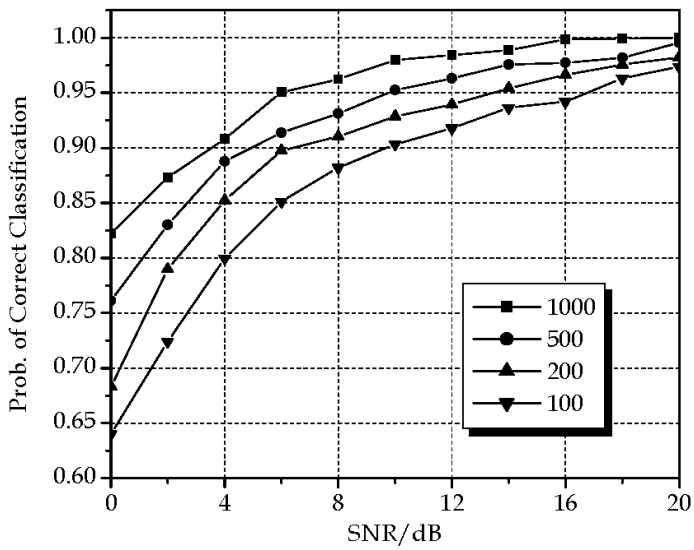

Fig. 13. The performance of SVM classifier with different data length over WRAN channel.
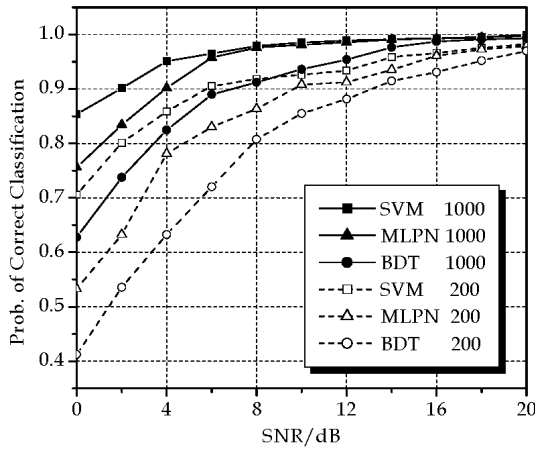
Fig. 14. Comparison between three classifiers with different data length over Gaussian channel.

6. Conclusion and Future Work

In this chapter, we proposed a novel approach combining the spectral correlation features and SVM for signal classification in cognitive radio environment. Four spectral correlation characteristic parameters were chosen as feature vector of SVM classifier. Simulation results show that the overall success rate is above 92.83% with data length of 1000 when SNR is equal to 4dB. Compared to existing methods, the proposed approach is more effective in the case of low SNR and limited training numbers. Future work in the area of signal classification for cognitive radio systems will involve the analysis of higher order spectrum correlation features of more communication signals. Based on these features, a multi-class SVM classifier can be used to improve the accuracy of classification and reduces the computational complexity. In addition, the classifier performance will be tested via simulations using several different channel models.

7. Acknowledgements

8. References

Mitola, J. & Maguire, G. Q. (1999). Cognitive radio: making software radios more personal, *IEEE Pers. Commun.*, vol. 6, no. 4, pp. 13-18
Mitola, J. (2001). Cognitive Radio - An Integrated Agent Architecture for Software Defined Radio, *PhD Thesis,* Royal Institute of Technology (KTH), Kista, Sweden
FCC. (2003). *Establishment of Interference Temperature Metric to Quantify and Manage Interference and to Expand Available Unlicensed Operation in Certain Fixed Mobile and*

Satellite Frequency Bands, ET Docket 03-289, Notice of Inquiry and Proposed Rulemaking.

Haykin, S. (2005). Cognitive Radio: Brain-Empowered Wireless Communications, *IEEE Journal on Selected Area in Communications,* vol.23, no 2, pp. 201-220

Akyildiz, I. F.; Lee, W. Y.; Vuran, M. C. & Mohanty, S. (2006). NeXt generation/dynamic spectrum access/cognitive radio wireless networks: a survey, *Computer Networks Journal(Elsevier),* vol. 50, no. 13, pp. 2127-2159

Ghasemi, A. & Sousa, E. S. (2005). Collaborative spectrum sensing for opportunistic access in fading environment, *Proc. of IEEE DySPAN 2005,* pp. 131-136

Sahai, A.; Hoven, N. & Tandra, R. (2004). Some fundamental limits in cognitive radio, *Proc. of Allerton Conf. on Commun., Control and Computing,* pp. 121-127

Digham, F. F.; Alouini, M. S. & Simon, M. K. (2003). On the energy detection of unknown signals over fading channels, *Proc. of IEEE ICC,* Anchorage, USA, pp. 3575-3579

Tang, H. (2005). Some physical layer issues of wide-band cognitive radio system, *Proc. of IEEE DySPAN 2005,* pp. 151-159

Cabric, D.; Mishra, S. M. & Brodersen, R. W. (2004). Implementation issues in spectrum sensing for cognitive radios, *Proc. of the 38th. Asilomar Conference on Signals, Systems, and Computers,* pp. 772 - 776

Fehske, A.; Gaeddert, J. D. & Reed, J. H. (2005). A new approach to signal classification using spectral correlation and neural networks, *Proc. of IEEE DySPAN 2005,* pp. 144-150

Bennett, W.R. (1958). *Statistics of Regenerative Digital Transmission,* Bell Syst. Tech. J, 37: 1501-1542

Gardner,W. & Franks, L.   (1975). Characterization of Cyclostationary Random Signal Processes, *IEEE Trans on Information Theory,* vol. 21, pp. 4-14

Polydoros, A. & Kim, K. (1990). On the Detection and Classification of Quadrature Digital Modulations in Broadband Noise, *IEEE Trans. Commun,* vol. 38, pp. 1199-1211

Sapiano, P.C. & Martin, J. D. (1996). Maximum Likelihooh PSK Classifier, *Proc. of IEEE Military Communications Conference,* vol. 3, pp.1010-1014

Sills, J. A. (1999). Maximum-likelihood Modulation Classification for PSK/QAM, *Proc. of IEEE Military Communications Conference,* vol. 1, pp. 217-220

Wei, W. & Mendel, J. M. (2000). Maximum-likelihood Classification for Digital Amplitude Phase Modulations, *IEEE Transactions on Communications,* vol. 48, no. 2, pp. 189-193

Hang, G. D.; Cai, B. & Wu, J. X. (2001). Spectral Correlation Method for Modulation Recognition," *Journal of Systems Engineering and Electronics,* vol. 23, no.3, pp. 34-37

Nandi, A. K. & Azzouz, E. E. (1995). Automatic Analogue Modulation Recognition, *Signal Processing,* vol. 46, pp. 211-222

Azzouz, E.E. & Nandi, A. K. (1995). Automatic Identification of Digital Modulation Types, Signal Processing, vol. 47, pp. 55-69

Azzouz, E.E. & Nandi, A. K. (1996). Procedure for Automatic Recognition of Analogue and Digital Modulations, *IEEE Proceedings on Communications,* vol. 143, no. 5, pp. 259-266

Azzouz, E.E. & Nandi, A. K. (1996). *Automatic Modulation Recognition,* Elsevier Science Ltd, pp. 241-273

Nandi, A. K. & Azzouz, E. E. (1998). Algorithms for Automatic Modulation Recognition of Communication Signals, *IEEE Trans. on Communications,* vol. 46, no. 4, pp. 431-436

Vapnik, V. N. (1995). The Nature of Statistical Learning Theory, Springer-Verlag, New York

Han, G. (2003). Research on Blind Signal Detection Techniques in Adaptive Single- and Multi-carrier Modulation, *PhD Thesis,* Xidian University, Xi'an, China

Gardner, W.A. (1987). Spectral Correlation of Modulated Signals:PART I –Analog Modulation, *IEEE Trans on Commun.,* vol. 35, no. 6, pp. 584-594

Gardner, W.A. (1987). Spectral Correlation of Modulated Signals: PART II   - Digital Modulation, *IEEE Trans on Commun.,* vol. 35, no. 6, pp. 595-601

Gardner, W. & Franks, L. (1975). Characterization of Cyclostationary Random Signal Processes, *IEEE Trans. on Information Theory,* vol. 21, pp.4-14

Gao, Y. (2007). Modulation recognition and high dynamic synchronization based on cyclic spectral density, *PhD Thesis,* Harbin Institute of Technology, Mar 2007.

Gardner, W. A. & Spooner, C. M. (1988). Cyclic Spectral Analysis for Signal Detection and Modulation Recognition, *Proceedings of the Conference on Military Communications (MILCOM),* San Diego, CA, pp. 419-423

Sofer, E. & Chouinard, G. (2005). *WRAN Channel Modeling,* IEEE802.22-05/0055r7

Wang, Y. (2007). Research on Identification of Modulation Signals, *Master Thesis,* Xidian University, Xi'an

Gurman, D. B. Hwang, E. S. & Vizcaino, J. M. (2008). *Modulation Identification Using Classic Pattern Recognition Techniques,* www.andrew.cmu.edu/user/danielgu/Pat_Rec_ Final _Paper _Gurman_Vizcaino_Hwang.pdf

Boser, B.; Guyon, I. & Vapnik, V. (1992). A training algorithm for optimal margin classifiers, *Proc. of Fifth Annual Workshop on Computational Learning Theory,* Pittsburgh: ACM Press

Cortes, C. & Vapnik, V. (1995). Support vector networks. *Machine Learning,* vol. 20, pp. 273 - 297

Scholkopf, B. & Burges, C. & Vapnik, V. (1995). Extracting support data for a given task, *Proc. of First Intl. Conf. on Knowledge Discovery & Data Mining,* AAA I Press, pp. 262-267

Burges, J. C. (1997). *A Tutorial on Support Vector Machines for Pattern Recognition.* Bell Laboratories, Lucent Technologies

Devroye, L.; Gyrofi, L. & Lugosi, L. (1996). *A Probabistic Theory of Pattern Recognition,* Springer, New York:

Daniel, J. S. & James, A. B. (2000). Support vector machine techniques for nonlinear equalization, *IEEE Trans. on Signal Processing,* vol. 11, no. 48, pp. 3271-3227

Cristianini, N.; Shawe-Taylor, J.; Elisseeff, A. & Kandola, J. (2002). On Kernel Target Alignment, *Proc. of Neural Information Processing Systems,* Cambridge, MA: MIT Press, pp. 367-373

Sherrod, P. H. (2008). *DTREG Predictive Modeling Software Manual,* http: //www.dtreg.com/ DTREG . pdf

Chapelle, O.; Vapnik,V. N. & Bacsquest, O. et al. (2002). Choosing multiple parameters for support vector machines, *Machine Learning,* vol. 46, pp. 131-159

Lanckriet, G.; Cristianini, N.; Bartlett, Ghaoui, P. L. & Jordan, M. (2002). *Learning the Kernel Matrix with Semi-Definite Programming,* Technical Report No. UCB/CSD-02-1206, University of California, Berkeley

Gupta, M. M. (2003). *Static and Dynamic Neural Networks: from Fundamentals to Advanced Theory,* Wiley, New York

Traffic Pattern Prediction Based Spectrum Sharing for Cognitive Radios

Xiukui Li and Seyed A. (Reza) Zekavat

Michigan Technological University, Houghton, MI 49931
USA
{xiuli,rezaz}@mtu.edu

1. Introduction

In this chapter, we introduce different traffic prediction techniques and discuss the process of evaluating channel availability through predicting traffic pattern of primary users for cognitive radios. When cognitive and non-cognitive users share the licensed spectrum, compared with secondary users (cognitive users), primary users have higher priority in using licensed channels. Therefore, whenever a primary user is detected, secondary users must vacate the relevant channels or decrease their transmitted power to reduce the interference on primary users. However, in some situations, due to the activities of primary users, secondary users may need to switch to other available channels, terminate communication, or reduce the transmitted power frequently. This leads to temporal connection loss of secondary users. In addition, if secondary users can not vacate a channel in a timely manner, it would interfere with primary users. To reduce the temporal connection loss and interference on primary users, secondary users should avoid using the channels which would be claimed by primary users with a high probability within a given time period. A solution to this problem is to enable secondary users to evaluate the channel availability through predicting the traffic pattern of primary users. This chapter discusses the methods of evaluating the probability of channel availability for secondary users, and examines the probability that there is a successful secondary user's transmission at one time instance. In addition, the cooperative prediction is briefly introduced. Finally, the applications of traffic pattern prediction technique to spectrum sharing are discussed.

2. Traffic Model

Traffic data patterns can be classified as (Haykin, 2005): 1) deterministic patterns: for example, each primary user (e.g., TV transmitter) is assigned a fixed time slot for transmission, and when it is switched off, the frequency band is vacated; 2) stochastic patterns: for example, the arrival times of data packets are modeled as a Poisson process, while the service times are modeled as exponentially distributed, depending on whether the data are of packet-switched or circuit-switched. Note that, in common channel signaling network, the exponential distribution drastically underestimates the proportion of short calls that do not last longer than the mean holding time (Bolotin, 1994). In general, the traffic stochastic parameters vary slowly. Hence, they can be estimated using historical data. The traffic model built on historical data

enables secondary users to predict the future traffic pattern of primary users (Li & Zekavat, 2008).

An overview of traffic modeling is provided in (Frost & Melamed, 1994). It discusses Markov modulated traffic models, autoregressive traffic models and self-similar traffic models, etc. In addition, the authors in (Adas, 1997) discuss different traffic models in telecommunication networks.

2.1 Traffic Model for Voice Communication

A speech process can be modeled as a two-state Markov chain which alternates between talk spurt and silent periods (Li, 1990); (Gruber, 1981). The authors in (Hong & Rappaport, 1986) propose a traffic model for cellular mobile radio telephone systems. The basic system model in (Hong & Rappaport, 1986) assumes that the new call origination rate is uniformly distributed over the mobile service area, and the channel holding time is approximated to a negative exponential distribution.

Considering a one-dimensional mobile system with cells in series (e.g., in highways), the authors in (Pavlidou, 1994) uses two-dimensional state diagrams to analyze the traffic in the mixed media cellular system. It assumes four Poisson arrival streams are entering each cell, which are originating new voice calls, originating new data packets, hand-off voice calls and hand-off data packets. The authors in (Leung et al., 1994) also consider the circumstance of a one-way, semi-infinite highway. With the assumption that there are an infinite number of channels available, they present a deterministic fluid model and a stochastic traffic model for a wireless network along the highway.

2.2 Traffic Model for Video Data

The authors in (Dawood & Ghanbari, 1999) provide a summary for traffic models of video data. (Lucantoni et al., 1994) proposes to model a single video source as a Markov renewal process whose states represent different bit rates. Some other models including Markov Modulated Fluid Flow (MMFF) model (Maglaris et al., 1988), Markov Modulated Poisson Process (MMPP) (Skelly, 1993), and AutoRegressive AR(1) stochastic model (Maglaris et al., 1988) are proposed to address the basic characteristics of the variable bit rate traffic. The MMFF and MMPP are suitable for queueing analysis of packet switched networks. The AR(1) stochastic model primarily characterizes the inter-frame source bit-rate variations and correlation.

For variable bit-rate traffic, the authors in (Knightly & Zhang, 1997) introduce a new deterministic traffic model called deterministic bounding interval-length dependent (D-BIND) to capture the multiplexing properties of bursty streams. For a large-scale satellite network simulation, (Ryu, 1999) proposes: 1) a discrete autoregressive process for MBone ("multicast backbone") video source modeling; 2) the superposition of fractal renewal processes (Sup-FRP) model for Web request arrivals, and, 3) a generalized shot-noise-driven Poisson point process (GSNDP) for aggregate traffic flow modeling.

2.3 Traffic Model for General Packet Data

Based on an analysis of Internet protocols for data communication, (Anderlind & Zander, 1997) proposes a simple model for future data traffic in wireless radio networks. Model parameters are selected to describe traffic from the Worldwide Web (WWW) access and from distributed file systems. A multilayer Markov model is considered in (Filipiak, 1992) for arrivals of calls, bursts, and packets to fast packet switching system, where the multilayer refers to call layer, packet layer and burst layer.

3. Traffic Prediction Techniques

Different methods and models are proposed to forecast the traffic of different networks. Generally, network traffic has a mix of self-similarity, Short Range and Long Range Dependencies (SRD and LRD) (e.g., (Kleinrock, 1993); (Paxon & Floyd, 1995); (Leland et al., 1994); (Jiang et al., 2001)). There exist concentrated periods of low activity and high activity (i.e., burstiness) in the network traffic (Feng, 2006).

An accurate predictor needs to capture the traffic characteristics such as SRD and LRD, self-similarity and nonstationarity (Sadek & Khotanzad, 2004a). The prediction models can he categorized as stationary and non-stationary ones (Adas, 1997). The stationary models include traditional models such as Poisson, Markov and autoregressive (AR), and self-similar models such as fractional autoregressive integrated moving average (FARIMA). Nonstationary models include artificial neural network (ANN) such as fuzzy neural network (FNN). These models can be applied to predict different types of traffic data in Ethernet, Internet, etc. (Sadek et al., 2003); (Shu et al., 1999); (Hall & Mars, 2000). In the following subsections, some prediction techniques are introduced including ARIMA based traffic forecasting, application of neural network in traffic forecasting, least mean square based traffic forecasting, etc.

3.1 ARIMA based Traffic Forecasting

Traditional models, such as autoregressive (AR) and autoregressive moving average (ARMA), can be used to predict the high-speed network traffic data (Adas, 1997); (Sang & Li, 2002), and they can capture the short range dependent (SRD) characteristic of traffic. For example, considering the traffic data from a regional emergency communication center, (Chen & Trajkovic, 2004) classifies network users into user clusters, and then predicts the network traffic using the SARIMA (seasonal autoregressive integrated moving-average) model. SARIMA is discussed in detail in (Brockwell & Davis, 2002). For a campus-wide wireless network (IEEE 802.11 infrastructure), (Papadopouli, 2005) uses the Yule-Walker method of ARIMA model to forecast the traffic at each wireless access point (AP) in an hourly timescale.

If a time series is non-stationary, but the first difference of the time series is stationary, ARIMA models can be used to characterize the dynamics of such a process. (Basu, 1996) indicates that appropriately differenced time-series generated from Internet traffic traces can be modeled as Auto-Regressive-Moving-Average (ARMA) processes. (Krithikaivasan et al., 2004) also proposes to use the ARIMA models to predict network traffic. Although ARIMA models can be used to model a class of non-stationary data, however, they cannot be applied to predict the network traffic which possesses the Long Range Dependent (LRD) characteristics.

In ARIMA models, the difference parameter d is restricted to integer values. In Fractional Autoregressive Integrated Moving Average models (FARIMA), this difference parameter can be a fraction. FARIMA model has the ability to capture both SRD and LRD characteristics of traffic (Corradi et al., 2001); (Shu et al., 1999). (Shu et al., 1999) also proposes using fractional ARIMA (FARIMA) to capture the self-similarity of network traffic. However, FARIMA model is time-consuming (Feng, 2006). (Sadek & Khotanzad, 2004a) discusses a two-stage predictor. It combines two different models, namely FARIMA and FNN. FARIMA captures the self-similarity, and FNN captures the non-stationarity. The combination of FARIMA and FNN enhances the prediction accuracy.

The authors in (Papagiannaki et al., 2005) model the evolution of IP backbone traffic at large time scales for long-term forecasting of Internet traffic. The aggregate demand between any two adjacent points is modeled as a multiple linear regression model. Short-term (e.g., sec-

onds or minutes) forecasting of Internet traffic is addressed in (Basu, 1999); (Sang, 2000); (Papadopouli et al., 2006).

The authors in (Randhawa & Hardy, 1998) model the VBR sources as AutoRegressive AR(1) Modulated Deterministic Arrival process which characterizes the inter-frame as well as the intra-frame bit rate variations, and they model the call arrival process as conventional birth-death Markov Process. The future traffic is predicted using a combination of linear prediction and transient state analysis of birth-death Markov Process.

3.2 Application of Neural Network in Traffic Forecasting

The authors in (Zhao et al., 2004) propose two neural network models for traffic forecasting in two situations. One addresses the forecasting of hourly traffic of the next day based on past observations, and the other focuses on the peak load prediction of the next day. Based on Time Series Forecasting (TSF), (Cortez et al., 2006) uses a Neural Network Ensemble (NNE) to predict the TCP/IP traffic.

Neural network based traffic prediction approach is complicated to implement. The accuracy and applicability of the neural network approach in traffic prediction is limited (Hall & Mars, 2000). Artificial Neural Network (ANN) can capture the non-linear nature of network traffic and the relationship between the output and input theoretically (Hansegawa et al., 2001); (Khotanzad & Sadek, 2003); (Lobejko, 1996), however, it might suffer from over-fitting (Doulamis et al., 2003).

The machine learning technique called support vector machine (SVM) can be applied to pattern recognition and other applications such as regression estimation. (Feng, 2006) employs the SVM to forecast the traffic in WLANs. It studies the issues of one-step-ahead prediction and multi-step-ahead prediction without any assumption on the statistical property of actual WLAN traffic.

3.3 Least Mean Square based Traffic Forecasting

Normalized Least Mean Square (NLMS) based prediction approaches are proposed for on-line variable bit-rate video traffic prediction. They do not require any prior knowledge of the video statistics (Adas, 1998). The authors in (Liu & Mao, 2005) propose a time-domain NLMS based prediction scheme, and a wavelet-domain NLMS based adaptive prediction scheme for video traffic prediction which exploits the redundant information in the wavelet transform coefficients for more accurate prediction (traffic is decomposed into wavelet coefficients and scaling coefficients at different timescales).

3.4 Other Traffic Forecasting Algorithms

The authors in (Kohandani et al., 2006) present a new forecasting technique called extended structural model (ESM) which is derived from the basic structural model (BSM). This technique replaces the constant parameters in BSM with variables and use the steepest descent search algorithm to find the values for these variables to minimize the mean absolute percentage error (MAPE).

Based on classical queueing theory, (Filipiak & Chemouil, 1987) derives discrete-time stochastic models and propose methods based on those models to forecast the number of congested trunk groups. The authors in (Liang, 2002) show that the ad hoc wireless network traffic is self-similar, and they apply a fuzzy logic system to ad hoc wireless network traffic forecasting. Fuzzy logic systems (FLSs) have been widely used in time-series forecasting (e.g., (Liang & Mendel, 2000)).

Traffic prediction techniques	Advantage	Disadvantage	Applications
ARIMA	Can capture the short range dependent (SRD) characteristic; can model a class of non-stationary data.	Cannot capture the Long Range Dependent (LRD) characteristic; the difference parameter d is restricted to integer values.	WLAN, Internet traffic, etc.
FARIMA	Can capture both SRD and LRD characteristics; can capture the self-similarity; the difference parameter d can be a fraction.	Time-consuming	Ad hoc wireless networks; Ethernet, etc.
Neutral Network based	Can capture the non-stationarity; can capture the non-linear nature of network traffic.	Complicated to implement; the accuracy and applicability is limited.	Internet traffic (video traffic), etc.
Least Mean Square based	Do not require any prior knowledge of the traffic statistics.	Slow convergence with the time-domain NLMS based algorithms; High complexity with the wavelet-domain NLMS based algorithms.	Internet traffic (video traffic), etc.

Table 1. Comparison of traffic prediction techniques.

The authors in (Sadek & Khotanzad, 2004b) present a parameter estimation procedure for the k-factor GARMA model (a generalized form of FARIMA). They use an adaptive prediction scheme to enable the model to capture the non-stationary characteristic and deal with any possible growth in the future traffic. The results indicate that the performance of the k-factor GARMA model outperforms the traditional AR model. GARMA models can be used to model a time series with SRD and LRD characteristics, and they also can model the periodicity of a time series with fewer parameters than ARMA models. The authors in (Ramachandran & Bhethanabotla, 2000) use the GARMA framework to measure the periodicities of Ethernet traffic.

The advantages and disadvantages of some traffic prediction techniques are compared in Table 1.

4. Probability of Channel Availability

Currently, most of the spectrum is assigned to GSM and CDMA networks (Brodersen, 2004). For CDMA networks (e.g., CDMA 2000, UMTS), the code division access makes the coexistence of primary users and secondary users difficult unless FDMA-CDMA systems are used (Pezeshk & Zekavat, 2003). In that case, secondary users may use the available spectrum only if the entire CDMA channels are available. Here, it is assumed that secondary users can find the idle channels for communication which are licensed to GSM networks.

4.1 Channels in GSM Networks

GSM uses TDMA techniques to provide multiple access for mobile users (Rappaport, 2002). In GSM, the available forward and reverse frequency bands are divided into 200kHz wide channels called ARFCNs (Absolute Radio Frequency Band Numbers). ARFCN is time shared between 8 subscribers using TDMA (there are eight time slots per TDMA frame). The combination of a time slot number and an ARFCN constitutes a physical channel for both the forward and reverse link. Each physical channel in a GSM system can be mapped into different logical channels at different times. There are two types of GSM logical channels, called traffic channels (TCHs) and control channels (CCHs).

In GSM systems, there are control messages transmitted over control channels between subscribers and base stations even if no call is in progress. There are three main control channels: the broadcast channel (BCH), the common control channel (CCCH), and the dedicated control channel (DCCH). The BCH and CCCH forward control channels are allocated time slot 0 (TS0). In the same frame, the time slot 1, TS1, through 7, TS7, still can carry regular traffic. The DCCH may exist in any time slot and on any ARFCN except TS0 of the BCH ARFCN. However, there are three types of DCCHs in GSM: (i) the stand-alone dedicated control channels (SDCCHs) which can be considered as intermediate and temporary channels for accepting a call from the BCH and holding the traffic while waiting for the base stations to allocate a TCH channel; (ii) the Slow, and, (iii) the Fast Associated Control Channels (SACCHs and FACCHs) which are used for supervisory data transmission between mobiles and base stations during a call. In summary, for the three control channels, BCH and CCCH only use TS0 in each frame and DCCHs are call related. Therefore, if there is no call in progress, secondary user can find an available channel among TS1-TS7 to use.

To detect an unused physical channel, a secondary user needs to tune to a forward channel (ARFCN) and synchronize with the base station. For a TDMA frame, a secondary user can detect whether the time slots TS1-TS7 are used or not by analyzing the data bursts. Here, secondary users are assumed to have the capability to remove the effect of dummy burst which are used as filler information for unused time slots on the forward link.

4.2 Channel Availability Evaluation

Generally, multiple channels are licensed to primary users. Secondary users can select any idle channel for communication. However, once a secondary user detects the appearance of primary users, it should vacate the channel immediately to allow primary users to continue using that channel. Therefore, the activities of secondary users do not affect the traffic distribution of primary users. Primary users keep their normal activities regardless of the existence of secondary users. Here, the number of channels licensed to primary users is assumed to be constant. In addition, with a fair scheduling policy, primary users utilize all licensed channel equally.

Traffic pattern prediction enables secondary users to estimate the channel utilization in a near future. For voice communications, two crucial factors in the traffic pattern are call arrival rate and call holding time. To estimate the utilization of one channel, secondary users can predict or estimate the call arrival rate and call holding time of primary users that use this channel. Then, according to the prediction and/or estimation results, secondary users are able to evaluate the probability that the channel would be available for a given time period. By comparing the evaluated probability with some threshold, secondary users can decide whether to use this channel.
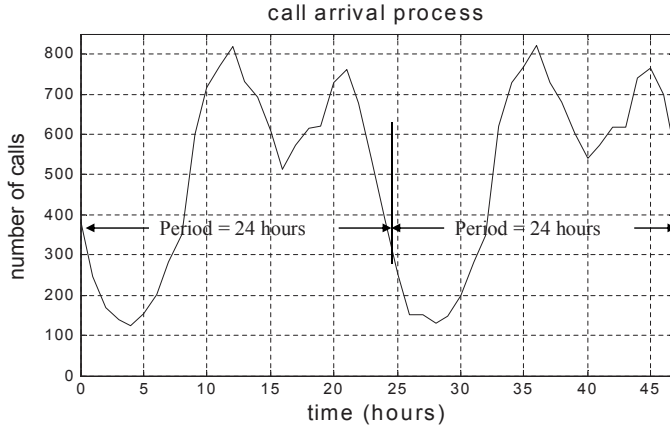
Fig. 1. Call arrival process.

In general, the traffic of voice communication in traditional wireless networks is periodic with a specific period T, e.g., $T = 24$ hours (one day), and it follows similar pattern in each period (Shu et al., 2003). An example of the call arrival process is shown in Fig. 1.

The call arrivals of primary users can be considered to follow non-homogeneous Poisson process $\{\mathcal{A}(t), t \geq 0\}$ (Snyder & Miller, 1991). The rate parameter for the process $\{\mathcal{A}(t)\}$ is $\lambda(t)$. $\lambda(t)$ may change over time. The expected call arrival rate between the time t_1 and t_2 is:

$$\lambda_{t_1,t_2} = \int_{t_1}^{t_2} \lambda(t)dt. \tag{1}$$

Thus, the number of call arrivals within the time interval $(t, t + \tau]$ follows a Poisson distribution with the parameter $\lambda_{t,t+\tau}$, i.e.,

$$\mathbb{P}\left\{(\mathcal{A}(t+\tau) - \mathcal{A}(t)) = k\right\} = \frac{e^{-\lambda_{t,t+\tau}}(\lambda_{t,t+\tau})^k}{k!}, k = 0, 1, \cdots \tag{2}$$

Evenly dividing one traffic period into 24 time intervals $(t_n, t_{n+1}]$ $(n = 0, 1, \cdots, 23)$, then, the time duration T_d $(T_d = t_{n+1} - t_n)$ for one time interval is one hour. For call arrival rate, a common metric employed in the telecommunication industry is the hourly number of calls (Chen & Trajkovic, 2004). Thus, the rate parameter $\lambda(t)$ can be assumed to maintain constant value $\frac{\lambda_n}{T_d}$ in each time interval $(t_n, t_{n+1}]$, i.e.,

$$\lambda(t) = \frac{\lambda_n}{T_d}, t \in (t_n, t_{n+1}] \tag{3}$$

where, λ_n is the total number of call arrivals in the time interval $(t_n, t_{n+1}]$.

Considering $t_n < t < t + \tau \leq t_{n+1}$, i.e., t and $t + \tau$ are within the same time interval $(t_n, t_{n+1}]$, the expected call arrival rate within τ is:

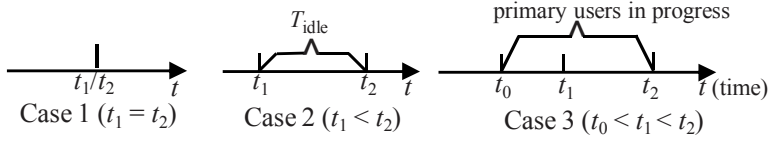$$\lambda_{t,t+\tau} = \frac{\lambda_n}{T_d}\tau. \tag{4}$$

Fig. 2. Three cases for channel availability evaluation.

Similarly, using (1) and (3), $\lambda_{t,t+\tau}$ can be obtained when t and $t + \tau$ are within different time intervals or within different periods.

When a secondary user finds an idle channel and intends to start transmission over this channel, first, it predicts the number of primary user call arrivals in the current time interval $(t_n, t_{n+1}]$. Using (4), the secondary user obtains the call arrival rate of primary users, $\frac{\lambda_n}{T_d}\tau_{ch}$, within its oncoming call holding time τ_{ch} (the call holding time of its next call). Then, the secondary user evaluates the probability, P_{na}, that no primary user would occupy the channel within the call holding time τ_{ch}. According to (2), P_{na} corresponds to:

$$P_{na} = \frac{e^{-\frac{\lambda_n}{T_d}\tau_{ch}}\left(\frac{\lambda_n}{T_d}\tau_{ch}\right)^0}{0!} = e^{-\frac{\lambda_n}{T_d}\tau_{ch}}. \tag{5}$$

The oncoming call holding time τ_{ch} for a secondary user is a random variable and it is hard to predict. τ_{ch} in (5) can be replaced with the average call holding time $\overline{T}$ of the secondary user. For a secondary user, $\overline{T}$ can be calculated based on the cumulative total call holding time T_a and the total number of calls $\mathcal{N}_c$ within a time duration, i.e., $\overline{T} = \frac{T_a}{\mathcal{N}_c}$.

To evaluate the probability that a channel is available for a given time period, secondary users should consider three scenarios discussed in the following.

Case 1 and 2 (see Fig. 2): Primary users end transmission at time t_1, and a secondary user starts transmission at t_2, $t_n < t_1 \le t_2 \le t_{n+1}$.

In these two cases, the secondary user only needs to evaluate the probability P_i that no primary user arrives within $\overline{T}$, $i = 1, 2$. According to (5), P_i corresponds to:

$$P_i = e^{-\frac{\lambda_n}{T_d}\overline{T}}, i = 1, 2. \tag{6}$$

Case 3 (see Fig. 2): One primary user starts transmission over a channel at time t_0, and it ends transmission at t_2. A secondary user intends to start transmission at t_1 over this channel (no other channels are available), $t_n < t_0 < t_1 < t_2 \le t_{n+1}$, i.e., at time t_1, the primary user call is still in progress. In this case, the secondary user is assumed to be capable of suspending its call for some time duration T_w ($T_w > 0$). Otherwise, the call of secondary users would drop (see (Li & Zekavat, 2009) for details).

In summary, to evaluate the probability that the channel would be available within $\overline{T}$, secondary users should be able to predict the number of primary user call arrivals in the corresponding time interval. In addition, secondary users should be able to obtain the PDF of call holding time distribution of primary users (for Case 3).

4.3 Call Arrival Prediction Algorithms

In Section 4.2, the rate parameter $\lambda(t)$ for the process $\{\mathcal{A}(t)\}$ is assumed to be constant, $\frac{\lambda_n}{T_d}$, in the time interval $(t_n, t_{n+1}]$ ($T_d = t_{n+1} - t_n$, $n = 0, 1, \cdots, 23$), and λ_n is the total number of call arrivals in the time interval $(t_n, t_{n+1}]$. Therefore, to estimate the call arrival rate of primary

users within a given time period, secondary users need to predict the number of call arrivals in the corresponding time interval. Denoting the time interval $(t_n, t_{n+1}]$ in the $(m+1)^{th}$ period as $(t_n + mT, t_{n+1} + mT]$, and the corresponding number of call arrivals in this time interval as λ_{n+mT}, the set of observations of the number of call arrivals in different time intervals of different periods can be considered as a discrete-time series $\{\lambda_t\}$ ($t = 0, 1, \cdots$). Thus, secondary users can predict the call arrivals of primary users by using the known observations of the number of primary user call arrivals in the past. Here, we use *SARIMA* model as an example to discuss one-step prediction of the number of call arrivals in a time interval, i.e., predicting the number of call arrivals λ_{l+1} given the known observations of call arrivals $\lambda_i, i \in \{1 \cdots l\}$.

$\{\lambda_t\}$ is a seasonal *ARIMA* $(p,d,q) \times (P,D,Q)_s$ process with period s if the differenced series $Y_t = (1-B)^d (1-B^s)^D \lambda_t$ is a causal *ARMA* process defined by:

$$
\begin{aligned}
&(1 - \phi_1 B - \cdots - \phi_p B^p)\left(1 - \Phi_1 B^s - \cdots - \Phi_P (B^s)^P\right) Y_t \\
&= (1 + \theta_1 B + \cdots + \theta_q B^q)\left(1 + \Theta_1 B^s + \cdots + \Theta_Q (B^s)^Q\right) Z_t,
\end{aligned}
\tag{7}
$$

where, p and P are the non-seasonal and seasonal autoregressive orders, respectively; q and Q are the non-seasonal and seasonal moving average orders, respectively, and d and D are the numbers of the regular and seasonal differences required. In addition, $\phi_1, \cdots, \phi_p, \Phi_1, \cdots, \Phi_P$, $\theta_1, \cdots, \theta_q, \Theta_1, \cdots, \Theta_Q$ are coefficient parameters. B is the backward operation, i.e., $B\lambda_t = \lambda_{t-1}$, and $Z_t \sim N(0, \sigma^2)$ (Brockwell & Davis, 2002).

With the simulated field data of $\{\lambda_t\}$, the autocorrelation of $\{Y_t\}$ with different d and D is calculated and it is found that $d = 1$ and $D = 1$ make the process $\{Y_t\}$ stationary. With period $s = 24$, the differenced observation Y_t corresponds to:

$$
Y_t = (1-B)(1-B^{24})\lambda_t = \lambda_t - \lambda_{t-1} - \lambda_{t-24} + \lambda_{t-25}.
\tag{8}
$$

Rearranging (8),

$$
\lambda_t = Y_t + \lambda_{t-1} + \lambda_{t-24} - \lambda_{t-25}.
\tag{9}
$$

Considering h-step prediction, and setting $t = l + h$:

$$
\lambda_{l+h} = Y_{l+h} + \lambda_{l+h-1} + \lambda_{l+h-24} - \lambda_{l+h-25}.
\tag{10}
$$

Using P_l to denote the best linear predictor, according to (10),

$$
P_l \lambda_{l+h} = P_l Y_{l+h} + P_l \lambda_{l+h-1} + P_l \lambda_{l+h-24} - P_l \lambda_{l+h-25}.
\tag{11}
$$

After calculating $P_l Y_{l+h}$ of Y_{l+h}, the prediction $P_l \lambda_{l+h}$ of λ_{l+h} can be computed recursively by noting that $P_l \lambda_{l+1-j} = \lambda_{l+1-j}$ for $j \geq 1$. For one-step prediction ($h = 1$), according to (11), $P_l \lambda_{l+1}$ corresponds to:

$$
P_l \lambda_{l+1} = P_l Y_{l+1} + \lambda_l + \lambda_{l-23} - \lambda_{l-24}.
\tag{12}
$$

The linear prediction of *ARMA* process $\{Y_t\}$ ($P_l Y_{l+h}$) can be implemented by the innovations algorithm (Brockwell & Davis, 2002).

Some algorithms discussed in Section 3 can also be used to implement the call arrival prediction. For other traffic types in various applications such as image and video, different algorithms can be selected to forecast the future traffic.

If the current time is within the time interval $(t_{n+1}, t_{n+2}]$, given the predicted result $\hat{\lambda}_{n+1}$, for Case 1 and 2 discussed in Section 4.2, the probability P_i that a channel would be available to the secondary user corresponds to:

$$P_i = e^{-\frac{\hat{\lambda}_{n+1}}{3600}\overline{T}}, i = 1, 2, \tag{13}$$

where, $\hat{\lambda}_{n+1}$ is the predicted result of λ_{n+1}, and $\overline{T}$ is the average call holding time of the secondary user.

4.4 Decision of Channel Availability

Introducing traffic prediction techniques to cognitive radio systems would impact the communication performance of both primary and secondary users. To maintain a trade-off between performance measurements including channel switching rate and call blocking rate of secondary users, interference on primary users and spectrum reuse efficiency, secondary users can set a threshold P_{th} ($P_{th} \in [0, 1]$) to determine whether to use a channel. If the probability P_i ($i = 1, 2$) is not less than this threshold, i.e.,

$$P_i \geq P_{th}, \tag{14}$$

then, a secondary user can proceed to use the channel. Otherwise, it would abandon using the channel.

The probability threshold P_{th} in (14) is determined by the requirements on performance measurements, whereas performance measurements are affected by the probabilities of false alarm and missed detection. Here, false alarm refers to the condition that a secondary user judges that primary users would appear whereas, actually, no primary user appears; missed detection refers to the condition that a secondary user judges that no primary user would appear whereas, actually, primary users appear.

To balance the performance in terms of the call blocking rate of secondary users, spectrum reuse efficiency and the interference on primary users, the probability threshold should be selected such that the occurrence rate of false alarm is equal to that of missed detection (Li & Zekavat, 2009).

Defining P_{avg} as:

$$P_{avg} = \frac{\sum P_i}{\mathcal{N}}, \tag{15}$$

where, $\sum P_i$ is the summation of the evaluated probabilities of channel availability over all $\mathcal{N}$ decisions (assuming, for one channel, a secondary user makes $\mathcal{N}$ decisions on channel availability).

Based on the discussion in (Li & Zekavat, 2009),

$$\mathbb{P}\{P_i \geq P_{th}\} = P_{avg}. \tag{16}$$

Considering P_i is uniformly distributed between $(1 - \alpha)P_{avg}$ and $(1 + \alpha)P_{avg}$ ($0 \leq \alpha \leq 1$, α is used to vary the range), according to (16),

$$P_{th} = \left(\frac{\sum P_i}{\mathcal{N}} \right) \left(1 + \alpha - 2\alpha \left(\frac{\sum P_i}{\mathcal{N}} \right) \right), 0 \leq \alpha \leq 1. \tag{17}$$

where, $\sum P_i$ is introduced in (15).

Taking advantage of the intelligence of cognitive radio, secondary users can dynamically select an appropriate α based on the evaluation results of probability of channel availability.

If P_i is assumed to be uniformly distributed between $P_i^{(max)}$ and $P_i^{(min)}$, and the PDF is $\frac{1}{P_i^{(max)} - P_i^{(min)}}$, secondary users can record the $P_i^{(max)}$ and $P_i^{(min)}$ for each time interval. Then, according to (16), P_{th} corresponds to:

$$P_{th} = P_i^{(max)} - \frac{\sum P_i}{\mathcal{N}} \left(P_i^{(max)} - P_i^{(min)} \right). \tag{18}$$

Within a time interval, for one channel, the secondary user can set the threshold P_{th} according to (17) or (18) by using the past evaluated probability results. Note that, within a time interval, the selection of P_{th} is impacted by the number of primary user call arrivals. Therefore, in different time intervals, P_{th} should be set dynamically. The impact of traffic prediction on communication performance is evaluated in (Li & Zekavat, 2009) based on simulations.

5. Probability of secondary users' successful transmissions

Before starting a transmission, a secondary user needs to check the channel availability. Here, if a channel is available, it means that currently the channel is idle, and the probability that the channel would not be occupied by primary users within $\overline{T}$ ($\overline{T}$ is the average call holding time of the secondary user) is not less than the probability threshold. Thus, at time t, the probability P_a that at least one channel is available to secondary users corresponds to (considering total m channels are licensed to primary users):

$$P_a = 1 - \prod_{k=1}^{m} \left[1 - P_c^{(k)} \cdot \mathbb{P}\{P_i^{(k)} \geq P_{th}^{(k)}\} \right], \tag{19}$$

where, $P_c^{(k)}$ is the probability that channel k is idle at time t; $P_i^{(k)}$ is the evaluated probability that channel k would not be occupied by primary users within $\overline{T}$, and $P_{th}^{(k)}$ is the probability threshold set for channel k. When the hourly number of call arrivals, λ_n, increases, $P_c^{(k)}$ would decrease. If, at time t, a secondary user has a transmission request, then, the probability P_{br} that this request would be blocked corresponds to:

$$P_{br} = 1 - P_a = \prod_{k=1}^{m} \left[1 - P_c^{(k)} \cdot \mathbb{P}\{P_i^{(k)} \geq P_{th}^{(k)}\} \right]. \tag{20}$$

To examine the probability that there is a successful secondary user's transmission at time t, P_b is assumed to be the probability that there is a secondary user's transmission request at time t. Then, the probability P_s that, at time t, there is a successful secondary user's transmission corresponds to:

$$P_s = P_a \cdot P_b = \left\{ 1 - \prod_{k=1}^{m} \left[1 - P_c^{(k)} \cdot \mathbb{P}\{P_i^{(k)} \geq P_{th}^{(k)}\} \right] \right\} \cdot P_b. \tag{21}$$

Generally, the occupancy of the channels is correlated. Here, for simplicity, P_c is assumed equal for all channels, i.e., $P_c^{(1)} = P_c^{(2)} = \cdots = P_c^{(m)}$. With another assumption that $\mathbb{P}\{P_i^{(k)} \geq$
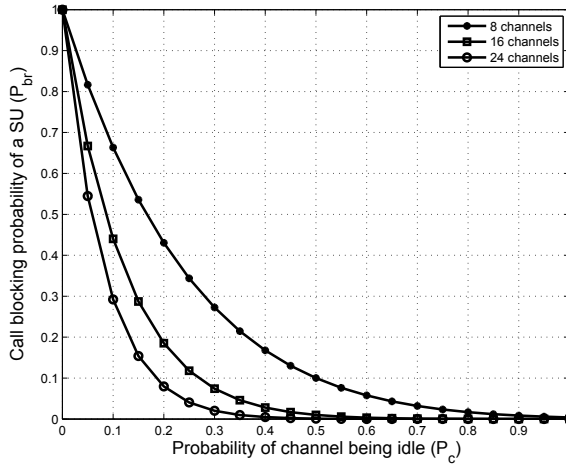
Fig. 3. Call blocking probability of a Secondary User (SU).

$P_{th}^{(k)}\} = 0.5$, according to (20), the call blocking probability of a secondary user, P_{br}, with respect to different P_c is sketched in Fig. 3. It can be observed that, (a) when P_c increases, i.e., λ_n decreases, P_{br} decreases, and, (b) when the number of available channels increases, P_{br} decreases.

According to (21), the probability P_s with respect to different P_b and P_c is sketched in Fig. 4 ($m = 8$). The probability P_s is low when either P_b or P_c is low (if the competition among secondary users in using the channel is considered, P_s would be lower). This is due to the fact that, when P_b is low, at time t, even if channels are available, nonetheless, secondary users may not have transmission requests; when P_c is low, i.e., λ_n is high, the transmission requests from secondary users would mostly be blocked (this can be illustrated by Fig. 5). In other words, requesting channels from secondary users and channels becoming available do not occur simultaneously. This implies that, in some situations, even if channels which are licensed to primary users are underutilized statistically, it is still hard for secondary users to obtain available channels because channel access by primary users and channel request from secondary users are both random processes.

6. Cooperative Prediction

To improve the accuracy of call arrival prediction, a secondary user may collect the prediction results of primary users' call arrivals from other secondary users, and obtain the average value of the predicted call arrivals. This is called cooperative prediction. Secondary users which are involved in the cooperative prediction should be able to monitor the same channel(s). Furthermore, a common channel (mostly, with low bandwidth) is required for those secondary users to share the prediction results. After collecting $\mathcal{M}$ prediction results of λ_{n+1} from other $\mathcal{M}$ secondary users, the secondary user may calculate the average value $\overline{\lambda}_{n+1}$ to evaluate the
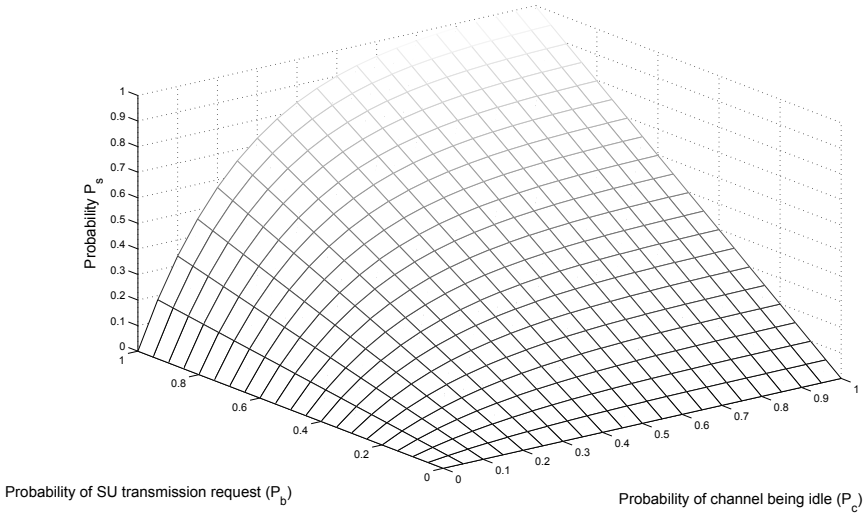
Fig. 4. Probability that there is a successful secondary user's transmission at time t.
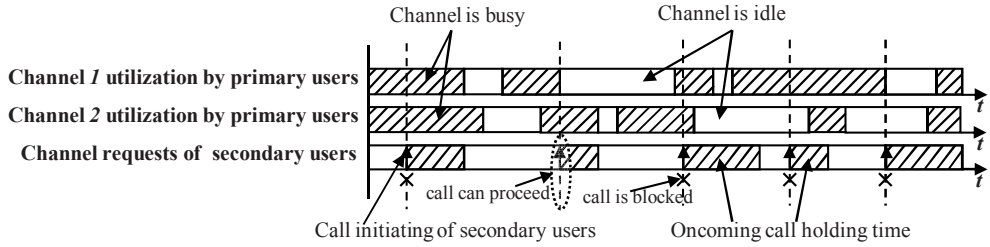

Fig. 5. Channel utilization and secondary users' channel requests.

probability of channel availability. $\overline{\lambda}_{n+1}$ can be calculated as:

$$\overline{\lambda}_{n+1} = \frac{\sum_{i=1}^{\mathcal{M}+1} \hat{\lambda}_{n+1}^{(i)}}{\mathcal{M}+1}, \tag{22}$$

where, $\hat{\lambda}_{n+1}^{(i)}$ is the prediction result of λ_{n+1} from secondary user i.

If the average call holding time $\overline{T}$ is the same for those involved secondary users, another cooperation technique can be employed (Unnikrishnan & Veeravalli, 2008); (Aalo & Viswanathan, 1992). Considering two hypotheses H_1 and H_0:

$$\begin{cases} H_1: & P_i \geq P_{th} \\ H_0: & P_i < P_{th}, \end{cases} \tag{23}$$

for secondary user k, the decision U_k corresponds to:

$$U_k = \begin{cases} 1, & \text{if } H_1 \text{ is true} \\ -1. & \text{if } H_0 \text{ is true} \end{cases} \tag{24}$$
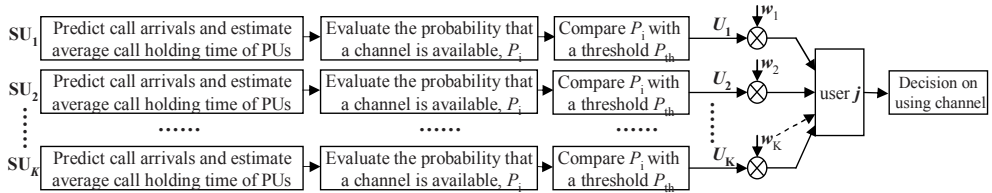
Fig. 6. Cooperative prediction (SU: Secondary User, PU: Primary User).

After receiving decisions in the form of U_i from other K secondary users, one secondary user can calculate the summation of all U_i's. If the summation result is greater than 0, i.e.,

$$\sum_{i=1}^{K+1} U_i > 0, \tag{25}$$

then, the secondary user can proceed to use the channel. Otherwise, it should abandon using the channel. Fig. 6 illustrates the whole process, in which $w_1, \cdots, w_K$ are the weights that can be simply considered as one or tuned based on field data.

7. Traffic Prediction based Spectrum Sharing

Cognitive Radio (CR) techniques enable the inter-vendor and user-central spectrum sharing (Zekavat & Li, 2005); (Li & Zekavat, 2007). The approach of finding available spectrum based on traffic pattern prediction can be applied to different cognitive radio based spectrum sharing techniques.

7.1 Spectrum Sharing across Multiple Service Providers

In business zones, for an infrastructure (e.g., base station) of a service provider, when the number of active users is greater than the maximal number of users that it can accommodate (i.e., the infrastructure is in overloaded status due to channel scarcity), the incoming users would be blocked. However, at the same time, nearby infrastructures of other service providers might be in the underloaded status. Available channels licensed to these service providers or other organizations can be used by the overloaded infrastructure (assuming multiple service providers coexist in the same area). The difference in traffic load across multiple service providers might be due to the diversity of services that service providers offer to their customers, e.g., mobile communication, wireless Internet for laptops, etc.

Generally, there are two approaches for the overloaded infrastructure to find available channels which are licensed to other service providers: (a) the overloaded infrastructure asks neighboring service providers for available channels; (b) the overloaded infrastructure is equipped with CR, and it senses the available channels within its coverage area. However, if the cell radii of service providers involved are different, and/or the infrastructures of different service providers are not located at the same positions, then, both approaches could lead to co-channel interference.

To reduce the co-channel interference and remove the need of equipping CR in each infrastructure of different service providers, a CR network consisting of multiple fixed CR nodes can be deployed to support the spectrum sharing. These CR nodes are distributed regularly within the area of interest, and each CR node is a dedicated node that senses the surrounding environment and monitors the channel usage within its sensing range. Those channels might
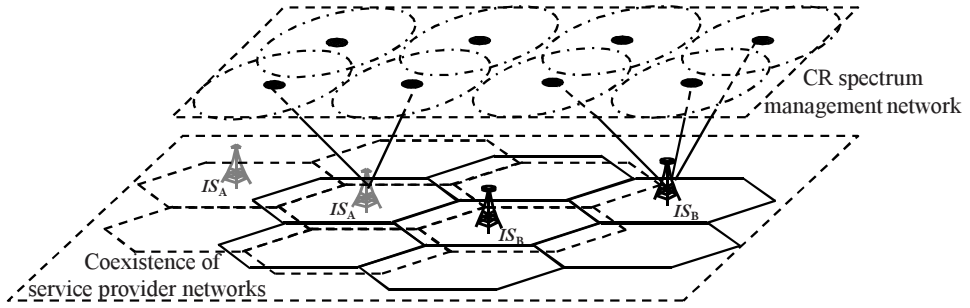
Fig. 7. A CR spectrum management network coexists with multiple cell based wireless networks.

be licensed to different service providers. To avoid co-channel interference, CR nodes which monitor the channel usage in the cell of interest and the corresponding co-channel cells cooperate to provide channel availability information for the overloaded infrastructures of service providers. CR nodes are connected together via wire or they communicate wirelessly to form a network. The CR network can be called a spectrum management network, and coexists with the wireless networks operated by different service providers (see Fig. 7). Thus, when one or more infrastructures of a service provider are overloaded, they request the channel availability information from those surrounding CR nodes. Then, overloaded infrastructures process the information to select the optimum channels based on the channel associated metrics such as interference level and the probability of channel being available for a certain time duration. A CR node would determine channel i is available if: 1) The instantaneous RF (Radio Frequency) energy (plus noise) in this channel, $I_{int}^{(i)}$, is less than the tolerable interference limit, $I_{tol}^{(i)}$, of this channel, i.e., $I_{int}^{(i)} < I_{tol}^{(i)}$, and, 2) The probability $P^{(i)}$ that this channel would be available (not occupied by its licensed users) for a given time period is not less than a threshold $P_{th}^{(i)}$, i.e., $P^{(i)} \geq P_{th}^{(i)}$. $P^{(i)}$ is evaluated by CR nodes using the discussed traffic pattern prediction technique. Thus, it reduces the probability that the selected channels are claimed by their licensed users when these channels are being used by the overloaded infrastructure.

8. Conclusion

In this chapter, traffic models and traffic prediction techniques which can be implemented in secondary (cognitive) users are discussed. Methods are introduced for secondary users to evaluate the probability of channel being available for a given time period in different situations. Comparing the evaluated probability of channel availability with a threshold, secondary users can determine whether to use a channel. The threshold maintains a trade-off between the channel switching rate and call blocking rate of secondary users, the interference on primary users and spectrum reuse efficiency.

In addition, the probability that there is a successful secondary user's transmission at one time instance is examined, and cooperative prediction techniques are briefly introduced. Finally, the application of the traffic pattern prediction technique in spectrum sharing is discussed.

9. References

Aalo V. & Viswanathan R. (1992). Asymptotic performance of a distributed detection system in correlated Gaussian noise. *IEEE Trans. Signal Processing*, vol. 40, pp. 211-213, Jan. 1992.

Adas A. (1997). Traffic models in broadband networks, *IEEE Communications Magazine*. vol. 35, no. 7, pp. 82-89, July 1997.

Adas A. (1998). Using adaptive linear prediction to support real-time vbr video under rcbr network service model. *IEEE/ACM Transactions Networking*, vol. 6, no. 5, pp. 635-644, 1998.

Anderlind E. & Zander J. (1997). A traffic model for non-real-time data users in a wireless radio network. *IEEE Commun. Letters*, vol. 1, no. 2, pp. 37-39, Mar 1997.

Basu S.; Mukherjee A. & Klivansky S. (1996). Time series models for internet traffic, *Proceedings of IEEE INFOCOM'96*, vol. 2, pp. 611-620, Mar 1996.

Basu S. & Mukherjee A. (1999). Time series models for internet traffic, *Proceedings of 24th Conf. Local Computer Networks*, pp. 164-171, Oct. 1999.

Bolotin V. A. (1994). Modeling call holding time distributions for CCS network design and performance analysis. *IEEE J. Select. Areas Commun.*, vol. 12, Issue 3, pp. 433-438, April 1994.

Brockwell P. J. & Davis R. A. (2002). *Introduction to Time Series and Forecasting*, second ed., Springer-Verlag New York, 2002.

Brodersen R. W.; Wolisz A.; Cabric D.; Mishra S. M. & Willkomm D. (2004). 2004 White Paper: CORVUS-A Cognitive Radio Approach for Usage of Virtual Unlicensed Spectrum, available online http://bwrc.eecs.berkeley.edu/Research/MCMA/.

Chen H. & Trajkovic L. (2004). Trunked radio systems: traffic prediction based on user clusters. *Wireless Communication Systems*, pp. 76-80, Sept. 2004.

Corradi M.; Garroppo R. G.; Giordano S. & Pagano M. (2001). Analysis of f-ARIMA processes in the modeling of broadband traffic, *Proceedings of ICC'01*, vol. 3, pp. 964-968, 2001.

Cortez P.; Rio M.; Rocha M. & Sousa P. (2006). Internet Traffic Forecasting using Neural Networks, *International Joint Conference on Neural Networks*, pp. 2635 - 2642, July 2006.

Dawood A. M. & Ghanbari M. (1999). Content-based MPEG video traffic modeling. *IEEE Trans. Multimedia*, vol. 1, no. 1, pp.77-87, Mar 1999.

Doulamis A. D.; Doulamis N. D. & Kollias S. D. (2003). An adaptable neural-network model for recursive nonlinear traffic prediction and modeling of MPEG video sources, *IEEE Trans. on Neural Networks*, vol. 14, no. 1, pp. 150-166, Jan. 2003.

Feng H.; Shu Y.; Wang S. & Ma M. (2006). SVM-Based Models for Predicting WLAN Traffic, *Proceedings of IEEE ICC 2006*, vol. 2, pp. 597-602, June 2006.

Filipiak J. & Chemouil P. (1987). Modeling and Prediction of Traffic Fluctuations in Telephone Networks. *IEEE Trans. Commun.*, vol. 35, no. 9, pp. 931-941, September 1987.

Filipiak J. (1992). Accuracy of traffic modeling in fast packet switching. *IEEE Trans. Communications*, vol. 40, no. 4, pp. 835-846, April 1992.

Frost V. S. & Melamed B. (1994). Traffic modeling for telecommunications networks. *IEEE Commun. Magazine*, vol. 32, no. 3, pp. 70-81, Mar 1994.

Gruber J. G. (1981). Delay related issues in integrated voice and data networks. *IEEE Trans. Commun.*, vol. 29, no. 6, pp. 786-800, June 1981.

Hall J. & Mars P. (2000). Limitations of artificial neural networks for traffic prediction in broadband networks. *IEE Proc. Communications*, vol. 147, no. 2, pp. 114-118, 2000.

Hansegawa M.; Wu G. & Mizuno M. (2001). Applications of nonlinear prediction methods to the Internet traffic, *Proceedings of ISCAS'01*, vol. 3, pp. 169-172, May 2001.

Haykin S. (2005). Cognitive radio: brain-empowered wireless communications. *IEEE J. Select. Areas Commun.*, vol. 23, pp. 201-220, Feb. 2005.

Hong D. & Rappaport S. S. (1986). Traffic model and performance analysis for cellular mobile radio telephone systems with prioritized and nonprioritized handoff procedures. *IEEE Trans. Vehicular Technology*, vol. 35, no. 3, pp. 77-92, Aug 1986.

Jiang M.; Nikolic M.; Hardy S. & Trajkovic L. (2001). Impact of self-similarity on wireless data network performance, *Proceedings of ICC 2001*, pp. 477-481, June 11-14, 2001.

Khotanzad A. & Sadek N. (2003). Multi-scale high-speed network traffic prediction using combination of neural network, *Proceedings of IJCNN'03*, vol. 2, pp. 1071-1075, July 2003.

Kleinrock L. (1993). On the Modeling and Analysis of Computer Networks. *Proceedings of IEEE*, vol. 81, no. 8, pp. 1179-1191, 1993.

Knightly E. W. & Zhang H. (1997). D-BIND: an accurate traffic model for providing QoS guarantees to VBR traffic. *IEEE/ACM Trans. Networking*, vol. 5, no. 2, pp. 219-231, April 1997.

Kohandani F.; McAvoy D. W. & Khandani A. K. (2006). Wireless airtime traffic estimation using a state space model, *Proceedings of CNSR 2006*, pp. 8, May 2006.

Krithikaivasan B.; Deka1 K. & Medhi D. (2004). Adaptive bandwidth provisioning envelope based on discrete temporal network measurements, *Proceedings of INFOCOM'04*, Hong Kong, pp. 1786-1796, Mar. 2004.

Leland W.; Taqqu M.; Willinger W. & Wilson D. (1994). On the self-similar nature of Ethernet traffic (extended version). *IEEE/ACM Transactions on Networking*, vol. 2, no. 1, pp. 1-15, 1994.

Leung K. K.; Massey W. A. & Whitt W. (1994). Traffic models for wireless communication networks. *IEEE J. Select. Areas Commun.*, vol. 12, no. 8, pp. 1353-1364, Oct 1994.

Li S. Q. (1990). Traffic characterization for integrated services networks, *IEEE Trans. Commun.*, vol. 38, no. 8, pp. 1231-1243, Aug. 1990.

Li X. & Zekavat S. A. (2007). Inter-Vendor Dynamic Spectrum Sharing: Feasibility Study and Performance Evaluation, *Proceedings of IEEE DySPAN 2007*, pp. 412-415, April 2007.

Li X. & Zekavat S. A. (2008). Traffic Pattern Prediction and Performance Investigation for Cognitive Radio Systems, *Proceedings of IEEE WCNC Conference, 2008*, pp. 884-889, March 2008.

Li X. & Zekavat S. A. (2009). Cognitive Radio Based Spectrum Sharing: Evaluating Channel Availability via Traffic Pattern Prediction. *Journals of Communications and Networks, Special Issue on Cognitive Radio: A path in the evolution of public wireless networks*, vol. 11, no. 2, pp. 104-114, April 2009.

Liang Q. & Mendel J. M. (2000). Interval type-2 fuzzy logic systems: Theory and design, *IEEE Trans. Fuzzy Systems*, vol. 8, pp. 535-551, Oct. 2000.

Liang Q. (2002). Ad hoc wireless network traffic-self-similarity and forecasting. *IEEE Commun. Letters*, vol. 6, no. 7, pp. 297-299, Jul 2002.

Liu H. & Mao G. (2005). Prediction Algorithms for Real-Time Variable-Bit-Rate Video, *Asia-Pacific Conference on Communications*, pp. 664-668, Oct. 2005.

Lobejko W. (1996). Traffic prediction by neural approach, *Proceedings of MILCOM'96*, vol. 2, pp. 21-24, Oct. 1996.

Lucantoni D. M.; Neuts M. F. & Reibman A. R. (1994). Methods for performance evaluation of VBR video traffic models. *IEEE/ACM Transs Networking*, vol. 2, no. 2, pp. 176-180, April 1994.

Maglaris B.; Anastassiou D.; Sen P.; Karlsson G. & Robbins J. (1988), Performance Models of Statistical Multiplexing in Packet Video Communications. *IEEE Trans. Communication*, vol. 36, pp. 834-844, July 1988.

Papadopouli M.; Shen H.; Raftopoulos E.; Ploumidis M. & Hernandez-Campos F. (2005). Short-term traffic forecasting in a campus-wide wireless network, *Proceedings of IEEE PIMRC 2005*, vol. 3, pp. 1446-1452, Sept. 2005.

Papadopouli M.; Raftopoulos E. & Shen H. (2006). Evaluation of short-term traffic forecasting algorithms in wireless networks. *Next Generation Internet Design and Engineering, 2006*, pp. 102-109, 2006.

Papagiannaki K.; Taft N.; Zhang Z. & Diot C. (2005). Long-term forecasting of Internet backbone traffic, *IEEE Trans. Neural Networks*, vol. 16, no. 5, pp. 1110-1124, Sept. 2005.

Pavlidou F. N. (1994). Two-dimensional traffic models for cellular mobile systems. *IEEE Trans. Commun.*, vol. 42, no. 234, pp. 1505-1511, Feb/Mar/Apr 1994.

Paxon V. & Floyd S. (1995). Wide Area Traffic: The Failure of Poisson Modeling. *IEEE/ACM Trans. on Networking*, pp. 292-298, 1995.

Pezeshk A. & Zekavat S. A. (2003). Inter-vendor spectrum sharing in DS-CDMA and MC-CDMA systems, *Proceedings of IEEE 37th Asilomar conference on Signals, Systems an Computers*, Nov. 9-12, 2003.

Ramachandran R. & Bhethanabotla V. N. (2000). Generalized autoregressive moving average modeling of the Bellcore data, *Proceedings of 25th Annual IEEE Conference on Local Computer Networks*, pp. 654-661, Nov. 2000.

Randhawa T. S. & Hardy R. H. S. (1998). Estimation and prediction of VBR traffic in high-speed networks using LMS filters, *Proceedings of IEEE ICC'98*, vol. 1, pp. 253-258, June 1998.

Rappaport T. S. (2002). *Wireless communications*, Prentice Hall PTR, 2nd edn. 2002.

Ryu B. (1999). Modeling and simulation of broadband satellite networks. II. Traffic modeling. *IEEE Commun. Magazine*, vol. 37, no. 7, pp. 48-56, July 1999.

Sadek N.; Kbotanzad A. & Chen T. (2003). ATM dynamic bandwidth allocation using F-ARIMA prediction model, *Proceedings of International Conference on Computer Communications and Networks*, Dallas, TX, pp. 359-363, 2003.

Sadek N. & Khotanzad A. (2004a). Dynamic bandwidth allocation using a two-stage fuzzy neural network based traffic predictor, *Proceedings of IEEE International Joint Conference on Neural Networks*, vol. 3, pp. 2407-2412, July 2004.

Sadek N. & Khotanzad A. (2004b). Multi-scale high-speed network traffic prediction using k-factor Gegenbauer ARMA model, *Proceedings of IEEE ICC'04*, vol. 4, pp. 2148-2152, June 2004.

Sang A. & Li S. (2000). A predictability analysis of network traffic, *Proceedings of INFOCOM*, pp. 342-351, Mar. 2000.

Sang A. & Li S. (2002). A predictability analysis of network traffic. *Computer networks*, vol. 39, pp. 329-345, 2002.

Shu Y.; Jin Z.; Zhang L.; Wang L. & Yang O. W. W. (1999). Traffic prediction using FARIMA models, *Proceedings of International Conference on Communications*, vol. 2, pp. 891-895, 1999.

Shu Y.; Yu M.; Liu J. & Yang O. W. W. (2003). Wireless traffic modeling and prediction using seasonal ARIMA models, *Proceedings of IEEE ICC'03*, vol. 3, pp. 1675-1679, May 2003.

Skelly P.; Schwartz M. & Dixit S. (1993). A Histogram-Based Model for Video Traffic Behavior in an ATM Multiplexer. *IEEE/ACM Trans. Networking*, vol. 1, pp. 446-458, August 1993.

Snyder D. & Miller M. (1991). *Random point processes in time and space*, second ed., Springer-Verlag New York, 1991.

Zekavat S. A. & Li X. (2005). User-Central Wireless System: Ultimate Dynamic Channel Allocation, *Proceedings of IEEE DySPAN'05*, pp. 82-87, Nov. 2005.

Zhao G. F.; Tang H.; Xu W. B. & Zhang Y. H. (2004). Application of neural network for traffic forecasting in telecom networks, *Proceedings of International Conference on Machine Learning and Cybernetics*, vol. 4, pp. 2607-2611, Aug. 2004.

Unnikrishnan J. & Veeravalli V. V. (2008). Cooperative sensing for primary detection in cognitive radio. *IEEE J. Selected Topics in Signal Processing*, vol. 2, pp. 18-27, Feb. 2008.

Cognitive Radio Dynamic Access Techniques for Mutual Interference Reduction and Efficient Spectrum Utilization

I. Budiarjo, M.K.Lakshmanan and H. Nikookar
International Research Centre for Telecommunications and Radar
Department of Electrical Engineering, Mathematics and Computer Science
Delft University of Technology, Mekelweg 4 2628CD, Delft,
The Netherlands

1. Introduction

A Background

Advances in wireless technologies have spawned the development of a host of new and innovative wireless applications and services. With each passing day there is demand for more wireless services even when the popularity of existing applications is on the rise. As a result the clamor for valuable resources such as transmission spectrum has reached a shrill note. Paradoxically studies after studies conclude that large swathes of licensed frequency bands remain unused for most of the time and that spectrum congestion is more due to inadequate access techniques rather than non-availability (Brodersen et al., 2004); (Haykin, 2005). This has sparked a debate in the Telecommunications circles on the need to revamp existing spectrum regulatory policies and introduce newer approaches. One such initiative is the idea of Cognitive Radio, a new paradigm that promises opportunistic utilization of unused spectrum and efficient spectrum management. In (Haykin, 2005) Haykin has defined Cognitive Radio as an intelligent wireless communication system that is cognizant of its environment, learns from it and adapts its internal states to statistical variations in the incoming RF stimuli by making changes in certain operating parameters in real time with objectives of highly reliable communications whenever and wherever needed, and efficient utilization of the radio spectrum. Modulation scheme, transmit power, channel coding, and carrier frequency are examples of parameters to be exploited in cognitive radio. The modulation scheme is chosen in such a way that the data is transmitted reliably using the least possible spectrum; in other words the modulation method must be spectrally efficient. Spectral efficiency is influenced by the noise and propagation condition. The latter varies with time due to environmental changes, hence the modulation scheme should be able to adapt to the channel propagation variation. It should also support multi node communication, as several nodes exist in the Cognitive Radio network. Spectral utilization is

optimized by allowing rental (i.e., unlicensed) users to transmit and receive data over portions of spectra when primary (i.e., licensed) users are inactive. This is done in a way that the rental users (RUs) are invisible to the licensed users (LUs). In such a setting, the LUs are ordinary mobile terminals and their associated base stations. They thus do not possess much intelligence. The RUs, on the other hand, should possess the intelligence of sensing the spectrum and use whatever resources are available when they need them. At the same time, the RUs should give up the spectrum when an LU begins transmission.

Among several modulation schemes, Orthogonal Frequency Division Multiplexing (OFDM) is a suitable scheme that fulfils the requirements (Moseley, 2004). The cyclic prefix is inserted before the transmitted data that makes the scheme resistant to multipath fading, hence equalization is not required at the receiver. Multiple access can be applied by frequency division multiple access (FDMA) due to the carriers orthogonality, by time division multiple access (TDMA) where each node can access different OFDM symbols in a time frame, and multi carrier CDMA where node access is differentiated by otrthogonal codes.

Recently a method called Spectrum Pooling is reported (Weiss et al., 2004). In spectrum pooling public access to spectrum is provided without sacrificing the transmission quality of the actual license users. Multicarrier communication for CR has been suggested in (Weiss et al., 2004). The rationale is that any CR system needs to sense the spectrum, and this involves some sort of spectral analysis. Since FFT can be used for the spectral analysis (Cabric et al., 2004) and at the same time it can act as the demodulator of an OFDM, it has been suggested as a proper candidate for multicarrier-based CR systems. A number of short-comings of OFDM in its application in CR have been noted in (Weiss et al., 2004) and solutions to them have been proposed. The carriers located at the LU's band are deactivated, The sidelobes are reduced by deactivating more carriers adjacent to LU's band. Further sidelobes reduction is attained by applying windowing to the OFDM signal in time domain. If the number of available carriers for RU transmission is inadequate, due to lot of deactivations, then frequency hopping to another band is applied (Hoeksema et al., 2005). In (Jamin & Mahoenen, 2004) and (Negash & Nikookar, 2000) an interest of replacing Fourier transform by wavelet transform in multicarrier OFDM system is presented. The scheme gives better intercarrier interference (ICI) and inter symbol interference (ISI) reduction. Single carrier CR method using adaptive waveform is proposed in (Chakravarthy et al., 2005) and (Lee, 2002). In this method the transmitted signal is shaped in a way that its band does not contain the frequencies occupied by the LUs. The method takes the advantage of Fourier or wavelet transform.

B Theme and Organization of the Chapter

Adaptive OFDM and its combination with spectrum pooling for improving the quality of service (QoS) and efficient use of the spectrum are reviewed. Wave shaping and beamforming as alternative methods of efficient utilization of the spectrum are addressed and their importance in CR is stressed. MIMO system as an added value to enhance the QoS of CR is described. The chapter is structured as follows: Multi carrier adaptive OFDM with Fourier and Wavelet basis functions and its combination with spectrum pooling are analyzed in section 2. Wave Shaping with single carrier modulation for CR is explained in Section 3. Section 4 shows the role of MIMO transmission in enhancing the CR performance.

Beamforming as spatial interference avoidance technique in CR is discussed in Section 5. A summary of the chapter is given in Section 6.

2. Cognitive Radio with Adaptive Multicarrier Modulation and Spectrum Pooling

A strategy which is called *spectrum pooling* is proposed in (Weiss et al., 2004) where the public access is enabled to these bands without giving significant interference to the actual license owners. Spectrum pooling enables public access to spectral ranges of licensed frequency bands which are seldom used by overlaying a secondary rental user (RU) to an existing licensed user (LU). The LUs are radio systems authenticated to operate under licensed spectral bands. The RUs are intelligent CR systems that actively scan the landscape and opportunistically utilize available and unused resources. The RU relinquishes control over the resources (here spectrum) as and when the LU starts using them. To identify and utilize unused bands, the frequency bands of various radio systems (including licensed and rental users) are combined to obtain a common spectral pool. Cohabitation of LU and RU systems is actualized by shaping the transmission waveform of the RU in a way that it utilizes the unoccupied time-frequency gaps of the LU.

The pioneering work on the subject was conducted in (Weiss et al., 2004) which devised a spectrum pooling scheme using multi-carrier modulation (MCM) where individual sub-carriers adjacent to the occupied licensed user spectrum are deactivated for frequencies which are occupied by a licensed user. Fig. 1 illustrates a typical scenario. The combination of adaptive MCM and spectrum pooling can serve as a robust method to achieve a good quality of communication and efficient use of the spectrum. In the next subsections Fourier based (OFDM) and Wavelet based MCM with its application in Spectrum Pooling are explored.
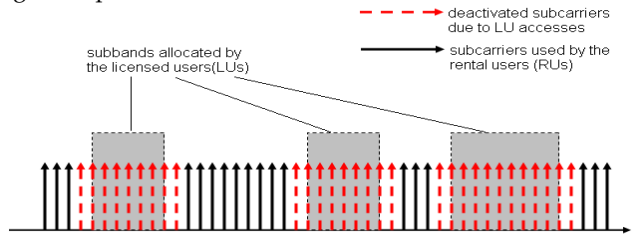

Fig. 1. Illustration of Spectrum pooling block diagram.

A Fourier Based Multicarrier OFDM

An obvious MCM candidate for CR systems is OFDM. With advantages of flexibility, ease of implementation and elegance in operation, OFDM is naturally suitable for CR system design. The short-coming of the OFDM solution is the large sidelobes of the frequency response of filters that characterize the channel associated with each of subcarrier. The sidelobes amplitude of the Fourier based OFDM can be observed from its power spectral density (PSD),

$$PSD(f) = \frac{1}{N_{FFT}} \left| \sum_{m=0}^{N_{FFT}-1} \sqrt{p_m}\, X_m \int_{-(1+\alpha)\frac{T_u}{2}}^{(1+\alpha)\frac{T_u}{2}} g(t) e^{-j2\pi(f-f_m)t} dt \right|^2 \tag{1}$$

where N_{FFT} is the FFT point, T_u is the useful signal duration, $g(t)$ is the window function, a is the roll off factor of the window, f_m is the frequency on carrier m. p_m and X_m are the allocated power and the symbol from QAM or PSK mapping on subcarrier m, respectively.

The large side lobes result in significant interference among the subcarriers that originate from different RUs and between LUs and RUs. Equation (1) shows that parameters like allocated power (p_m), symbol (X_m) amplitude and window ($g(t)$) can be set to resolve this problem. Other elements which can be varied to improve the performance of RUs are the channel coding parameters and carrier frequency. The utilization of the parameters to reduce the PSD in the LU's band and elements to improve the quality of service (QoS) of RUs will be further explored in the next following sub subsections.

A.1 Spectrum Shaping by Time Domain Windowing

In (Weiss et al., 2004) extension of each OFDM block with long cyclic prefix and suffix samples and application of some windowing to reduce the side-lobes of the subcarrier channels are suggested. Obviously, this solution is at the cost of bandwidth loss because excessive time should be allocated to cyclic extensions that otherwise could be used for data transmission. Further in (Weiss et al., 2004), interference reduction method is extended by deactivating subcarriers located adjacent to the licensed system (LS) which provides a kind of shield to the LUs. Due to the channel propagation influence, in OFDM bit errors are typically concentrated in a set of severely faded subcarriers. This fact indicates that there is possibility that the quality of service can not be maintained. Meanwhile, as mentioned before, the carriers resource is limited due to deactivation of RU carriers on the occupied LU band. Having adaptive OFDM in CR is thus a solution. The heavily faded and noisy subcarriers are excluded from transmission to improve the overall bit error rate (BER) of the system, and the loss of throughput is counteracted by applying higher order modulation modes to the subcarriers which have better signal to noise ratios (SNR) or BER.

Commonly used window type is the raised cosine window, which is defined as (Weiss et al., 2004):

$$g(t) = \begin{cases} \dfrac{1}{T_u} & 0 \le |t| \le \dfrac{T_u(1-\alpha)}{2} \\ \dfrac{1}{2T_u}\left\{1+\cos\left[\dfrac{\pi}{\alpha T_u}\left(|t| - \dfrac{T_u(1-\alpha)}{2}\right)\right]\right\} & \dfrac{T_u(1-\alpha)}{2} \le |t| \le \dfrac{T_u(1+\alpha)}{2} \\ 0 & Otherwise \end{cases} \tag{2}$$

where a is the roll off factor, and T_u is the useful OFDM symbol interval (without guard interval). The shape of the window in the region of $\dfrac{T_u(1-\alpha)}{2} \le |t| \le \dfrac{T_u(1+\alpha)}{2}$ determines how rapid the OFDM spectrum goes down to zero. The PSD will hit zero on the frequencies in the interval of $1/T_u$. Bartlett window is defined by equation (3), (Nikookar & Prasad, 1997b)

while *Better than Raised Cosine* (BTRC) Window which gives the lowest sidelobes among the three windows is defined by equation (4) (Tan & Beaulieu, 2004).

$$g(t) = \begin{cases} \dfrac{1}{T_u} & 0 \le |t| \le \dfrac{T_u(1-\alpha)}{2} \\ \dfrac{1}{2T_u} - \dfrac{1}{T_u}\left[\dfrac{|t|}{\alpha T_u} - \dfrac{1}{2\alpha}\right] & \dfrac{T_u(1-\alpha)}{2} \le |t| \le \dfrac{T_u(1+\alpha)}{2} \\ 0 & Otherwise \end{cases} \tag{3}$$

$$g(t) = \begin{cases} \dfrac{1}{T_u}, & 0 \le |t| \le \dfrac{T_u(1-\alpha)}{2} \\ \dfrac{1}{T_u}e^{(-2\ln 2/\alpha T_U)[|t|-(T_U(1-\alpha)/2)]} & \dfrac{T_u(1-\alpha)}{2} \le |t| \le \dfrac{T_u}{2} \\ \dfrac{1}{T_u}\left[1 - e^{\left(\frac{-2\ln 2}{\alpha T_u}\right)\left[|t|-\frac{T_u(1-\alpha)}{2}\right]}\right], & \dfrac{T_u}{2} \le |t| \le \dfrac{T_u(1+\alpha)}{2} \\ 0 & Otherwise \end{cases} \tag{4}$$

A window with lower sidelobes than BTRC which is called the *Flipped- Inverse Hyperbolic Secant* (farcsech) window is proposed in (Assalini & Tonello, 2004). The window is designed according to (5),

$$g(t) = \begin{cases} \dfrac{1}{T_u}, & 0 \le |t| \le \dfrac{T_u(1-\alpha)}{2} \\ \dfrac{1}{T_u}\left[1 - \dfrac{1}{\alpha T_u \gamma}arcsech\left(\dfrac{1}{\alpha T_u}\left(\dfrac{T_u(1+\alpha)}{2} - |t|\right)\right)\right], & \dfrac{T_u(1-\alpha)}{2} \le |t| \le \dfrac{T_u}{2} \\ \dfrac{1}{T_u}\left[\dfrac{1}{\alpha T_u \gamma}arcsech\left(\dfrac{1}{\alpha T_u}\left(|t| - \dfrac{T_u(1-\alpha)}{2}\right)\right)\right], & \dfrac{T_u}{2} \le |t| \le \dfrac{T_u(1+\alpha)}{2} \\ 0 & Otherwise \end{cases} \tag{5}$$

where $\gamma = 2\ln(\sqrt{3}+2)/(\alpha T_u)$. According to (Muschalik, 1996) the duration of OFDM signal should be $2T_u$ in order to complete a total $2N_{FFT}$ samples to preserve the orthogonality, and zeros are added on the region outside the window $g(t)$. The window designs by equations (2)-(5) affect the transmitted signal, as the consequence an error floor will be formed. Fig. 2(a) shows an example Bartlett window design according to (3) which affects the transmitted signal.

It is emphasized in (Witrisal, 2002) that the applied window must not influence the signal during its effective period to avoid error floor. In order to fulfil this requirement the window forms are expanded, i.e., new T_u is becoming $2T_u$, and α is restricted in the range of $0 \le \alpha \le$ *(1- T_{GI}/ T_u)/2*, where T_{GI} is the guard interval duration. The orthogonality is preserved since the PSD hits zeros in the interval of $1/(2T_u)$ at the cost of longer duration of OFDM signal $(2T_u (\alpha +1))$. As a solution the window is truncated to fit $2T_u$ OFDM duration,

as depicted in Fig. 2(b). The orthogonality is preserved by applying the rectangular receiver filter with duration T_u , implemented by DFT (Witrisal, 2002). In this way there is more freedom in choosing window in the region of $T_u(1-\alpha)\leq|t|\leq T_u$ as long as its PSD on the symbol boundaries is low, e.g Gaussian window or half sine window (Nikookar & Prasad, 1997(a)).
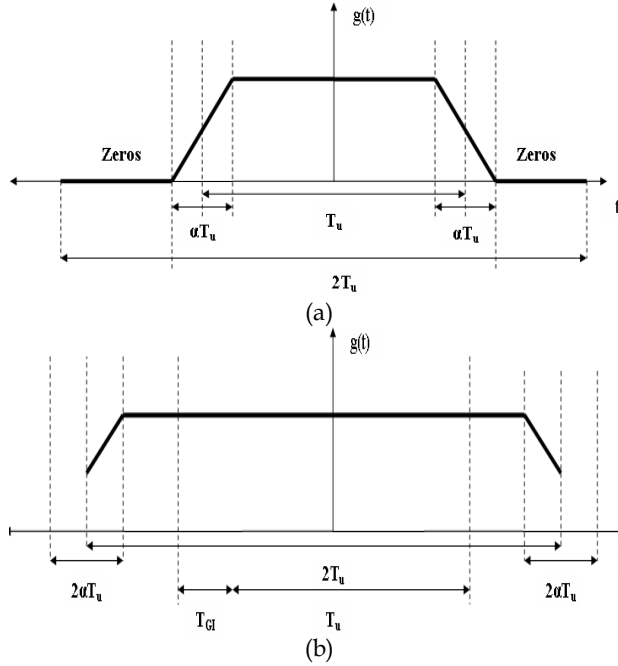

Fig. 2. (a) Window design with influence to the transmitted signal (b) Window design without influence to the transmitted signal.

It should be emphasized that applying only windowing is not enough to reduce interference to LU. A more powerful technique to suppress OFDM sidelobes is carriers deactivation. Deactivation of carriers adjacent to the licensed band provides flexible guard bands which will make the PSD sidelobes of RU's OFDM signal on the licensed band lower. Applying only one of the methods may not be sufficient, therefore combining both methods (windowing and adaptive deactivation of carriers adjacent to licensed bands) is suggested to make the PSD sidelobes even lower in the region of LU's band. The loss of throughput due to carriers deactivation and long OFDM symbol duration can be compensated by applying adaptive bits allocation. The next subsubsection will explore the application of adaptive bit allocation in CR system including its available algorithms.

Recently in (Harris & Kjeldsen, 2006) overlapped OFDM symbol transmission with long symbol duration is proposed to counteract the throughput loss. The scheme is described in Fig. 3 . The delay between one OFDM symbol to another should be designed in such a way that no intersymbol interference occurs, which means the next OFDM symbol should start after the end of the previous useful data part of an OFDM symbol. If zeros are inserted at

the prefix of suffix (outside the T_{GI} and T_U area as depicted in Fig. 2(b)), the useful data of the next OFDM symbol should start after the last useful data of the previous OFDM symbol.
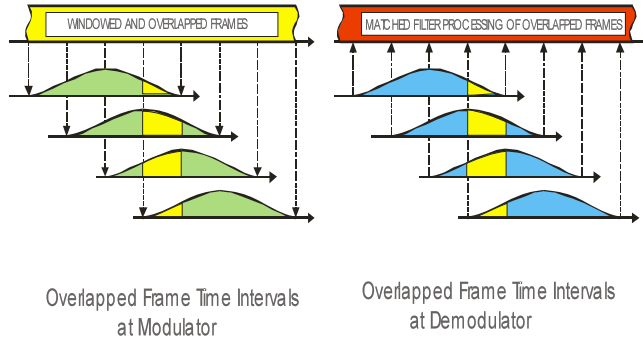

Fig. 3. Overlapped OFDM symbol with zero ISI.

A.2 OFDM Sidelobes Reduction Techniques

In this section a technique to reduce sidelobes by manipulating the composition of X_m in (1) is described. The results in (Yamaguchi, 2004); (Brandes et al., 2005a); (Brandes et al., 2005b); (Brandes et al., 2006); (Berthold et al., 2007); (Cosovic et al., 2005); (Cosovic et al., 2006); (Cosovic & Mazzoni, 2006); (Cosovic and Mazzoni, 2007); (Mahmoud & Arslan, 2008); (Padagarai et al., 2008a); (Padagarai et al., 2008b) show that a significant sidelobes reduction gain can be obtained.

A.2.1 Sidelobe Suppression by Cancellation Carriers Insertion

In (Yamaguchi, 2004);(Brandes et al., 2005a);(Brandes et al., 2006) instead of having deactivated carriers adjacent to the LU band, these carriers are used to cancel out the sidelobes on the LU band. The concept is to find the proper amplitude of the cancellation carriers in such a way that the resulting sidelobes be as low as possible. The optimization can be formulated as:

$$|X(f)|^2 = \frac{1}{N_{FFT}} \left| \sum_{\substack{m=1 \\ m \neq cc}}^{N_{FFT}} \sqrt{p_m}(X_m) \int_{-(1+\alpha)T_u}^{(1+\alpha)T_u} g(t)e^{-j2\pi(f-f_m)t}dt + \sum_{\substack{cc=cc1 \\ cc \neq m}}^{ccN} \sqrt{p_{cc}}(X_{cc}) \int_{-(1+\alpha)T_u}^{(1+\alpha)T_u} g(t)e^{-j2\pi(f-f_{cc})t}dt \right|^2 \tag{6}$$

$$subject \quad to \quad |X(f_{LU})|^2 < P_{Threshold,} \quad and \quad \sum_{m=1}^{N_{FFT}} p_m = P_T$$

where ccN is the total number of sidelobes cancellation carriers, and Xc_{cc} is the symbol on the cancellation carrier position with the range within $cc1,\ldots,ccN$. The power constraint for all subcarriers including the cancellation carriers remains thus, no extra energy is added. The possible high power required by the cancellation carriers can be compensated from the unused power due to the carriers deactivation.

We introduce an alternative optimization problem by simplifying it to set of ccN linear equations to solve ccN cancellation carrier amplitude values. Fig. 4 provides a simple example of how the cancellation carriers with this new optimization problem work. The

spectrum is derived from 9 subcarriers (including the cancellation carriers) and uses a rectangular window which produces a spectrum from the sum of the sinc functions. One cancellation carrier is located at the left edge of the RU band adjacent to the LU band and another one is at the right edge. They are signed by the dark vertical arrows. If we take a look at the OFDM spectrum with carriers deactivation (one carrier at the left and one carrier at the right edge) denoted by the solid curve, we can see that the maximum sidelobe occurs at the frequency point half of carrier spacing from the deactivated carrier which, in Fig. 4 is indicated by the coordinate position (X:-4.5,Y:-0.1564). The cancellation carriers can be weighted in such a way that the sum of their spectras will nullify the OFDM spectrum at that particular coordinate position (e.g. by having a cancellation carrier spectrum coordinate $(X_{c1}=X_{c2}=-4.5, Y_{c1}=0.1477, Y_{c2}=0.0087)$).
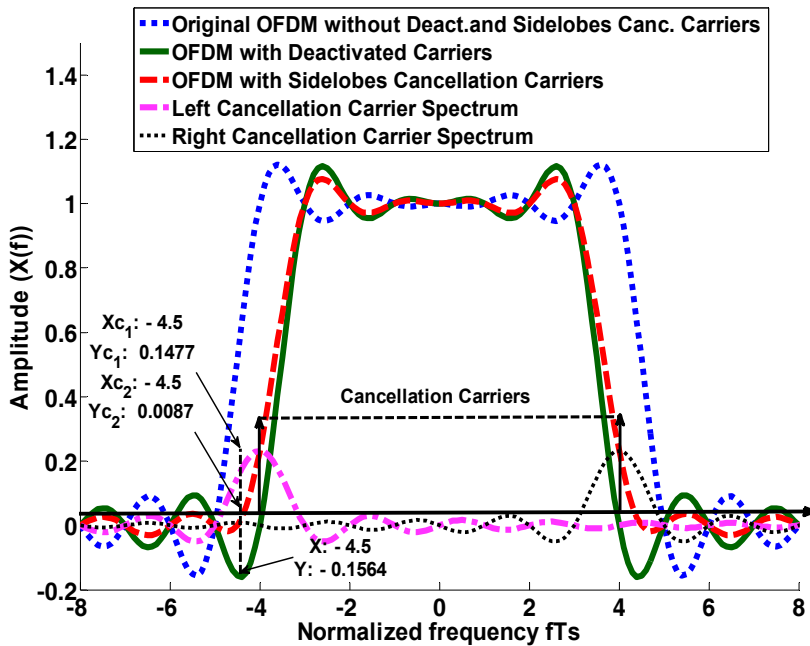

Fig. 4. The OFDM spectrum amplitude with the effect of the deactivated carriers or cancellation carriers.

The same technique is also applied to the sidelobes reduction on the right side. The resulting spectrum derived from using the cancellation carriers technique is depicted by the dashed lines. The sidelobes are lower than the sidelobe of the OFDM with the deactivated carriers.
The cancellation carriers can be combined with windowing (Brandes et al., 2005b);(Berthold et al., 2007);. Different windows provide different sidelobe levels. Combining a window that gives very low sidelobes (e.g. the farchsech window) with the cancellation carriers technique will lead to a very significant sidelobe (interference) reduction in the LU band. The results in (Brandes et al., 2005b) have shown that the combination of 2 cancellation carriers on each

side adjacent to LU band and raised cosine window with roll off factor (*a*) 0.2 gives more than 30 dB sidelobes reduction compared to the conventional OFDM.

A.2.2 Sidelobe Suppression by Subcarrier Weighting

Subcarrier weighting is introduced in (Cosovic et al., 2005); (Cosovic et al., 2006);. The weighting can be seen as varying the allocation power p_m on each carrier m which, obviously, according to (1) will affect the spectrum amplitude. By selecting the appropriate weight for each of the carriers the desired low sidelobes can be obtained. The sidelobe reduction by means of carriers weighting is attained at the expense of the bit error rate (BER) versus signal to noise ratio (SNR) degradation (Nikookar & Prasad, 2000), since the weighting will give unequal amounts of transmission power on each of the OFDM carriers. The weighting method was originally used in (Nikookar & Prasad, 2000) to counteract the high peak to average power ratio (PAPR) problem in OFDM.

A.2.3 Sidelobe Suppression by Multiple Choice Sequence

In (Cosovic & Mazzoni, 2006) a multiple choice sequence is introduced, where the multiple choice sequence abiding by the symbol constellation approach, the interleaving approach and the phase approach are discussed. The symbol constellation approach is applied by mapping the symbol X_m in (1) into $X'_m (l)$ which is still derived from one of the symbols set after the QAM or PSK mapping that was previously used to produce X_m. The index l ranges from $l=1,2,…,L$, where L is the number of possible alternative values. This means that the bits can be mapped into L possible symbols. By putting the N_{FFT} symbols of $X'_m (l)$ into (1), and observing from all L possible PSDs, the one with the lowest sidelobe is then chosen for the transmission. An example of simple mapping from X_m into $X'_m (l)$ is given in (Cosovic and Mazzoni, 2006). The *CS* points of QAM or PSK are numbered as $0,1,….,CS$-1, hence the symbol X_m is assigned to the number I_m ($I_m \in \{0,1,...,CS-1\}$) according to its point position. The $X'_m (l)$ will be the symbol value on the new number $I_m(l)$ which is derived from the addition of the assigned number I_m with the random number $D_m(l)$ which has L different possible values (Cosovic & Mazzoni, 2006),

$$I_m(l) = I_m + D_m(l) \qquad\qquad (7)$$

By transmitting the information about the set index information l, the receiver can decode the received signal and reconstruct the original X_m symbol. By means of the interleaving approach, the N_{FFT} X_m symbols are interleaved by L different interleaver rules,

$$X'(l)=\Pi(l)X, \qquad\qquad (8)$$

where $\Pi(l)$ is the permutation matrix that will do the interleaving, X is the symbol vector, and X' is the symbol vector after interleaving. The matrix $\Pi(l)$ has L different possible matrix values. The permutation matrix $\Pi(l)$ will be available at the transmitter and the receiver. Among L different possible OFDM spectrum using a set of symbols $X'(l)$, the one with the lowest sidelobes is chosen for the transmission. The receiver will use the deinterleaver permutation matrix $\Pi^{-1}(l)$ to recover the transmitted data upon receiving the side information index l. In the phase approach each phase of symbol X_m is shifted by random

phase $\theta_m(l)$ with L different possible values. The phase shifted symbol $X'_m(l)$ becomes (Cosovic & Mazzoni, 2006),

$$X'_m(l) = X_m e^{j\theta_m(l)} \tag{9}$$

where $\theta_m(l)$ is in the range of between 0-2π. In the vector notation, the phase shifted signal $X'(l)$ is derived from point to point multiplication between the original signal X and the complex number vector derived from $e^{j\theta(l)}$. As in the constellation and interleaver approach, in the phase approach, among L possible sets of vector $\theta(l)$, the one that produces the lowest sidelobes is chosen and applied to the symbol vector X. The index information l is transmitted so that the receiver can reconstruct the symbol vector X. Apart from the random shift, random amplitude shift can also be applied, but in such cases the BER performance will be affected. Another approach is to only apply the multiple choice sequence technique to the carriers adjacent to the LU band, since these carriers mostly affect the sidelobes magnitude (Cosovic & Mazzoni, 2006),.

A.2.4 Sidelobe Suppression by Adaptive Symbol Transition

Sidelobes Suppression by time domain approach has been introduced in (Mahmoud & Arslan, 2008). Time domain symbol transition is inserted between two OFDM symbol, and the symbol transition value is obtained in such a way that the signal spectrum or the Fourier transform of the two consecutive time domain OFDM symbol (including the symbol transition) will have low sidelobes. The objective is to minimize the spectrum resulting from two consecutive time domain OFDM symbol without the symbol transition added by the spectrum contributed from the symbol transition on the LU band (Mahmoud & Arslan, 2008). If the OFDM based cognitive radio signal contribution on the LU is described as :

$$I = F_{LU} \begin{bmatrix} y^n \\ a^n \\ y^{n+1} \end{bmatrix} \tag{10}$$

where I is the signal part of the cognitive radio signal in LU band, F_{LU} is the Fourier transformation matrix for the signal in LU band, y^n is the nth time domain OFDM symbol including the guard interval, a^n is the additive symbol transition, and y^{n+1} is the following OFDM symbol, then the optimization problem will be to find the vector a^n in such a way that I will be as small as possible. The symbol transition scheme is depicted in Fig. 5.

The symbol transition will not introduce any inter symbol interference (ISI) but the transmission duration becomes longer, which means that the bit rate is reduced. The search for the time domain symbol transition values are applied by the linear least squares problem with a quadratic inequality constraint.
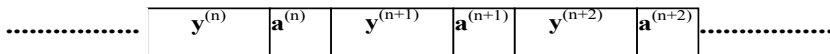

Fig. 5. Adaptive symbol transition scheme.

A.2.5 Sidelobe Suppression by Frequency Domain Additive Signal

In (Cosovic & Mazzoni, 2007) sidelobe suppression for OFDM based cognitive radio can be achieved by mapping the data symbol using QPSK modulation mode and add the frequency domain symbol vector **X** with random values vector **d** in such a way that the resulting PSD in the LU region be as small as possible. The optimization problem becomes as described in (11),

$$|X(f)|^2 = \frac{1}{N_{FFT}} \left| \sum_{m=1}^{N_{FFT}} \sqrt{p_m} (X_m + d_m) \int_{-(1+\alpha)T_u}^{(1+\alpha)T_u} g(t) e^{-j2\pi(f-f_m)t} dt \right|^2 \tag{11}$$

$$subject \quad to \quad |X(f_{LU})|^2 < P_{Threshold}, \quad and \quad \sum_{m=1}^{N_{FFT}} p_m = P_T$$

where d_m is the additive signal on subcarrier n. If $X'_m = X_m + d_m$ then in order not to destroy the cognitive radio BER performance and keep the total allocated power constant, the value of d_m is restricted to a circle with magnitude $|d_m| \le 0.7$ (Cosovic & Mazzoni, 2007). In this way, according to the QPSK gray mapping, the value of X'_m will not give decision error at the receiver, except in the severely noise and faded channel. Simulation results in (Cosovic & Mazzoni, 2007) have shown about 15 dB sidelobes reduction could be achieved with the cost of 2.8 dB SNR degradation as the upper limit value of $|d_m|$ is 0.7.

A.2.6 Sidelobe Suppression by Constellation Expansion

The most recent technique in OFDM sidelobe suppression has been proposed in (Padagarai et al., 2008a), where the M-PSK modulation mapping is expanded to 2xM-PSK mapping in order to have lower OFDM sidelobe. The original symbol after M-PSK mapping which is located at point a of M available PSK symbol points is transformed into two possible new points a_1 and a_2, where $a_1=-a_2$. By combining all possible new OFDM symbol set on the 2x M-PSK points through all OFDM carriers, the one set which gives the smallest sidelobes will be chosen for the transmission. No additional information is required by the receiver for data detection, since the receiver has the knowledge of all possible a_1 and a_2 positions. Therefore, the receiver can detect the data by calculating the minimum Euclidean distance between the received symbol and the reference points on 2xM- PSK mapping, and accordingly translates the result to the M PSK mapping. Fig. 6 shows an example on how the points from QPSK are expanded to 8-PSK mapping. In this way there are two possibilities of mapping a data symbol, and accordingly the constellation expansion will add extra randomness to X_n in (1) with various possible sidelobes magnitude.

The one which gives the smallest sidelobes will be chosen to be transmitted. The simulation results in (Pagadarai et al., 2008a) have shown that the sidelobes suppression technique by the constellation expansion from QPSK to 8-PSK gives only slight BER degradation. With Constellation Expansion technique about 9 dB sidelobes suppression can be attained (Padagarai et al., 2008a). The work has been extended in (Padagarai et al., 2008b) by combining the constellation expansion technique with the cancellation carriers inserted on the carriers adjacent to the LU band. Extra sidelobes reduction has been achieved without having high value of cancellation carrers magnitude, due to prior sidelobes reduction by the constellation expansion technique. According to (Padagarai et al., 2008b) the combination

between constellation expansion and one sidelobes cancellation carrier on each side adjacent to LU band gives around 16 dB sidelobes reduction.
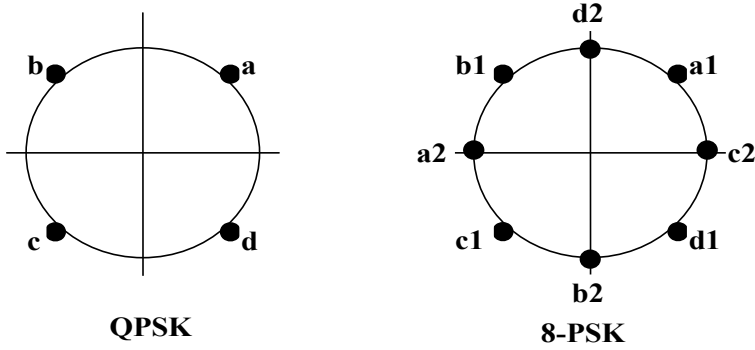

Fig. 6. Constellation Expansion from QPSK to 8-PSK mapping.

A.3. Adaptive Bit Allocation (Bit Loading)

Referring to (1) the sidelobes to LU's band can be lowered by applying the low constellation size (e.g BPSK) to the symbols adjacent to LU's band. Another advantage of the low constellation size is its resistance to noise and interference compared to high constellation size. Hence, although the sidelobes of LU to RU's band exists, the BER of the RU's carriers adjacent to LU's band can be better preserved.

Adaptive bit loading is one of adaptive OFDM methods which allocates bits to subcarriers efficiently by taking estimated channel fading and noise variance information, and accordingly selecting the subcarriers modulation modes By utilizing the unused power due to carriers deactivation on the subband occupied by the LUs and allocating them to some of the other carriers intelligently, according to the adaptive bit loading algorithms, reliable communications without sacrificing the legacy system can be achieved. Transmitting more bits on a carrier requires more power. This power can be derived from the unallocated ones.

As a consequence of applying adaptive bit loading, the receiver needs information about the number of bits allocated on each carrier in order to decode the signal properly. This information can be derived blindly by examining the received signal (Hanzo et al., 2002); (Reddy et al., 2003); or from dedicated signaling symbols. An example of the signaling symbol is depicted in Fig. 7, where the allowable constellation size is restricted to the set of {0, 2,4,16 and 64}, (Budiarjo et al., 2007a). The inner rectangle which is the decision area for 0 modulation mode (no symbol is transmitted) can be made smaller up to a distance of about *3d/4* from the other reference symbols, where *d* is the distance between outer points and the no symbol point shown in Fig. 7. This is due to the fact that a decision error (e.g no symbol transmitted but BPSK or other modulation mode is decided) is only caused by the noise (no contribution from channel fading). Fischer-Hueber algorithm allocates the bits with the intention to minimize the BER on each carrier. It tries to guarantee equal SNR on every carrier. The bits are assigned as (Barreto and Furrer, 2001):

$$R_m = \frac{1}{K}\left[R_T + \log_2\left(\prod_{n=0}^{K-1}\frac{\sigma_n^2}{\Phi_n}\right) - K\log_2\left(\frac{\sigma_m^2}{\Phi_m}\right)\right], \tag{12}$$

where R_m is the number of allocated bits on carrier m, K is the number of active carriers, and R_T the total target bits per OFDM symbol. In the beginning value of K is equal to N_{FFT}, after applying (12) on each carrier the nonpositive values of R_ms are replaced by zeros, and the number of active carriers K is substracted by the number of carriers with nonpositive R_m values. The loop of calculating (12) continues until all R_ms are nonnegative. Each of the R_m is quantized into an integer number R_{Qm}. If the sum of the R_{Qm} is not equal to the target rate, addition or subtraction of bit is applied. It starts from the carrier with the lowest deviation of R_{Qm} to its R_m (for the addition) or the carrier with the highest deviation (for the subtraction. The addition or subtraction is iteratively applied bit per bit until the target rate is fulfilled. Another adaptive bit allocation algorithm is the Chow algorithm introduced in (Chow et al., 1995). The goal is to maximize the allowable noise margin while the minimum desired BER is still achieved. The bits are assigned as follows (Chow et al., 1995):

$$R_m = \log_2 \left[1 + \frac{p_m \Phi_m}{\sigma_m^2 \cdot (\Gamma + \gamma_{m\,arg\,in})} \right],$$ (13)

where γ_{margin} is the additional amount of noise (in dB) that the system can tolerate, while still achieving the minimum desired bit error rate requirement.
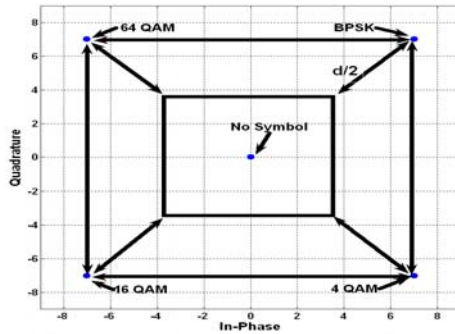

Fig. 7. An example of signaling symbol points for adaptive bit loading with allowable M= {0, 2, 4, 16, and 64} (Budiarjo et al., 2007a).

The SNR gap Γ is a constant which estimates the difference between channel capacity and the actual capacity usage by the transmission scheme, which is defined as:

$$\Gamma = \frac{1}{3} \left[Q^{-1} \left(\frac{SER}{4} \right) \right]^2 .$$ (14)

SER is the expected symbol error rate, and $Q(x)$ is the function $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{\frac{-t^2}{2}} dt$.

Besides allocating the bits carrierwise, the bits can be allocated groupwise according to the average channel gain of the carriers within a certain group. This method reduces the

signaling symbols required to determine the modulation mode on each carrier, as the required modulation information of a group of N_{cg} carriers is the same in a signaling symbol.

A Simple Blockwise Loading Algorithm (SBLA) is proposed in (Gruenheid et al., 2001). The algorithm divides the subcarriers into N_B blocks of N_{cb} adjacent subcarriers. The same modulation level is assigned to subcarriers within one block. The SNRs of neighboring sub carriers are highly correlated; hence no significant performance degradations are expected by introducing blocks of subcarriers, which are smaller than the coherence bandwidth of the channel. The choice of the modulation level for the specific block of subcarriers is based on the mean SNR of the block. Initially, the SNR grid is predetermined as e.g depicted in Fig. 8 (Gruenheid et al., 2000). The grid contains a list of differences of the SNR thresholds, required to switch between the modulations levels of the supported PSK/QAM modulation alphabets for a specific block. The middle point (center of 8-PSK) is placed on the overall SNR average of OFDM frame.
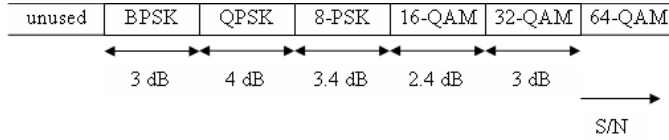

Fig. 8. SNR grid for Simple Blockwise Loading Algorithm (SBLA) (Gruenheid et al., 2000).

In (Lei and Zhang, 2003) the Fischer Hueber algorithm is modified to determine the allocated bits on a block of subcarriers. The bits are distributed as:

$$R_i = \frac{R_T}{N'_B} + \frac{1}{N'_B} \cdot \log_2 \left(\frac{\left(|H_i|^2 \right)^{N'_B}}{\prod_{k \in \psi} |H_k|^2} \right) \tag{15}$$

where R_i is the allocated bits to carriers block i, N'_B is the number of active blocks, H_i is the average channel gain in block i, and ψ is the set of active carriers block.

Like in the subcarrierwise Fischer –Hueber algorithm, after several iterations, the value of all bit allocations are non negative, but if the target rate is not yet achieved, the addition or substraction of bit to each carrier based on the deviation between the quantized bit allocation and unquantized bit allocation for Chow, SBLA and groupwise Fischer algorithm, is applied until the target rate is satisfied.

In (Budiarjo et al., 2005) these bit loading algorithms are evaluated and compared by utilizing the 2x1-D Wiener filter as the channel estimation technique (Budiarjo et al., 2006a). In (Budiarjo et al., 2006b) the performance of combination of adaptive bit loading algorithms and windows with the design according to Fig. 2.b are evaluated. Simulation results in (Budiarjo et al., 2006a) and (Budiarjo et al., 2006b) show that carrierwise bit loading outperforms the groupwise bit loading, since the method maps the bit allocation to each carrier more accurately. If throughput is a critical issue, the groupwise bit loading is a better candidate than carrierwise bit loading. This is because in adaptive OFDM, the signaling symbols required by the receiver to recognize the modulation mode of each carrier, reduces the throughput.

A.4. Adaptive Coding and Frequency Hopping

An element which does not influence the level of interference to LU's band but is needed to boost or improve the quality of service (QoS) of CR system is the channel coding. As it can be noticed, none of the parameters shown in (6) refers to channel coding element. Most of references such as (Mutti et al., 2004); (Jetlund et al., 2002); (Vucetic, 1991); (Keller et al., 2000); discuss the adaptation of channel coding by varying the rate according to the estimated channel gain. The coding rate level is decreased as the channel fading power is increased (Mutti et al., 2004).

In the situation where due to LU's access and bad channel condition, too many carriers are deactivated, and therefore the frequency hopping is suggested. The scanning module will give information to the front end to switch transmission to a certain carrier frequency where available spectrum in that region is adequate to achieve the target quality of service (QoS). The performance of OFDM based CR using Wiener filter channel estimation with frequent frequency hopping at GSM 900 MHz channel model is evaluated in (Budiarjo et al., 2007c), where the channel models are based on ETSI GSM specification for rural area and urban area power delay profiles.

B. Wavelet Based Multicarrier OFDM

In traditional implementations of Multicarrier OFDM (MC-OFDM), the generation and modulation of the sub-channels are accomplished digitally using Fourier bases. In this subsection another modulation technique is introduced. The conventional Fourier-based complex exponential carriers of OFDM are replaced with orthonormal wavelet packet (WP) bases for use in cognitive radio environments. The WP bases are derived from perfect reconstruction two-band FIR filter bank solutions (Jamin and Mahoenen, 2004); (Negash & Nikookar, 2000); (Bouwel et al., 2000); (Jensen & Harbo, 2001).

Cohabitation of the WP-MCM based CR systems with existing licensed users is actualized by shaping its transmission waveform and by adaptively activating or deactivating sub-carriers in a way that it utilizes the unoccupied time-frequency gaps of the LU. The idea is to dynamically sculpt the transmission signal so that it has no or very little time-frequency components competing with the LU. This way the CR can seamlessly blend with the LU operation. The steps to generate the WP-MCM transmission signal are elucidated in Fig. 9. The level and the number of bases generated is given as $M = 2^J$. A level 3 tree structure generates up to 8 wavelet packet carriers and duals each.

B.1. Generation of Wavelet Packet Bases and Their Duals   through Two Channel Filter Bank Analysis

The WP-MCM transmission is realized by replacing orthonormal complex exponential basis functions, as used in OFDM systems, with orthonormal wavelet packet basis functions. The wavelet packet bases and their dual bases are derived from perfect reconstruction two-band FIR filter bank solutions from multistage tree-structured paraunitary filter banks derived by cascading 2-channel filter banks.

B.1.1 Generation of Wavelet Packet Sub-Carrier Bases

The wavelet packet sub-carriers (to be used at the transmitter end) are generated through a multichannel filterbank consisting of cascaded two-channel filterbanks applying the

synthesis filters (*H'* and *G'*). The process known as the synthesis procedure consists of binary interpolation (up-sampling) by 2, filtering and recombination at each level.

To demonstrate the process of generation of wavelet packet bases, here we consider a cascaded level-3 tree structure as shown in Fig. 10(a). Such an arrangement can give rise to eight wavelet packet bases. This procedure is simplified further by applying the identities shown in Fig. 10(b). These expressions called the noble identities are popularly applied in the implementation of multi-rate systems. And in Fig. 10(c) the equivalent eight-channel system after the application of the noble identities is shown. In general, to generate *M* bases or sub-carrier waveforms, a level-*J* tree has to be constructed.

B.1.2 Generation of Wavelet Packet Dual Bases

The wavelet packet duals (to be used at the receiver end) are obtained from multichannel filter bank analysis too, though the processes are reversed. The duals are obtained from the analysis filters (*H* and *G*) through the analysis procedure which consists of filtering, decimation (downsampling) by 2 and decomposition at each stage. Fig. 11 illustrates the generation of 8 wavelet packet duals from a level-3 tree cascaded filter bank.
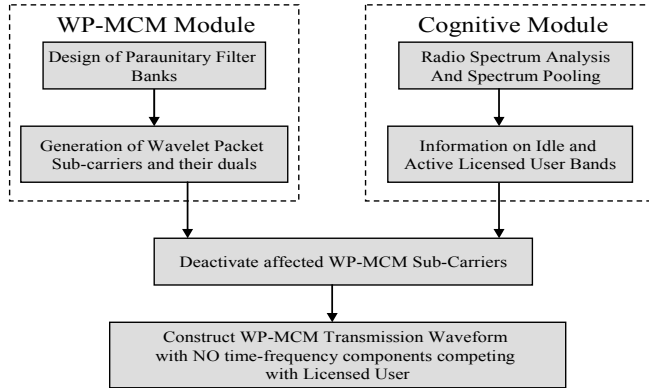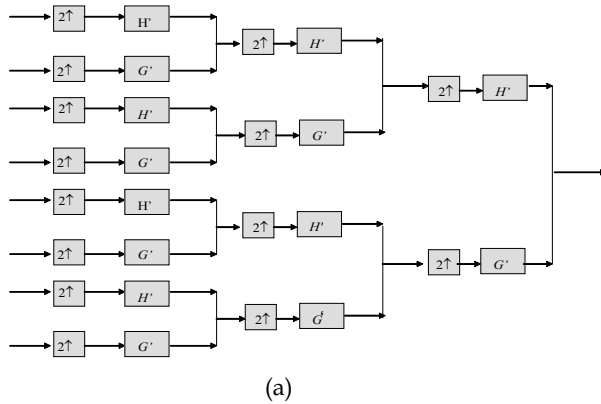

Fig. 9. Fundamental Blocks of WP-MCM based Cognitive Radio system.
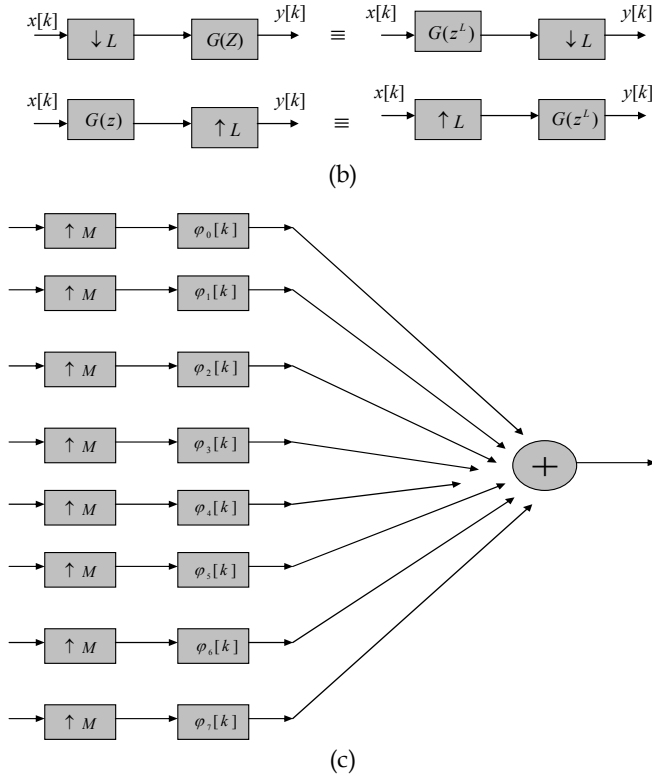

(a)

(b)

(c)

Fig. 10. Generation of wavelets. (a) A level-3 tree gives 8 wavelet packet bases. The up arrows represent interpolation by 2. H' and G' denote the frequency responses of the low and high pass reconstruction filters, respectively; (b) Noble identities; (c) Equivalent 8-channel system after applying the noble identities, where φ' is equivalent reconstruction filter as results of the convolution combination H' and G' according to the tree lines. $M=8$ for the considered example.

B.1.3 Wavelet Design Considerations for WP-MCM application

Choosing the right wavelet is a delicate and at times even an overwhelming issue. In theory any time and frequency limited function can be utilized. However in practice, the wavelet bases cannot be arbitrarily chosen and instead have to satisfy a number of requirements.

With regard to the applicability to WP-MCM systems, the desirable properties may be listed as follows:

- The bases must be time and band limited
- The bases and their duals must be orthogonal to one another to enable perfect reconstruction.
- The bases must be orthogonal to one another in order to have unique demodulation.
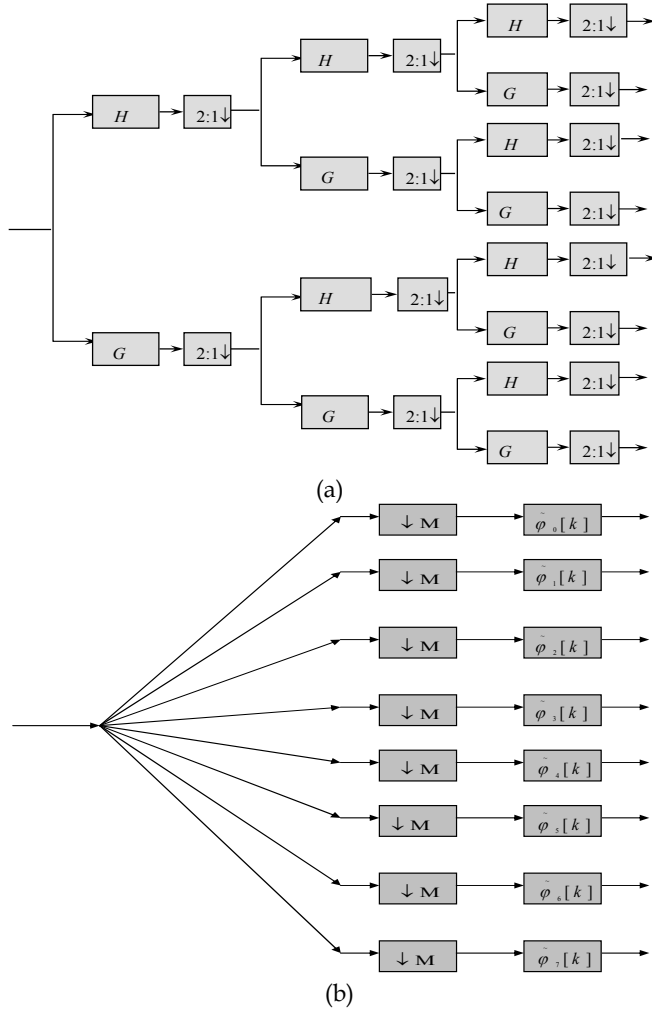
(a)

(b)

Fig. 11. Generation of wavelet Duals. (a) A level-3 tree gives 8 wavelet packet dual bases. The down arrows represent decimation by 2. H and G denote the frequency responses of the low and high pass decomposition filters, respectively; (b) Equivalent 8-channel system after applying the noble identities, where φ is equivalent decomposition filter as results of the convolution combination H and G according to the tree lines $M=8$ for the considered example.

Considering these requirements, among several available wavelet such as: Coifflets, Daubechies, Haar, Symlets, the most suitable wavelet is the family of maximally frequency filter banks derived using a modified Remez exchange algorithm. It provides degree of freedom to shape its sidelobes by varying its regularity order, transition bandwidth and the filter length. The longer the filter length, the more degree of freedom is available in

designing the frequency selectivity and also the regularity order of the signal. The smaller the transition bandwidth the more frequency selective the filter will be. The regularity order corresponds to the smoothness of the signal sidelobes. The higher the regularity order, the smoother or the lower the sidelobes of the signal will be. More on this subject can be found in (Rioul and Duhamel, 1994).

In (Lakshmanan et al., 2007) WP-MCM with Remez exchange algorithm as CR system is evaluated and compared with Fourier based MC-OFDM CR system in the presence of LU signal in Additive White Gaussian Noise (AWGN) channel. The results show within a certain Remez parameters value, WP-MCM outperforms the Fourier based MC-OFDM.

C. Interpolated Tree Orthogonal Multiplexing

In the WP-MCM technique the filter banks perform the dual role of shaping the spectrum as well as interpolating in time series. A more intuitive method is to separate the two processes to have a greater control over the characteristics of the carriers. This method called the Interpolated Tree Orthogonal Multiplexing (ITOM) was introduced in (Harris & Kjeldsen, 2006) by constructing a binary wavelet packet tree and performing the spectral shaping external to the tree structure. The procedure is depicted in Fig. 12 from which we can notice that up-sampled shaping filters precede the input ports of the wavelet packet tree structure. Notching over the desired spectral interval is achieved by vacating one or more of the input branches.

Fig. 13 shows an example of the ITOM mechanism. We may note from Fig. 13(a) and (b) that the enabled and disabled carries neatly fit into the spectral gaps of one-another illustration the superiority of the ITOM procedure towards spectrum shaping.
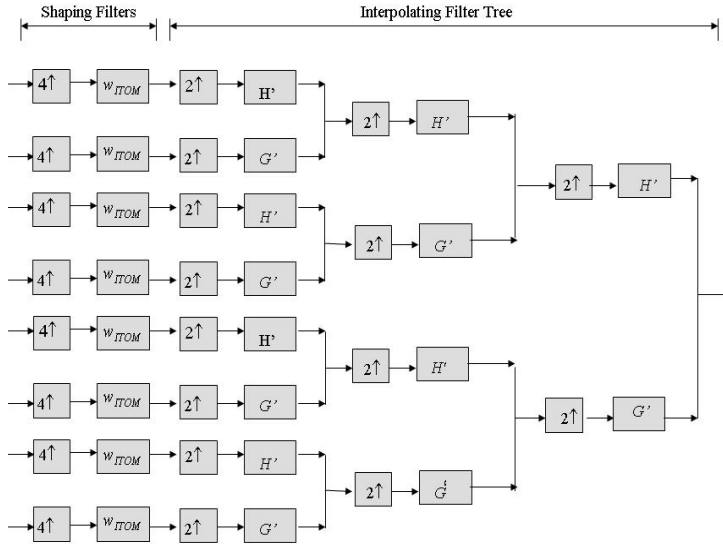

Fig. 12. Modulator of ITOM Interpolated Tree Structure (Harris & Kjeldsen, 2006). In the figure w_{ITOM} is the ITOM shaping filters.
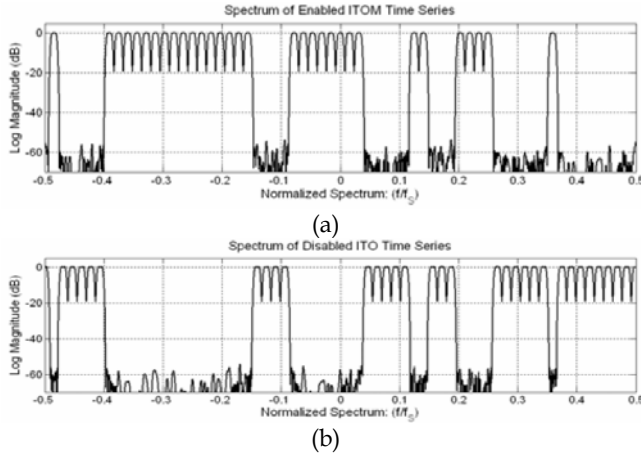
(a)

(b)

Fig. 13. Illustration of ITOM operation: (a) Spectra of Enabled Spectral Bands of 64-point ITOM; (b) Spectra of Disabled Spectral Bands of 64-point ITOM (Harris & Kjeldsen, 2006).

3. Single Carrier Cognitive Radio with Adaptive Waveform

Another method of utilizing the spectrum efficiently is shaping of the waveform of the transmitted signal dynamically over a spectral region, in order not to interfere with the occupied spectrum. Transform Domain Communication System (TDCS) is introduced in (Chakravarthy et al., 2005) as a candidate technique for Cognitive Radio. In principle the transmitted signal is shaped in such a way that its bandwidth does not contain the LUs band. In (Lee, 2002) the Fourier transform as a tool in shaping the spectrum in TDCS is replaced by the wavelet basis function. The technique is called Wavelet Domain Communication System. The two techniques will be further explored in the following subsections.

A Transform Domain Communication System (TDCS)

The TDCS transmitter system is described in Fig. 14(a) where single carrier is used. A clean spectrum information is derived from $A = \{A_1, A_2, …, A_{NFFT}\}$, where amplitudes of interfering frequency components exceeding a threshold are set to zero, while the others are set to one. Multi-valued complex pseudorandom phase vector ($e^{j\theta}$) is multiplied element by element with A to produce the spectral vector B_b. By multiplication with the scaling factor C, a vector B is produced. After IFFT operation, a time domain basis function $b(t)$ is derived, which subsequently is stored and modulated by using pulse amplitude modulation (PAM), pulse position modulation (PPM) or cyclic code shift keying (CCSK) where the bits are represented by the cyclic shifted pulse in time. The impact of this zeros insertion can be described from the power spectrum density (PSD) equation of TDCS,

$$PSD(f) = \left| \frac{1}{N_{FFT}} \sum_{m=0}^{N_{FFT}-1} A_m e^{j\theta_m} \int_{-(1+\alpha)\frac{T_u}{2}}^{(1+\alpha)\frac{T_u}{2}} p(t) e^{-j2\pi(f-f_m)t} dt \right|^2 \tag{16}$$

where T_u is the useful signal duration, $p(t)$ is the window function, a is the roll off factor of the window, and θ_m is the phase on subband m produced by the random phase module. If $p(t)$ is a rectangular window and α is zero, the area within the integral can be replaced by $T_u sinc(\pi f - \pi f_m)T_u)$. Zero amplitude at the carrier m will make the PSD on that carrier position becoming zero, and due to the orthogonality among carriers the power contributions from the other carriers are also zero. The equation is almost similar to (1), the differences are the power p_m and symbol amplitude A_m in (1) are replaced by the spectrum information A_m and the random phase θ_m.

TDCS gives more degree of freedom in choosing window $p(t)$ to lower the sidelobes of its spectrum. Unlike OFDM, the spectrum sidelobes of the window of TDCS do not have to be zero on the frequency spacing interval (in OFDM this is described as the orthogonality between carriers), as long as its sidelobes are very small. Since data detection at the receiver is performed by correlating the received signal with reference signal, as described in Fig. 14(b), the designed non-orthogonal spectrum of the transmitted signal will not affect the data detection.

In an attempt to increase the bit rate of TDCS, in this section we describe the proposed embedded symbol to TDCS-CCSK. The embedded symbol can be derived from PAM, QAM or PSK modulation. The proposed transmitter and receiver architectures of the system are depicted in Fig. 15.

There are two bit sources available. The first one is for the embedded symbol, and the second one is for the CCSK modulation. At the transmitter one embedded symbol will be multiplied with the vector from the point to point multiplication between the vector produced by the random phase θ and the vector produced by the spectrum magnitude A. Further, the legacy processes in producing TDCS signal are applied. At the receiver, the CCSK data detection is applied first. After removing the fading effect in frequency domain, the CCSK data detection is applied by de-correlating the time domain received signal with the time domain reference signal. De-correlation can be simplified by taking the maximum absolute value of inverse Fourier transform of the product of the frequency domain conjugate received signal and reference signal.

The new TDCS bit rate becomes:

$$R_{NEWTDCS} = \frac{\log_2 CS \cdot CS_{EM}}{T_s'} \tag{17}$$

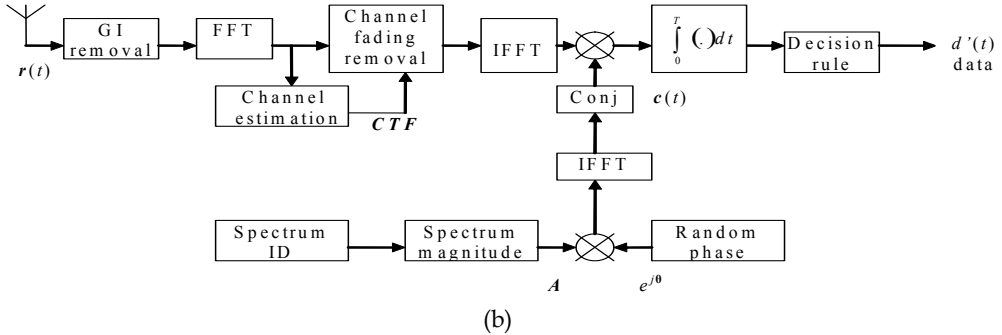where CS_{EM} is the embedded symbol constellation size.

(a)

(b)
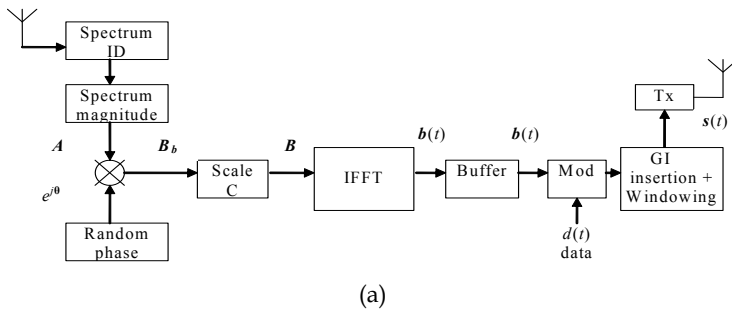
Fig. 14. TDCS (a) Transmitter and (b) Receiver.

From the spectrum point of view, the PSD of TDCS-CCSK with extra embedded symbol becomes (Budiarjo et al., 2008a); (Budiarjo et al., 2008b) :

$$PSD\,(f) = \left| \frac{1}{N_{FFT}} Se^{\,j\varphi} \sum_{k=0}^{N_{FFT}-1} A_k\, e^{\,j\theta_k} \int_{-(1+\alpha)\frac{T}{2}}^{(1+\alpha)\frac{T}{2}} p(t) e^{-\,j2\pi(f-f_k)t} dt \right|^2 \qquad (18)$$

By observing (18) it is reasonable to consider PSK as the most appropriate embedded symbol modulation mode among the three mapping types (PSK, PAM, and QAM). The PSK symbols are spread in a unit circle with magnitude of one. Therefore embedding the PSK symbol to the TDCS with CCSK modulation will not change the TDCS transmission power, while embedding PAM and QAM requires more power. The extra amount of power required will depend on the constellation size of the PAM/QAM embedded symbol. Moreover, the embedded symbol will affect the sidelobes of the proposed TDCS. Because PSK symbol magnitude is one, according to (18), the sidelobes amplitude of the proposed TDCS will not be changed. The PAM and QAM symbols with constellation size bigger than 2 have average magnitude larger than one therefore the sidelobes amplitude of the proposed TDCS will be affected. The PAM modulation in this case will be the worse option since for the same constellation size (larger than 2) its maximum amplitude is bigger than the QAM mode. Increasing the embedded symbol's constellation size will enhance the symbol- to-noise ratio on each of the TDCS subband, and symbol-to noise ratio enhancement will lead
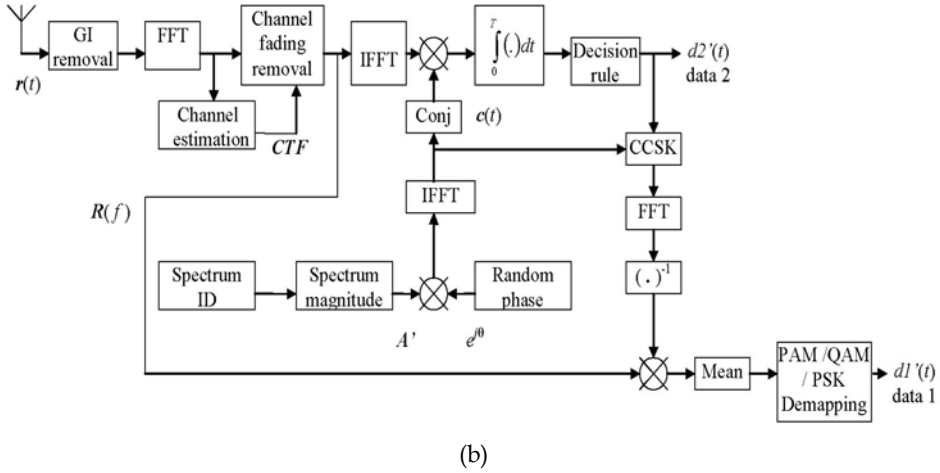

(a)

(b)
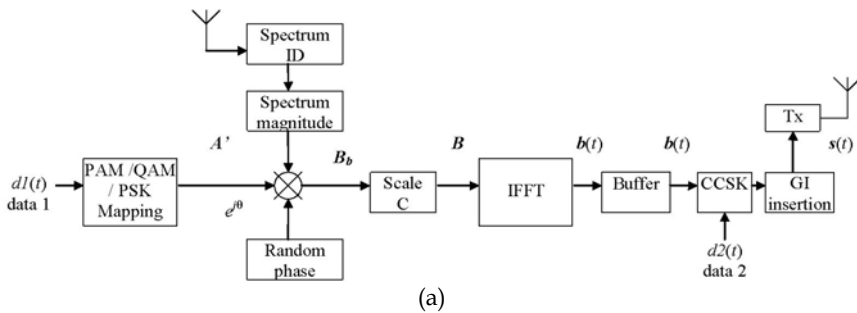
Fig. 15. (a) Transmitter and (b) Receiver architecture of TDCS with extra embedded symbol (Budiarjo et al., 2008a); (Budiarjo et al., 2008b).

to the symbol probability of error improvement and eventually to the bit error rate improvement. Increasing the constellation size means allocating more bits to the transmitted signal and as the consequence more energy is practically allocated. The amount of allocated energy is linearly comparable with the number of allocated bits. Therefore, the SNR gain obtained by increasing the CS_{EM} compared to the conventional TDCS-CCSK corresponds to

$$SNR_{Gain} = 10\log_{10}\left(\log_2(CS_{EM})\right)dB \qquad (19)$$

where $\log_2(CS_{EM})$ refers to the number of bits allocated to the embedded symbol.

B Wavelet Domain Communication System (WDCS)

An offshoot of TDCS is the wavelet domain communication system (WDCS) (Lee, 2002) where the Fourier transform operation in TDCS is replaced by wavelet transformation. Wavelets are used in this scheme to identify and establish an interference-free spectrum. The advantages of using wavelet are (Lee, 2002):

1. Increased adaptation over a larger class of interfering
   Signals,
2. Finer high-frequency resolution,
3. Allow implementation of CS-ary orthogonal signaling

The WDCS uses a packet-based transform to estimate the electromagnetic spectrum (Lee, 2002). Through the use of adaptive thresholds and notches, sub-bands containing the interference are effectively canceled. From this estimate, a unique communication basis function A in the transform domain is generated so that no (or very little) energy-bearing information is contained in the areas occupied by primary users. These functions are then multiplied with a pseudo random(PR) phase vector $e^{j\theta}$ to generate B_b. The PR code is used to randomize the phase of the spectral components. The resulting complex spectrum is then

scaled *C* to provide the desired energy in the signal spectrum. A time domain version *b*(*t*) of the basis function is then obtained by performing an inverse wavelet transfom. The wavelet basis function can be Coifflets, Daubechies, Haar, Symlets or Remez filter as also be applied in WP-MCM. Finally, the basis function is modulated with data using PAM, or PPM, or CCSK modulation and then transmitted. The block diagram of the WDCS transmission process is shown in Fig. 16.
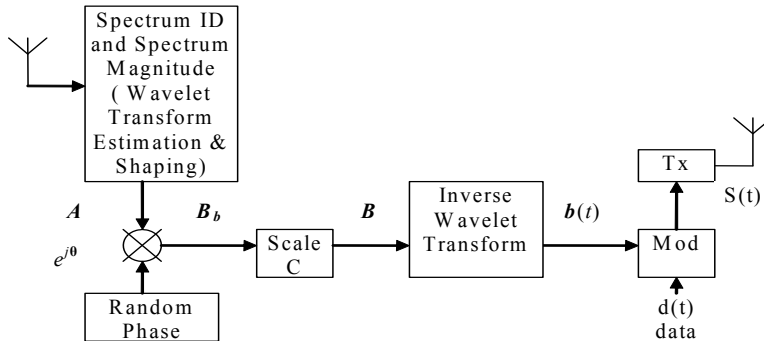
Fig. 16. WDCS Transmitter blocks (Lee, 2002).

At the receiver the detection is preceded by correlating the received signal with the reference signal in time domain. The receiver structure of WDCS is almost similar to the TDCS, the differences are the FFT and IFFT modules in TDCS that are replaced by discrete wavelet transform (DWT) and inverse discrete wavelet transform (IDWT), respectively.

4. MIMO for Cognitive Radio

Due to deactivation of carriers and long windowing duration upon the LU's occupancy, the CR bit rate is decreased. Multiple transmit antennas can be applied to CR for the purpose of transmitting different data on different antennas to increase the bit rate. The resource loss due to carriers deactivation on LU's band and bit rate loss due to long symbol duration are compensated. The transmitted power is distributed uniformly among the transmit antennas, so the total power is kept constant. In (Budiarjo et al., 2007b) CR with MC-OFDM and SC – TDCS on MIMO system using V-BLAST Receiver architecture has been evaluated. The results show the bit rate increases with the trade_off of more required pilots to estimate the channel transfer function of each transmit (Tx) and receive (Rx) antenna link. The composition of larger number of receive antenna compared to transmit antenna boosts the BER performance of the CR system. Especially for the TDCS the performance gain is even more significant than MC-OFDM for high SNR. Interesting results of (Budiarjo et al., 2007b) show that more SNR gain can be achieved by having more Rx antennas when compared to Tx antennas. Furthermore, this SNR gain is more significant when compared to the adaptive bit loading applied to OFDM with spectrum pooling (Budiarjo et al., 2006b).

5. Beamforming

Beamforming is formerly a multi access technique in mobile communications that takes the advantage of the position and angle of the other communications partners (nodes). This technique can increase the range between communications nodes, since the signal beam is concentrated (focused) on the intended communication partner hence no energy is wasted to the other direction. In CR system two scenarios can be considered in the application of beamforming: first is the interference avoidance to the LUs by not directing the signal beams to the LUs while occupying the LUs frequency, and second is the communications to multiple CR nodes without interference to each other (no multiple access interference).

In the case where no beamforming is applied, the RU can only occupy the LU's band if its distance to the licensed user is far so that it will not interfere with the LU's signal decoding process. Fig. 17(a) shows how the scenario should be when LU base station (BS) transmits data to LU mobile station (MS) which is located on the border of decodability region (dashed line). The RU can only transmit outside the "no talk zone" (solid line). The shaded region is called "protected region" where the signal decodability of the LUs is guaranteed (Sahai et al., 2004). Fig. 17(b) depicts the case when beamforming is applied by the RUs, the beams are directed to the intended RU partner, where communication to several RU nodes does not interfere to each other, and beams to the direction of LUs are nullified.

The important process in beamforming is the direction of arrival (DOA) estimation of the incoming signals. DOA estimation with Multiple Signal Classification (MUSIC) is introduced in (Schmidt, 1986), while Estimation of Signal Parameters by Rotational Invariance Techniques (ESPRIT) is proposed in (Paulraj et al., 1986). These DOA techniques give the angle directions from where the LUs signals are coming. An antenna array as well as a baseband model to estimate the direction of the intended partner is shown in Fig. 18 (a) and (b), respectively. If the plane wave incident on the array is from an angle (θ, ϕ) relative to the axis of the array, where θ is the angle of the wave incident with respect to the z axes, while ϕ is the angle between the x axis and the map of the wave incident on the x-y plane. If the array elements have uniform gain in all directions, and the spacing between array elements (Δx) is small, there will be no amplitude variations between signals received at different elements. All incident fields can be decomposed into a discrete number of plane waves. Other assumptions are well defined in (Liberti & Rappaport, 1999). The received signal will update the weights of the linear array elements and accordingly the maximum main beam of the array will be directed to the desired direction. By pointing the beam to specific angle where the LU signal comes and taking the Fourier transform of the received signal on that direction then the CR node with beamforming capability will have the information regarding the spectrum occupied on specific spatial direction. Accordingly this information is used for transmission of the CR system signal without interfering the LUs within spatial directions and certain frequency band. Investigation in (Liberti and Rappaport, 1999) shows that the utilization of beamforming by the LUs can enhance the coverage area of the CR system, and the propagation condition between the LUs and CR system affects the CR system's coverage area size. According to (Nishimori et al., 2006) for reusable CR coverage area greater than 70%, a signal to Interference Ratio (SIR) gain of about 14 dB is achieved by the CR system where the LU's beamwidth is 10 degrees in vertical plane compared to omni directional beam.

6. Summary

In this tutorial chapter techniques for CR were discussed. Spectrum sensing as the crucial element of cognition in terms of spectrum awareness was reviewed. Adaptive OFDM with Fourier and wavelet basis functions and the combination with spectrum pooling as well as the single carrier modulation with wave shaping technique using Fourier and wavelet transform were explained. MIMO technique as a QoS enhancement scheme for CR system was explained and beamforming as a spatial awareness technique in CR was studied.

Spectrum Pooling enables the access to spectrum without giving significant interference to the actual license owners. The high sidelobes of OFDM in spectrum pooling can be counteracted by deactivating carriers adjacent to the LUs (non contiguous OFDM), and applying a proper window to the signal in time domain. The loss due to carrier deactivation and long windowed symbol duration can be compensated by adaptive bit loading so the system still can achieve the target bit rate. Spectrum pooling provides a high degree of freedom in choosing the window to suppress the sidelobes of the signal since the signal orthogonality can be preserved at the receiver by applying the rectangular window implemented by DFT. If target bit rate is the main concern in adaptive bit loading, then groupwise bit loading is the choice, while if target error rate is the constraint, then carrierwise bit loading will be the solution.

Overlapping OFDM symbol transmission is another method in counteracting the loss of bit rate due to the long OFDM window duration. However the starting point of the consecutive OFDM symbols should be taken care of in order to avoid the ISI. Frequency hopping is the solution in very harsh channel environments where propagation condition and LU's access is very severe. Frequency hopping is only conducted when it is necessary.

Wavelet transform is considered to be a better modulation method compared to the Fourier transform method due to its flexibility in shaping the spectrum and its better ICI and ISI suppression. With the proper basis functions in the wavelet transform, the spectrum can be used efficiently; hence bit rate can be maintained or even increased. Furthermore, no guard interval is needed for the wavelet which is another boost for the throughput. A practical utilization of wavelet in CR system is by representing the RU signal as wavelet packet coefficients and by selectively removing those coefficients whose spectra footprint lies in the LU band. In all these techniques, the reliable information from the spectrum sensing unit is essential.

TDCS and WDCS are single carrier transmission techniques in utilizing the spectrum efficiently. The technique provides a degree of freedom in reducing its sidelobes which also directly reduces the interference to LU's band. In TDCS unlike in OFDM, the transmitted signal inter frequency orthogonality is not necessary since the data detection at the receiver is conducted by correlating the received signal with the reference signal. The drawback of the single carrier approach is the achievable data rate which is lower compared to OFDM.

MIMO system boosts the bit rate of CR system with additional computational costs of estimating all of the Tx and Rx channel links. For the improvement of BER of CR system, the usage of more receive antennas (than transmit antennas) is suggested.

Beamforming gives spectrum sharing capability to a wireless system. It also enhances the multiple access capability of the CR system. As beamforming technique can be used by the scanning module to detect the spectrum occupancy in a specific direction, it will also enhance the coverage area of a CR system.
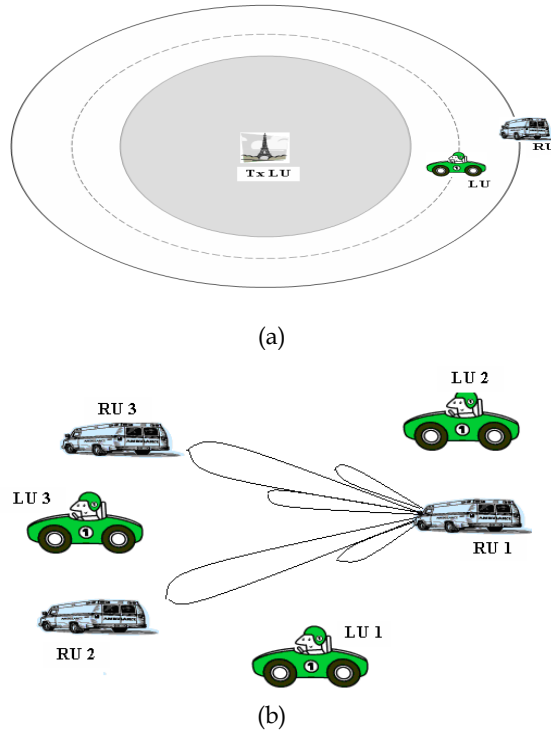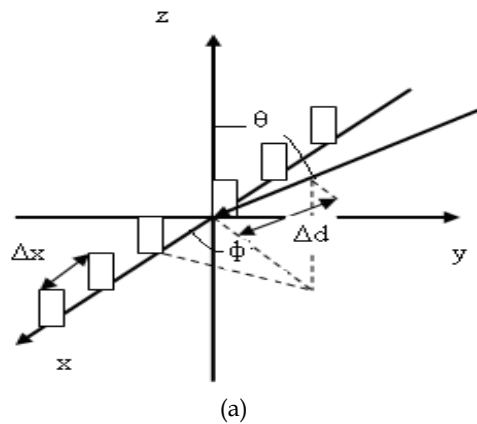
7. Acknowledgement

(a)

(b)

Fig. 17. (a) CR without beam forming (Sahai et al., 2004) (b) CR with beamforming.
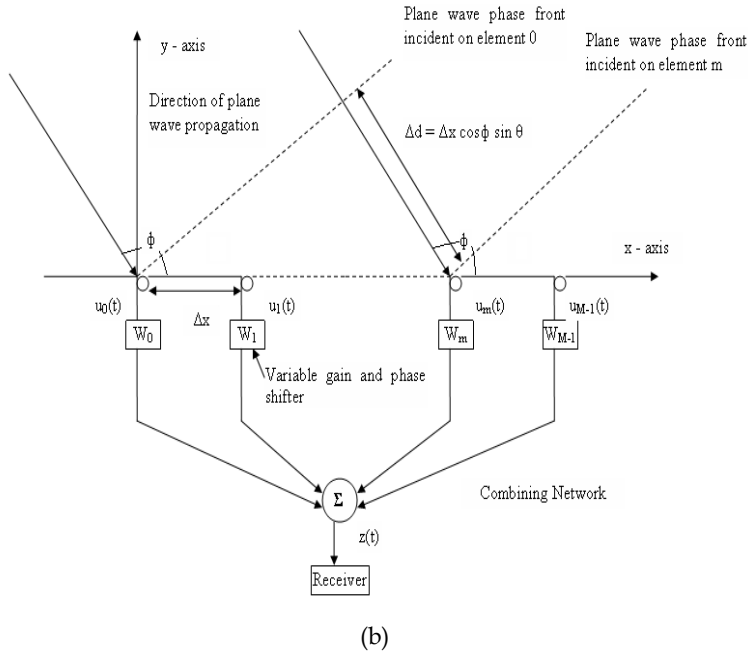

(a)

(b)

Fig. 18. (a) A linear equally spaced array antenna (b) Baseband model of a linear equally spaced array for estimating Direction of Arrival of a plane wave incident on the antenna elements (Liberti & Rappaport, 1999).

8. References

Assalini, A. & Tonello, A. M. (2004). "Improved Nyquist Pulses", *IEEE Communicaions Letters*, Vol. 8, No.2, (February 2004) pp 87-89

Barreto, A. N. & Furrer, S. (2001). " Adaptive Bit Loading for Wireless OFDM Systems", Proceedings of *IEEE International Symposium on Personal, Indoor, Mobile Radio Communications*, pp. 88–92, San Diego California, USA, October 2001

Berthold, U.; Jondral, F. K.; Brandes, S. & Schnell, M. (2007). OFDM Based Overlay Systems: A Promising Approach for Enhancing Spectral Efficiency, *IEEE Communications Magazine*, Vol. 45, No. 12, (Dec. 2007) pp. 52-58

Bouwel, C.V.; Potemans, J.; Schepers, S.; Nauwelaers, B. & Capelle, A.V. (2000). Wavelet Packet Based Multicarrier Modulation, *Proceedings of IEEE Benelux Symposium on Communications and Vehicular Technology*, Leuven, Belgium, October 2000

Brandes, S.; Cosovic, I. & Schnell, M. (2005a). Sidelobe Suppression in OFDM Systems by Insertion of Cancellation Carriers, *Proceedings of 62nd IEEE Vehicular Technology Conference (VTC)*, pp. 152-156, Dallas, USA, September 2005

Brandes, S.; Cosovic, I. & Schnell, M. (2005b). Reduction of Out-of-Band Radiation in OFDM Based Overlay Systems**,** *Proceedings of First IEEE International Symposium New Frontiers in Dynamic Spectrum Access Networks (DySPAN)*, pp. 662 – 665, Baltimore Maryland, USA, November 2005

Brandes, S.; Cosovic, I. & Schnell, M. (2006). Reduction of Out of Band Radiation in OFDM Systems by Insertion of Cancellation Carriers, *IEEE Comunications Letters*, Vol. 10, No. 6, (June 2006) pp. 420-422

Brodersen, R.W.; Wolisz, A.; Cabric, D.; Mishra, S. M. & Wilkomm, D. (2004). CORVUS: A Cognitive Radio Approach for Usage of Virtual Unlicensed Spectrum, White Paper, Dept. of Electrical Engineering and Computer Science University of Berkeley, July 2004

Budiarjo, I.; Nikookar, H. & Ligthart, L. P. (2005). An Overview of Adaptive OFDM in the Context of Cognitive Radio, *Proceedings of 12th IEEE Symposium on Communications and Vehicular Technology (SCVT)*, pp. 1-6, Enschede, the Netherlands , November 2005

Budiarjo, I.; Nikookar, H. & Ligthart, L. P. (2006a). Performance Evaluation of Adaptive Bit Loading For OFDM With Channel Estimation Using 2x1-D Wiener Filter, *Proceedings of Int. Sym. on Wireless Personal and Multimedia Communications*, pp. 1138-1142, San Diego CA, USA, September 2006

Budiarjo, I.; Nikookar, H. & Ligthart, L. P. (2006b). Combined Spectrum Pooling and Adaptive Bit loading for Cognitive Radio OFDM Based System, *Proceedings of 13th IEEE SCVT*, pp. 73-76,Liege, Belgium, November 2006

Budiarjo, I.; Heskamp, M.; Zhang, Q.; Nikookar, H. & Kokkeler, A. J. (2007a). *D.3.41 Radio system: Second Phase Verification andAdvances in Architecture and Algorithms*", Adaptive Ad-hoc Freeband (AAF), The Netherlands, July 2007

Budiarjo, I.; Nikookar, H. & Ligthart, L. P. (2007b). Cognitive Radio with Single Carrier TDCS and Multicarrier OFDM Approach with V-BLAST Receiver in Rayleigh Fading Channel, *Proceedings of 2nd International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CrownCom)*, pp. 140-146, Orlando, USA, August 2007

Budiarjo, I.; Nikookar, H. & Ligthart, L. P. (2007c). Performance Evaluation of OFDM Based Cognitive Radio System with Wiener Filter Channel Estimation Using Frequency Hopping GSM Channel Model at 900 MHz, *Proceedings of 10th European Conference on Wireless Technology (ECWT)*, pp. 74-77, Munich, Germany, October 2007

Budiarjo, I.; Nikookar, H. & Ligthart, L. P. (2008a). On the Utilization of Embedded Symbol for CCSK BER Improvement in TDCS Dynamic Spectrum Access, *Proceedings of IEEE European Wireless Technology Conference*, pp. 123-126, Amsterdam, The Netherlands, October 2008

Budiarjo, I.; Nikookar, H. & Ligthart, L. P. (2008b). Cognitive Radio Modulation Techniques: Reducing Mutual Interference, *IEEE Signal Processing Magazine*, Vol. 25, No. 6, (November 2008) pp. 24-34

Cabric, D.; Mishra, S. M. & Brodersen, R. W. (2004). Implementation Issues in Spectrum Sensing for Cognitive Radios, in *Proceedings of Asilomar Conference on Signals, Systems, and Computer*s, pp. 772 – 776, November 2004, Pacific Grove CA, USA

Cabric, D. & Brodersen, R. W. (2005). Physical Layer Design Issues Unique to Cognitive Radio Systems", *Proceedings of IEEE Personal Indoor and Mobile Radio Communications (PIMRC)*, pp. 759-763, September 2005, Berlin, Germany

Cabric, D.; Tkachenko, A. & Brodersen, R.W. (2006). Spectrum Sensing Measurements of Pilot, Energy, and Collaborative Detection, *Proceedings of IEEE Military Communications Conference (MILCOM)*, pp. 1-7, October 2006, Washingthon D.C., USA

Chakravarthy, V.; Shaw, A. K.; Temple, M. A. & Stephens, J. P. (2005). Cognitive Radio -An Adaptive Waveform with Spectrum Sharing Capability, *Proceedings of IEEE Wireless*

Communications and Networking Conference (WCNC), pp. 724-729, March 2005, New Orleans, USA

Chow, P. S., Cioffi, J. M. & Bingham, J. A. C. (1995). A Practical Discrete Multitone Transveiver Loading Algorithm for Data Transmission over Spectrally Shaped Channels, *IEEE Transactions on Communications,* Vol. 43, No.2, (February 1995) pp. 773–775

Cosovic, I.; Brandes, S & Schnell, M. (2005). A Technique for Sidelobe Suppression in OFDM System, *Proceedings of IEEE GLOBECOM*, pp. 204-208, St. Louis Missouri, USA, November 2005

Cosovic, I.; Brandes, S. & Schnell, M. (2006). Subcarrier Weighting : A Method for Sidelobe Suppression in OFDM Systems, *IEEE Communications Letters*, Vol. 10, No. 6, (June 2006) pp. 444-446.

Cosovic, I. & Mazzoni, T. (2006). Suppression of Sidelobes in OFDM systems by multiple choice sequences, *European Transactions on Telecommunications Special Issue on Multi carrier – Spread Spectrum Wiley InterScience*, Vol. 17, No. 6, (June 2006) pp. 623-630.

Cosovic, I. & Mazzoni, T. (2007) "Sidelobe Suppression in OFDM Spectrum Sharing Systems via Additive Signal Method", *Proceedings of 65th IEEE Vehicular Technology Conference (VTC)*, pp. 2692-2696, Dublin, Ireland, April 2007

Gruenheid, R.; Bolinth, E.; Rohling, H. & Aretz, K. (2000). Adaptive Modulation for the HIPERLAN/2 Air Interface, *Proceedings of 5th International OFDM Workshop*, Hamburg, Germany, September 2000

Gruenheid, R.; Bolinth, E. & Rohling, H. (2001). A Blockwise Loading Algorithm for The Adaptive Modulation Technique in OFDM Systems, *Proceedings of Vehicular Technology Conference*, pp. 948–951, Atlantic City New Jersey, USA, October 2001

Hanzo, L.; Wong, C. H. & Yee, M.S. (2002). *Adaptive Wireless Transceivers*, John Wiley and Sons Ltd, Chichester,England, 2002

Harris, F. & Kjeldsen, E. (2006). A Novel Interpolated Tree Orthogonal Multiplexing (ITOM) Scheme with Compact Time-Frequency Localization: an Introduction and Comparison to Wavelet Filter Banks and Polyphase Filter Banks", *Proceedings of Software Defined RadioTechnical Conference and Product Exposition*, Orlando, USA, November 2006

Haykin, S. (2005). Cognitive Radio: Brain Empowered Wireless Communications. *IEEE Journal on Selected Areas In Communications,* Vol. 23, No. 2, (February 2005) pp. 201–220

Hoeksema, F.W.; Heskamp, M.; Schiphorst, R. & Slump, C.H. (2005). A Node Architecture for Disaster Relief Networking, *Proceedings of 1st IEEE Int. Symposium on new Frontiers in Dynamic Spectrum Access Networks (DySPAN'05)*, pp. 577–584, November 2005, Baltimore Maryland, USA

Jamin, A. & Maehoenen, P. (2004). Wavelet Packet Modulation for Wireless Communications, *Wireless Communications and Mobile Computing magazine Wiley Interscience*, Vol. 5, Issue 2, (December 2004) pp 123-137

Jensen, A. & Cour-Harbo, A.L. (2001). *Ripples in Mathematics: The Discrete Wavelet Transform*, Springer, Heidelberg, 2001

Jetlund, O.; Oien, G. E.; Hole, K. J. & Markhus, V. (2002). Rate - Adaptive Coding and Modulation with LDPC Component Codes, *Proceedings of COST 273 TD(02) -108, 5th*

Management Committee Meeting COST Action 273, Towards Mobile Broadband Multimedia Networks, Lisbon, Portugal, September 2002

Kaemarungsi, K. & Krishnamurthy, P. (2003). On the Use of Adaptive OFDM to Preserve Energy in Ad Hoc Wireless Networks, *Proceedings of 13th MPRG / Virginia Tech Symposium on Wireless Personal Communications*, June 2003, Blacksburg Virginia, USA

Keller, T.; Liew, T. H. & Hanzo, L. (2000). Adaptive Rate RRNS Coded OFDM Transmission for Mobile Communication Channels, *Proceedings of 51 st IEEE Vehicular Technology Conference*, pp. 230-234,Tokyo, Japan, May 2000

Krongold, B. S.; Ramchandran, K. & Jones, D. L. (2000). Computationally Efficient Optimal Power Allocation Algorithms for Multicarrier Communications Systems, *IEEE Transactions on Communications,* Vol. 48, No. 1, (January 2000) pp. 23–27

Lakshmanan, M.K.; Budiarjo, I. & Nikookar, H. (2007). Maximally Frequency Selective Wavelet Packets Based Multi-Carrier Modulation Scheme for Cognitive Radio Systems, *Proceedings of 50th IEEE Global Communications Conference (Globecom)*, pp. 4185-4189, Washingthon DC, USA, November 2007

Lee, M. J. (2002). *Wavelet Domain Communication System (WDCS): Packet-Based Wavelet Spectral Estimation And M-ary Signaling*. Master thesis, Air Force Institute of Technology, Ohio, USA, March 2002

Lei, M. & Zhang, P. (2003). Subband and Power Loading for Adaptive OFDM, *Proceedings of Vehicular Technology Conference*, pp. 1482–1486, , Orlando Florida, USA, October 2003

Liberti Jr, J. C.; & Rappaport, T. S. (1999). *Smart Antennas For Wireless Communications*, Prentice Hall, NJ, USA, 1999

Mahmoud, H. A. & Arslan, H. (2008). "Sidelobe Suppression in OFDM –Based Spectrum Sharing Systems by Using Adaptive Symbol Transition", *IEEE Communications Letters*, Vol. 12, No. 2, (February 2008) pp. 133-135

Moseley, N. A. (2004). *Radio Resource Discovery for Ad-hoc Wireless Networking*, M.Sc Thesis, University of Twente, Enschede, Netherland, August 2004

Muschalik , C. (1996). Improving An OFDM Reception Using An Adaptive Nyquist Windowing, *IEEE Transactions on Consumer Electronics,* Vol.42, No.3, (August 1996) pp. 259-269

Mutti, C.; Dahlhaus, D. & Destefanis, D. (2004). Adaptive Coding based on LDPC Codes for OFDM Systems with HD Decoding, *Proceedings of 13th IST Mobile & wireless Communications Summit*, Lyon, France, June 2004

Negash, B. G. & Nikookar, H. (2000). Wavelet Based Multicarrier Transmission over Multipath Wireless Channel, *IEE Electronic Letters*, Vol. 36, No. 21, (October 2000) pp. 1787-1788

Nikookar, H. & Prasad, R. (1997a). Optimal Waveform Design for Multicarrier Transmission Over a Multipath Channel, *Proceedings of 47th IEEE Vehicular Technology Conference*, pp. 1812-1816, Phoenix Arizona, USA, May 1997

Nikookar, H. & Prasad, R. (1997b). Wave shaping of Multicarrier Signal for Data Transmission Over Wireless Channels, *Proceedings of 6th IEEE International Conference on Universal Personal Communications Record (ICUPC)*, pp 173–177, San Diego California, USA, October 1997

Nikookar, H. & Prasad, R. (2000). Weighted OFDM for Wireless Multipath Channels, *IEICE Transactions on Communications*, Vol. E83-B, No. 8, (August 2000) pp. 1864-1872

Nishimori, K.; Bottega, E.; Yomo, H.; Popovski, P.; Takatori, Y.; Prasad, R. & Kubota, S. (2006). Spatial Availability for Cognitive Radio System Under Directional Interference, *Proceedings of Int. Sym. on Wireless Personal and Multimedia Communications*, pp. 44-48, San Diego CA, USA, September 2006

Pagadarai, S.; Rajbanshi, R.; Wyglinski, A. M. & Minden, G. J. (2008a). Sidelobe Suppression for OFDM-Based Cognitive Radios Using Constellation Expansion, *Proceedings of IEEE Wireless Communicaitons and Networking Conference (WCNC)*, pp. 888-893, Las Vegas Nevada, USA, April 2008

Pagadarai, S.; Wyglinski, A. M. & Rajbanshi, R. (2008b). A Novel Sidelobe Suppression Technique for OFDM-Based Cognitive Radio Transmission", *Proceedings of IEEE Symposia on New Frontiers Dynamic Spectrum Access Networks (DySPAN)*, pp. 1-7, Chicago, USA, October 2008.

Paulraj, A.; Roy, R. & Kailath, R. (1986). A Subspace Rotation Approach to Signal Parameter Estimation, Proceedings *of IEEE* , Vol. 74, No. 7, (Jul. 1986) pp. 1044-1046

Reddy, S. B.; Yucek, T & Arslan, H. (2003). An Efficient Blind Modulation Detection Algorithm for Adaptive OFDM Systems, *Proceedings of IEEE Vehicular Technology Conference (VTC)*, pp. 1895–1899, Orlando, USA,   October 2003

Rioul, O. & Duhamel, P. (1994). A Remez Exchange Algorithm for Orthonormal Wavelets, *IEEE Transactions on Circuits Systems - II*, Vol. 41, No. 8, (August 1994) pp. 550-560

Sahai, A.; Hoven, N. & Tandra, R. (2004). Some Fundamental Limits on Cognitive Radio, *Proceedings of 42nd Allerton Conference on Communication, Control, and Computing,* Monticello IL, USA, October 2004

Schmidt, R. O. (1986). Multiple Emitter Location and Signal Parameter Estimation, *IEEE Transactions on Antennas and Propagation*, Vol. 34, No. 3, (March 1986) pp. 276-280

Tan, P. & Beaulieu, N. C. (2004). Reduced ICI in OFDM Systems Using the Better than Raised – Cosine Pulse, *IEEE Communications Letter,* Vol. 8, No. 3,(March 2004) pp. 135-137

Vucetic, B. (1991). An Adaptive Coding Scheme for Time Varying Channels, *IEEE Transactions on Communications*, Vol. 39, No. 5 , (May 1991) pp 653 – 663

Weiss, T.; Hillenbrand, J.; Krohn, A. & Jondral, F. K. (2004). Mutual Interference in OFDM based Spectrum Pooling Systems, *Proceedings of IEEE Vehicular Technology Conference (VTC)*, pp. 1873- 1877, Milan, Italy, May 2004

Witrisal, K. (2002). OFDM Air Interface Design for Multimedia Communications, *Ph.D. Thesis,* Delft University of Technology, The Netherlands, April 2002

Yamaguchi, H. (2004). Active Interference Cancellation Technique for MB-OFDM Cognitive Radio, *Proceedings of 34th European Microwave Conference*, pp. 1105-1108, Amsterdam, The Netherlands, October 2004

Robust Bases for Spectrum Pooling Systems on Wavelet Packet Multi-carrier Modulation MIMO Architecture

M.K.Lakshmanan, I. Budiarjo and H. Nikookar
International Research Centre for Telecommunications and Radar
Department of Electrical Engineering, Mathematics and Computer Science
Delft University of Technology, Mekelweg 4 2628CD, Delft,
The Netherlands

1. Introduction

1.1 Background

The growing popularity of wireless applications has placed enormous burden on valuable resources such as spectral bandwidth. This has brought about a major revamp of traditional resource allocation policies culminating in an explosion of research activity in the field of Cognitive Radio (CR). In this chapter we demonstrate the operation of a spectrum pooling system built from a wavelet packet Multi-carrier modulation Multiple Input Multiple Output (MIMO) scheme. The objective is to combine the promise of optimum utilization of radio resources by Cognitive Radio, the high data throughput without additional bandwidth/ power requirements offered by MIMO and the flexibility of wavelets, to develop intelligent wireless systems. Taking the example of effective spectrum utilization, we demonstrate the feasibility and efficiency of the proposed system through simulation studies. To enable the WP-MCM system to co-exist with other licensed users wavelet packet (WP) carriers in and near the region of the licensed user spectrum are dynamically deactivated. Various wavelets including the well-known families Daubechies, Coiflet, Symlet are applied and studied. The emphasis is on the design and development of optimal wavelet bases that have narrow and well confined spectral footprints. To this end filter banks that are maximally frequency selective are derived through a modified Remez exchange algorithm. Through simulation results the operation of the proposed system is demonstrated.

1.2 Theme of Work

The paradox of non-availability of spectrum even when large swathes of licensed spectrum is underutilized most of the time has incited an explosion of research activity in the field of Cognitive Radio CR) (Mitola, 2000); (Haykin, 2005) - wireless systems that intelligently adapt their transmission parameters in accordance with changing environments and opportunistically utilize radio resources. CR enables public access to spectral ranges of

licensed frequency bands which are seldom used by overlaying a secondary rental user (RU) to an existing licensed user (LU).

Multi-carrier modulation (MCM) has been mooted as a strong candidate for CR system design (Weiss and Jondral, 2004). By merely vacating a set of subcarriers, the spectrum of a MCM based CR can be easily and flexibly shaped to occupy spectral gaps without interfering with the LU. It has been shown that adaptive MCM based CR is a robust method to achieve good quality of communication and efficient use of the spectrum.

A further enhancement to the system performance can be brought about by building MIMO (Multiple Input - Multiple Output) systems (De Lima E.R. et al, 2004) using multiple antennas both at the transmitter and receiver with adaptive MCM arms. Through clever spatial processing, the MIMO offers significant increases in data throughput and link range without additional bandwidth or transmit power requirements.

A third dimension of system optimization can be gained by deriving the MCM carriers using wavelet bases instead of conventional Fourier bases (Bouwel et al, 2000); (Jamin and Mahonen, 2005); (Lindsey, 1997); (Negash and Nikookar, 2000). Unlike Fourier bases which are static sines/cosines, wavelets (Lakshmanan and Nikookar, 2006) offer flexibility and adaptability which can be tailored to satisfy an engineering demand.

In this work we attempt to combine the desirable features of CR, MIMO and wavelet features and demonstrate the operation of a Wavelet packet Multi-carrier modulation (WPMCM) based MIMO system in the context of a spectrum pooling setup. The wavelet packet subcarriers are derived from multistage tree-structured paraunitary filter banks (Lakshmanan et al, 2008); (Vaidyanathan, 1993); (Vetterli, 1995); (Daubechies, 1992).

MIMO systems are possible because of space-time coding (STC) algorithms. In this chapter a promising STC realization for MIMO systems called vertical Bell-labs Layered Space-Time (VBLAST) (Wolniansky P.W. et al, 1998) coding technique is applied. The effectiveness of the proposed system is demonstrated through simulation results.

Various wavelets including the well-known Daubechies, Coiflet, Symlet families are considered. A key goal in this process is to ensure that the carriers have narrow and well-confined spectral footprints that don't spill over to neighboring regions. This way the affected carriers can be easily identified and isolated to facilitate efficient spectrum notching. In this regard filter banks that are maximally frequency selective are derived through a modified Remez exchange algorithm (Rioul and Duhamel, 1994).

1.3 Organization of the Chapter

The chapter is organized as follows. In Section 2 the major elements of the proposed MIMO system and its implementation are laid out. Section 3 delves into the procedure to estimate the radio environment and accordingly shape the spectrum of the transmission waveform to utilize spectrum holes without hindering LU operation. The procedure to derive the wavelet packet signal waveforms for the MCM is detailed in Section 4. Section 5 explains the process to derive the best wavelet bases for WP-MCM. Section 6 gives the spatial processing technique used to detect data at the receiver. The simulation setup and results are discussed in Sections 7 and 8, respectively. Finally, the conclusions are drawn in Section 9.

2. System Blocks And Operation

The system model is illustrated in Figure 1. The major blocks of the system are the spectrum estimator, wavelet packet based adaptive multicarrier modulator and the MIMO setup.
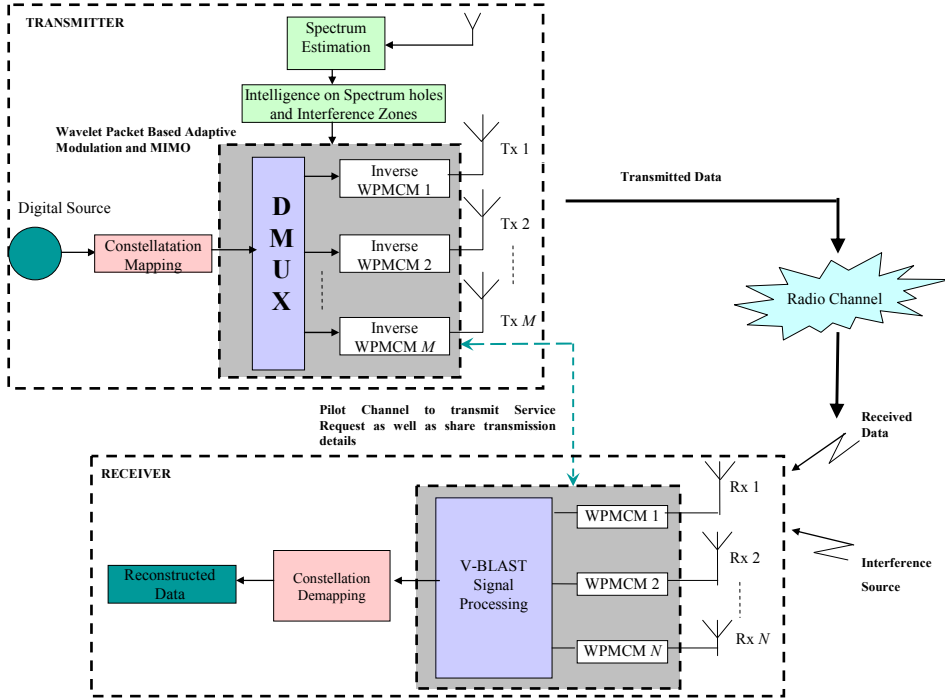

Fig. 1. System model of the Cognitive Radio. Transmitter: The major blocks include the spectrum and channel state estimator, adaptive pulse shaper (Transmission waveform shaping) and modulator. Receiver: The main components are the Signal base band processor and detector.

An incoming high-rate serial data stream is mapped and split into lower-rate parallel streams. The data in each parallel branch is upsampled and modulated with wavelet packet subcarriers. The modulated subcarriers are then added to obtain a single WPMCM symbol. This constitutes an inverse WPMCM (equivalent of IFFT in Orthogonal Frequency Division Multiplexing or OFDM) operation. The spectrum estimator gauges the radio spectrum scenario and performs radio scene analysis to detect the presence of interference regions and spectrum holes. This information is coded in the form of a spectrum vector containing ones (representing free bands) and zeros (representing occupied bands). Based on this vector the transmitter adapts the signal spectrum by activating or vacating WPMCM carriers and a transmission waveform that has little or no energy in the interference domains is derived.
A collection of such shaped WPMCM symbols is then demultiplexed into M substreams, and each substream is fed to a different transmit antenna in the same frequency band. The

transmitters' assemblage forms a vector-valued transmitter with the transmit power in each arm being proportional to $1/M$ so that the total radiated power is constant and independent of the number of transmitter arms M (De Lima E.R. et al, 2004).

The receiver consists of an array of N antennas each of which pick up the signals from all M transmit antennas. The received signals at each receiver antenna n from each of the m transmit antenna is be given as:

$$Y_n = \sum_{m=1}^{M} C_{nm} X_m + \eta_n \tag{1}$$

This can be represented in matrix form as:

$$\mathbf{Y} = \mathbf{C}\mathbf{X} + \eta \tag{2}$$

where $\mathbf{Y}=[Y_1, Y_2,\dots Y_N]$ is the received signal vector, $\mathbf{X}=[X_1, X_2,\dots X_M]^T$ is the transmitted signal vector, $\eta=[\eta_1, \eta_2,\dots \eta_N]^T$ is the noise vector and the channel matrix $\mathbf{C}$ and C_{nm} is the channel link between transmit antenna m and receive antenna n (Figure 2).

To detect the transmitted data, each of the substreams of the received signal vector is first demodulated by WPMCM (equivalent of FFT operation in OFDM). WPMCM involves de-convolving the signal substream with a sieve of receiver sub-carrier waveforms which are orthogonal duals of the sub-carriers used at the transmitter end. Estimation of information symbols through a VBLAST signal processing algorithm (described in Section 6) follows WPMCM demodulation.

The transmitter and receiver are kept cognizant at all times on the radio environment and the transmission signal characteristics through a pilot channel.
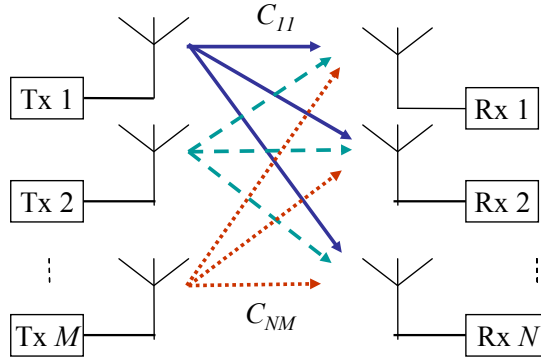
Fig. 2. MIMO Channel model. Tx1-TxM are the M transmitter antennas and Rx1-RxN are the N receiver antennas. C_{nm} is the channel link between transmit antenna m and receive antenna n.

3. Spectrum Estimation, Notching And Spectrum Shaping

3.1 Spectral Estimation.

The first and foremost task of any cognitive radio system is to gauge the wireless environments over wide frequency bands and identify spectrum holes and occupied bands. This is done so that the CR system can opportunistically claim unused bands and operate

invisibly without causing any distortion to other primary and licensed users. The challenge of spectrum sensing module is in identification and detection of primary user signals amidst harsh and noisy environs.

Spectral estimation in the proposed WPMCM based MIMO system is accomplished by a wavelet packet transformation involving filtering and decimating the samples of the radio environment (Lee et al., 2002). The first iteration of the signal decomposition (filtering and decimating) process divides the data into two sub-bands, the detailed and coarse sub-bands. Detailed subband coefficients are the result of passing the data through a highpass filter and decimating, or down-sampling, the filter output by a factor of two. Coarse sub-band coefficients are the result of lowpass filtering the data and decimating the filter output by a factor of two. The wavelet packet decomposition process continues by subsequently splitting and down sampling the low and high pass sub-components. This iterative decomposition is repeated until the wavelet packet tree structure has been fully expanded. Following iterative decomposition, through suitable threshold, a "notched" spectral magnitude vector is generated by setting the retained wavelet packet sub-band coefficients to one and those discarded to zero. The final output of the iterative decomposition process represents the magnitude of the spectral estimates.

3.2 Thresholding and Spectral Notching

The spectral information is coded in the form of a spectrum vector containing ones and zeros. The zeros correspond to bands which are occupied and the ones represent bands that are free (spectrum holes). The pattern of ones and zeros effectively characterizes the desired magnitude of the spectral estimate. The threshold is performed on a sub-band-by-sub-band basis whereby the power contained in each sub-band is independently compared to a predetermined threshold. Following the recommendations of (Lee et al., 2002) the threshold value is defined in terms of the noise power. When sub-band power exceeds the noise power by 20%, interference is declared present and all of the sub-band coefficients are set to a value of zero. If sub-band power does not exceed the threshold, all of the sub-band coefficients are retained (set to a value of one).

3.3 Transmission Waveform Shaping By Identification of Affected WPMCM Carriers

Based on the spectrum vector, sub-channels of the WPMCM system that lie in and around the spectrum of the LU are vacated to facilitate coexistence. This way the CR transmission signal is dynamically sculpted such that it has no or very little time-frequency components competing with the LU and the CR operation is made invisible to the LU.

4. Signal Waveforms for MCM

4.1 Background

The subcarrier signal waveforms in traditional MCM implementations, such as OFDM, are sine/cosine basis functions. In WPMCM the sub-carrier waveforms are derived from poly-channel tree structures built by cascading multiple two-channel filter banks like the one shown in Figure 3.

A two-channel filter bank consists of a set of 4 perfect reconstruction filters (2 high pass and 2 low pass) which allow the decomposition and reconstruction of a signal without

amplitude or phase or aliasing distortion (Vaidyanathan, 1993). The two-channel filter bank has the property of splitting the signal into two lower resolution versions – namely the coarse (low pass) and the detail (high pass). When the decomposition into coarse and detail components is continued iteratively, it leads to the generation of wavelet packet bases. When the perfect reconstruction filters used satisfy an additional property known as paraunitary condition (explained later), they lead to wavelet packet bases with impulse responses that are mutually orthogonal to one-another and to their duals.
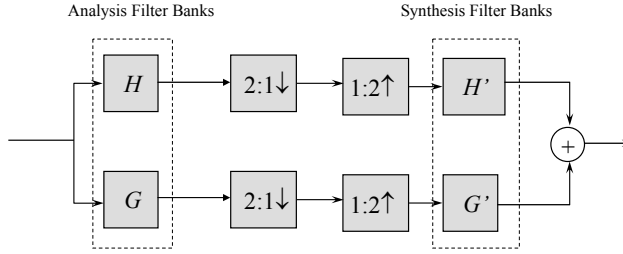
Fig. 3. Two channel filter bank analysis with analysis filters *H* and *G* (low and high pass, respectively) and synthesis filters *H'* and *G'* (low and high pass, respectively).

4.2 Generation of Wavelet Packet Sub-Carrier Bases
The wavelet packet sub-carriers (to be used at the transmitter end) are generated through a multichannel filterbank consisting of cascaded two-channel filters applying the synthesis filters (*H'* and *G'*). This represents an inverse discrete wavelet packet transformation or IDWPT and consists of binary interpolation (up-sampling) by 2, filtering and recombination at each level. The number of iterations *J* determines the number of subcarriers *M* generated and the relationship is given as $M \leq 2^{J}$. The time domain representation of the wavelet packet bases $\psi_i[k]$ is obtained through a simple convolution rule as given in (3).

$$\psi_i[k] = f(k) * f(k/2) * .... * f(k/2^{J-2}) * f(k/2^{J-1});$$
$$\text{where,} \, 0 \leq i \leq 2^{J} - 1 \tag{3}$$
$$\text{and,} \, f(k) = \begin{cases} h'(k), \text{for lowpass branches} \\ g'(k), \text{for highpass branches} \end{cases}$$

Here *h'* and *g'* stand for the impulse responses of the low and high pass synthesis filters, respectively.

4.3 Generation of Wavelet Packet Dual Bases
The wavelet packet duals (to be used at the receiver end) are obtained from multichannel filter bank analysis too, though the processes are reversed. The duals are obtained from the analysis filters (*H* and *G*) through the analysis procedure which consists of filtering, decimation (down sampling) by 2 and decomposition at each stage. This process is called discrete wavelet packet transformation or DWPT. First the signal is passed through a half-

band high and low pass filter. The half-band low pass filter removes all frequencies that are above half of the highest frequency, while the half-band high pass filter removes all frequencies that are below half of the highest frequency of the signal. Such a half-band filtering halves the resolution, but leaves the scale unchanged. The signal is then sub-sampled by two since half of the number of samples is redundant, according to the Nyquist's rule. This decomposition halves the time resolution since only half the number of samples then comes to characterize the entire signal. Conversely, it doubles the frequency resolution, since the frequency band of the signal spans only half the previous frequency band effectively reducing the uncertainty by half. This procedure is iteratively repeated till

the desired degree of resolution. The duals $\tilde{\psi}_i[k]$ are derived as:

(a)

(b)

Fig. 4. (a) Generation of wavelets. A level-3 tree gives 8 wavelet packet bases. The up arrows represent interpolation by 2. *H'* and *G'* denote the frequency responses of the low and 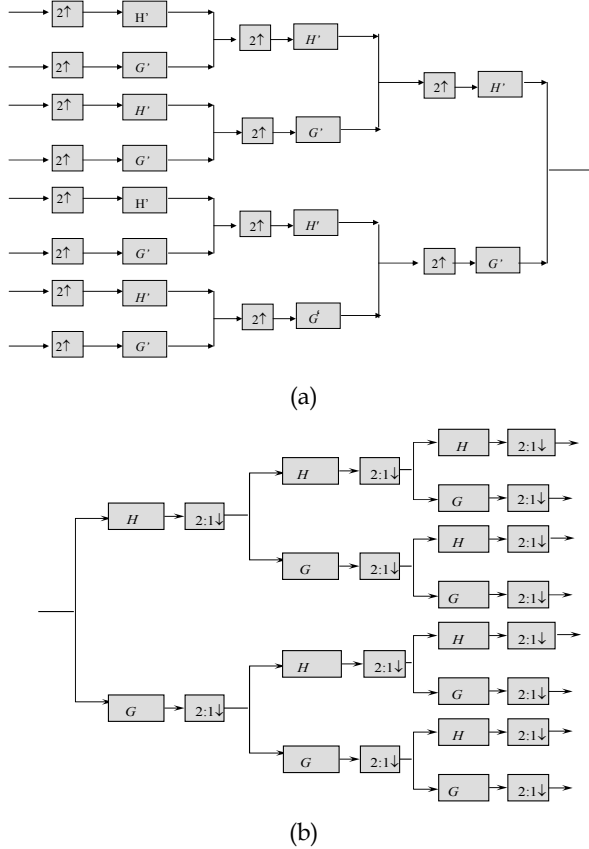high pass reconstruction filters, respectively; (b) Generation of wavelet Duals. A level-3 tree gives 8 wavelet packet dual bases. The down arrows represent decimation by 2. *H* and *G* denote the frequency responses of the low and high pass decomposition filters, respectively.

$$\tilde{\psi}_i[k] = f(k)*f(2k)*....*f(2^{J-2}k)*f(2^{J-1}k);$$
$$\text{where}, 0 \le i \le 2^J - 1 \tag{4}$$
$$\text{and}, f(k) = \begin{cases} h(k), \text{for lowpass branches} \\ g(k), \text{for highpass branches} \end{cases}$$

In the equation (4), h and g denote the impulse responses of the low and high pass analysis filters, respectively. Figure 4 illustrates the derivation of 8 wavelet packet bases and their duals from a cascaded level-3 tree structure.

5. Best Wavelet Packet Bases for WPMCM

5.1 Wavelet Theory
The attributes of a multicarrier modulation system greatly depends on the set of waveforms used. The property of the waveforms in turn depends on underlying filter banks used. Many considerations go into the design of a wavelet system including properties such as orthogonality, compact support, symmetry, and smoothness. Here we shall discuss a few important ones.

5.1.1 Paraunitary Condition
The paraunitary condition is essential for many reasons. Firstly, it is a prerequisite for generating orthonormal wavelets (Vaidyanathan, 1993). Second, it automatically ensures perfect reconstruction of the decomposed signal (Vaidyanathan, 1993) i.e., the original signal can be reconstructed without amplitude or phase or aliasing distortion, if the filter banks used satisfy the paraunitary condition. Only paraunitary filters are considered in this article and for such solution pairs the high pass and low pass filters share the relationship (Vaidyanathan, 1993) ; (Vetterli and Kovacevic, 1995); (Daubechies, 1992):

$$g[L-1-n] = (-1)^n h[n] \tag{5}$$

where L is the length of the filters. Further, paraunitary filters automatically satisfy the perfect reconstruction criterion (Vetterli and Kovacevic, 1995) with the decomposition and reconstruction filters being complex conjugate time reversed versions of one another i.e.

$$h'[n] = h^\dagger[-n] \text{ and } g'[n] = g^\dagger[-n] \tag{6}$$

Filters satisfying this condition are commonly used in signal processing, and are known as the Quadrature Mirror Filters (QMF). A nice import of these relations is that it is enough to design a single filter, either the low or high pass filter alone.

5.1.2 Compact support
This property ensures that the wavelet is of finite duration and the filter banks used to derive the wavelets have a finite number of non-zero coefficients (Burrus et al, 1998).

5.1.3 Regularity
This property is a measure of smoothness of the wavelet. The regularity condition requires that the wavelet be locally smooth and concentrated in both the time and frequency domains. It is normally quantified by the number of times a wavelet is continuously differentiable. The simplest regularity condition is the "flatness" constraint which is stated on the low pass filter. A LPF is said to satisfy Kth order flatness if its transfer function $H(z)$ contains K zeroes located at the Nyquist frequency ($z = -1$ or $\omega = \pi$). Parameter K is called the regularity order and for a filter of length L it satisfies the relation $0 \leq K \leq L/2$.

Wavelets are defined by the wavelet function $\psi(t)$ (i.e. the mother wavelet) and scaling function $\varphi(t)$ (also called father wavelet) in the time domain. Another way to determine the regularity of the wavelets is in terms of the number of vanishing moments of the wavelet and scaling functions (Burrus et al, 1998) and used the dual vanishing moments to determine the convergence rate of the multiresolution projections. The jth moments of the wavelet and scaling functions, $m_w(j)$ and $m_s(j)$, respectively, are defined in continuous time domain as follows:

$$m_w(j) = \int t^j \psi(t)dt \ \text{ and } \ m_s(j) = \int t^j \varphi(t)dt \tag{7}$$

5.2 Wavelet Families
In this work we shall largely deal with the Daubechies family and its variants.

5.2.1 Daubechies
The Daubechies are a family of orthonormal wavelets with compact support with highest degree of smoothness. It was derived by Ingrid Daubechies (Daubechies, 1992) who used all the degrees of freedom K to generate a wavelet family of maximum regularity for a given filter length L, or minimum L for a given regularity (Daubechies, 1992). This she did by imposing the maximum number of zero moments to the wavelet function in the vanishing moments' condition (equation 7).

5.2.2 Coiflet
Coiflets are a variation of the Daubechies wavelets. They are so named because it was derived by I. Daubechies at the behest of R. Coifman who suggested the construction of orthonormal wavelet basis with vanishing moment conditions for both wavelet and scaling functions (unlike Daubechies where only the wavelet functions have zero moments). The wavelet function has $2L$ moments equal to 0 and the scaling function has $2L$ -1 moments equal to 0.

5.2.3 Symlet

The symlet family of wavelets is another variant of the Daubechies family which are nearly-symmetrical (as opposed to being symmetrical). These modifications were also proposed by I. Daubechies and the properties of the two wavelet families are similar.

5.3 Choosing the Right Wavelet

In theory any time and frequency limited function can be utilized. However in practice, the wavelet bases cannot be arbitrarily chosen and instead have to satisfy a number of requirements. In general the choices to make can be with regard to the system of representation (continuous or discrete), properties of the wavelets desired (orthogonality/ biorthogonality, regularity/smoothness, frequency selectivity), the application in hand and the context of use (Burke, 1998). A framework that accounts for these requirements must first be defined and the wavelet selected in a principled approach through optimization of the wavelet design parameters.

5.4 Wavelet Design Considerations for WP-MCM application

With regard to the applicability to WP-MCM systems, the desirable properties may be listed as follows:

- The wavelet bases must be time-limited
- The bases must be well confined in frequency.
- The wavelet packet bases and their duals must be orthogonal (or at least linearly independent) to one another to enable perfect reconstruction.
- The bases must be orthogonal (or at least linearly independent) to one another in order to have unique demodulation.
- The bases must enable the system to handle channel effects and other distortions.
- The system must be easily realizable and must permit application of fast algorithms.

And in the filter bank domain the objective of the design procedure translates to construction of filters with the characteristics that they:

- have finite impulse response (FIR)
- are maximally frequency selective
- allow orthonormal expansion and perfect reconstruction of discrete-time signals
- satisfy the paraunitary condition
- satisfy a desired flatness/regularity condition

Amongst these properties the paraunitary and regularity properties are mandatory to the design of the filter banks. In addition to these properties the criterion that needs specific focus is the property of frequency selectivity.

5.5 Need for Maximally Frequency Selective Filter banks

In an ideal scenario the filter banks used to generate the wavelets have zero transition bands B, i.e., difference between pass and stop band frequencies (refer Figure 5). Under such an ideal scenario the wavelet packet bases derived from a level-i decomposition have confined spectral footprints with bandwidth $(1/2^i)$ times that of the Nyquist frequency. However, available wavelet families are derived from filter banks that have a wide transition band and

hence the resultant wavelet sub-carriers have a dispersed spectrum with footprints spilling into neighboring regions. The wider the transition bandwidth the greater the dispersion of the carrier's spectral footprint and therefore the greater the difficulty in isolating those sub-carriers that fall in the region of the licensed user. This greatly reduces the efficiency of the system. It is therefore important to design filter banks that have narrow transition bands.

5.6 Design of Maximally Frequency SelectiveFilter banks

The design procedure comprises of defining a low pass FIR filter, satisfying the regularity, paraunitary and frequency selectivity conditions, expressed in the form of an impulse response $h(n)$ or a transfer function $H(z)$ or a difference equation. For a filter of length L this is essentially solving L unknown filter coefficients from L linear equations. Of these L linear equations, $L/2$ equations come from the paraunitary constraint, K equations come from the regularity or flatness constraint and the remaining $L/2 – K$ conditions offer the room for maneuverability to establish the desired wavelet property such as frequency selectivity. The larger the value of $L/2 – K$, the greater the degree of freedom for frequency selectivity and the greater the loss in regularity. There is therefore a trade-off between frequency selectivity and regularity. Wavelets such as the Daubechies family are maximally flat with regularity order $K=L/2$ and hence they are not frequency selective.
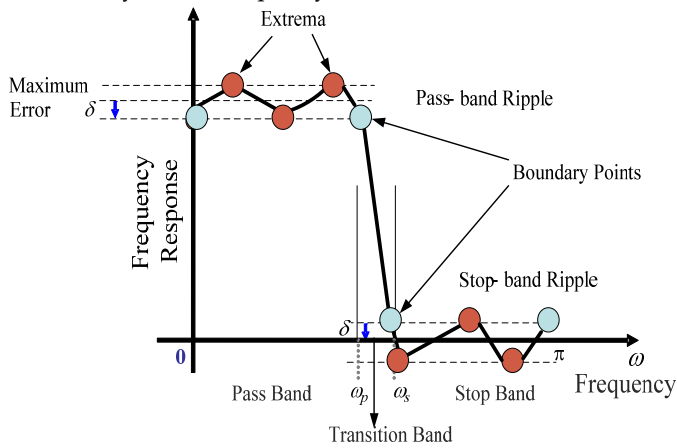

Fig. 5. Filter Characteristics in frequency. In the figure ω_p and ω_s, denote pass and stop band frequencies, respectively, $[0, \omega_p]$ is called the pass-band, $[\omega_s, \pi]$ is called the stop-band and $[\omega_p, \omega_s]$ is the transition band. δ is the tolerance or ripple.

To implement the frequency selective filters, the design parameters are stated in the frequency domain in terms of the desired magnitude response of the LPF as shown in Figure 5. In the figure ω_p and ω_s denote pass and stop band frequencies, respectively, $[0, \omega_p]$ is called the pass-band, $[\omega_s, \pi]$ is called the stop-band and $[\omega_p, \omega_s]$ is the transition band. δ is the tolerance or ripple. The design goal is to generate the filter with a desired transition band even while the maximum error δ in the pass/stop-band is minimized.

The fundamental theory on the design of frequency selective filter banks was developed by Rioul and Duhamel (Rioul and Duhamel, 1994). They devised the procedure to design maximally frequency selective filter banks under a given set of constraints using the Remez

exchange algorithm. The Remez exchange algorithm is an optimization algorithm that is commonly used in the design of FIR filters. It is popular because of its flexibility and computational efficiency. Also known as the Parks-McClellan algorithm, it works by converting the filter design problem into a problem of polynomial approximation (Oppenheim and Schafer, 1989). The algorithm is an application of the Chebyshev alternation theorem that constructs the polynomial of best approximation to a desired function under a set of constraints. Through a minimax approximation the scheme seeks to arrive at a lth order approximation polynomial function $A(x)$ that best approximates a desired filter polynomial function $D(x)$ (in our case the LPF $H(z)$ that satisfies the design specifications) in the given interval such that the absolute maximum error is minimized. The error is defined here as the weighted difference between the desired filter polynomial function and the approximation polynomial function and is given as

$$E(x) = W(x)(D(x) - A(x)) \tag{8}$$

$E(x)$ and $W(x)$ are respectively the error and weighting polynomial functions. All polynomial functions are of the form $\sum_{i=0}^{l} p_i x^i$, with coefficients p_i and degree of the polynomial function l. Chebyshev proved that such a polynomial $A(x)$ exists and that it is unique. He also gave the criteria for a polynomial to be a minimax polynomial. The algorithm states that in the interval of consideration, the necessary and sufficient condition that $A(x)$ is the unique mini-max polynomial solution of degree l is that there are at least (l + 2) points at which the error function $E(x)$ attains the absolute maximum value δ with alternating sign i.e.

$$E(x_i) = -E(x_{i+1}) = \pm \max_{x \in I} \{| E(x) |\} = (-1)^i \delta \tag{9}$$

for $x_1 < x_2 < ... < x_{l+2}$ in the desired interval I. Parks and McClellan proved that this approach could be used to derive a filter of a given length with minimal ripple. The right set of extremal points x_i is arrived through an iterative procedure. In each iteration an interpolation problem is solved and the reference set of extremal points is updated. Rioul and Duhamel deduced that $L/2 - K + 1$ extremal points in the pass-band are necessary and sufficient to characterise a unique and optimal solution (Rioul and Duhamel, 1994).

The procedure starts by choosing an arbitrary set of $L/2 - K + 1$ points in the given interval. These $L/2 - K + 1$ points help form $L/2 - K + 1$ linear equations. The filter coefficients are obtained by solving the $L/2 - K + 1$ linear equations in a way that the error at the $L/2 - K + 1$ points considered is equal in magnitude and alternating in sign. It cannot be guaranteed after the first step that solution satisfies the minimax condition for the error function. That is the magnitude of the error need not be the absolute maximum magnitude in the interval of consideration. In order to find the minimax solution, the second step of the algorithm seeks new set of $L/2 - K + 1$ points that approach the $L/2 - K + 1$ points of the minimax solution. The new set of is determined by locating those points where the slope of the error function $E(x)$ is zero. Once these points are identified, the old set of

$L/2 - K + 1$ points is exchanged with the new points. This process is iteratively performed till the desired set of points that satisfy the minimax solution is obtained. The algorithm is said to have converged when the set of extremal points remains unchanged. Once the right set of extremal points is identified, the optimum error and the filter can be obtained. The exact details of how these equations are solved to obtain the low pass filter H(z) using the modified Remez exchange algorithm can be found in (Rioul and Duhamel, 1994). From the low pass filter $H(z)$, the high pass filter $G(z)$ and the reconstruction filters ($H'(z)$ and $G'(z)$) can be obtained by applying (5) and (6).

6. VBLAST and Spatial Processing to Estimate Data

The transmitted signals arrive at the receiver antenna array as multiple streams with different spatial signatures and to estimate the data a suitable detection algorithm is necessary. In this work we consider VBLAST, a promising and elegant spatial coding technique for MIMO realizations (Wolniansky P.W. et al, 1998), for detection of transmitted data.

The V-BLAST algorithm detects data by a technique known as linear nulling or symbol cancellation. In linear nulling, each received sub-stream is considered to be the desired signal with the remaining sub-streams taken as interferers. Using symbol cancellation, interference from already-detected components of the transmitted vector is subtracted from the received signal vector. This results in a modified received vector with fewer interferers. To this end the received data streams are ordered on the basis of their strength. The substream with the strongest signal is detected and its contribution subtracted from the total received vector signal. The process is continued till all other substreams are identified. The entire process may be likened to decision feedback equalization (Wolniansky P.W. et al, 1998).

Assuming that the receiver has complete knowledge of the channel matrix **C**, the V-BLAST algorithm is implemented as follows:

1. Build a Moore pseudo-inverse matrix **P** of the channel matrix **C**,

$$\mathbf{P} = (\mathbf{C}^*\mathbf{C})^{-1}\mathbf{C}^* \qquad (10)$$

2. Find the row of P where its Euclidean norm is the smallest one,

$$k = \arg\min_j \| \mathbf{P}_j \| \qquad (11)$$

and j is the column of matrix **P**.

3. Take the row k of **P** as the nulling vector w,

$$\mathbf{w} = (\mathbf{P})_k \qquad (12)$$

4. Obtain the strongest transmit signal,

$$r_k = \mathbf{w}^*\mathbf{y} \qquad (13)$$

5. Estimate the transmitted symbol $\hat{s}_k$ by demapping r_k.

6. After detection of the strongest transmitted signal, its effect is cancelled from the received signal vector to reduce the detection complexity of the remaining transmit signals.

$$\mathbf{y} = \mathbf{y} - (\mathbf{C})_k * \{\text{Mapping}\}(\hat{s}_k) \qquad (14)$$

here k is the column index. The kth column of channel matrix $\mathbf{C}$ is then zeroed for the purpose of detection of the strongest transmitted signal on the next layer.

The procedure is continued until transmitted symbols in all layers are detected.

7. Simulation Setup

The CR is a WPMCM based MIMO system which operates with 128 equally spaced carriers that can be adaptively deactivated to shape the transmission spectrum. The 128 wavelet packet carriers are derived from a level-7 cascaded tree. The filter banks considered are daubechies, symlet, coiflet and the maximally frequency selective filters with parameters L=30, K=12, B=0.1. The LU is taken to occupy two equal bands with the cumulative bandwidth comparable to 32 carriers (1/4th) of the CR system and located centered on the 32nd and 96th carrier (each equal to around 16 CR carriers' bandwidth) of the CR spectral band, as shown in Figure 6.
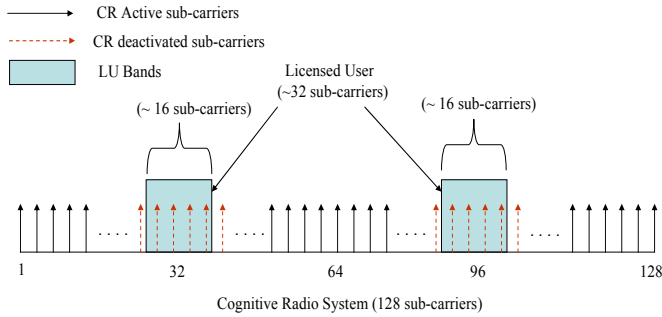

Fig. 6. Cognitive Radio (CR) and Licensed User (LU) Characteristics.

As stated earlier, the carriers of the CR system in and around the LU spectrum are dynamically deactivated to facilitate co-existence. Various MIMO configurations are considered to enhance the system functioning with all the MIMO arms transmitting the same data. The modulation scheme considered is binary phase shift keying (BPSK).To measure the effectiveness of the CR system, the Bit-error rate (BER) versus the Signal-to-Noise Ratio (SNR) is plotted. To simplify the evaluation of the system, the communication channel is assumed to be Additive White Gaussian Noise (AWGN). No multi-path interference is considered. The transceivers are taken to operate in unison maintaining full time/phase/frequency synchronization. Moreover, they are assumed to be stationary with negligible Doppler effects.

8. Simulation Results

8.1 Performance of CR Setup

To understand the performance of the CR setup, two set of results are evaluated – operation of CR under LU and operation of LU under CR. A well designed CR setup must allow both the CR and LU systems to coexist seamlessly without any compromise in performance of either system.

Figure 7(a) shows the BER performance curves of the WP-MCM MIMO based CR system in the scenarios: absence of LU (only AWGN), presence of LU without any carrier deactivation and presence of LU with carrier deactivation. The MIMO configuration considered is 2x2. The carriers were derived from maximally frequency selective filters with L=30, K=12 and B=0.1. From the plots it is quite clear that the presence of LU affects the CR performance. And when the CR transmission is communicated around the LU with carriers in and around the region of interference removed, the CR system recovers. Best results are obtained when a total of 36 (18 each on and around the two LU bands) or more of the CR carriers are vacated. Figure 7(b) shows the corresponding BER curves for the LU co-existing with CR. The results of the WPMCM based CR system with carrier deactivation are equally encouraging.

8.2 Influence of MIMO Configuration on WP-MCM MIMO Setup

We now present the performance of the CR system under various WPMCM MIMO configurations. The WPMCM MIMO based CR system operates under a LU with 40 of its 128 carriers in and around the region of LU deactivated. Figure 8 shows the related plots. To better understand the operation, we classify the MIMO configurations as follows:

a) First we consider the case with the same of number of antennas in the both the transmitter and receiver ends (Figure 8a). The formations considered are 1x1, 2x2, 3x3 and 4x4. The pattern of the BER curves is unambiguous. With increase in the number of transmit-receive pairs, the BER system functioning improves. This is due to the array gain.

b) Effect of receiver antennas – Increasing the number of receiver antennas significantly improves the system performance. This is illustrated in Figure 8b where the number of transmit antennas is kept constant at 2 while increasing the number of receive antennas to 3, 4, 5 and 6 respectively. For a BER of 10^{-4} the improvements with respect to the 2x2 configuration are 2 dB for 2x3, 3dB for 2x4, 4dB for 2x5 and 5dB for 2x6.

c) Figure 8c shows the curves for a few other MIMO configurations. In all MIMO configurations the performance of the CR setup is efficient and allows co-existence with LU.

8.3 Comparison of performance of WP-MCM and OFDM based CR systems

It will be interesting to see as to how the WP-MCM construction devised in this paper matches up to traditional OFDM implementations. Both the OFDM and WP-MCM scheme use the same set of transmission parameters – 128 carriers each with 40 carriers in the interference band removed. The wavelet used belongs to the family of maximally frequency selective filters. The curves (Figure 9) are for the MIMO configurations 1x1, 2x2 and 3x3. For all the cases considered the WPMCM setup performs comparably well with its OFDM counterpart.
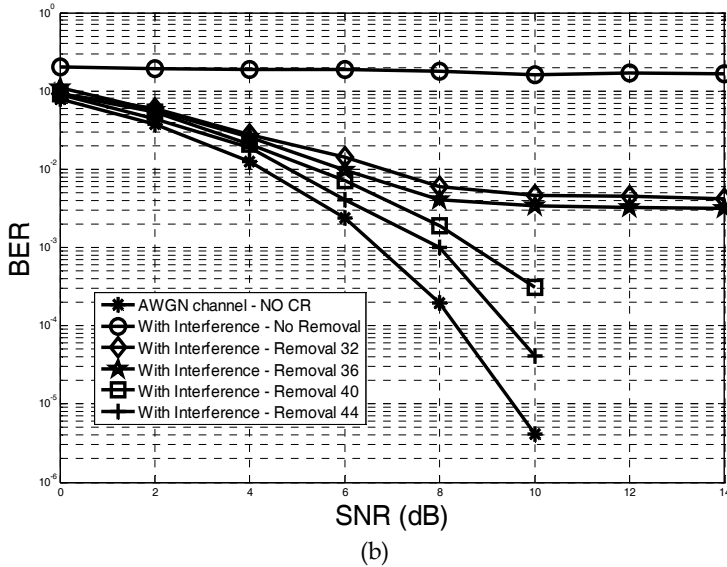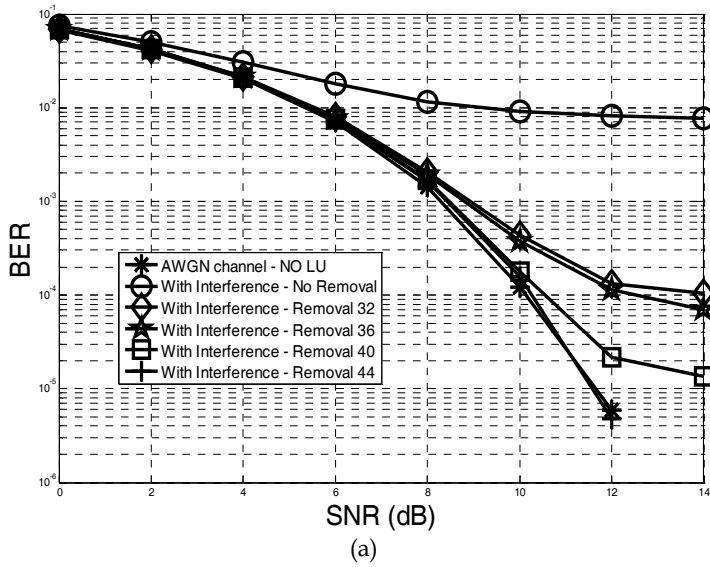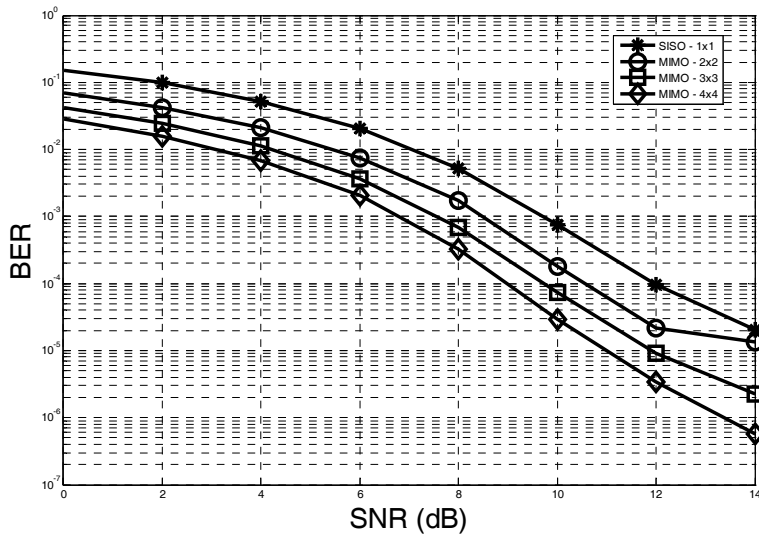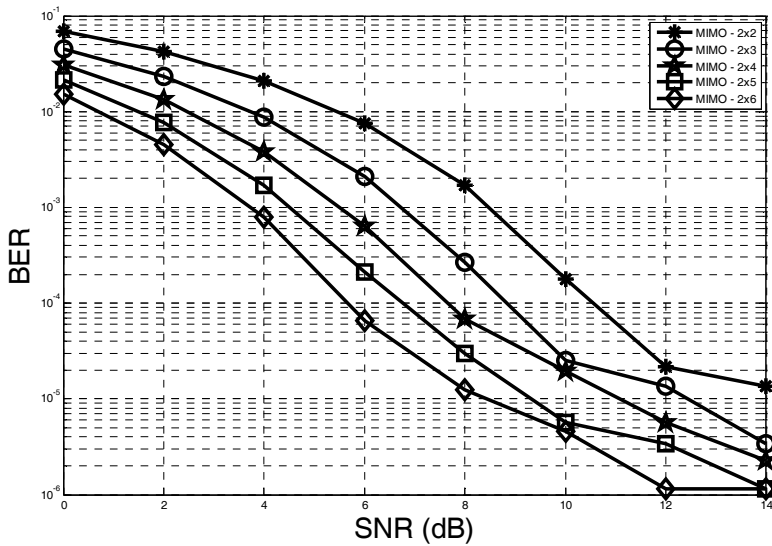
(a)

(b)

Fig. 7. Performance of the proposed WP-MCM MIMO based CR system. a) CR under LU; b) LU under CR. The carriers of CR in the region of the LU are removed to promote co-existence of CR and LU. The carriers were derived from Frequency selective filters with L=30, K=12 and B=0.1. The CR is 2x2 MIMO.

(a)

(b)

(c)

Fig. 8. Performance of WPMCM MIMO based CR system for various combinations of transmit-receive antennas. a) Balanced Case when the number of transmit and receive antennas are the same; b) Effect of receiver configuration. c) A few other MIMO configurations; MIMO stands for Multiple input-multiple output, SISO for single input-single output and SIMO for single input-multiple output. The carriers were derived from Frequency selective filters with L=30, K=12 and B=0.1. The CR is 2x2 MIMO.

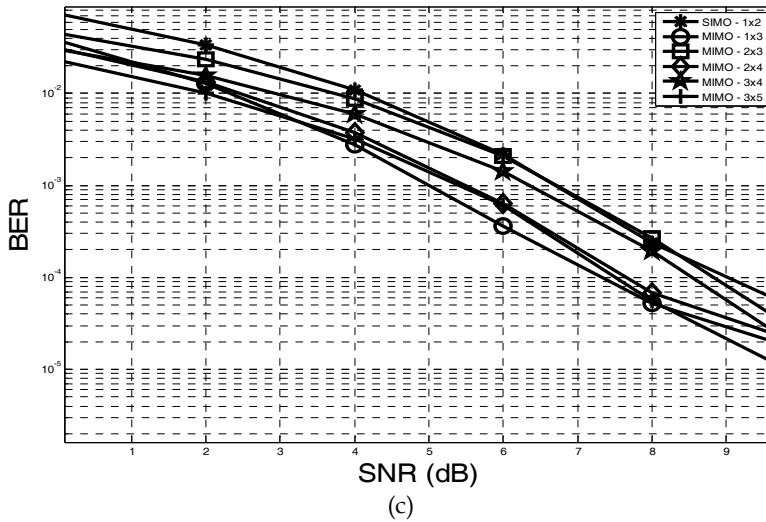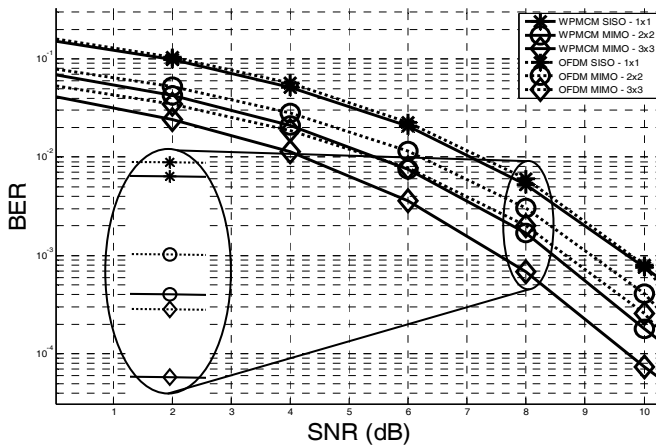Fig. 9. OFDM versus WP-MCM based CR. BER performance comparison of OFDM and WP-MCM based CR systems in the presence of LU. MIMO stands for Multiple input-multiple output, SISO for single input-single output.
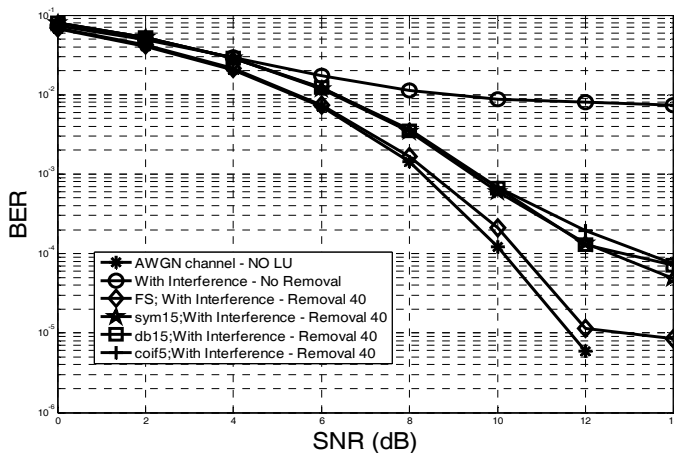
Fig. 10. Comparison of different wavelets: BER performance of the WPMCM MIMO based CR system in the presence of LU. The carriers of CR in the region of the LU are removed to enable it to co-exist with LU. FS denotes the maximally frequency selective wavelet, db denotes Daubechies, sym denotes symlet and coif stands for coiflet. The MIMO used is 2x2.

Name	Orthonormal	Length	Regularity
Daubechies	Yes	30	15
Coiflet	Yes	30	15
Symlet	Yes	30	15
Frequency Selective	Yes	30	12

Table 1. Filter Bank Characteristics

8.4 Performance comparison of the system when using the maximally frequency selective filter bank with respect to conventional wavelets of similar lengths.

Lastly we equate the performance improvements brought by maximally frequency selective filter bank with a few well known wavelets. The wavelets considered are Daubechies-15, Coiflet-5 and Symlet-15. All of these filters satisfy the paraunitary condition and hence give orthonormal bases. More on these wavelets can be found in table 1. For the fairness of comparison, all the filter banks are taken to be of the same length (L=30). The system performance has a direct correlation to the frequency selectivity of the filter bank used to derive the carriers. This effect is exemplified in Figure 10 where the BER performance curves of the CR under LU systems have been plotted. The MIMO used is 2x2.

The maximally frequency selective filter banks have the narrowest transition band and hence they easily surpass the other wavelet families (by up to 3dB). This essentially translates to less carrier removals and better SNR gain.

9. Summary

A novel wavelet packet based multi-carrier modulation system that attempts to blend the benefits of Cognitive Radio, and MIMO was presented. Cohabitation of the WP-MCM based CR system with LU was made possible by dynamically activating/deactivating the CR carriers in a way that the CR and LU systems don't have any competing time-frequency components. The carriers of the WP-MCM system were generated by multistage tree-structured paraunitary filter banks. Wavelets including the Daubechies, Coiflet and Symlet families were used for the study. The emphasis was on deriving optimal maximally frequency selective wavelet packet bases that best suit applicability to spectrum shaping in CR systems. Through simulation studies the usefulness and potential of WP-MCM for developing CR systems was demonstrated.

In addition to the CR and MIMO features, using the wavelet packet bases provided a third dimension of system optimization. Unlike conventional Fourier bases which are static in nature, wavelets offer multitude of variations which can be customised to the requirement in hand. In this article, the design criterion is set to avoid interference to licensed primary users. This can be easily furthered by altering the design criterion to include other requirements such as reduction of ISI/ICI or PAPR.

Comparing with traditional OFDM implementations, the performance of the proposed system in the presence of a LU matches quite well.

In conclusion, the performance results of the simulation studies make us to conclude that the novel proposed system can be fruitfully used to construct adaptive intelligent cognitive systems.

10. References

Bouwel C.V., Potemans J., Schepers S., Nauwelaers B. and Capelle A.V., "Wavelet Packet Based Multicarrier Modulation", *Proc. IEEE Benelux Symposium on Communications and Vehicular Technology*, Leuven, Belgium, October 2000.

Burke. B, The World According to Wavelets: The Story of a Mathematical Technique in the Making, *Second Edition, A K Peters*, May 1998.

Burrus C. S., Gopinath R. A., Guo H., "Introduction to Wavelets and Wavelet Transforms, a Primer", *Upper Saddle River*, NJ (USA): Prentice Hall, 1998.

Daubechies I., Ten Lectures on Wavelets, *SIAM, Philadelphia*, 1992.

De Lima E.R., Flores S.J., Almenar V., Canet M.J., "Performance Enhancements in OFDM-WLAN Systems Using MIMO Access Techniques", *International Workshop on Telecommunications*, Santa Rita do Sapucai, MG, Brazil, August, 2004.

Haykin. S, "Cognitive Radio: Brain-Empowered Wireless Communications", *in IEEE Journal On Selected Areas In Communications*, No.2, vol 23, Feb 2005, pp. 201-220.

Jamin A. and Mahonen P., "Wavelet Packet Modulation for Wireless Communications", *Published In Wireless Communications & Mobile Computing Journal*, Vol. 5, Issue 2, pp. 123-137, John Wiley and Sons Ltd. March 2005.

Lakshmanan M.K. and Nikookar H., "A Review of Wavelets for Digital Wireless Communication", *Wireless Personal Communications*, Volume 37, Numbers 3-4, May 2006, pp. 387-420(34), Springer.

Lakshmanan M.K., Budiarjo I. and Nikookar H., "Wavelet Packet Multi-carrier Modulation MIMO Based Cognitive Radio Systems with VBLAST Receiver Architecture", *IEEE*

Wireless Communications and Networking Conference (*WCNC*), March-April 2008, Las Vegas, USA.

Lee M. J., Temple M. A., Claypoole R. L., and Raines R. A., "Wavelet domain communication system: Bit error sensitivity characterization for geographically separated transceivers," *Proc. MILCOM 2002*, Anaheim, CA, vol. 2, pp. 1378–82, October 2002.

Lindsey A., "Wavelet Packet Modulation for Orthogonally Transmultiplexed Communications," *IEEE Transaction On Signal Processing*, vol. 45, May 1997, pp. 1336–1339.

Mitola III J., "Cognitive Radio: An Integrated Agent Architecture for Software Defined Radio", *Doctoral Dissertation*, May 2000.

Negash B.G., and Nikookar H., "Wavelet-based Multicarrier Transmission over Multipath Wireless Channels", *IEE Electronics Letters*, Vol. 36 No. 21, pp. 1787–1788, October 2000.

Oppenheim A. V. and Schafer R. W., Discrete-Time Signal Processing, *Englewood Cliffs*, NJ, Prentice Hall, 1989.

Rioul O. and Duhamel P., "A Remez Exchange Algorithm for Orthonormal Wavelets", *IEEE Trans. Circuits Systems - II, vol. 41, no. 8,* pp. 550-560, August 1994.

Vaidyanathan P. P., Multirate Systems and Filter Banks, *Englewood Cliffs, NJ, Prentice-Hall, Inc., 1993*.

Vetterli M. and Kovacevic I., Wavelets and Subband Coding, *Prentice Hall PTR, Englewood Cliffs, New Jersey, 1995*.

Weiss T.A. and Jondral F.K., "Spectrum Pooling: An Innovative Strategy for the Enhancement of Spectrum Efficiency", *IEEE Communications Magazine*, 42:S8–S14, March 2004.

Wolniansky P.W. et al., "V-BLAST: An Architecture for Realizing Very High Data Rates Over the Rich Scattering Wireless Channel", *Proc. ISSE*, Pisa, Italy,pp 295-300, Oct 1998.

A Competitive Approach for Designing Multi-Antenna Cognitive Access Networks[*]

E. Baccarelli, M. Biagi, N. Cordeschi,
T. Patriarca and V. Polli
"Sapienza", University of Rome
Info-Com Dept.
Italy

1. Introduction to the cognitive MIMO radio paradigm

Cognitive radio has recently shown its great potential for 4th-Generation wireless local access networks (i.e., 4GWLANs) and attracted wordwide interest in academics, industry and standardization activities (Glisic, 2006). The US Federal Communication Commission (FCC, 2003) has issued a notice of public rule-making and order regarding cognitive radio technologies. The Defense Advanced Research Project Agency (DARPA, 2001) launched the 'next generation (XG) communication program' to develop new technologies to dynamically manage the radio spectrum. The US Army has also been researching 'adaptive spectrum exploitation' (ASE) for real-time spectrum management in the battlefield (Lee, 2001). The US National Science Foundation (NSF) has recently launched the 'Programmable Wireless Networking' (ProWin) program focused on adaptive, agile, cognitive radios and networking. On the standardization side, the IEEE 802.22 working group is elaborating on the use of TV bands for spectral-agile wireless regional access networks (WRANs) (Cordiero et al., 2005; IEEE 802.22, 2005). Basically, cognitive radio is capable of ascertaining its operating environment and adapting to real-time conditions of its (localized) wireless channel, including the ability to sense spectrum-usage by neighbouring devices, change operating frequency band, adjust radiated power and modify transmission parameters (Mitola, 1999; Haykin, 2005; Mangold et al., 2004; Zheng & Cao, 2005). Such an intelligent and agile use of the spectrum resource is expected to improve the spectrum-access capability of emerging 4GWLANs (Glisic, 2006; Butler & Webb, 2006; IEEE 802.22, 2005; Chow et al., 2007).

However, wireless access capacity may be also improved significantly by resorting to multi-antenna (i.e., MIMO) architectures that allow exploiting the spatial dimension of the

underlying access system via SDMA (Shad et al., 2001; Paulraj et al., 2003). From this point of view, the MIMO paradigm offers, indeed, several important capabilities (such as spatial degree of freedom, spatial diversity, multiple-access interference (MAI) mitigation) that may be suitably exploited for increasing the overall access efficiency of 4GWLANs (Glisic, 2006; Paulraj et al., 2003). However, in order to fully avail of these spatial capabilities offered by the MIMO paradigm in a flexible and agile way, a still open key-question concerns the design of asynchronous, scalable and self-configuring SDMA policies able to meet the QoS-requirements advanced by (battery-powered nomadic) users.

In principle, this flexibility may be attained by merging the cognitive paradigm with the MIMO one, thus giving rise to the cognitive MIMO radio paradigm. In fact, the cognitive concept directly leads to the development of distributed, scalable, and flexible radio access networking architectures by resorting to an opportunistic and (possibly) competitive real-time management of the available radio-resource, that, in turns, may be suitably modelled via the formal tool of Game Theory (Fudenberg & Tirole, 1991, Chaps 16, 24).

1.1 Previous works on competitive cognitive-inspired access

Specifically, distributed power-control in wireless access networks by exploiting Game Theory is tackled in (MacKenzie & Wicker, 2001). A game-theoretic approach is also pursued in (Palomar et al., 2003) for dealing with power-allocation in single-user MIMO systems in the presence of imperfect channel estimation. The paper (Altman & Altman, 2003) examines the general application of the super-modular games to the distributed power-control, while the contribution in (Goodman & Mandayam, 2000) considers an application scenario wherein the accessing radios attempt to maximize a suitable function of their throughput and leads to the conclusion that the underlying game converges to a Nash Equilibrium (NE). Later, (Saraydar et al., 2002) point out that the game considered in (Goodman & Mandayam, 2000) is, indeed, a super-modular game. The contribution in (Sung & Wong, 2003) focuses on a (somewhat modified) version of the single-cell distributed power-control game that is based on the concept of nonlinear group pricing. Applications of Game Theory to fair-efficient admission control are presented in (Virapanicharoen & Benjpolakaul, 2004), while the whole topic of competitive flow-control and distributed routing is discussed in (Kannan & Iyengar, 2004; Altman et al. 2002).

Game theoretic approaches for agile spectrum sharing have been recently developed in (Fattahi et al.,2007; Xing et al., 2007; Wang & Brown, 2007). Specifically, (Fattahi et al., 2007) focus on the design of a coordinated mechanism for allocating time and frequency slots to competing users that are capable to self-reconfigure their terminals via cross-layer strategies. (Xing et al., 2007) explore the price dynamics of a competitive market composed by (multiple) self-interested service providers that compete for potential customers by offering heterogeneous spectrum-agile networking technologies at different costs. (Wang & Brown, 2007) develop a public safety and commercial spectrum-sharing strategy that resorts to both agile network pricing and call admission control for the competitive maximization of the commercial revenue under low (e.g., upper limited) blocking-probability to public safety calls.

Overall, all the above cited works consider applications scenarios characterized by either single-antenna radios (i.e., they do not consider the spatial dimension), or single-user MIMO systems (i.e., they do not deal with the access problem), thus leaving open the question about potential access capability of the cognitive MIMO radio paradigm.

1.2 Proposed contributions and application scenarios

This chapter focuses on the competitive access in WLAN systems operating in an infrastructure-mode, where power-limited multi-antenna non-cooperative cognitive self-configuring radios attempt to join in uplink to a (possibly multi-antenna) access point (AP) (Fig. 1(a)). The target pursued by each radio is to maximize its own access information-throughput in the presence of both (Rayleigh-distributed) flat-fading and MAI induced by the other accessing terminals. For achieving this goal in a fully distributed and asynchronous way (i.e., in a non-cooperative way), each radio exploits its cognitive capability to 'learn' (i.e., estimate) both its own current MIMO faded uplink and the covariance matrix of the MAI induced by the other accessing radios. Thus, on the basis of these learned (i.e., acquired) context information, each radio proceeds to self-reconfigure its access policy by performing suitable power-allocation and (statistical) spatial-shaping of the radiated signals. This learning/self-configuring action is autonomously iterated by the accessing radios according to the general rules of the non-cooperative strategic games, until the overall Access Network (AN) converges to a stable operating state (e.g., the NE of the underlying game). We investigate the performance of this competitive and cognitive access policy under both best-effort and contracted-QoS access strategies.

Furthermore, we also consider the case when even the AP is cognitive, thus meaning that it attempts to learn and subtract the MAI contribution affecting the signal received by each accessing radio. Overall, the key-results of this work may be so summarized.

- First, the access strategy we present exploits both the spatial-dimension offered by the physical MIMO platform and the cognitive capability of the AN to maximize (in a competitive sense) the access throughput of each accessing radio. Interestingly, this goal is achieved without requiring radio synchronisation and/or peer-to-peer information-passing among the radio terminals (i.e., in a fully asynchronous and distributed way).

- Second, the proposed access strategy allows to implement both best-effort and contracted-QoS access policies, and it may also provide for multiple QoS classes. Interestingly enough, we anticipate that, when the QoS-requirements advanced by the radios are no sustainable by the AN, thus, the proposed access algorithm is able to self-move the operating point of the whole AN (e.g., the operating NE point) to the nearest sustainable one. This means that, besides the radio terminals, the overall AN we go to develop is, indeed, cognitive and self-configuring. Thus, it may be considered as an instance of active access network (Glisic, 2006; Chap.16).

- Finally, several numerical results corroborate the conclusion that the proposed cognitive and self-configuring AN is able to outperform conventional CSMACA/ OFDMA/CDMA/TDMA-based no cognitive access systems in terms of both aggregate access throughput and access capacity, especially when the MAI experienced at the AP side is strong.
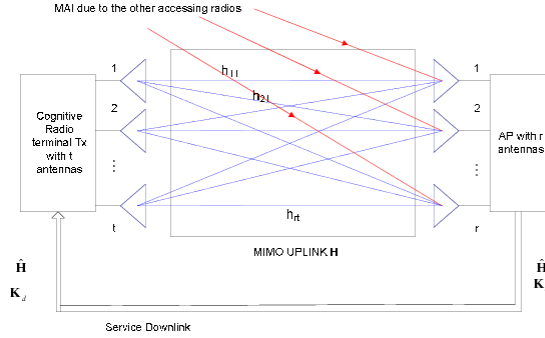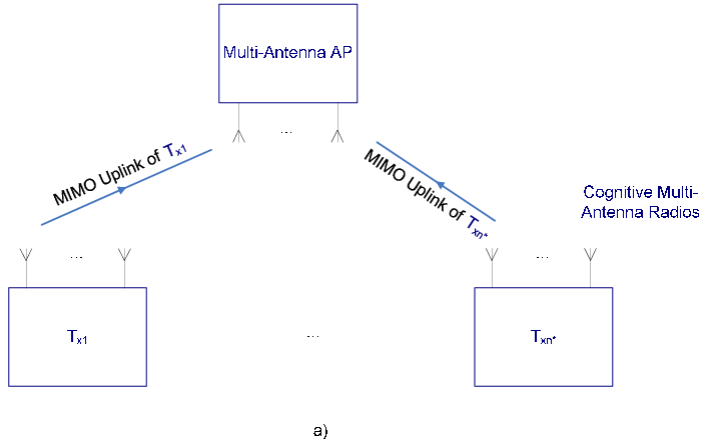
Fig. 1. (a) The considered multi-antenna local AN (b) Zoomed block diagram of the up-down links joining the radio terminal Tx to the AP

1.3 Organization of the chapter

The remainder of this work is organized as follows. After the system modelling of Section 2, Section 3 deals with the evaluation of the conveyed information throughput in wireless access networks affected by MAI. In Section 4 the optimized power-allocation and interference mitigation algorithms are presented. Thus, after shortly reviewing in Section 5 some Game Theory essentials, in Section 6 we propose a game-inspired access policy. Actual performance of the proposed access strategy in terms of access throughput and access capacity is tested in Section 6, while some conclusive remarks are drawn in Section 7.

About the adopted notation, we anticipate that, capital letters indicate matrices, lower-case underlined symbols denote vectors, while characters overlined by arrow denote block-matrices and block-vectors. Furthermore, apexes $*$, T , $\dagger$, mean conjugation, transposition and conjugate-transposition respectively, while lowercase letters will be used for scalar

quantities. Furthermore, $\det[A]$ and $Tra[\mathbf{A}]$ mean determinant and trace of the matrix $A \triangleq [\underline{a}_1 \ldots \underline{a}_m]$. Finally, $\mathbf{I}_m$ is the $(m \times m)$ identity matrix, $\|\mathbf{A}\|_E$ is the Euclidean norm of the matrix $\mathbf{A}$, $\mathbf{A} \otimes \mathbf{B}$ is the Kronecker product of the matrix $\mathbf{A}$ by matrix $\mathbf{B}$, $\underline{0}_m$ is the m-dimensional zero-vector, lg denotes natural logarithm and $\delta(m,n)$ is the Kronecker delta.

2. The considered SDMA network-model

As previously anticipated, the application scenario we consider is the MIMO uplink of a packet-oriented WLAN, where $n^* \geq 2$ multi-antenna battery-powered radios attempt to simultaneously access to a (possibly multi-antenna) AP (Fig. 1(a)). Since the accessing radios are assumed non-cooperative, we begin to focus on the MIMO uplink joining a single radio Tx to the AP, thus considering the signals radiated by all other $(n^*\text{-}1)$ accessing radios as (additive) MAI. The resulting (complex base-band equivalent) point-to-point MAI-impaired MIMO uplink is sketched in Fig. 1(b). Simply stated, this uplink is composed by an accessing radio Tx (equipped with $t \geq 1$ antennas) communicating to the AP (equipped with $r \geq 1$ antennas) via a MIMO channel impaired by both Rayleigh flat fading and additive MAI. Path gain h_{ji} from the transmit antenna i to the receive one j in Fig. 1(b) may be modelled as a complex zero-mean unit-variance proper complex random variable (r.v.) (Baccarelli & Biagi, 2003; Paulraj et al., 2003), and, for sufficiently spaced-apart antennas, the path gains $\{h_{ji} \in \mathbb{C}, \ 1 \leq j \leq r, \ 1 \leq i \leq t\}$ may be considered mutually uncorrelated (Paulraj et al., 2003). Furthermore, these gains h_{ji} may be also assumed time-invariant over $T \geq 1$ signalling periods, after which they change to new statistically independent values held for another T signalling period, and so on. The resulting 'block-fading' model well captures the main features of several frequency-hopping or packet based interleaved 4G systems, where each transmitted packet is detected independently of any other (Paulraj et al., 2003). Since the statistic features of the MAI in Fig. 1(b) depend on both WLAN topology and signals radiated by all accessing radios, it is reasonable to assume that these statistics remain constant over (at least) the whole packet duration. However, since both path gains h_{ji} and MAI statistics may change from a packet to another, we assume that radios and AP are not aware of them at the beginning of each transmitted packet, but they may exploit their cognitive capabilities to learn (i.e., acquire) them. For this purpose, according to the packet-oriented structure of the considered WLAN, we assume that the (coded and modulated) data-streams radiated in uplink by the transmit antennas of Fig. 1(b) are split into packets composed by $T \geq 1$ slots, where the first T_{L1} slots are used by the receiver (i.e., the AP) to 'learn' the MAI statistics (Section 2.1), the second T_{L2} slots are employed to 'learn' the path gains of the forward MIMO channel (Section 2.2), and the last $T_{pay} = T\text{-}T_{L1}\text{-}T_{L2}$ slots convey payload data (Section 2.3).

2.1 First learning phase
During the first learning phase, no signal is radiated by the Tx radio of Fig. 1(b), so as to allow the AP to learn the statistics of the MAI induced by all other accessing radios (Fig. 1(a)). Thus, the r-dimensional (complex column) vector collecting the outputs of the receive antennas at the AP side over the n-th slot of this first learning phase may be modelled as

$$\underline{\dot{y}}(n) \triangleq \underline{\dot{d}}(n) \equiv \underline{\dot{v}}(n) + \underline{\dot{w}}(n), \quad 1 \le n \le T_{L1}. \tag{1}$$

In (1), $\underline{\dot{w}}(n)$ accounts for the receiver thermal noise, so that $\{ \underline{\dot{w}}(n) \}$ may be modelled as a zero-mean, proper complex, spatially and temporally white sequence, with covariance matrix equal to

$$E\left\{ \underline{\dot{w}}(n)\left(\underline{\dot{w}}(n)\right)^{\dagger} \right\} = \mathcal{N}_0 \boldsymbol{I}_r \delta(m,n), \tag{1.1}$$

where $\mathcal{N}_0$ (*Watt / Hz*) is the thermal noise level. The component $\{ \underline{\dot{v}}(n) \}$ in (1) accounts for the MAI induced by all other accessing radios, and thus it may be suitably modelled as a zero mean, temporally-white, spatially-colored Gaussian sequence (Paulraj et al., 2003), with covariance matrix $\boldsymbol{K}_v \triangleq E\left\{ \underline{\dot{v}}(n)\left(\underline{\dot{v}}(n)\right)^{\dagger} \right\}$ constant over (at least) a packet-transmission interval. However, since $\boldsymbol{K}_v$ may change from a packet to another, it is reasonable to assume that both Tx and AP in Fig. 1(b) are not aware of the overall disturbance covariance matrix

$$\boldsymbol{K}_d \triangleq E\left\{ \underline{\dot{y}}(n)\left(\underline{\dot{y}}(n)\right)^{\dagger} \right\} \equiv \boldsymbol{K}_v + \mathcal{N}_0 \boldsymbol{I}_r, \tag{2}$$

at the beginning of each transmitted packet. Besides, since during this first learning phase the signal $\underline{\dot{y}}(n)$ in (1) received at the AP equates the MAI one $\underline{\dot{d}}(n)$, the AP may exploit its cognitive capabilities to learn $\boldsymbol{K}_d$. A very simple way to accomplish this task is suggested by the law of large numbers (Poor, 1994). By fact, this last guarantees that an unbiased and consistent (i.e., asymptotically exact) estimate $\check{\boldsymbol{K}}_d$ may be obtained via the following sample-average:

$$\check{\boldsymbol{K}}_d = \frac{1}{T_{L1}} \sum_{n=1}^{T_{L1}} E\left\{ \underline{\dot{y}}(n)\left(\underline{\dot{y}}(n)\right)^{\dagger} \right\} . \tag{3}$$

As already pointed out in (Baccarelli et al., 2007), we anticipate that the performance effects of (possible) mismatches between actual $\boldsymbol{K}_d$ and its estimate $\check{\boldsymbol{K}}_d$ are not so critical, so that, in the sequel, we directly assume $\check{\boldsymbol{K}}_d = \boldsymbol{K}_d$. However, this assumption will be relaxed in Section 6.5, where the effects of possible mismatches are numerically tested.

2.2 Second learning phase

Goal of the second learning phase is to allow the accessing radio to perform the optimized shaping of the (deterministic) pilot streams $\{ \tilde{x}_i(n), \ T_{L1}+1 \le n \le T_{L1}+T_{L2} , \ 1 \le i \le t \}$ to be used to estimate the (a priori unknown) $(r \times t)$ uplink gains h_{ji}. Thus, the (sampled) signal $\tilde{y}_i(n)$ measured at the output of the *j*-th receive antenna of the AP node during this second learning phase may be modelled as (Paulraj et al., 2003)

$$\tilde{y}_j(n) = \frac{1}{\sqrt{t}} \sum_{i=1}^{t} h_{ji} \tilde{x}_i(n) + \tilde{d}_j(n), \quad T_{L1} + 1 \le n \le T_{L1} + T_{L2}, \tag{4}$$

where the overall disturbance $\tilde{d}_j(n)$ accounts for both MAI and thermal noise and exhibits the same statistics previously reported in (2). Therefore, the $(T_{L2} \times r)$ observations in (4) may be recast into the $(T_{L2} \times r)$ matrix $\tilde{Y} \triangleq [\underline{\tilde{y}}_1 \ldots \underline{\tilde{y}}_r]$ given by

$$\tilde{Y} = \frac{1}{\sqrt{t}} \tilde{X} H + \tilde{D}, \tag{5}$$

where $\tilde{X} \triangleq [\underline{\tilde{x}}_1 \ldots \underline{\tilde{x}}_t]$ is the $(T_{L2} \times t)$ matrix of the radiated pilot symbols in (4), $H \triangleq [\underline{h}_1 \ldots \underline{h}_r]$ is the $(t \times r)$ matrix composed by the uplink gains $\{h_{ji}\}$ to be learned, and the $(T_{L2} \times r)$ matrix $\tilde{D} \triangleq [\underline{\tilde{d}}_1 \ldots \underline{\tilde{d}}_r]$ collects the MAI-plus-noise terms $\tilde{d}_j(n)$ in (4).

Afterwards, observations $\tilde{Y}$ in (5) are employed by the AP to 'learn' H via the computation of the corresponding minimum mean squared error (MMSE) matrix estimate $\hat{H} \triangleq E\{H | \tilde{Y}\}$. Thus, at step $n = T_{L1} + T_{L2}$ (i.e., at the end of the second learning phase), this matrix estimate $\hat{H}$ is communicated back to the accessing radio via the service downlink of Fig. 1(b). A detailed analysis of the structure and performance of the estimator computing the set $\hat{h}_{ji} \triangleq E\{h_{ji} | \tilde{Y}\}$ of the $(r \times t)$ MMSE estimates of the MIMO uplink H in (5) may be found in (Baccarelli et al., 2007; Biagi, 2006). For our purposes, it suffices to stress that the pilot matrix $\tilde{X}$ minimizing the total average squared estimation error:

$$\sigma_{tot}^2 \triangleq \sum_{j=1}^{r} \sum_{i=1}^{t} E\{\|\hat{h}_{ji} - h_{ji}\|^2\} \equiv \sum_{j=1}^{r} \sum_{i=1}^{t} E\{\|\varepsilon_{ji}\|^2\}, \tag{5.1}$$

must meet the following relationship (Baccarelli & Biagi, 2003; Baccarelli et al., 2007):

$$K_d^{-1} \otimes \tilde{X}^\dagger \tilde{X} \equiv a I_{rt}, \tag{6}$$

where

$$a \triangleq \frac{T_{L2} \tilde{P}}{r} Tra\{K_d^{-1}\}, \tag{7}$$

and $\tilde{P}$ *(Watt)* in (7) is the average power available at Tx radio for the transmission of the pilot signals in (4). Furthermore, it may be also proved (Baccarelli et al., 2007) that, when $\tilde{X}$ meets the condition in (6), the resulting MMSE estimation errors $\{\varepsilon_{ji} \triangleq \hat{h}_{ji} - h_{ji}\}$ are mutually uncorrelated, zero-mean, and equidistributed proper complex Gaussian r.v.s, with variance

$$\sigma_\varepsilon^2 \triangleq E\{\| \varepsilon_{ji} \|^2\} \equiv (1 + a / t)^{-1}. \tag{8}$$

2.3 Payload phase

On the basis of the available K_d and $\hat{H}$ matrices and actual packet to be transmitted, the Tx radio of Fig. 1(b) suitable shapes the signal streams $\{\phi_i(n) \in \mathbb{C}^1,\ T_{L1} + T_{L2} + 1 \le n \le T,\ 1 \le i \le t\}$ to be radiated during the payload phase. Hence, the corresponding (sampled) signals $\{y_j(n) \in \mathbb{C}^1,\ T_{L1} + T_{L2} + 1 \le n \le T,\ 1 \le j \le r\}$ measured at the outputs of the receive antennas at the AP side may be modelled (Paulraj et al., 2003; Baccarelli et al., 2007) as

$$y_j(n) = \sqrt{\frac{\beta}{t}} \sum_{i=1}^{t} h_{ji} \phi_i(n) + d_j(n), \qquad T_{L1} + T_{L2} + 1 \le n \le T, \tag{9}$$

where the sequence $d_j(n) \triangleq v_j(n) + w_j(n),\ 1 \le j \le r$, accounts for the overall disturbances (e.g., MAI plus thermal noise) experienced during the payload phase, $\beta \triangleq D^{-z}$ takes care of the path loss over a distance of D meters and z is the corresponding path-loss exponent. Therefore, after assuming that the transmitted streams meet the (usual) average power constraint

$$\frac{1}{t} \sum_{i=1}^{t} E\{\| \phi_i(n) \|^2\} \le P , \tag{10}$$

the resulting SINR γ_j measured at the output of the j-th receive antenna at the AP side during the payload phase equates (equations (9), (10))

$$\gamma_j = aP/(\mathcal{N}_o + c_{jj}), \quad 1 \le j \le r, \tag{11}$$

where c_{jj} is the (j,j)-th entry of the MAI covariance matrix K_d in (2). Moreover, from (9) we also deduce that the $(r \times 1)$ column vector $\underline{y}(n) \triangleq [y_1(n) \dots y_r(n)]^T$ collecting the outputs of the r receive antennas over the n-th payload slot is linked to the $(t \times 1)$ column vector $\underline{\phi}(n) \triangleq [\phi_1(n) \dots \phi_t(n)]^T$ of the corresponding signals radiated by the accessing radio as in

$$\underline{y}(n) = \sqrt{\frac{\beta}{t}} H^T \underline{\phi}(n) + \underline{d}(n), \quad T_{L1} + T_{L2} + 1 \le n \le T, \tag{12}$$

where $\{\underline{d}(n) \triangleq [d_1(n) \dots d_r(n)]^T,\ T_{L1} + T_{L2} + 1 \le n \le T\}$ is the temporally-white spatially-colored Gaussian sequence of disturbances with spatial covariance matrix still given by K_d. Furthermore, directly from (10) it follows that the $(t \times t)$ spatial covariance matrix

$R_\phi \triangleq E\{\underline{\phi}(n)\,\underline{\phi}(n)^\dagger\}$ of the t-dimensional signal vector radiated during each by the Tx radio of Fig. 1(b) must meet the following power constraint:

$$Tra[R_\phi] \equiv E\{\underline{\phi}(n)^\dagger\,\underline{\phi}(n)\} \le tP. \tag{13}$$

Finally, after stacking the T_{pay} observed vectors in (12) into the corresponding $(T_{pay}r \times 1)$ block vector $\vec{y} \triangleq [\underline{y}^T(T_{L1} + T_{L2} + 1)\ldots\underline{y}^T(T)]^T$, we may compact the T_{pay} relationships (12) into the following one:

$$\vec{y} = \sqrt{\frac{\beta}{t}}[I_{T_{pay}} \otimes H]^T\,\vec{\phi} + \vec{d}, \tag{14}$$

where the (block) covariance matrix of the corresponding disturbance (block) vector in (14) $\vec{d} \triangleq [\underline{d}^T(T_{L1} + T_{L2} + 1)\ldots\underline{d}^T(T)]^T$ is given by

$$E\{\vec{d}(\vec{d})^\dagger\} = I_{T_{pay}} \otimes K_d, \tag{15}$$

while the average Euclidean squared norm of the block vector $\vec{\phi} \triangleq [\underline{\phi}^T(T_{L1} + T_{L2} + 1)\ldots\underline{\phi}^T(T)]^T$ of the transmitted random signals is constrained as (equation (10))

$$E\{\vec{\phi}^\dagger\vec{\phi}\} \le T_{pay}\,tP. \tag{16}$$

3. Self-reconfiguration of the cognitive MIMO-radios

We pass now to introduce a suitable performance metric that allows the accessing radio of Fig. 1(b) to both 'learn' its current achieved performance and, possibly, self-reconfigure its access strategy for improving it. Specifically, the performance metric we consider is the Shannon capacity of the MIMO uplink, formally defined as

$$\mathcal{C}(\hat{H}) \triangleq \sup_{\vec{\phi}:E\{\vec{\phi}^\dagger\vec{\phi}\}\le tT_{pay}P} \frac{1}{T_{pay}}I(\vec{y};\vec{\phi}\,|\,\hat{H}), \quad (nat/slot) \tag{17}$$

where $I(\vec{y};\vec{\phi}\,|\,\hat{H})$ is the mutual information (Gallagher, 1968) conveyed by the MIMO uplink when $\hat{H}$ is the (possibly imperfect) estimate of the uplink gains available at the transmit radio. Unfortunately, barring the limit cases of perfect channel state information (where $\hat{H} = H$) and no channel state information ($\hat{H} = 0$), the probability density function (pdf) of the input signals $\vec{\phi}$ achieving the supremum in (17) is currently unknown (Baccarelli & Biagi, 2004; Paulraj et al., 2003; Baccarelli 2007). However, it is known that Gaussian distributed input signals reach the supremum in (17) both when $\hat{H} = H$, as well as for imperfect channels estimates (i.e., $\hat{H} \ne H$) when the length T_{pay} of the payload phase

(largely) exceeds the number of transmit antennas. Therefore, justified by the above considerations, in the sequel, we assume that the T_{pay} components $\{ \underline{\phi}(n),\ T_{L1} + T_{L2} + 1 \le n \le T \}$ in (12) of the overall signal vector $\vec{\underline{\phi}}$ in (14) are uncorrelated proper complex Gaussian vectors, with correlation matrix R_ϕ meeting (13). The corresponding throughput

$$\mathcal{R}(\hat{H}) \triangleq \sup_{Tra[\,R_\phi\,] \le tP} \frac{1}{T_{pay}} I(\vec{\underline{y}}; \vec{\underline{\phi}} \mid \hat{H}), \quad (nat/slot) \tag{18}$$

conveyed by MIMO uplink is the performance metric considered by the accessing cognitive radio Tx. About the analytic evaluation of $\mathcal{R}(\hat{H})$ in (18), in (Baccarelli & Biagi, 2003; Biagi, 2006; Baccarelli et al., 2007) the following result is provided.

Proposition 1

Under the above reported assumptions about the considered access system, the $\hat{H}$-depending information- throughput $\mathcal{R}(\hat{H})$ in (18) may be evaluated in closed form as in

$$I(\vec{\underline{y}}; \vec{\underline{\phi}} \mid \hat{H}) = T_{pay} lg\ det\left(I_r + \frac{\beta}{t} K_d^{-1/2} \hat{H}^T R_\phi \hat{H}^* K_d^{-1/2} + \beta \sigma_\varepsilon^2 P K_d^{-1} \right)$$
$$- lg\ det\left(I_{rt} + \frac{\beta \sigma_\varepsilon^2 T_{pay}}{t} \left(K_d^{-1} \right)^* \otimes R_\phi \right), \tag{19}$$

when at least one of the following conditions is met:
1. both T_{pay} and t are large;

2. $\hat{H}$ approaches H;
3. all SINRs γ_j in (11) vanish.   □

3.1 Optimized cognitive access policy

In order to compute the supremum in (18), we must proceed to carry out the power-constrained maximization of $\mathcal{R}(\hat{H})$. For this purpose, let us indicate with

$$K_d = U_d \Lambda_d U_d^\dagger, \tag{20}$$

the singular value decomposition (SVD) of the MAI spatial covariance matrix K_d, where

$$\Lambda_d \triangleq diag\{ \mu_1, \dots, \mu_r \}, \tag{21}$$

is the corresponding $(r \times r)$ diagonal matrix of the magnitude-ordered singular values of K_d. Thus, after introducing the $(t \times r)$ dummy matrix

$$A \triangleq \hat{H}^* K_d^{-1/2} U_d, \tag{22}$$

accounting for the combined effects of the imperfect channel estimate $\hat{H}$ and the spatial MAI K_d, let us denote by

$$A = U_A D_A V_A^\dagger, \tag{23}$$

the corresponding SVD, where U_A and V_A are unitary matrices, while

$$D_A = \begin{pmatrix} diag\{k_1,\ldots,k_s\} & 0_{s\times r-s} \\ 0_{t-s\times s} & 0_{t-s\times r-s} \end{pmatrix}, \tag{24}$$

is the $(t \times r)$ diagonal matrix collecting the $s \triangleq min\{r,t\}$ magnitude-ordered singular values $k_1 \geq k_2 \geq \ldots \geq k_s > 0$ of the matrix A. Finally, for future convenience, let us also introduce the following dummy positions:

$$\alpha_m \triangleq \frac{\mu_m k_m^2}{t(\mu_m + P\sigma_\varepsilon^2)}, \quad 1 \leq m \leq s; \quad \lambda_l \triangleq \frac{\sigma_\varepsilon^2 T_{pay}}{t\mu_l}, \quad 1 \leq l \leq r. \tag{25}$$

In this way, it can be proved (Baccarelli & Biagi, 2003; Biagi, 2006; Baccarelli et al., 2007) that an application of the Kuhn-Tucker conditions (Gallagher, 1968) leads to the following optimized access strategy for the cognitive Tx radio.

Proposition 2 (Optimized access strategy)
Let the assumptions reported in *Proposition 1* be fulfilled. Therefore, for $m=s+1,\ldots,t$ the powers $\{P^*(m)\}$ achieving the supremum in (18) vanish, while for $m=1,\ldots,s$ they equate

$$P^*(m) = 0, \quad \text{when} \quad k_m^2 \leq \left(1 + \frac{\sigma_\varepsilon^2 P}{\mu_m}\right)\left(\frac{t}{\rho} + \sigma_\varepsilon^2 Tra\{K_d^{-1}\}\right), \tag{26}$$

$$P^*(m) = \frac{1}{2\lambda_{min}}\left\{\lambda_{min}L - 1 + \sqrt{(\lambda_{min}L)^2 + 4\lambda_{min}\left(\rho - \frac{1}{\alpha_m} - \frac{r\rho\lambda_m}{\alpha_m T_{pay}}\right)}\right\},$$

$$\text{when} \quad k_m^2 > \left(1 + \frac{\sigma_\varepsilon^2 P}{\mu_m}\right)\left(\frac{t}{\rho} + \sigma_\varepsilon^2 Tra\{K_d^{-1}\}\right) \tag{27}$$

where $\lambda_{min} \triangleq min\{\lambda_l, l=1,\ldots,r\}$ and $L \triangleq (1 - (r/T_{pay}))\rho - (1/\alpha_m)$. Furthermore, the non-negative scalar parameter ρ in (26), (27) is set to meet the following power-constraint:

$$\sum_{m \in \mathcal{I}(\rho)} P^*(m) \equiv P, \tag{28}$$

where

$$\mathcal{I}(\rho) \triangleq \left\{ m = 1, \ldots, s : k_m^2 > \left(1 + \frac{\sigma_\varepsilon^2 P}{\mu_m} \right) \left(\frac{t}{\rho} + \sigma_\varepsilon^2 Tra\left\{ K_d^{-1} \right\} \right) \right\},$$ (29)

is the (ρ-dependent) set of m-indexes fulfilling the inequality in (27). Finally, the corresponding optimized spatial correlation matrix $R_\phi(opt)$ of the radiated signals is aligned along the right eigenvectors of the matrix $\mathbf{A}$ in (22) as in

$$R_\phi(opt) = U_A \, diag\left\{ P^*(1), \ldots, P^*(s), \underline{0}_{t-s} \right\} U_A^\dagger,$$ (30)

so that for the resulting conveyed throughput $\mathcal{R}(\hat{H})$ we have

$$\mathcal{R}(\hat{H}) = \sum_{m=1}^{r} lg\left(1 + \frac{\sigma_\varepsilon^2 P^*(m)}{\mu_m} \right) + \sum_{m=1}^{s} \left(lg\left(1 + \alpha_m P^*(m) \right) \right) \; (nat/slot). \qquad \square$$ (30.1)

3.2 MAI-learning and MAI-mitigation at the AP

Before proceeding to detect the transmitted packet, the AP may attempt to estimate the MAI component $\underline{d}(n)$ affecting the received signal $\underline{y}(n)$ in (12), so as to subtract the computed MAI estimate $\underline{\hat{d}}(n)$ from the received signal $\underline{y}(n)$. Since both K_d and $\hat{H}$ are available at the AP, this last is also aware of $R_\phi(opt)$ (equation (30)), so that it may exploit both $\hat{H}$ and $R_\phi(opt)$ to improve the MAI estimate's accuracy. Specifically, being $\phi(n)$ and $\underline{d}(n)$ Gaussian distributed, the MMSE estimate $\underline{\hat{d}}(n) \triangleq E\{\underline{d}(n) | \underline{y}(n)\}$ of the MAI is linear with respect to the observation $\underline{y}(n)$ (Poor, 1994), so that we may write

$$\underline{\hat{d}}(n) = F\underline{y}(n).$$ (31)

In turns, the $(r \times r)$ matrix $\mathbf{F}$ in (31) may be computed via an application of the (usual) Orthogonal Principle (Poor, 1994), that allows us to write

$$E\{\left(\hat{d}(n) - d(n) \right)\left(y(n) \right)^\dagger | \hat{H}, R_\phi(opt)\} = \mathbf{0}_{r \times r},$$ (32)

where the conditioning in (32) accounts for the availability of $\hat{H}$ and $R_\phi(opt)$ at the AP side (e.g., the AP is cognitive). Hence, by developing the expectations in (32), after some (tedious) algebra, we arrive at the following final expression for the matrix $\mathbf{F}$ in (31):

$$F = \left\{ \frac{\beta}{t} \left[\hat{H}^T R_\phi(opt) \hat{H}^* + Pt\sigma_\varepsilon^2 \, I_r \right] K_d^{-1} + I_r \right\}^{-1}.$$ (33)

The MSE performance of the estimator in (31) is dictated by the corresponding error covariance matrix $K_{d\varepsilon} \triangleq E\{\left(\hat{\underline{d}}(n) - \underline{d}(n)\right)\hat{\underline{d}}(n) - \underline{d}(n)^{\dagger} \mid \hat{H}, R_{\phi}(opt)\}$ that equates (equations (31), (33))

$$K_{d\varepsilon} = [I_r - F]K_d. \tag{34}$$

Roughly speaking, equation (34) measures the performance improvement arising from the exploitation of the cognitive capability of the AP node. In particular, equation (33) points out that F vanishes for vanishing K_d, while F approaches I_r for large K_d. As a consequence, on the basis of equation (34), we conclude that actual effectiveness of the MAI estimation/subtraction procedure implemented by the AP tends to improve in the presence of strong MAI.

4. Some Game Theory essentials

In order to model the dynamic behaviour of the uplink-state of the access system of Fig. 1(a), we resort to the formal tool of the Game Theory. We shortly recall that a non-cooperative and strategic game $G \triangleq \langle \mathbb{N}, \mathbb{A}, \{u_g\} \rangle$ has three components (MacKenzie & Wicker, 2001; Fudenberg & Tirole, 1991): a finite set $\mathbb{N} \triangleq \{1, 2, \ldots, n^*\}$ of players, a set $A_g, g \in \mathbb{N}$, of admissible actions for the g-th player and a set of utility functions. Specifically, after denoting as $\mathbb{A} \triangleq A_1 \times A_2 \times \ldots \times A_{n^*}$ the space of action profiles, let us indicate as $u_g : \mathbb{A} \rightarrow \mathbb{R}$ the utility function of the g-th player. Therefore, after indicating by $\underline{a} \in \mathbb{A}$ an action profile, by $a_g \in A_g$ the action of g-th player in $\underline{a}$ and by $\underline{a}_{-g}$ the actions in $\underline{a}$ of the other $(n^* - 1)$ players, we can say that $u_g(\underline{a}) \equiv u_g(a_g, \underline{a}_{-g})$ maps[2] each action profile $\underline{a}$ into a real number. In particular, in a strategic non-cooperative game each player chooses a suitable action $a_g^{\bullet}$ from his action set A_g to maximize its own utility function, according to the following game rule:

$$a_g^{\bullet} \equiv \max_{a_g \in A_g} u_g(a_g, \underline{a}_{-g}). \tag{35}$$

Therefore, since there is no cooperation among the players, it is important to ensure the dynamic stability of the overall game. A concept which relates to this issue is the so-called Nash Equilibrium (NE). Simply stated, a NE is an action profile $\underline{a}^*$ at which no player may gain by unilaterally deviating from. So, a NE is a stable operating point of the game, because

---

[2] The notation $u_g(a_g, \underline{a}_{-g})$ emphasizes that the g-th player controls only own action a_g but his achieved utility depends also on the actions $\underline{a}_{-g}$ taken by all remaining players (MacKenzie & Wicker, 2001; Fudenberg & Tirole, 1991).

no player can obtain any profit from a change in his strategy. Formally stated, a NE is an action profile $\underline{a}^*$ such that for all $a_g \in A_g$ the following inequality is met:

$$u_g(a_g^*, \underline{a}_{-g}^*) \geq u_g(a_g, \underline{a}_{-g}^*), \quad \forall a_g \in A_g, \; \forall g \in \mathbb{N}. \tag{36}$$

5. The proposed cognitive access game

Let us focus now on the uplink in Fig. 1(a) of a WLAN composed by n^* mutually interfering multi-antenna accessing radios. The ultimate task of the g-th radio terminal is to maximize the information throughput $\mathcal{R}(g)$ $(nat/slot)$, sustained by the corresponding uplink $T_g \to AP$ via suitable power-allocation and shaping of the transmitted signals and interference cancellation at the AP. Since the signals radiated by g-th radio induce MAI at the AP side and we assume that the radios do not exchange information (the accessing terminals do not cooperate), we may model the interaction among the accessing cognitive radios as a non-cooperative strategic game. Specifically, in the considered access scenario, the players' set $\mathbb{N}$ is composed of n^* accessing radios, while the set of actions A_g available to the g-th player is the set of all covariance matrices $R_\phi^{(g)}$ meeting the power constraint (2), so we can pose

$$A_g \equiv \left\{ R_\phi^{(g)} : 0 \leq Tra\left[R_\phi^{(g)} \right] \leq t_g P^{(g)} \right\} \quad g = 1, \ldots n^*, \tag{37}$$

where $P^{(g)}$ $(Watt)$ is the power available at the g-th radio and t_g is the number of transmit antennas equipping the g-th radio. This means that the generic action a_g of T_g consists in the transmission of a Gaussian distributed payload sequence with covariance matrix $R_\phi^{(g)}$. Furthermore, the utility function $u_g(\cdot)$ for the g-th access link is the corresponding conveyed rate, so that we can write (equation (9))

$$
\begin{aligned}
u_g(\underline{a}) &\triangleq u_g\left(R_\phi^{(1)}, \ldots, R_\phi^{(g)}, \ldots R_\phi^{(n^*)} \right) \equiv \frac{1}{T_{Pay}} I\left(\underline{\vec{y}}^{(g)}; \underline{\vec{\phi}}^{(g)} \mid \hat{H}_g \right) \\
&\equiv \lg \det\left(I_r + \frac{\beta_g}{t_g}\left(K_d^{(g)} \right)^{-1/2} \hat{H}_g^T R_\phi^{(g)} \hat{H}_g^*\left(K_d^{(g)} \right)^{-1/2} + \beta_g \sigma_\varepsilon^2(g) P^{(g)}\left(K_d^{(g)} \right)^{-1} \right) \\
&\quad - \frac{1}{T_{Pay}} \lg \det\left(I_{rt_g} + \frac{\beta_g \sigma_\varepsilon^2(g) T_{pay}}{t_g}\left(\left(K_d^{(g)} \right)^{-1} \right)^* \otimes R_\phi^{(g)} \right),
\end{aligned} \tag{38}
$$

where the g-th MAI covariance matrix $K_d^{(g)}$ depends on the spatial covariance matrices $\left\{ R_\phi^{(i)}, i \neq g \right\}$ of the signals radiated by all other accessing radios (Fig. 1(b))[3]. About the rule of the game, each player (i.e., accessing cognitive radio T_g) chooses the action $R_\phi^{(g)\bullet}$ which maximize the throughput conveyed by its own access link, so we can write

$$R_\phi^{(g)\bullet} \equiv arg \max_{R_\phi^{(g)} \in A_g} \left\{ \frac{1}{T_{Pay}} I\left(\underline{\bar{y}}^{(g)}; \underline{\bar{\phi}}^{(g)} \mid \hat{H}_g \right) \right\}, \quad g = 1, \ldots, n^*. \tag{39}$$

5.1 Competitive optimal SDMA

We pass now to detail the algorithm for the competitively optimal access under 'contracted-QoS' and best-effort policies when the AP of Fig. 1(a) performs MAI estimation/subtraction. Before proceeding, we point out that the concept of 'contracted-QoS' we go to introduce relies on multiple QoS classes defined in terms of requested, or desired, access throughput. Therefore, the resulting access algorithm we report in Table 1 attempts to achieve the target throughput $T\mathcal{R}_A^{(Z)}$ desired by the g-th radio; and if this throughput is not achievable due to the MAI induced by the other accessing terminals, the algorithm tries to achieve the next lower QoS-class $T\mathcal{R}_A^{(Z-1)}$.

Furthermore, from the outset it also follows that the best-effort policy is an instance of the 'contracted-QoS' one, where the number of allowed QoS classes approach infinity. The algorithm for achieving the competitively maximal access throughput over the uplink is detailed in Table 1. It must be iteratively run by all accessing radios of Fig. 1(a). $T\mathcal{R}_A^{(Z)}$ (nat/slot) in Table 1 indicates the target throughput defining the Z-th QoS class.

0	Set the target throughput $T\mathcal{R}_A^{(Z)}$ of the Z-th QoS class
1	Initialize $R_\phi^{(g)}$ (new):= $R_\phi^{(g)}$ (old)=$[\ \mathbf{0}_{t_g \times t_g}\ ]$
2	fl(g)=1
3	$\mathcal{R}(g) = 0$
4	$a^{(g)} \triangleq (\tilde{P} T_{tr}/r_g) Tra\left[(K_d^{(g)})^{-1} \right]$
5	$\sigma_\varepsilon^2(g) \triangleq (1 + a^{(g)}/t_g)^{-1}$
6	Compute and sort the r eigenvalues of $K_d^{(g)}$
7	Compute the SVD of $\hat{H}_g^* K_d^{(g)-1} \hat{H}_g^T$
8	Sort the $s \triangleq min(r,t)$ eigenvalues $\left\{ k_1^{(g)2}, \ldots, k_s^{(g)2} \right\}$ of $K_d^{(g)}$
9	$\alpha_m^{(g)} \triangleq \mu_m^{(g)} k_m^{(g)2}/t_g \left(\mu_m^{(g)} + P^{(g)} \sigma_\varepsilon^2(g) \right)$
10	$\beta_l^{(g)} \triangleq \sigma_\varepsilon^2(g) T_{pay} / \mu_l^{(g)} t_g$

---

[3] In the sequel, we adopt the index g for indicating the system parameters of the g-th uplink $T_g \rightarrow AP$ (Fig. 1(a)).

11 $\qquad \mu_{min}^{(g)} \triangleq \min\limits_{1\le l\le r}\left\{\mu_m^{(g)}\right\};\ \beta_{max}^{(g)} \triangleq \dfrac{\sigma_\varepsilon^2(g)T_{pay}}{\mu_{min}^{(g)}t_g}$

12 $\qquad$ If $k_m^{(g)2}\ge\left(\mu_m^{(g)}+P^{(g)}\sigma_\varepsilon^2(g)\right)\dfrac{\sigma_\varepsilon^2(g)\sqrt{rT_{pay}}}{\mu_{min}^{(g)}\mu_m^{(g)}}$ for all m and fl(g)=1

{

13 $\qquad$ Set $\rho^{(g)}:=0$ and $\rho^{(g)}:=0\quad \mathcal{J}\left(\rho^{(g)}\right):=0$

14 $\qquad$ Set the step size Δ

15 $\qquad$ While $\sum\limits_{m\in\mathcal{J}\left(\rho^{(g)}\right)}p^{*(g)}(m)<P$ do

}

16 $\qquad$ Update $\rho^{(g)}=\rho^{(g)}+\Delta$

17 $\qquad$ Update the set $\mathcal{J}\left(\rho^{(g)}\right)$ via equation (29)

18 $\qquad$ Compute the powers and the covariance matrix via equation (26), (27), (30)

19 $\qquad$ Set $\Psi^{(g)}:=R_\phi^{(g)}(new)-R_\phi^{(g)}(old)$

20 $\qquad$ If $\left(\|\Psi\|_E^2\le 0.05\left\|R_{\phi^{(g)}}^{(g)}(old)\right\|_E^2\right)$

21 $\qquad$ then fl(g)= 0, else fl(g)=1;

22 $\qquad R_\phi^{(g)}(old):=R_\phi^{(g)}(new)$

}

23 $\qquad$ Evaluate $\mathcal{R}(g)$ via (30.1) for the g-th link

24 $\qquad$ If $\mathcal{R}(g)=T\mathcal{R}_A^{(Z)}$ stop; else

25 $\qquad$ If $\mathcal{R}(g)=T\mathcal{R}_A^{(Z)}$ reduce the radiate power $p^{(g)}$ and go to Step 1; else

26 $\qquad$ If $\mathcal{R}(g)>T\mathcal{R}_A^{(Z)}$ lower the target class to (Z-1) and go to Step 1.

}

Table 1. Pseudocode implementing the competitively optimal cognitive access on the g-th uplink under the 'contracted-QoS' policy

5.2 Asynchronous implementation of the access game and Nash Equilibria

After assuming that the access algorithm of Table 1 is iteratively run (in a non-cooperative and possibly asynchronous way) by all accessing radios of Fig. 1(a), let be $V_g\triangleq\left\{t_{g1},t_{g2},\dots\right\}$ the set of times at which g-th radio executes this algorithm. Thus, after indicating by $V\triangleq\left\{\tau_1,\tau_2,\dots\right\}$ the resulting overall set of update times $V_1\bigcup V_2\bigcup\dots\bigcup V_{n^*}$ sorted in the increasing order ($\tau_i\le\tau_{i+1}$), the asynchronous and distributed implementation of the considered access game works according to the pseudocode of Table 2.

Hence, key-questions about the access game of Table 2 regard the existence, uniqueness and achievability of the corresponding NE. The following *Proposition 3* (Biagi, 2006) gives insight about these questions.

For all k such that $\tau_k \in V$

{

For all terminals $g \in \mathbb{N}$ such that $\tau_k \in V_g$

{

Evaluate the MAI matrix $K_d^{(g)}$

Run the algorithm of Table 1 so

to compute [$\underline{\bar{\phi}}^{(g)\bullet}$] and [$R_\phi^{(g)\bullet}$]

radiate the signal vector [$\underline{\bar{\phi}}^{(g)\bullet}$]

Table 2. Pseudocode implementing the proposed access game

Proposition 3
By referring to the access game of Table 2, let us assume that the following conditions are met:

$$k_m^{(g)\,2} > \left(1 + \frac{\sigma_\varepsilon(g)^2\,P^{(g)}}{\mu_m(g)} \right)\left(\frac{t_g}{\rho} + \sigma_\varepsilon(g)^2 Tra\left[K_d^{(g)-1} \right] \right) \tag{40}$$

$$r \geq t_g, \quad 1 \leq g \leq n* \tag{41}$$

$$T_{Pay} \gg T_g > 1 \ \ and\,/\,or \ \ \sigma_\varepsilon(g)^2 \to 0, \quad 1 \leq g \leq n*. \tag{42}$$

Thus, the NE of the access game of Table 2 exists, is unique and it may be reached by moving from any starting point.

6. Cognitive MIMO access performance

Since WLANs are characterized by topology-depending time-varying MAI, in principle, an effective SDMA protocol should be able to adaptively exploit (in a combined way) the above mentioned flexibility characteristics of the MIMO physical layer by operating in a scalable and asynchronous way. Currently, Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) is 'de-facto' the MAC protocol considered for WLANs. Interestingly enough, a simple extension of CSMA/CA for MIMO links can be designed to provide an $s = min\,(t,r)$ - fold improvement in throughput performance compared to a single-input single-output (SISO) network. We refer to this simple extension as CSMA/CA(s). Essentially, CSMA/CA(s) operates in the same fashion as the conventional CSMA/CA, except that all transmissions are performed using s independent parallel streams, so as to attain a spatial multiplexing gain. Such a protocol is still collision-free, and, when compared to default single-antenna CSMA/CA, it is able to achieve s times the throughput performance than the latter. Thus, while this s-fold improvement guaranteed by the collision-free CSMA/CA(s) is, indeed, quite appealing, the key question we proceed to tackle is: Is it possible for a more 'smart' SDMA scheme to attain a better throughput performance by working in a scalable and asynchronous way? We anticipate that the main conclusion we arrive at, is, indeed yes. Roughly speaking, the key-rationale behind this conclusion is that CSMA/CA(s) is still a

collision-free MAC scheme, so that it is not able to fully exploit the advantages arising from the above mentioned flexible MAI-mitigation capability of the MIMO physical layer.

To corroborate this claim, we simulated a Rayleigh-flat faded WLAN composed of $n^* = 2$ accessing radios located at a same distance of $D_1 = D_2 = 100 \ (mt)$ from the AP. The fading phenomena impairing the uplinks $T_1 \rightarrow AP$ and $T_2 \rightarrow AP$ are equidistributed and mutually independent. No obstacles are assumed between the accessing radios and the AP, so that MAI and signal average power levels measured only depend on the path-loss. All the reported numerical results refer to a path-loss exponent of 2.5, while the average received signal-to-noise ratio (SNR) is 20 dB. Furthermore, unless otherwise stated, the MMSE estimates $\left\{ \hat{h}_{ji}^{(g)}, \ g = 1,2 \right\}$ of the uplink gains available at the radios are assumed exact (i.e., error free).

6.1 The achievable throughput regions

In this operating conditions, the (two-dimensional) set of accessing rates sustainable by the AN may be (formally) described by resorting to the concept of achievable rate-region (Baccarelli et al., 2005). In a nutshell, given a statistical description of the uplinks $T_1 \rightarrow AP$, $T_2 \rightarrow AP$, and under an assigned set of constraints on the available radio resources (e.g., available power, bandwidth, noise level, and so on), the corresponding achievable rate region of the access network is the closure of all 2-ple ($\mathcal{R}_1$, $\mathcal{R}_2$) of the access rates that may be simultaneously sustained by the network.

Unfortunately, apart from some partial contributions, no closed-form analytical formulas are available yet for the evaluation of the rate-region of MAI-impaired access networks (Baccarelli et al., 2005). Forced by this consideration, we resort to numerical tests and Fig. 2 reports the rate-regions we have numerically obtained for some different access strategies. The $n^* = 2$ radios of the simulated access network of Fig. 2 are equipped with $t_1 = t_2 = 4$ transmit antennas, while $r = 4$ receive antennas are available at the AP. The innermost $\mathcal{A}$-labelled region of Fig. 2 reports the sustainable access rates when the power available at the radios is evenly split over the transmit antennas (i.e. $R_\phi^{(1)} = R_\phi^{(2)} = \dfrac{P}{4} I_4$ in equation (39)) and,

in addition, no channel estimate are available for MAI mitigation ($\hat{H} = 0$ in equation (33), so that the AP node is no cognitive). In this case, when both the radios attempt to access to the AP, the resulting maximal throughput is quite low and it is limited up to $\mathcal{R}_1 = \mathcal{R}_2 = 11 \ (bit/slot)$. However, when the AP performs reliable MAI estimation and subtraction (i.e. $\hat{H} \equiv H$ in equation (33) and thus the AP node is perfectly cognitive), we obtain the (larger) rate-region $\mathcal{B}$ of Fig. 2, that covers the above $\mathcal{A}$-region and allows radios to simultaneously access at a maximal throughput of $\mathcal{R}_1 = \mathcal{R}_2 = 16.5 \ (bit/slot)$.
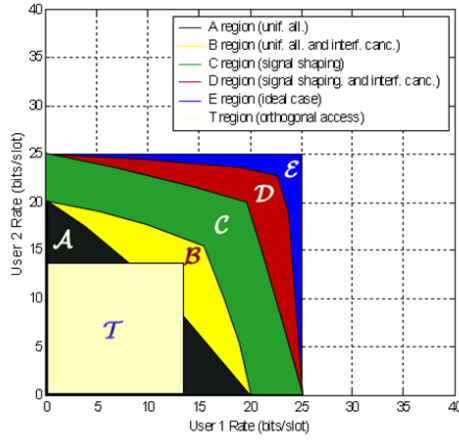
Fig. 2. Rate regions for some different accessing policies and systems setting

When optimized power-allocation and signal-shaping are also performed at the radios (that is, $R_\phi^{(g)}$ are computed as in equation (39)) but no reliable MAI-mitigation is performed by the AP, we obtain the rate-region $\mathcal{C}$ of Fig. 2 that allows radios to simultaneously access at a (maximal) throughput of $\mathcal{R}_1 = \mathcal{R}_2 = 21 \; (bit/slot)$. A comparison of $\mathcal{D}$ and $\mathcal{E}$ regions of Fig. 2 gives insight on the ultimate effectiveness of the competitive SDMA game played by the cognitive radios and the MAI mitigation performed by the cognitive AP. Specifically, the $\mathcal{D}$ region of Fig. 2 collects the 2-ple ($\mathcal{R}_1$, $\mathcal{R}_2$) of achievable access rates when both the competitive SDMA game of Table 2 is carried out by the radios and a reliable MAI estimation/subtraction is performed by the AP. In this case, the maximal access throughput sustained by the AN approaches $\mathcal{R}_1 = \mathcal{R}_2 = 23.5 \; (bit/slot)$. This value differs, indeed, less than 6% from the ultimate one of $\mathcal{R}_1 = \mathcal{R}_2 = 25 \; (bit/slot)$ bounding the corresponding $\mathcal{E}$ region. This last refers to the ideal case when MAI-cancellation is perfect (i.e. $\underline{\hat{d}}(n) \equiv \underline{d}(n)$ in equation (31)) and both radios are allowed to transmit over the whole working-time of the system.

About the degrading effects induced by possibly imperfect MIMO uplink estimates, these last have been investigated in depth, for example, in (Baccarelli & Biagi, 2003; Baccarelli & Biagi, 2004). So, in this context we limit to remark that the $\mathcal{A}$-region of Fig. 2 represents the sustainable access rates when both radios and AP are fully unaware about actual values of the MIMO uplink gains (in other words, they are not cognitive and $\hat{H}^{(1)} = \hat{H}^{(2)} = 0$), while the corresponding $\mathcal{D}$-region captures the achievable access rates when both radios and AP are equipped with perfect MIMO uplink estimates (i.e. $\hat{H}^{(1)} = H^{(1)}$ and $\hat{H}^{(2)} = H^{(2)}$). A direct comparison of these regions leads to the conclusion that the penalty in the access throughput arising from full unawareness about MIMO uplink-gains may be, indeed, no negligible and of the order of 5–6 (bit/slot) per accessing radio. Obviously, this conclusion is somewhat mitigated by considering that long throughput-consuming learning phases may be required in order to acquire reliable uplink estimates (Section 3.2). The interested reader

may refer to the works in (Baccarelli & Biagi, 2003; Baccarelli & Biagi, 2004; Baccarelli et al., 2007) and references therein for a detailed description of the optimal estimation accuracy - vs. - throughput trade off.

6.2 Cognitive SDMA Game-vs-CSMA/CA(*s*): a throughput comparison

The effectiveness of the proposed competitive SDMA policy for cognitive ANs may be also appreciated by comparing the above mentioned $\mathcal{D}$-region against the one of Fig. 2. Specifically, this $\mathcal{T}$-region of Fig. 2 reports the 2-ple ($\mathcal{R}_1$, $\mathcal{R}_2$) of achievable access rates when the access policy followed by the (no cognitive) multi-antenna radios is the CSMA/CA(*s*). In this regard, we explicitly stress that, to guarantee both collision-free and fair access, in the carried out numerical tests the CSMA/CA(*s*) policy we implemented schedules a single uplink at a time, and transmits over the scheduled uplink at the maximum allowed power *P* and over a $1/n^* = 1/2$-fraction of the overall system working time. Furthermore, the throughput loss due to the (possible) exchange of RTS/CTS packets is not accounted for. Thus, being the implemented CSMA/CA(*s*) scheme collision-free (that is, fully MAI-free), the corresponding $\mathcal{T}$-region of Fig. 2 is the largest attainable by the CSMA/CA(*s*) policy. Now, an examination of Fig. 2 leads to the conclusion that the access region achieved by the CSMA/CA(*s*) policy is strictly included in the $\mathcal{D}$-region obtained by the proposed competitive cognitive-based SDMA policy, at least in the considered MAI-limited access scenario. Since the implemented CSMA/CA(*s*) strategy is, indeed, an example of multi-antenna orthogonal access policy, we conclude that the rate-region of any multi-antenna orthogonal access strategy (such as TDMA, FDMA, CDMA) overlaps the $\mathcal{T}$-region of Fig. 2. This supports the superiority of the proposed competitively optimal cognitive access strategy over the orthogonal no-cognitive ones, at least when the spatial-dimension offered by the AN may be efficiently exploited to perform both signal-beam forming at the accessing (cognitive) radios and MAI-mitigation at the (cognitive) AP.

6.3 Cognitive access capacity

The following Figs 3, 4 and 5 give insight into the performance of the proposed cognitive SDMA game with MAI-mitigation in terms of access capacity (i.e., in terms of maximum number of radios capable to simultaneously access the network at a common target throughput). The plots of Fig. 3 report the (numerically evaluated) system capacity at access throughput of 10, 15, 20 (bit/slot) when all radios are equipped with $t = 4$ transmit antennas, while the AP performs MAI-mitigation via a number of receive antennas ranging from one to eight. An examination of these plots leads to two main conclusions. The first (quite obvious) one is that the access capacity decreases for increasing values of the target access throughput (the target QoS-level) demanded by the radios. The second (less trivial) one is that, at any target QoS level, the slope of the capacity curves of Fig. 3 grows for increasing *r* up to $r = t = 4$, and then it starves for $r > t = 4$. Due to the iterative nature of the proposed access strategy of Table 2, we are not able to give formal (e.g., analytical) support to this behaviour. However, roughly speaking, it may be understood by noting that, under mild assumptions, the Shannon capacity (in bit/slot) of a point-to-point flat-faded MIMO channel scales as the $min(r, t)$ (Paulraj et al., 2003), so that no slope increment in the capacity curves is expected when $r > t$. However, by increasing *r* beyond *t*, MAI mitigation

capability of the AP still grows, so that the capacity curves of Fig. 3 increases (at a nearly constant rate) for growing values of r.
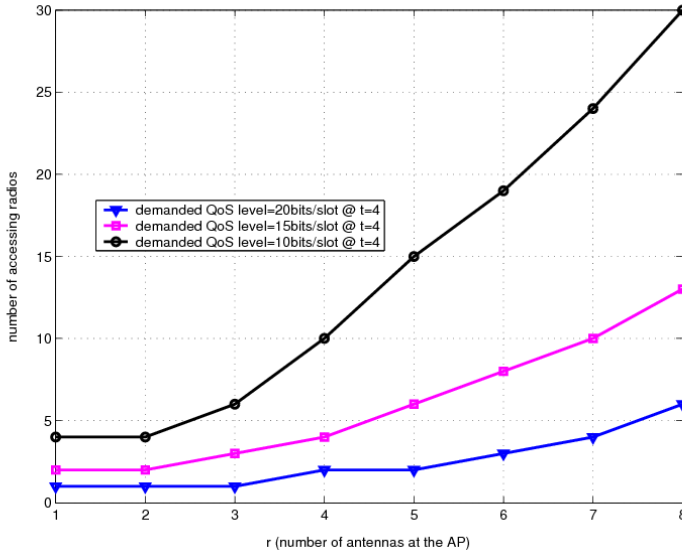

Fig. 3. Number of radios accessing the net under the guaranteed QoS policy for various r values

Similar conclusions may be drawn after an examination of the plots of Fig. 4, that capture the effects induced by the number t of transmit antennas equipping each accessing radio. Interestingly, a comparison of the curves of Fig. 3 and 4 shows that, at any target access throughput, the capacity of the access system with $t = m$ transmit antennas and $r = n$ receive ones outperforms the corresponding capacity of the 'dual' access system with $r = m$ and $t = n$. This behaviour supports the conclusion that, at least in the considered operating scenarios, the increment in the access capacity arising from more accurate transmit beam forming outperforms the corresponding capacity improvement due to more reliable MAI-mitigation. Finally, Fig. 5 reports the system capacity under the above mentioned 'contracted-QoS' access policy (Table 1), when the radios accept access throughput within 10% of the desired one. A comparison of the plots of Fig. 5 corroborates the conclusion that the 'QoS-contracted' access policy may be valuable in terms of system capacity, especially when the MAI-mitigation capability of the AP is limited (e.g., when the number r of the receive antennas is less than that of the transmit ones).
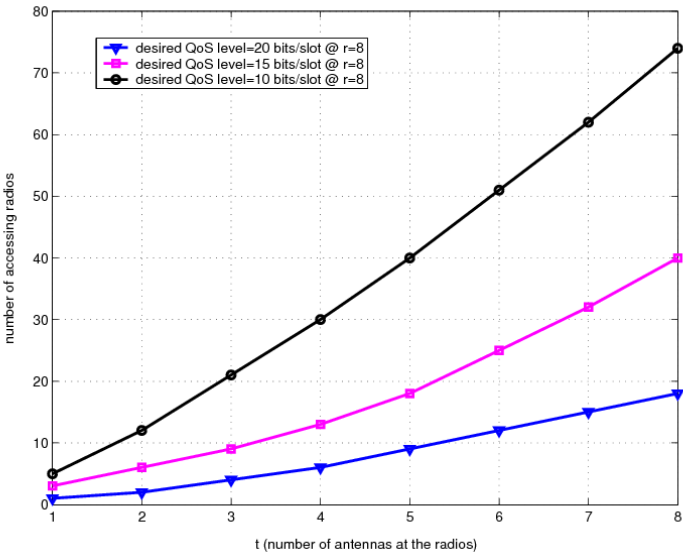
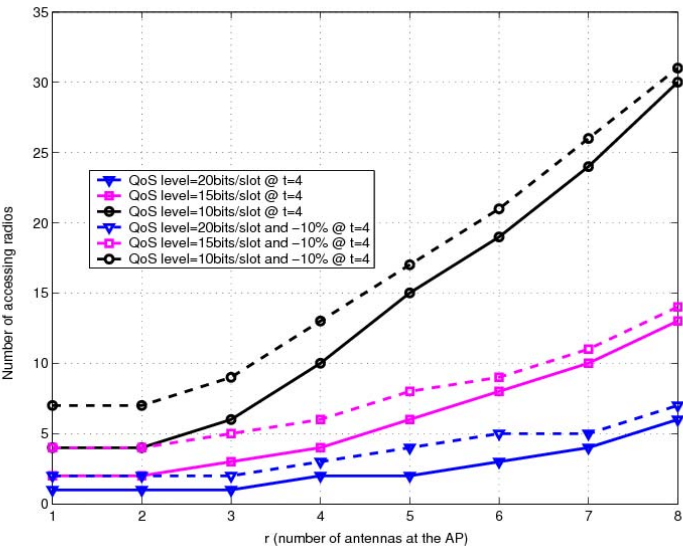Fig. 4. Number of radios accessing the net under the guaranteed QoS policy for various *t* values

Fig. 5. Number of radios accessing the net under the 'contracted-QoS' policy for various *r* values

6.4 Self-convergence of the cognitive SDMA Game toward the nearest sustainable operating point

In actual application scenarios, the accessing radios are not aware in advance about the sustainable access regions of Fig. 2, neither these last may be analytically evaluated in closed-form (Baccarelli et al., 2005). Therefore, a key-question concerns the self-convergence of the operating point of the SDMA cognitive game of Table 2, when the initial access throughput $\left(\mathcal{R}_1^{(0)}, \mathcal{R}_2^{(0)}\right)$ demanded by the radios falls out of the corresponding access region of Fig. 6. It can be proved (Biagi, 2006, Appendix 3) that, under the best-effort policy, the operating point of the SDMA game of Table 2 is able to self-move from $\left(\mathcal{R}_1^{(0)}, \mathcal{R}_2^{(0)}\right)$ and to converge to the boundary point of the underlying access region at the minimum Euclidean distance from that starting point (see the dotted arrow of Fig. 6). Likewise, under the 'contracted-QoS' policy, the operating point self-moves from $\left(\mathcal{R}_1^{(0)}, \mathcal{R}_2^{(0)}\right)$ to the point of the QoS grid at the minimum Euclidean distance (see the dashed grid of Fig. 6).
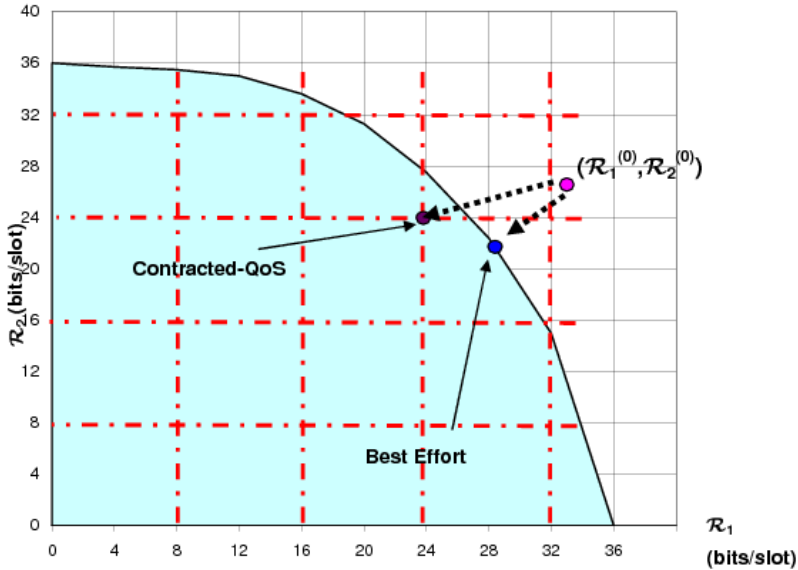
Fig. 6. Access region of the proposed cognitive SDMA game of Table 2 with MAI-mitigation ($t_1=t_2=4$ and $r=4$)

6.5 Mismatches in the estimation of the MAI covariance matrix

To test the sensitivity of the proposed cognitive SDMA game on errors possibly affecting the estimated MAI covariance matrices $\hat{K}_d^{(g)}$, $g=1,...n*$, we have perturbed actual matrices $K_d^{(g)}$, $g=1,...n*$, via randomly-generated $(r \times r)$ matrices $G^{(g)}$, $g=1,...n*$, which are composed by zero-mean proper complex unit-variance uncorrelated Gaussian elements. In doing so, we generated

$$\hat{K}_d^{(g)} = K_d^{(g)} + \sqrt{\frac{\left\|K_d^{(g)}\right\|_E^2}{r^2}} \sqrt{\delta} G^{(g)}, \quad g = 1,\ldots n^*, \tag{43}$$

where $\delta \triangleq E\left\{\left\|K_d^{(g)} - \hat{K}_d^{(g)}\right\|_E^2\right\} / \left\|K_d^{(g)}\right\|_E^2$ was set according to the desired average squared estimation errors affecting $\hat{K}_d^{(g)}$. Thus, after replacing the actual covariance matrices $K_d^{(g)}$ with the corresponding perturbed ones $\hat{K}_d^{(g)}$, we have played once again the SDMA Game of Table 2 for the same access scenario previously described in Section 6.1. The resulting access region (labelled as $\mathcal{D}^*$) is reported in Fig. 7, together with the corresponding $\mathcal{D}$-region previously drawn in Fig. 2 for the case of perfect MAI covariance-matrix estimates. An examination of the regions of Fig. 7 points out that the resulting access throughput-loss is limited up to 4%, even for δ values as high as 0.05.
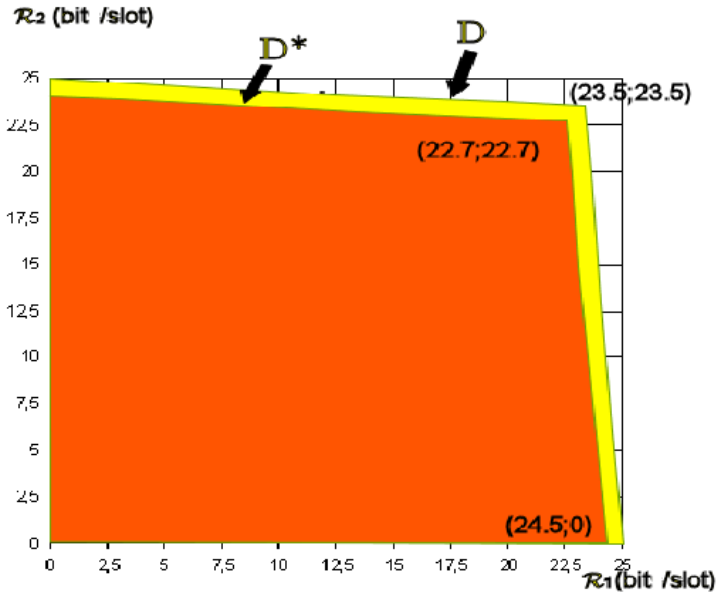
Fig. 7. Access regions for perfectly estimated MAI-covariance matrices ($\mathcal{D}$) and with estimation errors ($\mathcal{D}^*$)

7. Conclusion

In this work, we attempted to give a first insight into the possible capacity improvement in the local radio access arising from the exploitation of the cognitive MIMO radio paradigm. The game-inspired approach we have proposed exploits the cognitive capabilities of the overall AN to dynamically perform both adaptive spatial-beamforming at the radios and MAI-mitigation at the AP. The reported numerical results support the conclusion that the

cognitive SDMA policy is able to outperform the more conventional no cognitive strategies based on orthogonal access, requiring neither peer-to-peer cooperation among the accessing radios nor radio synchronization.

8. References

Altman, E.; Basar, T. & Srikant, R. (2002). Nash equilibria for combined flow control and routing in networks: asymptotic behaviour for a large number of users, *IEEE Trans.on Automatic Control*, Vol. 47, No. 6, June, pp.917–930.

Altman, E. & Altman, Z. (2003). S-modular games and power control in wireless networks, *IEEE Trans. on Automatic Control*, Vol. 48, No. 5, May, pp.839–842.

Baccarelli, E. & Biagi, M. (2003). Optimized power allocation and signal shaping for interference-limited multi-antenna 'ad-hoc networks', *Proc. of PWC 2003*, Sept., pp.138–152.

Baccarelli, E. & Biagi, M. (2004). Power-allocation policy and optimized design of multiple-antenna systems with imperfect channel estimation, *IEEE Trans. on Vehicular Technology*, Vol. 53, No.1, January, pp.136–145.

Baccarelli, E.; Biagi, M.; Pelizzoni, C.; Cordeschi, N. & Garzia, F. (2005). When does interference not reduce capacity in multi-antenna networks?, *Proc. of IEEE Signal Processing Advances in Wireless Communications*, pp.298–302.

Baccarelli, E.; Biagi, M.; Pelizzoni, C. & Cordeschi, N. (2007). Optimized power-allocation for multi-antenna systems impaired by multiple access interference and imperfect channel-estimation, *IEEE Trans. on Vehic. Techn.*, Vol. 56, No. 5, Sept., pp.3089–3015.

Biagi, M. (2006). Cross-layer design of multi-antenna 4G-WLAN, *PhD dissertation* available at http://infocom.uniroma1.it/~biagi.

Butler, J.M. & Webb, W.T. (2006). An implementation of spectrum usage rights for liberalization of the radio spectrum, *Journal of Communication and Networks*, Vol. 8, No. 2, June, pp.163–168.

Chow, C.T.; Shankar, S.; Kim, H. & Shin, K.G. (2007). What and how much to fain by spectrum agility?, *IEEE Journal of Sel. Areas in Comm.*, Vol. 25, No. 3, Apr., pp.576–588.

Cordiero, C.; Challapali, K.; Birru, D. & Shankar, S. (2005). IEEE 802.22: the first world wide standard for cognitive radio, *IEEE DySPAN Conf. Proceedings*.

DARPA (2001). Program available (vision RFC) at http://www.darpa.mil/ato/programs/XG/rfc vision.pdf.

Fattahi, A.; Fu, F.; Schaar, M.v.d. & Paganini, F. (2007). Mechanism-based resource allocation for multimedia transmission over spectrum agile wireless networks, Apr., *IEEE JSAC*, Vol. 25, No. 3, pp.601–612.

FCC Notice of Proposed Rulemaking and Order (2003). Facilitating opportunities for flexible, efficient and reliable spectrum use employing cognitive radio technologies, *ET Docket*, Dec., No. 03–108.

Fudenberg, D. & Tirole, J. (1991). *Game Theory*, MIT Press.

Gallagher, R.G. (1968). *Information Theory and Reliable Communication*, Wiley.

Glisic, S. (2006) .*Advanced Wireless Networks: 4G Technologies*, Wiley.

Goodman, D. & Mandayam, N. (2000). Power-control for wireless data, *IEEE Personal Communications*, April, pp.48–54.

Haykin, S. (2005). Cognitive radio: brain-empowered wireless communications, *IEEE Journal of Select. Areas in Comm.*, Feb., Vol. 23, No. 1, pp.201–220.

IEEE 802.22 (2005). working group available at website http://www.ieee802.org/22.

Kannan, R. & Iyengar, S.S. (2004). Game-theoretic models for reliable path-length and energy constrained routing with data aggregation in wireless sensor networks, *IEEE Journal on Sel. Areas in Comm.*, June, Vol. 22, No. 6, pp.1141–1150.

Lee, P.K. (2001). Joint frequency hopping and adaptive spectrum exploitation, *IEEE MILCOM 2001*, Vol. 1, pp.566–570.

MacKenzie, A.B. & Wicker, S.B. (2001). Game Theory in communications: motivation, explanation and application to power control, *GLOBECOM 2001*, pp.821–825.

Mangold, S.; Zhong, Z.; Challapali, K. & Chow, C.T. (2004). Spectrum agile radio: radio resource measurements for opportunistic spectrum usage, *Proc. of GLOBECOM '04*, Dec., Vol. 6, pp.3467–3471.

Mitola, J. (1999). Cognitive radio for flexible mobile multimedia communications, *Proc. of MoMuC'99*, Nov., pp.3–10.

Palomar, D.P.; Cioffi, J.M. & Lagunas, M.A. (2003). Uniform power-allocation in MIMO channels: a game-theoretic approach, *IEEE Trans. on Inform. Theory*, July, Vol. 49, No. 7, pp.1707–1727.

Paulraj, A.; Nabar, E. & Gore, D. (2003). *Introduction to Space-Time Wireless Communications*, Cambridge University Press.

Poor, H.V. (1994). An introduction to signal detection and estimation, *Springer Texts in Electrical Engineering*.

Saraydar, C.; Mandayanan, N.B. & Goodman, D.J. (2002). Efficient power control via pricing in wireless data networks, *IEEE Trans. on Comm.*, Feb., Vol. 50, No. 2, pp.291–303.

Shad, F.; Todd, T.D.; Kezys, V. & Litva, J. (2001). Dynamic slot allocation (DSA) in indoor SDMA/TDMA using a smart antenna base station, *IEEE/ACM Trans. on Networking*, Feb., Vol. 9, No. 1, pp.69–81.

Sung, C. & Wong, W. (2003). A non-cooperative power-control frame for multirate CDMA data networks, *IEEE Trans on Wireless Comm.*, Jan., Vol. 2, No. 1, pp.186–219.

Virapanicharoen, J. & Benjpolakaul, W. (2004). Fair-efficient guard bandwidth coefficients selection in call admission control for mobile communications using game-theoretic framework, *IEEE ICC 2004*, June, Vol. 1, pp.80–84.

Wang, Q. & Brown, T.X. (2007). Public safety and commercial spectrum sharing via network pricing and admission control, *IEEE Journal of Sel. Areas in Comm.*, April, Vol. 25, No. 3, pp.622–632.

Xing, Y.; Chandramouli, R. & Cordeiro, C. (2007). Price dynamics in competitive agile spectrum access markets, *IEEE Journal of Sel. Areas in Comm.*, April, Vol. 25, No. 3, pp.613–621.

Zheng, H. & Cao, L. (2005). Device-centric spectrum management, *Proc. of IEEE Intern. Symp. on Dynamic Spectr. Acc. Net. (SySPAN) 2005*, Nov., pp.56–65.

Achievable Throughput Comparison of Sensing-Based and Interference-Constrained Transmissions in Cognitive Radio Networks

Gosan Noh and Daesik Hong
Yonsei University
Korea

1. Introduction

Cognitive radio has been introduced in order to solve the spectrum scarcity problem (Haykin 2005). Although having limited radio resources, we need lots of spectrum to deploy newly developed wireless applications. Meanwhile, a large portion of allocated spectrum is identified as unused by an actual radio spectrum measurement. Thus, a new access scheme which allows spectrum sharing between different wireless systems referred as dynamic spectrum access is required (Zhao & Sadler 2007). Dynamic spectrum access is based on cognitive radio technology which enables learning from and adapting to the external radio environment.

As a way of spectrum sharing between licensed and unlicensed users, spectrum overlay approach is considered. In spectrum overlay networks, a primary network has a license for the exclusive use of the allocated spectrum. In contrast, a secondary network has lower access priority. While the primary users access the licensed spectrum wherever and whenever they want, the secondary users access the spectrum on condition that the transmission of the primary network is sufficiently protected, i.e., the interference to the primary network should be less than a predefined threshold.

There are two ways for satisfying the protection condition: Sensing-based and interference-constrained transmissions. In sensing-based transmission, as a means of avoiding interference, the secondary users sense the spectrum before they start transmission (Liang et al. 2007; Kim et al.). Only if the secondary users detect the white space, they can access the spectrum. Even during the transmission, if the primary user uses the spectrum, the secondary user has to stop transmission. In interference-constrained transmission, on the contrary, the secondary users are allowed to transmit during the primary user transmission (Gastpar 2007). However, the transmit power should be adjusted not to interfere the primary user transmission.

In this chapter, we compare the two transmission schemes in terms of achievable throughput of the secondary user and provide a criterion for the transmission mode selection of sensing-based transmission and interference-constrained transmission.

The rest of this chapter is organized as follows. Section 2 presents the system model for the sensing-based and interference-constrained transmissions. Throughput analyses for both of the schemes are followed in Section 3. The simulation results are shown in Section 4. In Section 5, the conclusion is drawn.

2. System Model

A coexistence scenario of spectrum overlay between primary and secondary networks is depicted in Fig. 1. Primary and secondary networks are deployed in the overlapping regions and use the same frequency band. In Fig. 1, there are pairs of transmitter and receiver for the primary and secondary networks. We define channel gains between each pair of transmitter and receiver as e (from the primary transmitter to the primary receiver) and g (from the secondary transmitter to the secondary receiver). We also define the channel gains representing the interference between primary and secondary networks as f (from the secondary transmitter to the primary receiver) and h (from the primary transmitter to the secondary receiver).
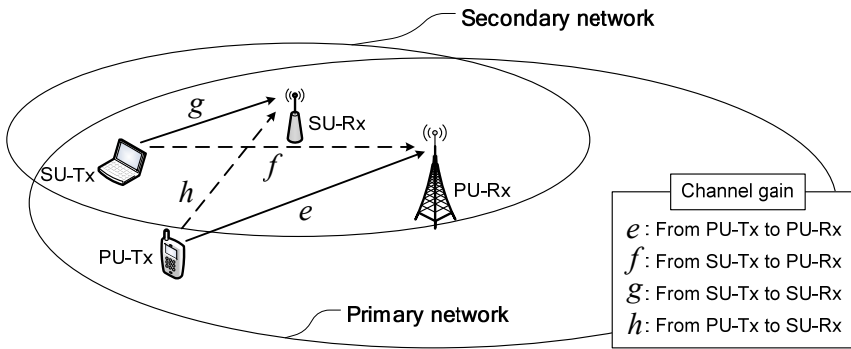

Fig. 1. A coexistence scenario of primary and secondary networks. Primary and secondary networks are over-deployed and interfere with each other.

The basic concept of spectrum overlay approach is to unlock licensed spectrum to secondary users. Secondary users can access the licensed spectrum on condition that their interference to primary users is limited. Possible solutions for this problem are sensing-based transmission and interference-constrained transmission. The difference of two schemes is based on the support for the simultaneous transmission.

In sensing-based transmission, the secondary user should sense the existence of the primary user in the licensed spectrum before transmission. Through the transmission, whenever the primary user accesses the licensed spectrum, the secondary user is required to stop transmitting and vacate the spectrum. Each secondary user frame is divided into a sensing slot and a data transmission slot. The time length that the secondary user senses the spectrum is sensing duration n_D and the time length between the consecutive sensing durations is sensing period n_P (Noh 2008).

In interference-constrained transmission, the secondary user can access while the primary user uses the spectrum. However, the power of the secondary user should be lessen so as to the QoS of the primary user is guaranteed. Without loss of generality, we employ a restriction that the interference power of the secondary transmitter measured at the primary receiver should be lower than a predetermined threshold.

3. Throughput Analysis

In this section, the achievable throughput of the secondary user by means of both sensing-based and interference-constrained transmission is presented and then compared in terms of location of terminals and the acceptable power level of the primary user.

3.1 Sensing-Based Transmission
The first thing to derive the secondary user throughput based on sensing-based transmission is evaluating the sensing accuracy. During the sensing duration, $r(n)$ denotes the primary user signal received by the secondary user. Then, the detection problem can be written as

$$\mathcal{H}_0 : r(n) = w(n)$$
$$\mathcal{H}_1 : r(n) = p(n) + w(n)$$

(1)

for $n = 0, \ldots, n_D - 1$. $\mathcal{H}_0$ denotes the noise only hypothesis and $\mathcal{H}_1$ denotes the primary user signal plus noise hypothesis. $p(n)$ is the primary user signal and $w(n)$ is the noise (Liang et al. 2007). The primary user signal $p(n)$ is assumed to be a zero-mean, complex-valued, circularly symmetric, white Gaussian random process with a variance σ_p^2. The complex-valued, circularly symmetric AWGN $w(n)$ with a variance σ_w^2 is independent of the primary user signal.
In order to detect the existence of a random signal such as p(n), an energy detector is applicable (Kay 1998). The decision rule is written as

$$T = \sum_{n=0}^{n_D-1} |r(n)|^2 \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \eta$$

(2)

where T is the test statistic and η is the threshold.
Two probabilities are defined in conjunction with the detection procedure: false alarm probability and detection probability. False alarm probability P_{FA} is the probability to decide $\mathcal{H}_1$ but $\mathcal{H}_0$ is true. Detection probability P_D is the probability to decide $\mathcal{H}_1$ and it is true.

False alarm probability P_{FA} is given by

$$P_{FA} = \Pr\{T > \eta; \mathcal{H}_0\} = Q\left(\frac{\eta}{2\sqrt{n_D}\sigma_w^2} - \sqrt{n_D}\right) \tag{3}$$

where $Q(\cdot)$ is the right-tail probability for the Gaussian random variable (Kay 1998). Detection probability P_D is also given by

$$P_D = \Pr\{T > \eta; \mathcal{H}_1\} = Q\left(\frac{\sigma_w^2}{\sigma_p^2 + \sigma_w^2}\left(Q^{-1}(P_{FA}) - \frac{\sqrt{n_D}\sigma_p^2}{\sigma_w^2}\right)\right) \tag{4}$$

By rearranging (3) and (4) and canceling η, we have following relations:

$$P_{FA} = Q\left(\sqrt{n_D}\gamma_p + (1+\gamma_p)Q^{-1}(P_D)\right) \tag{5}$$

$$P_D = Q\left(\frac{1}{1+\gamma_p}\left(Q^{-1}(P_{FA}) - \sqrt{n_D}\gamma_p\right)\right) \tag{6}$$

where γ_p represents the signal-to-noise ratio (SNR) of the primary user, i.e., $\gamma_p = \sigma_p^2/\sigma_w^2$. There is a tradeoff relationship between the false alarm probability and the detection probability. Lower false alarm probability gives the secondary user more chances to access the channel and higher detection probability guarantees the transmission of the primary user more strongly. However, realizing both the two goals are impossible. Hence, by using the tradeoff relationship above, the secondary user throughput can be derived. The secondary user succeeds in transmitting its data when there is no false alarm, where no false alarm means the primary user does not use the spectrum and the secondary user knows this information (Liang et al. 2007). The secondary user throughput of the sensing-based transmission is given by

$$C_{SB} = P_{idle}\frac{n_P - n_D}{n_P}\left(1 - P_{FA}\right)C_0 \tag{7}$$

such that $(1-P_D)\,|\,f\,|^2\,P_s^{\max} \leq \lambda$, where λ is the interference threshold at the primary receiver. This constraint means that the expected interference from the secondary transmitter experienced at the primary receiver should be lower than a predetermined threshold. This constraint is required to protect the transmission of the primary user.

However, in practical situations, we set as $(1 - P_D) |f|^2 P_s^{max} = \lambda$ so as to maximize the secondary user throughput. The reason comes from the fact that the monotonic increase of the detection probability (6). Thus we set the detection probability as $P_D = 1 - \lambda / |f|^2 P_s^{max}$. P_{idle} is the probability that the primary user does not uses the spectrum, which is given as: $P_{idle} = \mu / (\lambda + \mu)$. $(n_P - n_D) / n_P$ represents the ratio of the transmission time to the total time. C_0 denotes the capacity of the secondary user without any interference. C_0 is given by

$$C_0 = \log_2(1 + \gamma_s)$$

(8)

where γ_s is the SNR of the secondary user.

3.2 Interference-Constrained Transmission

The secondary user throughput based on interference-constrained transmission is presented. Interference-constrained transmission allows the secondary user simultaneous transmission with the primary user. In this case, however, the primary user transmission also should be protected by means of power control of the secondary user. If the interference of the secondary user experienced by the primary receiver is strong, the power of the secondary transmitter should be lessened. Thus, the secondary user power is controlled so as to its interference at the primary receiver is lower than a predetermined threshold, i.e., $|f|^2 P_s \leq \lambda$ (Gastpar 2007). P_s is the transmit power of the secondary user. Hence, the secondary user throughput of the interference-constrained transmission is given by

$$C_{IC} = \log_2 \left(1 + \frac{|g|^2 P_s}{|h|^2 \sigma_p^2 + \sigma_w^2} \right)$$

(9)

such that $|f|^2 P_s \leq \lambda$.

In order to maximize the secondary user throughput, the optimal strategy is that the interference power of the secondary user meets the threshold, i.e., $|f|^2 P_s = \lambda$. By this principle, the throughput (9) is rewritten as

$$C_{IC} = \log_2 \left(1 + \frac{|g|^2 \lambda / |f|^2}{|h|^2 \sigma_p^2 + \sigma_w^2} \right)$$

(10)

Notice that there is an interference term in (10) which is comes from the primary transmitter and degrades the secondary user throughput.

Parameter	Value
λ	0.2 kpkts/s
μ	0.6 kpkts/s
n_D	30 samples
n_P	300 samples
W	1 MHz

Table 1. Common simulation parameters.

4. Simulation Results

In this section, the throughput performances of both the sensing-based and interference-constrained transmissions are evaluated by computer simulations. In a primary network, we assume that there is a primary transmitter and a primary receiver. Similarly, in a secondary network, there is a secondary transmitter and a secondary receiver.

We assume a cellular scenario where the primary network and the secondary network are over-deployed. Thus, we are interested in the distances between two terminals and base stations. The distance from the secondary transmitter to the primary receiver is denoted by l_f. l_f affects of the interference from the secondary user to the primary user. The distance from the secondary transmitter to the secondary receiver is denoted by l_g. l_g determines the signal quality of the secondary user. The distance from the primary transmitter to the secondary receiver is denoted by l_h. l_h affects of the interference from the primary user to the secondary user.

We consider the path loss and short-term fading as a channel model. Path loss exponent is assumed to be 4. In addition, Rayleigh block fading is assumed, where each of channel gain is circularly symmetric complex-Gaussian random variable. Hence, the channel gain f is decomposed into the path loss component and the Rayleigh fading component, i.e., $f = (l_f/l_0)^{-\alpha/2} \hat{f}$, where l_0 denotes a unit distance and $\hat{f}$ denotes a zero mean complex-Gaussian random variable or unit variance. Similarly, g and h are also decomposed into $g = (l_g/l_0)^{-\alpha/2} \hat{g}$ and $h = (l_h/l_0)^{-\alpha/2} \hat{h}$, respectively.

The primary user follows Markovian traffic, with arrival rate $\lambda = 0.2$ kpkts/s and service rate $\mu = 0.6$ kpkts/s. Thus, the time portion that the primary user is in idle state is given as $P_{idle} = 0.75$. For sensing-based transmission, we assumed an energy detector with the sensing duration $n_D = 30$ samples and the sensing period $n_P = 300$ samples. The system bandwidth is given by $W = 1$ MHz. We assume a full-traffic model, where the secondary user always has traffic to transmit. These common parameters are summarized in Table 1.

Under the sensing-based transmission, the secondary user throughput as a function of the SNR of the primary user is depicted in Fig. 2. It is observed that the secondary user throughput increases with the increase with the SNR of the primary user. As the SNR increases, the energy detection becomes more accurate and enhances the throughput of sensing-based transmission. In addition, The secondary user throughput increases with λ. With higher λ, the interference limit of the primary user is relaxed, the secondary user can transmit with higher power, and the throughput is enhanced.

Oppositely, under the interference-constrained transmission, the secondary user throughput as a function of the SNR of the primary user is shown in Fig. 3. Notice that the secondary user throughput decreases with the increase with the SNR of the primary user. Differently from the sensing-based transmission, as the SNR of the primary user increases, the interference from the primary transmitter greatly degrades the secondary user throughput as in (10). Similar to that of the sensing-based transmission, it is observed that the secondary user throughput increases with λ.

We then compare the throughput performance of the sensing-based transmission with the interference-constrained transmission in Fig. 4, Fig. 5, and Fig. 6.

Fig. 4 depicts the secondary user throughput of both the sensing-based and the interference-constrained transmission as a function of the interference threshold of the primary receiver. The secondary user throughput increases with the increase of λ in both schemes. However, there is a crossing point at about $\lambda = 0.05$. In low λ region, the sensing-based transmission achieves higher throughput, but in the high λ region, the interference-constrained transmission outperforms the other.
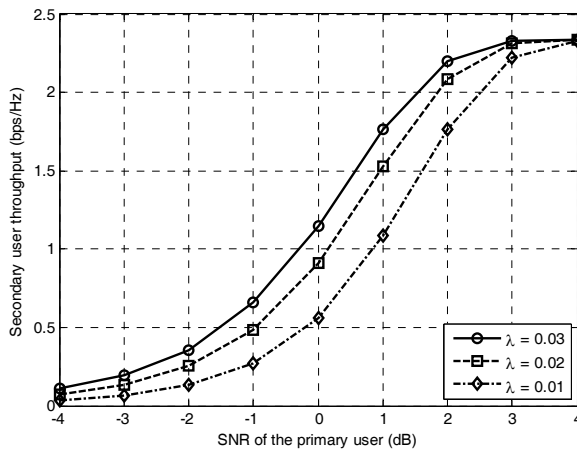

Fig. 2. Secondary user throughput of sensing-based transmission vs. the SNR of the primary user. The throughput of the secondary user increases with the SNR of the primary user and the interference threshold for the primary user.
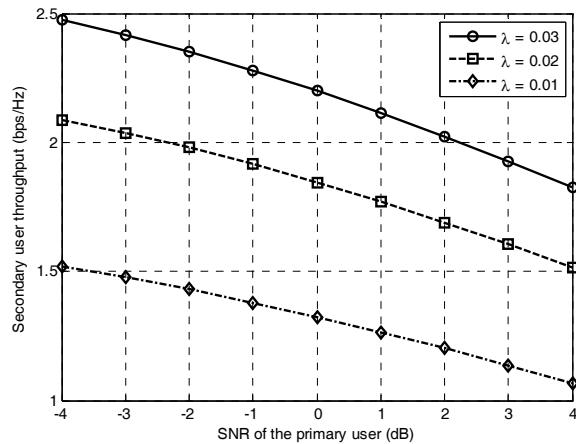
Fig. 3. Secondary user throughput of interference-constrained transmission vs. the SNR of the primary user. The throughput of the secondary user decreases with the SNR of the primary user but increases with the interference threshold for the primary user.
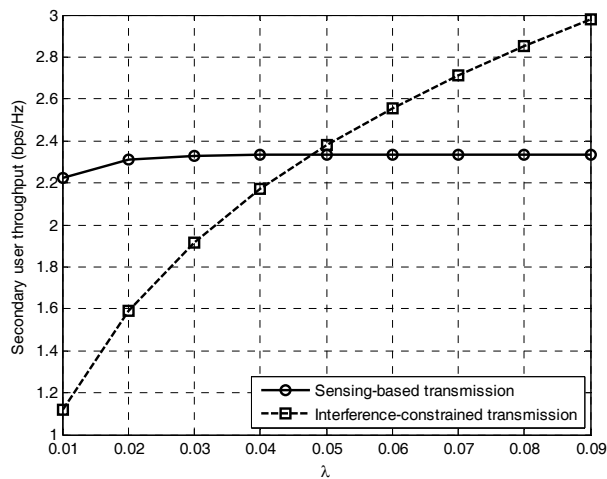
Fig. 4. Throughput comparison of the sensing-based and interference-constrained transmissions vs. the interference threshold of the primary receiver. The throughput of the sensing-based scheme is saturated as the threshold gets higher without saturation.

Fig. 5. Throughput comparison of the sensing-based and interference-constrained transmissions vs. the distance ratio l_f/l_g . The throughput is saturated in the sensing-based scheme while the throughput is sharply increasing in the interference-constrained scheme.
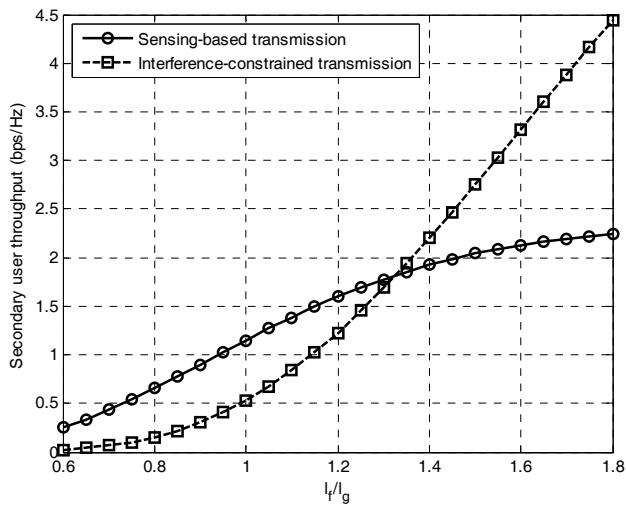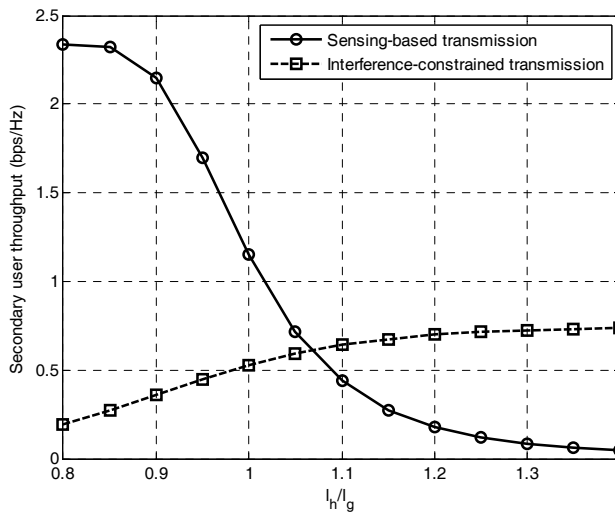

Fig. 6. Throughput comparison of the sensing-based and interference-constrained transmissions vs. the distance ratio l_h/l_g . The throughput by the sensing-based scheme decreases while that by the interference-constrained scheme increases.

Fig. 5 shows the secondary user throughput as a function of the distance ratio l_f/l_g. As l_f/l_g increases, the primary receiver becomes farther from the secondary transmitter. Then, the spectrum opportunity of the secondary user increases and the throughput is also increased. However, the throughput is saturated in the sensing-based scheme while the throughput is sharply increasing in the interference-constrained scheme. Thus, the effect of the interference threshold is more sensitive to the interference-constrained transmission.

Fig. 6 shows the secondary user throughput as a function of the distance ratio l_h/l_g. Notice that as l_h/l_g increases, the throughput by the sensing-based scheme decreases while that by the interference-constrained scheme increases. If the primary transmitter goes farther from the secondary receiver, it becomes difficult to sense the primary user and the throughput degrades in the sensing-based scheme but the interference decreases and the throughput is enhanced in the interference-constrained scheme.

7. Conclusion

We have discussed the achievable throughput of both the sensing-based and interference-constrained transmission in cognitive radio networks. The derivations of both schemes are presented and their throughput performances have been compared in various environments via computer simulations. In conclusion, the sensing-based scheme is advantageous when the interference threshold is tight and the primary and secondary users are relatively close. Oppositely, the interference-constrained scheme is better when the interference threshold is loose and the primary transmitter and receiver are located far from the secondary user.

8. References

Gastpar, M. (2007). On capacity under receive and spatial spectrum-sharing constraints. *IEEE Trans. Wireless Communications,* Vol. 6, No. 2, pp. 649-658, Feb. 2007

Haykin, S. (2005). Cognitive radio: Brain-empowered wireless communications. *IEEE Communications Mag.,* Vol. 23, No. 2, pp. 201–220, Feb. 2005

Kay, S. M. (1998). *Fundamentals of Statistical Signal Processing: Detection Theory,* Prentice-Hall, Upper Saddle River, NJ

Kim, S. ; Lee, J. ; Wang, H . & Hong, D. Sensing Performance of Energy Detector with Correlated Multiple Antennas. *IEEE Signal Processing Letters,* Accepted for publication.

Liang, Y. C-. ; Zeng, Y . ; Peh, E. & Hoang, A. T. (2008). Sensing-throughput tradeoff for cognitive radio networks. *IEEE Trans. Wireless Communications,* Vol. 7, No. 4, pp. 1326-1337, Apr. 2008

Noh, G. ; Lee, J. ; Wang, H. ; Kim, S. & Hong, D. (2008). A new spectrum sensing scheme using cyclic prefix for cognitive radio systems. *Proceedings of IEEE Vehicular Technology Conference,* pp. 1891-1895, May 2008

Zhao, Q. & Sadler, B. M. (2007). A survey of dynamic spectrum access: Signal processing, networking, and regulatory policy. *IEEE Signal Processing Mag.,* Vol. 24, No. 3, pp. 79-89, May 2007

Intermittent DCF: a MAC protocol for Cognitive Radios in Overlay Access Networks

Athanassios V. Adamis and Prof. Philip Constantinou
National Technical University of Athens
Greece

1. Introduction

Cognitive Radios have recently appeared as a solution to the RF spectrum underutilization problem, which results from the sparse use of wireless networks services at certain time periods and the static spectrum allocation policies. A plethora of measurement campaigns have pointed out that there is a lot of unused spectrum, referred to as "white space", even in the commercial broadcast and mobile networks frequency bands, which are considered to be crowded. Cognitive Radios - CR, that inherently can sense and adapt to their environment, can form Dynamic Spectrum Access - DSA networks by utilizing the white spaces and thus improve spectrum utilization. In Overlay Access - OA, the DSA network operates on spectrum bands currently licensed to and used by conventional radio systems that are called Primary systems. The DSA network, which is called the Secondary system, is allowed to access spectrum portions that currently remain unused by the Primary system, provided that it will respect the interference criteria of the licensees. New functions and capabilities, not needed by conventional networks, must be developed at DSA networks. A few of these are spectrum sensing methods, spectrum sharing algorithms and spectrum access techniques. Our work lies in the spectrum sharing function and more specifically at the MAC sublayer.

This chapter presents a modified version of the Distributed Coordination Function-DCF, originally found in the IEEE 802.11 (IEEE Standards, 1999), as the MAC protocol for OA networks. The overlay environment poses certain challenges to the MAC design. In order for the DSA network to be up to date with spectrum availability, it must perform periodic spectrum scan procedures. During scan procedures the MAC must be interrupted and hold its state. Moreover, it must synchronize the scan procedures across the DSA network terminals, in order for them to simultaneously cease transmission attempts and perform the sensing task. After the scan procedures, it must adapt to the new PHY layer provided transmission rate. Frequent scan procedures result in frequent interruptions. Fast changes in spectrum availability, caused by fluctuating Primary network traffic, necessitates frequent scan procedures and results in sharp changes in the transmission rate. All these form the demanding overlay environment, that an overlay MAC protocol must deal with. The

proposed modified DCF, called the Intermittent DCF – iDCF, is stable and robust in this demanding environment and moreover in low achievable transmission rates. DCF has been also proposed as the MAC protocol for OA in many DSA research works. Nevertheless, no modifications have been described on it. Moreover it is proposed mostly for accessing a dedicated control channel, or data channels in a multichannel environment of multi-transceiver terminals. We show that in the demanding overlay environment, an unmodified DCF can lead to heavy channel saturation, unfairness and pure performance. The proposed iDCF maintain fairness and performance.

Moreover, this chapter presents an analytical iDCF throughput and packet delay calculation method in the overlay environment. The analysis is based on existing Markov chains models and extends them in order to include the characteristics of the intermittent operation (Adamis et al., 2009). The predictions of the model are validated against simulations and are found to accurately capture many interesting features of the iDCF. The analytical model can be used by researchers for feasibility studies in various overlay scenarios and in admission control algorithms at QoS enabled future overlay networks. Using the model, we present a numerical example for the effect of the transmission rate, the scan interruptions frequency and the packet payload length on the iDCF performance.

2. The Distributed Overlay Access Environment

Many different coexistence scenarios between Secondary and Primary systems can be found in the literature. Worldwide research on overlay access technologies is still at the beginning and at this point it is not clear which scenarios are the most realistic and practical. We study the case in which the Primary system utilizes some kind of TDMA/FDMA combination access scheme (Berthold et al., 2007 b), and thus spectrum holes are created at idle Primary system channels.

The Secondary System is a distributed network of CR terminals and its main objective is to operate in the same spectral band with the Primary system. This spectrum coexistence of two radio systems, a Primary and a Secondary, poses certain rules and priorities in spectrum access. The ultimate goals of a Secondary System are (a) to access the spectrum respecting the interference tolerance metrics imposed by the Primary system, that is causing no harmful interference to it, and (b) to accomplish goal (a) without requiring neither any modification to the Primary system nor control information exchange with it. A lot of DSA research studies propose centralized architectures for Secondary systems and also engage information exchange mechanism between the coexisting systems. Nevertheless the road through distributed approaches and complete independence from the Primary system can lead to the implementation of autonomous, fast deployed, low cost CR networks for a variety of applications such as broadband internet access through self organizing mesh networks, rural access with reduced delay and cost and pervasive content access at home (Stirling, 2008). In the following subsections some important MAC design issues for the OA Environment are presented.

2.1 The Scan Period T
In order to identify the spectrum portions left unused by the Primary system and catch up with this knowledge, the CR terminals must perform spectrum scan procedures. These scan procedures are performed periodically with a *scan period T*. Communications services are

stopped during the scan and the PHY layer performs the spectrum sensing task. The *scan period* is a Secondary system parameter that directly affects its performance as it designates how frequently the communications services must be interrupted.
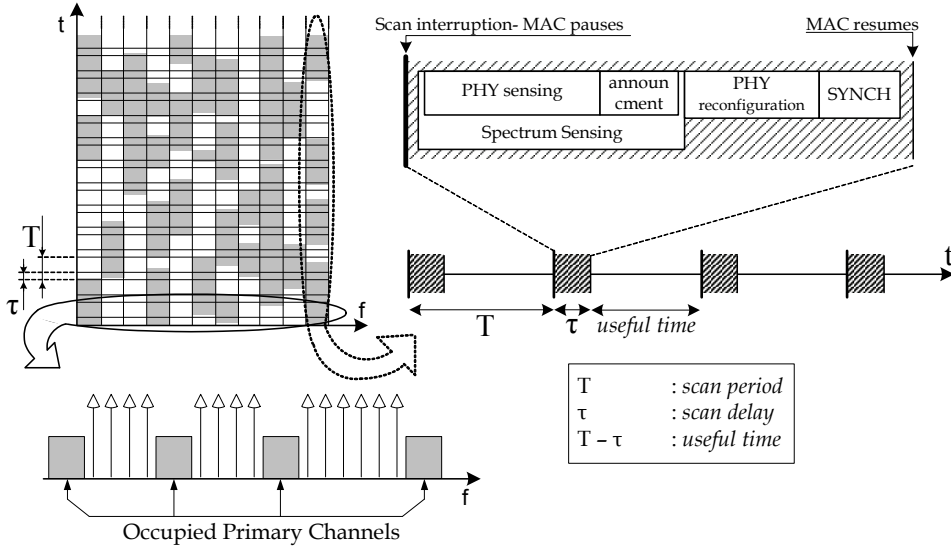

Fig. 1. Spectrum Pooling for Overlay Access Networks

The figure above presents in the time axes, the intermittent operation of Secondary networks having to perform spectrum scan procedures periodically. The scan procedures should be frequent enough, depending on the underlying Primary system, in such a way that the DSA network meets its design targets (Berthold & Jondral, 2007b) and respects the interference criteria of the Primary system. *T* may vary considerably. In the presently developing IEEE 802.22 standard (Cordeiro et al., 2005) and other approaches (Yuan et al., 2007) a maximum value is set to 30 min. for operation over TV spectrum, while for operation over a GSM Primary system (Papadimitratos, et al, 2005) it is set to 577μs. Certain approaches (Ghasemi & Sousa, 2008) propose that the *scan period* should be set for each licensed band by the spectrum regulator. In (Willkomm et al., 2008) the advantages of using dynamic selection of the *scan period*, as the secondary system roams over different Primary systems and operates on different times of a day and different days of a week, are depicted, using real world spectrum usage measurements. It is clear that an overlay MAC protocol should be robust and capable of operating at a wide *scan period* range.

2.2 The Scan Delay τ

The intermission, within which the sensing task takes place, has a total duration of τ as shown in Fig. 1. At the beginning of the interval τ the PHY layer performs spectrum sensing. The spectrum sensing duration may include a measurements announcement procedure between CR terminals in case of a cooperative sensing scheme. At the end of the sensing task the identified white spaces are treated as a single channel, common for all participating

terminals. PHY layer reconfigures to meet the new parameters and a communication channel is established. It is very crucial that the scan procedure is performed simultaneously by all participating CR terminals. This will ensure that all terminals will stop transmitting during the scan. Otherwise the sensing algorithm may not be able to distinguish between Primary and Secondary system transmissions. This crucial time synchronization is a task to be performed by the MAC layer, which is responsible to pause its operation and initiate the scan procedure. In our scenario we assume that the synchronization task is performed within the time interval τ and specifically at the end of it (Fig. 1): after a common channel has been established between terminals, some synchronization messages exchange mechanism similar to the one proposed in (Berthold et al., 2007 a) can be applied. For simplicity τ is called *scan delay* and the included functions scan procedure.

The mechanisms for spectrum sensing, measurements announcement, PHY layer reconfiguration and synchronization affect the total τ duration. Scan delay may also vary depending on the total scanned bandwidth, the sensing algorithm and the number of samples needed to achieve the required sensing performance. In other words different Primary systems, having different interference tolerance behaviour, demand different sensing algorithm performance in the Secondary network, which is obtained by different sensing times. An overlay MAC protocol should be robust and capable of operating at a wide *scan delay* range, too.

2.3 The PHY layer provided Transmission Rate

After the finish of the scan procedure, for a time interval equal to T-τ, till the next scan procedure, the identified spectrum holes are treated as a single channel, common for all participating CR terminals and the MAC performs the communications functions. This time interval between successive spectrum scans is called *useful time* in our study, as shown in Fig. 1. OFDM based spectrum pooling (Weiss & Jondral, 2004) is a transmission technology that can accommodate the adaptive and smooth filling of the spectrum white spaces. This kind of accessing the spectrum by transmitting on idle Primary system channels, shown in Fig. 1, is called Overlay Access and gives its name to the Overlay Access Networks.

The number of the available data subcarriers, and thus the transmission rate, depend on the primary channels current usage. After every scan procedure, the transmission rate may be different. Small *scan period*s, imposed by rapidly changing Primary system traffic characteristics, may result in rapid changes in the transmission rate. This type of the continuously adapting PHY layer provided transmission rate must be handled efficiently by the MAC layer.

A Secondary system MAC protocol, designed for such an OA mode, must be able to operate in a wide range of the varying *scan period*, *scan delay* and transmission rate. Next, we modify the DCF in order to be robust in the demanding environment described above.

3. The DCF in Overlay Access Research

Many OA studies propose the DCF as the MAC scheme to be used. In the majority, it is proposed as the scheme for accessing the Common Control Channel (CCC) (Hang & Xi, 2008); (Motamedi & Bahai, 2007). The CCC is a separate dedicated channel, which is continuously available to the CR terminals, as no Primary system operates on it. In these approaches the CCC is accessed using the DCF or a p-persistent CSMA scheme. The access

on CCC is not intermittent and thus no modifications on the DCF are required. In our proposed scenario the DCF is used for accessing the unique available common channel in the intermittent overlay mode. Moreover, the former approaches utilize at least two RF Front Ends, i.e. transceivers, one of which is always tuned to the CCC. Increase of the number of transceivers increases the reliability of the system and reduces certain delays, however it also increases the total cost. Our approach utilizes a single transceiver and all terminals use a single channel. This can be seen either as a single channel environment, or as a multi channel one with the focus given on the coexistence of multiple terminals in each channel with the use of DCF.

Other approaches with a single transceiver also use schemes similar to the DCF during the contention period (Juncheng et al., 2008). The difference from our approach is that they use it exclusively on the CCC. Moreover, to the best of our knowledge, no modifications on the DCF have been proposed that can handle its operation in the demanding overlay mode. In (Papadimitratos et al., 2005) the DCF protocol is used in an intermittent way over a GSM network with scan procedures at the beginning of each GSM-slot. This implementation is a modification of DCF, but only for overlay access on the GSM band and is applicable only to this Primary system. For example, the NAV counter for the virtual carrier sensing mechanism is set in number of GSM-slots needed for the data transmission. The backoff interval in this approach, is generated after every scan procedure, which can lead to heavy congestion in the CCC and unfairness. In our work though, the DCF is modified in a way that it can be robust in a variety of overlay scenarios.

4. The Intermittent DCF - iDCF implementation

The intermittent DCF performs data transmission functions between successive spectrum scan procedures, during the *useful time* shown in Fig. 1. The basic communications operation of the protocol, as in the standard protocol (IEEE Standards, 1999), is by successive Frame Exchange Sequences - FES which always take the form *rts-cts-data-ack*. The three way hand-shake mechanism is mandatory in the OA environment because of the substantially lower levels of interference it causes to the Primary system as compared to the direct *data-ack* FES form (Adamis et al., 2008). Each station contends for gaining access by having to sense the medium idle for a DCF interframe space - DIFS period and then by performing a backoff procedure for a random number of protocol slots. Upon backoff expiration it initiates a FES by sending the *rts* message. In case of a successful transmission it receives the *cts* message by the receiver and the FES continuous with the data transmission.

Except from the basic access scheme functions, which are identical for both the iDCF and the standard DCF, some functions had to be added or modified because of the intermittent operation mode in the OA environment, described earlier. We present these modifications in detail in the following sections.

4.1 Control and Data Packet Transmission

As described in Section 2.2 it is very important that all CR terminals cease transmissions when the scan procedure takes place. The iDCF is always aware of the remaining *useful time* until the upcoming scan procedure and it does not initiate a transmission if it will exceed the *useful time* boundary. There is a distinction in this iDCF behaviour between control packets and data packets. Control packets *cts* and *ack* cannot be fragmented. Upon a control packet

transmission, in case the remaining *useful time* is not adequate, the packet is sent after the end of the scan procedure in the beginning of the next *useful time* period. This way the remaining time until the scan procedure is wasted. The *rts* control packet is not examined here because the backoff modification described later in Section 4.2 guarantees that on backoff expiration, there will always be enough remaining *useful time* for an *rts* message transmission.
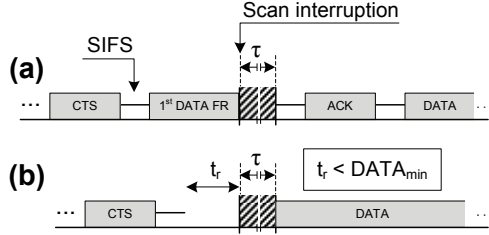

Fig. 2. Data Packet Transmission at Scan Procedure Interruption

When a data packet is ready to be sent and there is not enough remaining time for the entire packet, then it is fragmented and the first data fragment fills the remaining time, as shown in Fig. 2a. This fragmentation is performed only if the fragment to be transmitted contains more than L_{min} data bits. Fig. 2b presents the case in which the remaining time t_r is shorter than the one required to accommodate a data fragment containing L_{min} data bits**'**, denoted as DATA$_{min}$. L_{min} is a parameter that shows the least amount of data bits that can be sent in a single fragment. As in the standard protocol, each data fragment is followed by an *ack* message. The rest of the data bits are sent at subsequent fragments.

4.2 Backoff Freezing

The intermissions imposed by the scan procedures necessitate a modification in the way that the backoff counter freezes, when the scan procedure approaches, in order for the protocol fairness and priorities to be preserved. The backoff down counter must freeze when the last *rts* transmit opportunity t_{lTxOp} time point is crossed. The t_{lTxOp} is defined as the time point in which if an *rts* message is transmitted, it will fit exactly in the remaining *useful time* before the interruption and can be seen in the figure below. A maximum anticipated propagation delay δ is also considered in the above definition.

Fig. 3 presents the back off counter values of two terminals that contend for the medium. As soon as a terminal's backoff counter expires the terminal wins the contention and transmits the *rts* message. In Fig. 3a the backoff procedure is presented without modifications. Here, the expiration of both terminals backoff counters when the t_{lTxOp} and before the scan interruption resulted in a collision, because both terminals transmitted at the beginning of the next *useful time*. This type of collisions occurs frequently in the intermittent operation mode, especially for low transmission rates (i.e. long $RTS+\delta$ duration) and small scan periods. Fig. 3b presents the modified backoff procedure with the counter freeze after the t_{lTxOp} is crossed. Using the proposed modification the priorities are preserved and the terminal with the lower backoff value wins the contention.
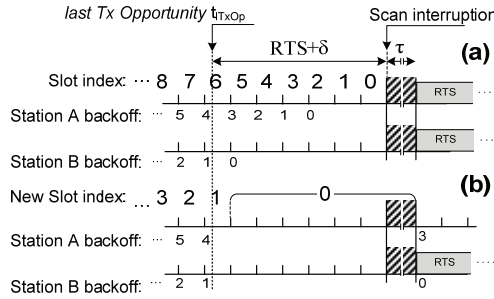
Fig. 3. Backoff freeze when a scan procedure interruption approaches.

The problem off the backoff expiration at time points that the transceiver is not able or allowed to carry out the transmission has been also identified in (Shu et al., 2007) for multi-mode terminals. There, by the name of the disruptive CSMA, it is encountered by a credit payback mechanism. The fact that in the iDCF the disruptions occur concurrently at all CR terminals, due to the scan procedure, leads to the previous simple and effective solution of freezing the backoff counter at an earlier point.

4.3 The Interframe Spaces - IFS

The inter frame spaces are time periods lying in between packet transmissions, during which the medium must be sensed idle. They exist for (a) leaving time to PHY and MAC layer to perform packet processing tasks, (b) anticipating a certain amount of propagation delay and (c) last but not least, priority setting among terminals through the IFS length differentiation, e.g. a terminal that adopts a shorter IFS before attempting transmissions has a larger probability of accessing the medium than one that adopts a longer IFS.

An IFS may be interrupted by an upcoming scan procedure before it is completed. When this occurs, the IFS is expanded to include the scan procedure and it is completed within the next *useful time*. Fig. 4 presents the different tasks performed within the IFS. Whether the IFS is extended by the *scan delay* duration τ, or more, depends on which task the interruption occurs at. Thus, if the interruption occurs at either a PHY or MAC layer processing task (D1, M1), the task is suspended, the scan procedure begins and the task is resumed after the scan procedure end. However, the transceiver switching task from receiving to transmitting mode (Rx>Tx), the anticipated air propagation delay task (AAPT) and the clear channel assessment procedure (CCA) are tasks that cannot be suspended and resumed. These are tasks that must be performed uninterrupted within a single *useful time* either before the scan procedure, or after it. If the interruption occurs at either of these IFS parts, these tasks are performed entirely in the upcoming *useful time*. Fig. 5 and Fig. 6 depict examples of a SIFS being interrupted at the PHY layer processing delay and a DIFS at the CCA task of the second slot.
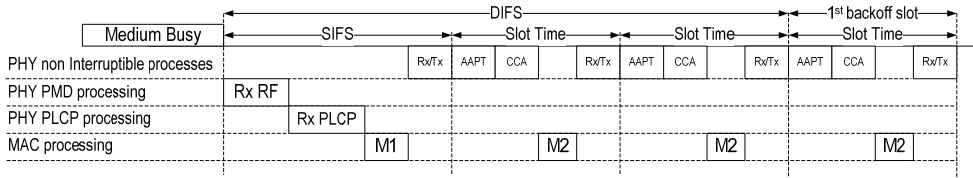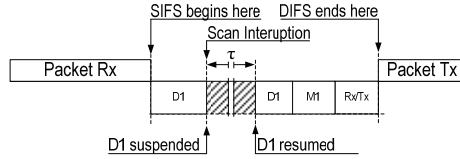
Fig. 4. IFS and Protocol Slot performed tasks

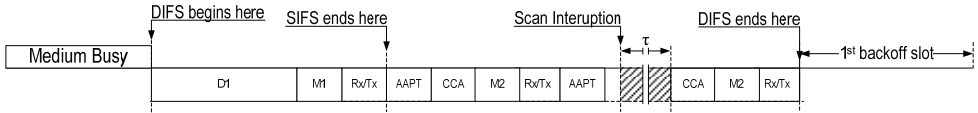Fig. 5. SIFS interrupted at PHY layer processing task

Fig. 6. DIFS interrupted at CCA task.

4.4 Dynamic Virtual Carrier Sensing

The virtual carrier sensing mechanism is implemented as a decreasing time counter called NAV (Network Allocation Vector), which holds a prognosis for how much longer the medium is expected to be busy due to the current FES. In the standard protocol, the NAV counter is set to the most recently received value, found in the DurationID packet field of the MAC header. This is calculated and set by the transmitting terminal. In the iDCF, a FES may take over more than one successive *useful times*. Moreover, the transmission rate after each scan procedure is unknown. Thus, the duration of an ongoing FES cannot always be calculated and inserted in the DurationID packet field and a different virtual carrier sensing mechanism must be implemented in the intermittent OA environment.

In iDCF the DurationID packet field instead of carrying direct information on the time needed until the end of the ongoing FES, it holds the data payload size L_P that the FES is going to transmit. In case of multiple MPDUs, either because the data payload size exceeds the maximum MPDU size, or because scan procedure interruptions force data fragmentation, DurationID holds the remaining payload size L_{PR} that has not been transmitted yet. DurationID is set to the values given in Table 1 for each case. Except from the FES transmitter and receiver, each terminal sets its NAV counter according to (a) the type of the received packet, which designates what is going to be transmitted next, (b) the received data payload size, (c) the current transmission rate and (d) the remaining *useful time*. Based on the above information the NAV counter is set accordingly to include the transmission of the next expected packet and the two idle SIFS periods before and after this packet. When the next expected packet cannot fit in the remaining *useful time* the NAV counter is not set until the transmission rate is updated in the beginning of the next *useful time*. Fig. 7 shows some examples of the NAV counter setting. The virtual carrier sensing

mechanism has to operate this way in iDCF, because it is not known how many scan procedures will take place before the FES completion and what will be the value of the transmission rate after each scan procedure.
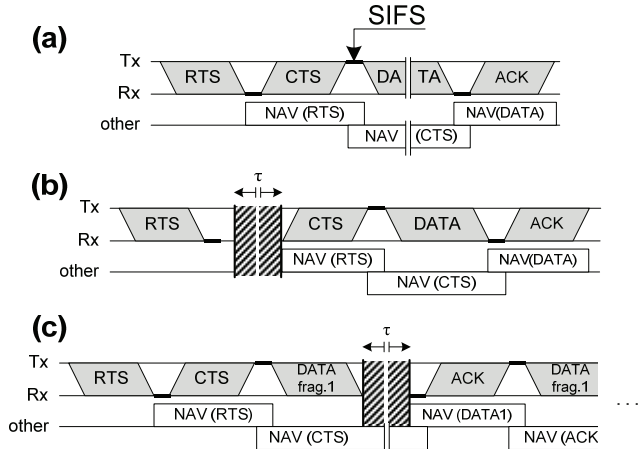

Fig. 7. Three examples of the Virtual Carrier Sensing Mechanism in the iDCF. (a) No scan procedure Interruption. (b) Interruption at *cts* packet transmission. (c) Interruption at *data* transmission forcing data fragmentation.

Packet Type	Multiple MPDU case		Single MPDU case	
	More Fragment	DurationID	More Fragment	Duration ID
rts	-	L_P	-	L_P
cts	-	L_P	-	L_P
data	1	L_{PR}	0	0
ack	-	L_{PR}	-	0

Table 1. Duration ID setting for each packet type and multiple MPDU case.

5. iDCF Analytical Model

A plethora of Markov chain models for the DCF performance analysis can be found in the literature since the first one was proposed in (Bianchi, 2000). Markovian models capture DCF features with good accuracy, which is proved by extensive simulations. The diversity of these models comes from different assumptions regarding the source traffic models, the wireless channel modelling and the levels of DCF details integrated into the Markov Chain. For example different models have been implemented using various fading channel models and ideal channel conditions, varying hidden terminal assumptions, saturated traffic conditions and various non saturation traffic models, finite and infinite number of packet retransmission attempts. Moreover a lot of Markovian models explore the DCF behaviour under different back-off algorithms and concern the optimal selection of the algorithm.

The common thing among these models analysis is that embedded points of the Markov chains are epochs where the backoff counter decrements. Between backoff counter decrements either a successful transmission, or a colliding transmission, or an idle protocol slot may take place. The Markov chain is used for calculation of certain probabilities such as the probability of successful transmission, colliding transmission and idleness in a randomly chosen slot. Then, using these probabilities and the slot durations, as calculated from the utilized DCF access scheme, throughput, packet delay and other system metrics equations are derived.

In the standard DCF analysis the idle slot, successful slot and collision slot durations are deterministic values σ, T_S and T_C respectively. In the iDCF, on the contrary, they are random variables t_{Idle}, t_S and t_C, depending on the position of these slots in time relatively to the scan procedure interruptions and may include several scan procedures. Our analytical model (Adamis et al., 2009) concerns the slots duration expectations calculation. It is used to extend the existing Markovian models in order to obtain performance metrics for the iDCF operation in the intermittent overlay access mode.

5.1 Successful and Collision Slot Duration

In the following equations a packet transmission time is denoted by capital letters. In order to calculate the expected duration of a successful FES, first we calculate the duration expectations of each FES part. Then:

$$E[t_S] = RTS + \delta + 3 \cdot E[t_{SIFS}] + 2 \cdot E[t_{RSP}] + E[t_{DATA}] + E[t_{DIFS}] \tag{1}$$

The rts packet transmission time is used directly without an expectation operator. The modification of the backoff procedure described earlier in Sect. 4.2 ensures that whenever an rts packet is to be transmitted there will always be enough time left before the upcoming scan interruption. Similarly the collision slot duration expectation is computed as:

$$E[t_C] = RTS + \delta + E[t_{DIFS}] \tag{2}$$

Our general key assumption in the following calculations is that every FES part is interrupted by scan procedures at any point with probability proportional to its length. This assumption leads to very accurate results. The IFS durations remain unchanged when the IFS is not interrupted by a scan procedure. When they are interrupted they expand to include the scan procedure as presented in Sect. 4.3. Since an IFS duration is mostly occupied by MAC and PHY processing times, which are suspended during a scan procedure, our model uses the following approximation:

$$E[t_{IFS}] = \frac{T_{IFS}}{T - \tau}(T_{IFS} + \tau) + \left(1 - \frac{T_{IFS}}{T - \tau}\right)T_{IFS} \ , \tag{3}$$

where the IFS is either SIFS or DIFS. As shown by the simulations this approximation has a minor effect to the analytical results.

The rsp packets - response packets cts and ack - expected duration can be found similarly by noting that when the upcoming scan interruption does not leave enough *useful time* for the

rsp transmission, the packet is sent directly after the end of the scan. In this case the remaining time till the beginning of the scan procedure is wasted. The wasted time is uniformly distributed in $(0, RSP+\delta)$. This yields:

$$E[t_{RSP}] = \frac{RSP + \delta}{T - \tau}\left(\frac{RSP + \delta}{2} + \tau + RSP + \delta\right) + \left(1 - \frac{RSP + \delta}{T - \tau}\right)(RSP + \delta) \tag{4}$$

The current analysis on the data packet expected transmission duration $E[t_{DATA}]$ is performed for a constant data payload size L_P. As presented in Sect. 4.1, when the data packet transmission is ready to start, either the whole packet is sent if there is enough remaining time, or fragmentation takes place if the fragment to be transmitted contains at least L_{min} payload bits. Each fragment is followed by a positive ack as the standard protocol dictates. The total sequence of data fragments and their associated ack packets, excluding the last ack or the unique one (in case a unique data packet is sent), is referred to as Data Fragment Sequence - DFS. The DFS may take different forms in time, depending on the *useful time* length relatively to the length of certain FES parts. Let x be the number of data bits sent at the first fragment of the transmission. $x \in \left[L_{min}, L_{ut}\right]$ where:

$$L_{ut} = \min\left(L_P, \left\lfloor (T - \tau - \delta - t_{oh}) \cdot R \right\rfloor - L_H\right) \tag{5}$$

is the number of data bits that can fit in a fragment, which occupies the entire *useful time*. In (5), if R is the transmission rate, an L bit packet transmission time is calculated as $t_{oh}+L/R$, where t_{oh} is the PHY layer time overhead, e.g. OFDM preamble. The bit overhead that comes from PHY and MAC layer headers is denoted as L_H. The DFS duration, denoted as $t_{DFS}(x)$ is a function of x and thus of the data transmission start position in time, relatively to the scan procedure. Moreover $t_{DFS}(x)$ depends on the *useful time* length relatively to the SIFS and ack transmission durations. Fig. 8 shows the possible DFS formulation cases and subcases in time.

We make the following observations/assumptions in these cases:

- All data fragments, except the last one, occupy the entire remaining *useful time*
- Which case, I, II or III, specifies the DFS form, depends on the SIFS and RSP duration. Which subcase, a or b, specifies the DFS form, depends on the transmission time of the minimum data fragment, $DATA_{min}$, which is controlled by the L_{min} parameter
- Decrease of the *useful time*, causes the transition of the DFS formulation from case I to III. Increase of L_{min} within a certain case, causes the transition of the DFS formulation from subcases a to b.

The inequalities that designate which form the DFS takes appear in Table 2. Except the first fragment, the entire data packet takes up $n_F(x)$ more successive fragments, containing L_{PF} data bits each. We denote as t_{av} the time within the useful time which is available for every data fragment, except the first one. Excluding the PHY time overhead and the anticipated propagation delay we obtain:

$$t_{av} = T - \tau - \delta - t_{oh} - \lambda \tag{6}$$

$$L_{PF} = \lfloor t_{av} \cdot R \rfloor - L_H \tag{7}$$

$$n_F(x) = \left\lceil \frac{L_P - x}{L_{PF}} \right\rceil \tag{8}$$
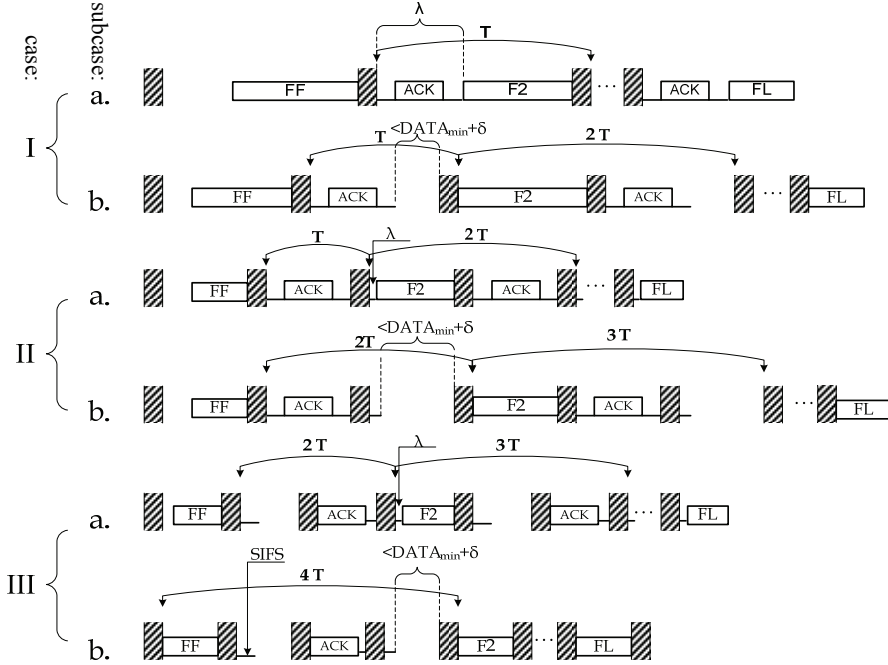
Fig. 8. DFS formulation in time. FF: first data fragment, Fn: nth data fragment, FL: last data fragment

Where λ is the part of the useful time before the fragment transmission occupied by an ack and/or SIFS and its value is given in Table 2 for every case/subcase. The DFS duration resulted from Fig. 8 is given by:

$$t_{DFS}(x) = \begin{cases} DATA_{FF}(x) + (n_T - 1)T + n_T T[n_F(x) - 1] + \tau + \lambda + DATA_{FL}(x), x < L_{FF} \\ DATA + \delta, x = L_P \end{cases} \tag{9}$$

n_T is the number of *scan period*s T needed for the transmission of one data fragment with the corresponding ack for cases Ib to IIIb. For case Ia, n_T is the number of scan periods needed for the transmission of a single data fragment and the ack of the previous fragment. The n_T value for every case is also given in Table 2. Note that the ack packet of the last (or the unique) data fragment is not included in (9) because it is considered separately in (1) as an $E[t_{RSP}]$, which is given by (4). $DATA_{FF}(x)$, $DATA_{FL}(x)$ are the transmission times for the first and last fragment containing $L_{FF}(x)$ and $L_{FL}(x)$ bits respectively, that are given by (10) and (11).

case	Inequality Satisfied	subcase	Inequality Satisfied	n_T	λ	Comments
I	$T-\tau \geq$ RSP+δ+2SIFS	A	$T-\tau \geq$ RSP+DATA$_{min}$ +2δ+2SIFS	1	RSP+δ+ 2SIFS	The ack and the subsequent data fragment transmission can fit in the same useful time
		b	$T-\tau <$ RSP+DATA$_{min}$ +2δ+2SIFS	2	0	There is not enough remaining time after the ack transmission and the SIFS for the data fragment transmission. It takes place at the beginning of the following useful time
I	RSP+δ+SIFS $\leq$ T-τ < RSP+δ+ 2SIFS	a	$2(T-\tau) \geq$ SP+DATA$_{min}$ +2δ+2SIFS	2	SIFS - (T-τ-RSP -δ-SIFS)	The SIFS that comes after the ack transmission is completed within the following useful time. After its completion, adequate time remains for a data fragment transmission
		b	$2(T-\tau)<$RSP +DATA$_{min}$+2δ+2SIFS	3	0	After the completion of the SIFS, that follows the ACK, there is not enough remaining time for the data fragment. The transmission will start at the beginning of the following useful time.
I	RSP+δ $\leq$ T-τ < RSP+δ+SIFS	a	$2(T-\tau) \geq$RSP+ DATA$_{min}$ +2δ+SIFS	3	SIFS - (T-τ-RSP -δ)	The SIFS, which comes after the data fragment transmission, does not leave enough time for the ack, which is transmitted at the upcoming useful time. The next data fragment can fit in the useful time after the one of the ack and after the SIFS completion.
		b	$2(T-\tau)<$RSP+ DATA$_{min}$ +2δ+SIFS	4	0	The ack is transmitted again two useful times after the data fragment. Upon completion of the SIFS that follows the ack, there is not enough time remaining for the data fragment, which is transmitted two useful times after the ack

Table 2. Case/Subcase inequalities, λ and n_T values.

$$L_{FL}(x) = L_H + x \tag{10}$$

$$L_{FL}(x) - L_H + L_P - x - \left[n_F(x) - 1\right]L_{PF} \tag{11}$$

The probabilities that a data packet will be interrupted a) at a point were no fragment can be sent P_{DIDm} b) at a later point were fragmentation can take place P_{DIDr} and c) nowhere P_{DNI}, are:

$$P_{DIDm} = \frac{(DATA_{\min} + \delta)}{T - \tau} \tag{12}$$

$$P_{DIDr} = \frac{\min(T - \tau, DATA + \delta) - (DATA_{\min} + \delta)}{T - \tau} \tag{13}$$

$$P_{DNI} = \frac{\max(T - \tau, DATA + \delta) - (DATA + \delta)}{T - \tau} \tag{14}$$

respectively. $DATA_{min}$ is the transmission time of the minimum permissible data fragment. When $DATA + \delta > T - \tau$ the last probability is zero and the rest two probabilities become complementary. Using (9), (12), (13) and (14) the expectation $E[t_{DATA}]$ can be expressed as:

$$E[t_{DATA}] = P_{DNI} \cdot (DATA + \delta) + P_{DIDr} \cdot E[t_{DFS}(x)] + P_{DIDm} \cdot \left[\frac{DATA_{\min} + \delta}{2} + \delta + t_{DFS}(x = L_{ut})\right] \tag{15}$$

From the general key assumption made earlier, x follows the uniform distribution. $E[t_{DFS}(x)]$ can be easily found by first arithmetically computing $E[n_F(x)]$ as:

$$E[n_F(x)] = \frac{1}{L_{ut} - L_{\min} + 1} \sum_{x = L_{\min}}^{L_{ut}} \left\lceil \frac{L_P - x}{L_{PF}} \right\rceil \tag{16}$$

The successful slot duration expectation can now be calculated from (1).

5.2 Idle Slot Duration

The expected duration of an idle slot $E[t_{Idle}]$ can be found by observing Fig. 3b. It depends on its position in time as related to the scan procedures interruptions, similar to the previous analysis for the FES parts. For example the slots having index 1, 2, 3,... have duration σ while the index 0 has duration $v_\sigma \sigma + \tau$. v_σ is the number of protocol slots that lay entirely after the t_{lTxOp} time point. Idle slots begin after DIFS periods and subsequently their duration depends on the relative DIFS positions. Let I be the total number of idle slots after a certain DIFS period, before a transmission starts. If W_N is the maximum contention window size then $I \in [1, W_N - 1]$. By calculating the integral over all possible DIFS positions y, $y \in [0, T - \tau]$, the expected idle slot duration given a certain I can be obtained as:

$$E[t_{Idle} \mid I = i] = \frac{1}{T - \tau} \int_0^{T - \tau} \frac{1}{i} \sum_{k=1}^{i} \left[T_{k^{th} slot}\right] dy \tag{17}$$

Where $T_{k^{th}slot}$ is the duration of the kth protocol idle slot after the DIFS period and the integrated quantity is the expected idle slot duration for a unique y DIFS position. The sum quantity within the integrated quantity in (17) is a function $F(y,i)$ of both the discrete variable i and the continuous variable y. In the standard DCF it would be a function of just the variable i as $F(i)=i\sigma$. By observing the duration of the idle slots for different DIFS positions we obtain the multi-branch function:

$$\sum_{k=1}^{i}\left[T_{k^{th}slot}\right]=F(y,i)=\begin{cases}i\sigma & \begin{cases}0<y\le T-\tau-(RTS+\delta)-(DIFS+i\sigma)\\ \quad or\ T-\tau-DIFS<y\le T-\tau\end{cases}\\ (i-1)\sigma+(v_\sigma\sigma+\tau) & \begin{cases}y\le T-\tau-(RTS+\delta)-DIFS\\ and\ y>T-\tau-(RTS+\delta)-(DIFS+i\sigma)\end{cases}\\ (i-1)\sigma+T-(y+DIFS) & \begin{cases}y\le T-\tau-DIFS\\ and\ y>T-\tau-(RTS+\delta)-DIFS\end{cases}\end{cases} \quad (18)$$

By breaking the integral into each branch and calculating the summations we finally obtain:

$$E\left[t_{Idle}\mid I=i\right]=\sigma+\sigma\frac{\tau+(v_\sigma-1)\sigma}{T-\tau}+\frac{(RTS+\delta)\left[\tau-\sigma+\frac{1}{2}(RTS+\delta)\right]}{i(T-\tau)} \quad (19)$$

For large *useful time*s the above quantity tends to σ, the value that is valid for the standard DCF, too. Finally, using the I pdf $P[I=i]=P_{idle}{}^{i}(1-P_{idle})$, the idle slot duration expectation can be calculated by (20):

$$E\left[t_{Idle}\right]=\sum_{i=1}^{W_N-1}P[I=i]\cdot E\left[t_{Idle}\mid I=i\right] \quad (20)$$

P_{idle} is the probability that a randomly chosen slot is idle and is calculated using the Markov chain extracted probabilities.

5.3 An example of application to a specific Markov Model

The analytical model for the calculation of the slots duration expectations is applied to an existing DCF Markovian model in order to obtain performance metrics for the iDCF. The existing Markovian model used was presented in (Kim et al., 2008). This model includes in great detail the DCF functionalities described in the 802.11 standard. Its main difference from other models is that it incorporates the immediate transmission of a packet, with no backoff procedure, when the packet arrives at an empty-queue station and the medium is detected idle for a DIFS period. The non-saturation traffic model, it uses, is described by packet flows with geometrical distribution of the number of packets in a flow, with mean $1/(1-\varphi)$. Idle period durations between flows are exponentially distributed with mean $1/\lambda_I$. The assumption of an ideal channel, which is also used in most of the proposed Markov models, is useful to isolate and study the effect of the intermittent operation on the iDCF performance.

In the Markov model (Kim et al., 2008) equations the slot durations T_S, T_C and T_{Idle} are replaced by the expectations $E[t_S]$, $E[t_C]$ and $E[t_{Idle}]$ respectively, calculated in the previous Sections. For example the probabilities P_a and P_b of new flow arrivals at a station, which has no packet to transmit – empty station, during an idle slot and a busy slot, respectively, are replaced by:

$$P_a = P_0 \cdot \left(1 - e^{-\lambda_l \cdot E[t_{Idle}]}\right) \tag{21}$$

$$P_b = P_1 \cdot \left(1 - e^{-\lambda_l \cdot E[t_S]}\right) + P* \cdot \left(1 - e^{-\lambda_l \cdot E[t_C]}\right) \tag{22}$$

where P_0, P_1 and $P*$ are the probabilities of a randomly chosen slot to include idleness, successful transmission and collision, as seen by an empty station and are given in (Kim et al., 2008). In this way the probabilities P_S of a successful transmission, P_C of a collision and P_{idle} of an idle slot occurrence are calculated according to the common method described in the DCF Markov Chain Models. Only that, in this context, their values depend not only on the engaged back-off algorithm, the number of contending terminals and the traffic model used. They also depend on the overlay intermittent mode parameters such as T and τ, which are silently incorporated in the Markov chain through the $E[T_S]$, $E[T_C]$ and $E[T_{Idle}]$ quantities. Throughput and mean ordinary packet delay can now be derived as:

$$S = \frac{P_S L_P}{P_{idle} E[t_{Idle}] + P_S E[t_S] + P_C E[t_C]} \tag{23}$$

$$
\begin{aligned}
E[Y] = & \sum_{i=0}^{N} p^i (1-p) \left(\sum_{m=0}^{i} E[Y_m] + i E[t_C] + E[t_S] \right) + \\
& \sum_{i=N+1}^{M} p^i (1-p) \left(\sum_{m=0}^{N} E[Y_m] + + (i-N) E[Y_N] + i E[t_C] + E[t_S] \right) + \\
& \left(1 - \sum_{i=0}^{M} p^i (1-p) \right) \left(\sum_{m=0}^{N} E[Y_m] + (M-N) E[Y_N] + (M+1) E[t_C] \right)
\end{aligned}
\tag{24}
$$

In (24) p is the conditional collision probability, N is the maximum back-off stage, M is the packet retransmission limit and Y_m is the delay of staying in the mth back-off stage. If W is the initial contention window size, $(2^m W$-$1)$ is the contention window size at the mth back-off stage and $E[Y_m]$ is given by:

$$E[Y_m] = \frac{2^m W - 1}{2} \left(P_{idle} E[t_{Idle}] + P_S E[t_S] + P_C E[t_C] \right) \tag{25}$$

Other performance metrics such as packet delay for the first packet of a flow, expected time to complete the transmission of a flow and packet loss probability presented in (Kim et al., 2008) can be derived accordingly.

5.4 Evaluation and Numerical Examples.

Numerical results obtained by the application of our model to the Kim et al. Markov chain, as described earlier, were validated with simulations. An OPNET discrete event network simulator was built, which models in detail all the DCF functional details and incorporates the iDCF modifications described in section 4. The numerical values used in both the analytical and the simulation models are presented in Table 3. PHY transmission rate, bit overhead, time overhead and guard interval correspond to 802.11a PHY parameters, for subcarrier spacing equal to 100 kHz and QPSK modulation per subcarrier. The required transmission rate in the simulator is controlled by the number of data subcarriers.

	Parameter	Value
MAC layer parameters	(N, M, W)	(6, 11, 16)
	(SIFS, DIFS, σ, δ)	(16, 34, 9, 1/3) µs
	(rsp, rts, data header)	(112, 160, 272) bit
PHY	t_{oh}	62.5 µs
	PHY bit overhead	22 bit
	R	var
Overlay Scan Procedure	T, τ	var, 200µs
Source Traffic	$n, \varphi, \lambda, L_P, L_{min}$	30,0.95,2,var, var

Table 3. Analytical & Simulation Model parameter set up. var: varying parameter.

In the following figures the analytical results, represented by the continuous line plots, closely follow the simulated results represented by discrete plots. By keeping the scan period T constant and having the scan delay τ varying a linear relationship between the *scan delay* and the saturation throughput is obtained which results from the linear relationship between τ and the *useful time*. The relation, though, between the throughput and the *scan period* is non-linear, if the *scan delay* is kept constant. The more frequent the scan procedures are performed, the stronger the effects of the intermittent operation.

In the following figures τ is set to 200µs and is kept constant. This choice of τ leaves enough *useful time* for packet transmissions in the entire examined T range. Fig. 9 presents the throughput versus the *scan period* for three different constant packet sizes. It is shown that the throughput of the iDCF strongly depends on the *scan period* that the system operates with. It heavily deteriorates for very frequent scan procedures, while as T gets larger values the throughput is approaching its upper bound specified by the standard DCF operation. For large values of the scan period, the effect of the intermittent operational mode on iDCF performance is negligible. Because of the DFS formulation, presented in Fig. 8, while T grows and the DFS switch over the cases III to I, the throughput regresses before it takes the final course towards higher values. This shows that in certain cases it is preferable not to transmit a data fragment before the upcoming *scan procedure,* but to transmit a longer fragment after the scan procedure. This decision is controlled via the L_{min} parameter.

Fig. 10 presents the obtained throughput as a function of the transmission rate R for various T values and for the standard DCF. Normal DCF's weakness in exploiting high transmission rates is depicted as the throughput decreases for higher values of R. This effect tends to disappear for extremely frequent scan interruptions in the iDCF. The obtained throughput versus the payload size L_P of the data packet is presented in Fig. 11. This figure shows that

although in the standard DCF the throughput increases for larger packet sizes, in iDCF this increase is not very notable as the *scan period* is getting shorter.
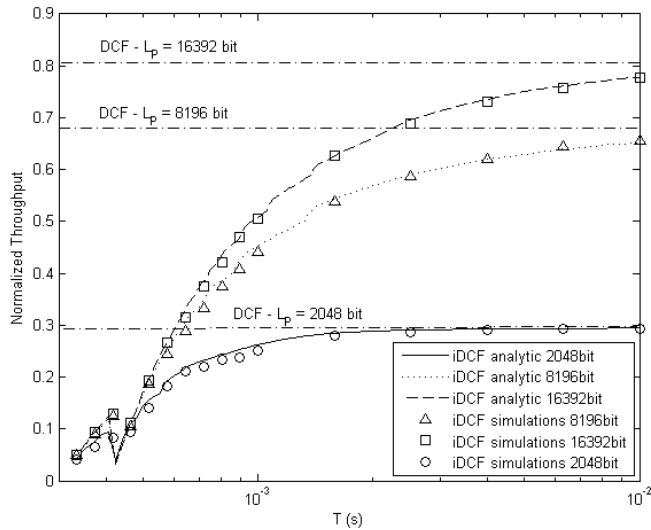

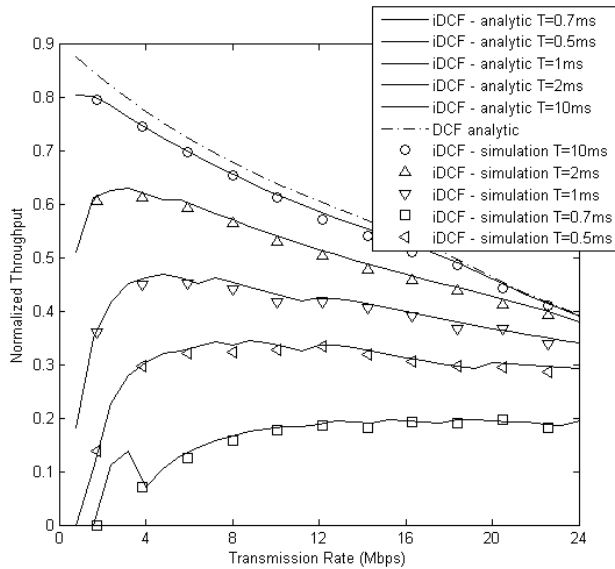Fig. 9. Throughput versus scan period. R=8Mbps, various L_P.

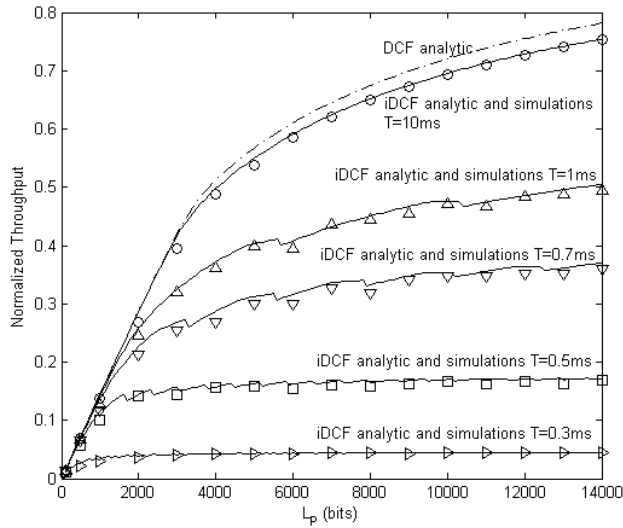Fig. 10. Throughput versus transmission rate. R=8Mbps, various T

Fig. 11. Throughput versus data Packet payload. R=8Mbps, various *T*

6. Conclusions and Further Work

In this chapter we presented the intermittent DCF as the MAC protocol for OA Networks. Furthermore an analytical method for calculating the expected performance metrics of iDCF was introduced. This model can be used in feasibility studies of realistic OA environments to provide realistic wireless access solutions and in Admission Control Algorithms of QoS enabled OA networks.

7. References

Adamis, V. & Constantinou, P. (2007). Performance Study of CSMA/CA Over Spectrum Pooling Environment for Cognitive Radios. *Proceedings of the 3rd IEEE International Conference on Wireless and Mobile Computing, Networking and Communications*, Oct. 2007, New York

Adamis, T., V.; Maliatsos, K., N. & Constantinou, P. (2008). Methods for Reducing Interference caused to Licensed Systems by Overlay-CSMA/CA Cognitive Radios, *Proceedings of the 3rd International Conference on Cognitive Radio Oriented Wireless Networks and Communications*, May 2008, Singapore

Adamis, T., V.; Maliatsos, K., N. & Constantinou, P. (2009). Throughput Analysis of the Intermittent DCF for Opportunistic Spectrum Access, Accepted, *Springer International Journal on Wireless Personal Communications (An International Journal on)*, Jul. 2009

Berthold, U.; Heimpel, H. & Jondral, F., K. (2007a). Coordination of Allocation Measurements in OFDM Based Ad Hoc Overlay Systems. *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, Apr. 2007, Honolulu

Berthold, U.; Jondral, F., K.; Brandes, S. & Schnell, M. (2007b). OFDM-Based Overlay
        Systems: A Promising Approach for Enhancing Spectral Efficiency. *IEEE
        Communications Magazine*, Vol. 45, No. 12, Dec. 2007, 52-58, 0163-6804

Bianchi, G. (2000). Performance Analysis of the IEEE 802.11 Distributed Coordination
        Function. *IEEE Journal on Selected Areas in Communications*, Vol. 8, No. 3, Mar. 2000,
        535-547, 0733-8716

Cordeiro, C; Challapali, K.; Birru, D. & Sai Shankar, N. (2005). IEEE 802.22: the first
        worldwide wireless standard based on cognitive radios. *Proceedings of the First IEEE
        International Symposium on New Frontiers in Dynamic Spectrum Access Networks*, Nov.
        2005, Baltimore

Ghasemi, A. & Sousa, E., S. (2008). Spectrum sensing in cognitive radio networks:
        requirements, challenges and design trade-offs. *IEEE Communications Magazine*,
        Vol. 46, No. 4, Apr. 2008, 32 - 39, 0163-6804

Hang S. & Xi Z. (2008). Cross-Layer Based Opportunistic MAC Protocols for QoS
        Provisionings Over Cognitive Radio Wireless Networks, *IEEE Journal on Selected
        Areas in Communications*, vol. 26, no.1, Jan. 2008, 118 – 129, 0733-8716

IEEE Standards. 1999. 802.11, Part 11: Wireless LAN Medium Access Control (MAC) and
        Physical Layer (PHY) Specifications, Nov. 1999, IEEE

Juncheng, J.; Qian, Z. & Xuemin, S. (2008). HC-MAC: A hardware-constrained cognitive
        MAC for efficient spectrum management, *IEEE Journal on Selected Areas in
        Communications*, vol. 26, no.1, Jan. 2008, 106-117, 0733-8716

Kim, T., O..;Kim, K., J. & Choi, B., D. (2008). Performance Analysis of IEEE 802.11 DCF and
        IEEE 802.11 EDCA in Non-Staturation Condition, *IEICE Transactions of
        Communications*, vol. E91-B, no.4, Apr. 2008, 1122-1131, 1745-1345

Motamedi, A. & Bahai, A. (2007). MAC Protocol Design for Spectrum-agile Wireless
        Networks: Stochastic Control Approach, *Proceedings of the Second IEEE International
        Symposium on New Frontiers in Dynamic Spectrum Access Networks*, Apr. 2007, Dublin

Papadimitratos, P.; Sankaranarayanan, S. & Mishra, A. (2005). A Bandwidth Sharing
        Approach to Improve Licensed Spectrum Utilization. *IEEE Communications
        Magazine*, Vol. 43, No. 12, Dec. 2005, supl.10 - supl.14, 0163-6804

Shu, J.; Yang, X. & Guo, X. (2007) Disruptive CSMA with credit payback (CP) protocols for
        multi radio network. *Proceedings of the 2nd International Conference on Cognitive Radio
        Oriented Wireless Networks and Communications*, Jul. 2007, Orlando

Stirling, A. (2008). White space coalition: The story so far. *Proceedings of The IET seminar on
        Cognitive Radio And Software Defined Radio: Technologies and Techniques*, Sep. 2008,
        London

Weiss, T., A.; & Jondral, F., K. (2004). Spectrum Pooling: An Innovative Strategy for the
        Enhancement of Spectrum Efficiency, *IEEE Communications Magazine*, vol.42, no.3,
        Mar. 2004, S8-14, 0163-6804

Willkomm, D.; Machiraju, S.; Bolot, J. & Wolisz, A. Primary users in cellular networks: A
        large-scale measurement study. *Proceedings of the Third IEEE International
        Symposium on New Frontiers in Dynamic Spectrum Access Networks*, Oct. 2008,
        Chicago

Yuan, Y.; Bahl, P.; Chandra, R.; Chou, P., A.; Ferrell, J., I.; Moscibroda, T.; Narlanka, S. &
        Yunnan W. (2007). KNOWS: Cognitive Radio Networks Over White Spaces.
        Proceedings of the Second IEEE DySPAN, Apr. 2007, Dublin

Adaptive Subband Selection in OFDM-BasedCognitive Radios

Pingzhou Tu, Xiaojing Huang
University of Wollongong, Australia

Eryk Dutkiewicz
Macquarie University, Australia

1. Introduction

The cognitive radio built on software-defined radio enables intelligent wireless communication systems (S. Haykin, 2005). Orthogonal frequency division multiplexing (OFDM) as an promising cognitive radio modulation technique provides great flexibility and adaptability in selecting system transmission bandwidth for interference avoidance among coexistent multiple wireless systems in a shared radio environment (G. Andrea, 2006).

An important issue in achieving system coexistence is to deal with the interference generated by multiple wireless systems. When they are deployed in the proximity of each other, share the same frequency band, and operate at the same time, they are more likely to interfere with each other.

Many parameters affect the inteference between different wireless systems, such as power levels, modulation types, amount of data traffic, and separation between antenanas. Medium access control (MAC) layer protocol can also play a very important role in the interfernece avoidance and affect the overal system performance.

Some other multiple access methods such as TDMA, FDMA, and CDMA are effective in solving the coexitence problem among multiple wireless devices in the same networks respectively, but how different wireless systems can coexist in the same frequency band is only considered resently. The first standards-making body to consider the multiple wireless system coexistence problem is the IEEE 802.15 Working Group. IEEE 802.15.2 was formed to develop a recommended practice for coexistence between Bluetooth/IEEE 802.15.1 and IEEE 802.11, in which two types of coexistence mechanisms, collaborative and noncollaborative ones, are developed. With collaborative mechanism, wireless personal area networks and wireless local area networks exchange information over a separate link between one another to minimize mutual interference. In noncollaborative systems, there are several methods used for interference suppression. However, these methods are only implementations of the **distributed coordination function (DCF)** channel access mechanism for IEEE 802.11 to suppress interference (T. Cooklev, 2004).

In this chapter, a novel adaptive subband selection method combined with an interleaved spread sepctrum OFDM (ISS-OFDM) technique is proposed to improve system coexistence in a shared spectrum environment such as the industrial, scientific, and medical (ISM) band. It uses noncollorative mechanism and is a physical layer technique for system coexistence. The system performance such as the transmitted signal peak to average power ratios (PAR) (S. B. Slimane, 2002) , and the system bit error rates (BER) (J. G. Proakis, 2000) before and after subband selection in different channel conditions are analyzed and simulated. The results show that the ISS-OFDM technique with adaptive subband selection can significantly improve the system performance when multiple systems coexist in a shared radio channel.

The contributions of this chapter are three-fold:

a. A novel complex exponential spreading (CES) method is developed to generate a spreading spectrum OFDM signal with multiple subbands.
b. The demodulation by using multiple fast Fourier transform (FFT) parallel operations is proposed to recover the transmitted information, which greatly reduces the receiver computational complexity and makes full use of signal diversity.
c. An adaptive subband selection transmission algorithm is proposed. Based on dynamically determined interference thresholds, system transmission band is adaptively selected to avoid interference among OFDM systems and other narrowband or ultra-wideband wireless systems.

This chapter consists of the following sections. Section 2 describes the proposed cognitive radio system model, and presents the ISS-OFDM technique by which the cognitive radio transmitter can generate signal with multiple subbands. The signal subbands used for system transmission can be decided based on channel sensing. Section 3 proposes the adaptive subband selection method and derives the interference power thresholds. Based on these thresholds, the cognitive radio receiver can decide if a subband should be selected or discarded in order to achieve the best system performance. The rule is that any subband with interference power level above an adaptively determined threshold will be discarded and the signals from weak or zero interference subbands will be decoded and combined to improve the signal-to-noise ratio. Section 4 provides simulation results on system performance, such as PAR and BER before and after subband selection in different channel conditions. Finally, Section 5 draws conclusions.

2. Cognitive Radio System Model

In this section, we describe the proposed baseband system model of the OFDM-based cognitive radios as displayed in Figure 1. Then, the generation and reception techniques of the ISS-OFDM signals are discussed. We assume that the data bit to symbol mapping scheme on the ISS-OFDM baseband system model is QPSK, the channel is frequency selective with slow fading, and the noise is AWGN with zero mean.

2.1 Transmitter Architecture

The function of the transmitter is to generate an ISS-OFDM signal with time and frequency diversities. At the transmitter, the complex QPSK data symbols to be transmitted are first divided into an N by 1 vector, where N is the number of subcarriers. This vector modulate N subcarries in parallel by using a spread spectrum modulation technique, called complex

exponential spreading (CES). Each data symbol modulates its corresponding subcarrier multiple times, and several different samples bearing the same data information can be generated. The replicas of the same data information are interleaved and then transmitted in different time slots, instead of adding the modulated samples together. After the spread spectrum modulation, an ISS-OFDM symbol is generated.
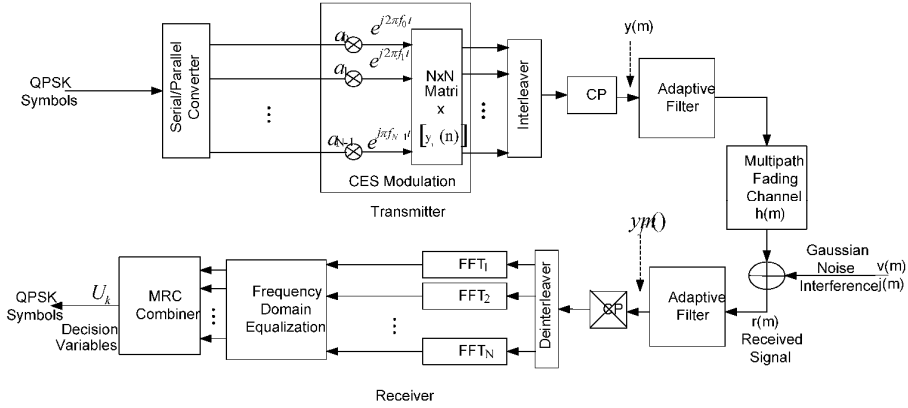

Fig. 1. Baseband system model.

2.1.1 Complex Exponential Spreading (CES) Modulator

For convenience, we will only describe the modulation process for one OFDM symbol. After serial to parallel conversion, each of the QPSK symbols modulates the corresponding subcarrier of the N subcarriers. If the ISS-OFDM symbol period is T_s, $f_i = i/T_s$ denotes the ith subcarrier frequency of the N orthogonal subcarriers, and a_i modulates the ith subcarrier at time $t = \frac{n}{N}T_s, n=0,1,...,N1$, where n is the sampling interval, the modulated symbol on the ith subcarrier and the nth time index is written as:

$$y_i(n) = a_i e^{j2\pi f_i t} = a_i e^{j2\pi ni/N} \ .$$

(1)

In ISS-OFDM symbol duration T_s, each element of the $N \times 1$ data symbol vector modulates the same corresponding subcarrier N times. N elements in the vector generate $N \times N$ matrix samples after modulation.

2.1.2 Interleaving

After the modulation, the $N \times N$ matrix samples are interleaved. Interleaving is used to obtain time diversity in a digital communications system without adding much overhead. The function of the interleaver is to spread the OFDM modulated signal samples out in time so that if there is deep fading or a noise burst, the signal samples from the modulated signal are not corrupted at the same time.

There are many forms of interleaving techniques to implement the permutation of the signal samples in the matrix of the CES modulator output. A periodical interleaver is the simplest method to permute the transmission sequence. For the simplicity of description, we adopt the periodic interleaving algorithm. For example, for the $I \times N$ matrix for one OFDM symbol, the interleaver writes the modulated samples for each of the same subcarrier row by row to form the $I \times N$ matrix and reads the samples from the matrix column by column and places the samples on a time axis, which is displayed in Figure 2.

Mathematically, if the $I \times N$ matrix is denoted as $[y_i(n)]$, $i = 0, 1, -, I$ -1, $n = 0, 1, -, N$ -1 , and I, N are integers. Let the period of the interleaving be N, after periodically interleaving, the interleaved serial sequence can be written as follows:

$$y(m) = y(nN + i) = y_i(n) = a_i e^{j2\pi f_i t} = a_i e^{j2\pi ni/N} \tag{2}$$

where $i = 0, 1, -, I$ -1, $n = 0, 1, -, N$ -1.

2.1.3 Generation of Multiple Subband OFDM Signal

In the following, we discuss interleaved results by using the periodic interleaving alogrithm. We implement the periodic interleaving by shifting N modulated subcarriers on a different time slots and adding them together. For instance, the ith subcarrier of the N subcarriers is shifted by $i \cdot \tau$, where τ is the sampling interval, $\tau = N$. Then, the N shifted subcarriers are added together to form one ISS-OFDM symbol with N^2 samples. The mth sample in the ISS-OFDM symbol, denoted by y(m) , m $= nN + i = 0,1, -, N^2$ -1 ., can be mathematically written as follows:

$$y(m) = \sum_{i=1}^{N} \sum_{n}^{N} y_i(n)\delta[m - i - n\,N],$$

where $\delta[m - nN - i] = \begin{cases} 1, & \text{for } m=nN+i \\ 0, & \text{for } m\neq nN+i \end{cases}$ is the unit impulse. If periodical interleaving is

performed, taking out y(m) for each $m = nN + i$, y (m) can be simplified in the form:

$y(m) = y(nN + i) = y_i(n), m = nN + i,$ and $n, i = 0, 1, ..., N - 1.$ Thus, N^2 samples are produced in the same symbol period T , and the number of samples is spread N times. The sampling rate is N^2 / T . an ISS-OFDM symbol is generated. Figure 3 shows the signal modulation and periodical interleaving process for generating an ISS-OFDM symbol in the case of subcarriers number $N = 4$. It is seen that the modulated samples on the ith ($i = 0, 1, 2, 3$) subcarrier are shifted i time intervals, and then they are added on the time axis, and the number of samples increases to N^2 from the original number N . In the frequency domain the ISS-OFDM signal can be written as:

$$y(k) = FFT(y(m)) = FFT(y(nN + i))$$

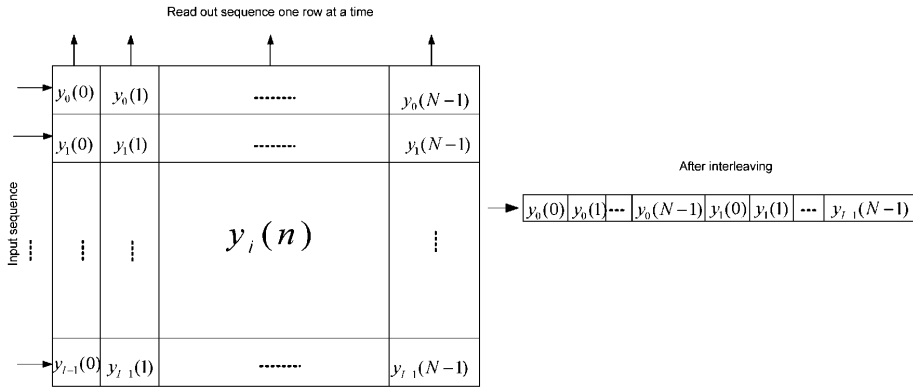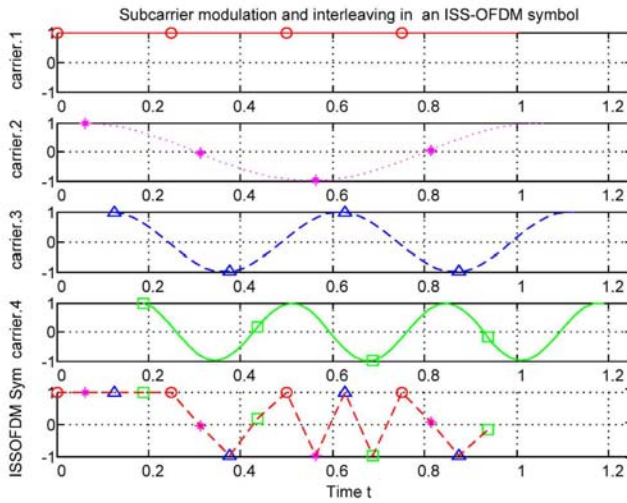$$= \sum_{nN+i=0}^{N^2 - 1} y(nN + i)e^{-j\frac{2\pi}{N^2}(nN+i)k} .$$

Fig. 2. Periodical interleaving.

Fig. 3. Modulation and periodical interleaving process with subcarrier number $N = 4$.

That is,

$$y(k) = \sum_{n=0}^{N-1}\sum_{i=0}^{N-1} y(nN+i) e^{-j\frac{2\pi}{N^2}(nN+i)k}.$$ (3)

Substituting $y(nN+i)$ by $y_i(n)$ from Equation (2) into Equation (4), $Y(k)$ has the form as:

$$Y(k) = \sum_{n=0}^{N-1} \sum_{i=0}^{N-1} y_i(n) e^{-j\frac{2\pi}{N^2}(nN+i)k}$$

$$= N \sum_{i=0}^{N-1} a_i e^{-j\frac{2\pi}{N^2}ki} \underbrace{\frac{1}{N} \sum_{n=0}^{N-1} e^{-j\frac{2\pi}{N}(k-i)n}}_{\delta((k-i)_N)} .$$

(4)

Replacing $\dfrac{1}{N} \sum\limits_{n=0}^{N-1} e^{-j\frac{2\pi}{N}(k-i)n}$ by $\delta((k-i)_N)$ in Equation (5), we have the equation as:

$$Y(k) = N \sum_{i=0}^{N-1} a_i e^{-j\frac{2\pi}{N^2}ki} \delta((k-i)_N) ,$$

(5)

Where $\delta((k-i)_N) = \begin{cases} 0, & k \neq pN+i \\ 1, & k = pN+i \end{cases}$ that means,

$$Y(pN+i) = Y_p(i)$$

$$= \begin{cases} 0, & k \neq pN+i \\ \\ Na_i e^{-j\frac{2\pi}{N^2}(pN+i)i}, & k = pN+i \end{cases} \qquad i,p = 0,1,\cdots,N-1.$$

(6)

Equation (7) indicates that each data symbol a_i is modulated on $p = N$ subcarriers, and the total spectrum of the signal is spread N times. Figure 4 shows the spectrum of the ISS-OFDM signal with subcarrier number $N = \mathbf{4}$. In this case, the signal spectrum of ISS-OFDM is spread 4 times and contains 4 subbands. Each of the subbands contains the whole data information of the baseband signal.

2.1.4 Cyclic Prefix and Pulse Shaping

After generation of each ISS-OFDM symbol, the cyclic prefix is inserted to the ISS-OFDM symbol. The cyclic prefix is a very crucial part which is used to combat the ISI and inter channel interference (ICI) introduced by multipath fading channels through which the signal is propagated. As we discussed above, the main problem of ISI is from the energy leaking from one symbol to the other neighbor symbol. Our objective is to confine one symbol energy from leaking. Cyclic prefix insertion can effectively solve the problem.

A pulse shaping technique is employed to further reduce ISI. In the simulation of the ISS-OFDM transmitted signal, pulse shaping is implemented by using a pulse that is shaped like a Sinc function.
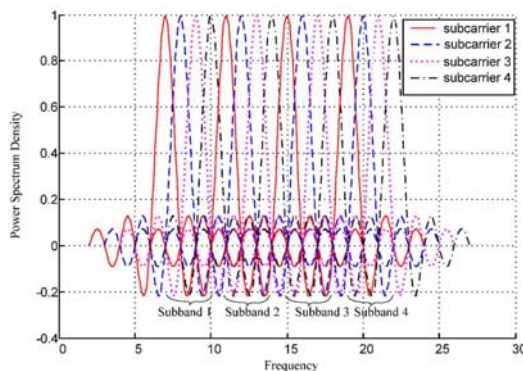
Fig. 4. Spectrum of ISS-OFDM symbol with subcarrier number $N = 4$**.**

2.1.5 Multiple Subband Signal Profile
After pulse shaping, the transmitted ISS-OFDM signal has a wide bandwidth. If the bandwidth of the pulse shaping filter is designed to equal to the bandwidth of the ISS-OFDM signal frequency band, the transmitted signal has N subbands and each of the subbands consists of N subcarriers as displayed in Figure 5.

If the bandwidth of the filter is designed to be equivalent to M out of N ($M < N$, N) is the number of subcarriers of the base band signal) subbands of the ISS-OFDM spread spectrum signal, only M subbands of the ISS-OFDM signal are transmitted, or only M subbands of the N subband ISS-OFDM signal are received, and the transmitted information will be recovered according to the M subbands. In the extreme case, if the filter pass band is equal to one subband of the ISS-OFDM signal, the ISS-OFDM system is equivalent to the conventional OFDM system.

2.2 Reception of Multiple Subband Signal
Before discussing the signal reception, we take a look at the channel impacts on the transmitted signal. We may assume that the transmitted ISS-OFDM signal bandwidth $B_{ISS\text{-}OFDM}$ is greater than the coherence bandwidth $(\Delta f)_c$, $B_{ISS\text{-}OFDM}$ $(\Delta f)_c$' thus the channel can be considered as a frequency-selective fading channel. The ISS-OFDM signal $y(m)$ (also represented by $Y(k)$ in the frequency domain) with spectrum spreading is
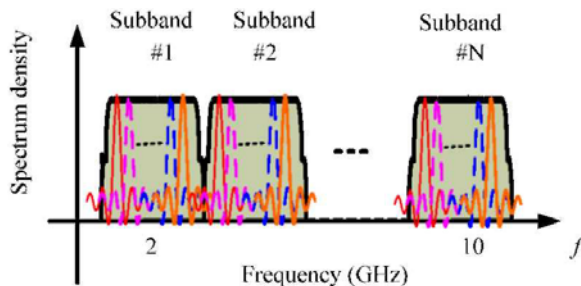

Fig. 5. Spectrum of ISS-OFDM signal.

transmitted into the frequency selective multipath fading channel with impulse response $h(m)$ (denoted as $H(k)$ in the frequency domain). Meanwhile, the AWGN noise v(m) and (denoted as $V(k)$ in the frequency domain) and interference $j(m)$(denoted as $J(k)$ in the frequency domain) is added to the signal. The channel model can be expressed in the form:

$$h(m) = \sum_{i=0}^{N^2-1} \alpha_i(m)\delta(m-m_i),$$

where a_i (m) is the multipath fading channel attenuation factor on the mth path and m_i is the propagation delay for the mth path. Correspondingly, the channel frequency response is written as:

$$H(k) = \sum_{m=0}^{N^2-1} h(m)e^{-j\frac{2\pi}{N^2}km}.$$

(7)

At receiver, the received equivalent low-pass signal is viewed as the convolution between $h(m)$ and $y(m)$ plus noise and interference as follows

$$r(m) = h(m) \otimes y(m) + v(m) + j(m),$$

(8)

where "$\otimes$" denotes convolution, v(m) represents the complex-valued white Gaussian noise process corrupting the signal and $j(m)$is the interference from the coexisting other systems. The receiver structure, as displayed in Figure 1, consists of the inverse process to the transmitter. The OFDM demodulation is realized by using N parallel FFTs, each of which is of size N. After CP removal, the serial data stream in the block of N^2 is demodulated by using N FFT operations. The output signal of the demodulator is equalized in parallel in the frequency domain, and then combined by using a maximal ratio combining (MRC) technique. The combined results are the decision variables which can be used for data detection, BER system performance computation and system performance analysis. To obtain the data vector for each FFT operation, the received time domain signal $r(m)$ is split into N groups, each group is demodulated by using an FFT operation with the size N. After the demodulation, the channel distortion needs to be compensated using a frequency domain minimum mean square error (MMSE) equalization technique. To use MMSE, we must find out how each parallel FFT output is related to the channel coefficient and the transmitted data symbol. The first step is to derive the mathematical expressions of the received signal in the time domain at the input of the receiver in terms of the channel frequency response and the transmitted data symbol.

2.2.1 Time domain Signal at the Receiver Input
For deriving the time domain signal model, we firstly gain its frequency domain signal. In the frequency domain, Equation (9) is equivalent to the output of the FFT demodulator:

$$R(k) = H(k) \cdot Y(k) + V(k) + J(k),$$

(9)

where R(k),Y(k),H(k),V(k), J(k) denote the frequency representations of the received signal *r(m)'* the transmitted signal *y(m)'* the channel impulse response *h(m)*, the noise v(m) and j(m), respectively. In the following section, the system transmitter and receiver architecture are introduced.

$$Y(k) = Y(pN+i) = \begin{cases} 0, & k \neq pN+i \\ Na_i e^{-j\frac{2\pi}{N^2}(pN+i)i}, & k = pN+i \end{cases}$$

$$(10)$$

Substituting *Y(k)* from Equation (11) into *R(k)*, the received signal in the frequency domain can be rewritten as follows:

$$R(k) = Na_i e^{-j\frac{2\pi}{N^2}ki} H(k) + V(k) + J(k), \qquad k = 0, \ 1, \cdots, \ N^2 - 1.$$

$$(11)$$

It can be observed from Equation (12) that the sample at the [th] position is repeated in every *N* samples of the N^2 samples of *R(k)*. Equation (11) also indicates that the energy of each data symbol a_i is distributed on the $(pN + i)$[th], *p* = 0, 1, — , *N* -1 , subcarriers. That means that the energy of a_i is distributed at the position *i* of each subband with *N*

subcarriers. Each subcarrier has a phase shift, which must be compensated by the equalization in the frequency domain before the signal can be combined. The effect of the equalization and combining is to compensate for the phase shift in the channel and to weight the signal by a factor that is proportional to the signal strength.

The signal in the time domain at the input of the receiver is the N^2 -point inverse discrete Foruier transform of *R(k),* i.e.,

$$r(m) = IFFT(R(k)) = \frac{1}{N^2} \sum_{k=0}^{N^2-1} \left(R(k) + V(k) + J(k) \right) e^{j\frac{2\pi}{N^2}km} .$$

$$(12)$$

Substituting *R(k)* from Equation (12) into Equation (13), and not considering *J(k)* impacts, the received signal in the time domain *r(m)* can be rewritten as:

$$r(m) = \frac{1}{N} \sum_{k=0}^{N^2-1} \sum_{i=0}^{N-1} a_i e^{-j\frac{2\pi}{N^2}k(i-m)} \delta((i-k)_N) H(k) + \frac{1}{N^2} \sum_{k=0}^{N^2-1} V(k) e^{j\frac{2\pi}{N^2}km} .$$

$$(13)$$

2.2.2 Receiver Signal Filtering
Since the transmitted signal consists of multiple subbands, each of which has the whole transmitted information of the baseband signal, the receiver can obtain the received signal with a different number of subbands by filtering the received signal using a filter with different parameters (B. C. Jung, 2007) . By adjusting the parameters of the filter according to

the different radio scenarios, the bandwidth of the signal used for demodulation can have different number of subbands.

Figure 6 displays an example of the signal spectrum before the receiver filter and after the receiver filter. The received signal contains N subbands, where N is the number of the subbands of the transmitted signal. After filtering, we can see that the received signals can be in different forms with a different number of subbands. Each subband contains the whole transmission information. If the pass band filter is configured to be equivalent to the bandwidth of 6 subbands of the transmitted signal as displayed in the green shadow, the signal used for demodulation has 6 subbands. The others will be filtered because of interference on the other subands, or because of channel bandwidth limitations. The filtered signal can be used to recover transmitted information, and receiver performance can satisfy different requirements in practical application scenarios.
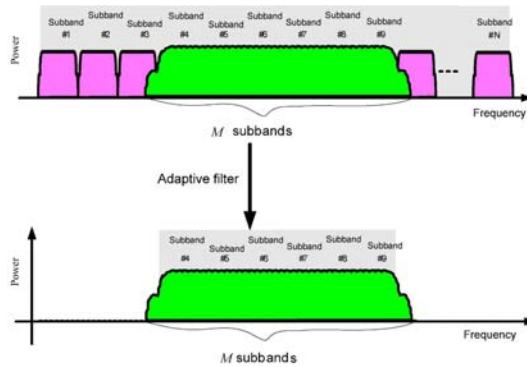

Fig. 6. Adaptive filtering at receiver.

Since only some subbands of the received signal are used for recovering the information, the process of filtering leads to a problem of signal energy loss. In consequence, it causes the reduction in power efficiency. Thus, this problem requires compromise between the selection of the subband number and signal to noise ratio. In the section on adaptive subband selection we will discuss the trade-off strategy.

2.2.3 Deinterleaver

Deinterleaver is to retrieve the order of the data sequence before the interleaving at the transmitter. Corresponding to the periodical interleaving algorithms used at the transmitter end, the deinterlever must be implemented by using the corresponding periodical algorithms.

2.2.4 Demodulation Using Multiple FFTs

After the deinterleaver, the signal in the time domain is split into N data streams, each of which can be demodulated using one FFT operation. Thus, to demodulate the N x N data samples simultaneously, N parallel FFTs, each in size N , are needed. In this way, the

demodulation of the received signal is completed by employing N parallel FFT operations. The deinterleaved signals can be written as:

$$r_i(n) = r(nN + i) = r(m), \quad i = 0, \ 1, \cdots, \ N - 1, \ n = 0, \ 1, \cdots, \ N - 1.$$

To derive the frequency domain representation of $r_i(n)$, we can perform the FFT operation on $r_i(n)$. The output of the demodulation can be obtained as follows:

$$R_i(k) = \sum_{n=0}^{N-1} r_i(n) e^{-j\frac{2\pi}{N}kn} = \sum_{n=0}^{N-1} r(nN + i) e^{-j\frac{2\pi}{N}kn} \ , \text{where} \ k = 0, \ 1, \cdots, \ N - 1.$$

Substituting $r(m)$ in Equation (14) into the above equation, $R_i(k)$ can be written as follows:

$$R_i(k) = \frac{1}{N} \sum_{n=0}^{N-1} \left[\sum_{k'=0}^{NN-1} \sum_{i'=0}^{N-1} a_{i'} e^{-j\frac{2\pi}{NN}k'(i'-nN-i)} \delta((k'-i')_N) H(k') + \frac{1}{N} \sum_{k'=0}^{NN-1} V(k') e^{j\frac{2\pi}{NN}k'(nN+i)} \right] e^{-j\frac{2\pi}{N}kn}$$

$$= \sum_{k'=0}^{N-1} \sum_{i'=0}^{N-1} a_{i'} e^{j\frac{2\pi}{NN}k'(i-i')} \delta((i'-k')_N) H(k') \underbrace{\frac{1}{N} \sum_{n=0}^{N-1} e^{j\frac{2\pi}{N}(k'-k)n}}_{\delta((k'-k)_N)}$$

$$+ \frac{1}{NN} \sum_{k'=0}^{NN-1} V(k') e^{j\frac{2\pi}{NN}k'i} \underbrace{\frac{1}{N} \sum_{n=0}^{N-1} e^{j\frac{2\pi}{N}(k'-k)n}}_{\delta((k'-k)_N)} .$$

Since k' - k must be integer multiples of N, i.e., $k' - k = qN$, we express $k' = qN + k$, and $R(k)$ is rewritten as

$$R_i(k) = \sum_{i'=0}^{N-1} a_{i'} \sum_{k'=0}^{NN-1} H(k') e^{j\frac{2\pi}{NN}k'(i-i')} \delta((i'-k')_N) \underbrace{\delta((k'-k)_N)}_{k'=qN+k} + \frac{1}{N} \sum_{k'=0}^{NN-1} V(k') \underbrace{\delta((k'-k)_N)}_{k'=qN+k} e^{j\frac{2\pi}{NN}k'i}$$

$$= \sum_{i'=0}^{N-1} a_{i'} \sum_{q=0}^{N-1} H(qN + k) e^{-j\frac{2\pi}{NN}(qN+i)(i'-i)} \underbrace{\delta((qN + k - i')_N)}_{i'=k} + \frac{1}{N} \sum_{q=0}^{N-1} V(qN + k) e^{j\frac{2\pi}{NN}(qN+i)i} .$$

Again, since $qN + k$ - i' must be integer multiples of N, we have $i = k$ and R (k) is finally expressed as

$$R_i(k) = a_k \sum_{q=0}^{N-1} H(qN + k) e^{-j\frac{2\pi}{NN}(qN+i)(k-i)} + \frac{1}{N} \sum_{q=0}^{N-1} V(qN + k) e^{j\frac{2\pi}{NN}(qN+k)i}$$

$$= a_k \underbrace{\sum_{q=0}^{N-1} e^{j\frac{2\pi}{N}q(i-k)} H(qN + k) e^{-j\frac{2\pi}{NN}i(k-i)}}_{C_{k,i}} + \underbrace{\frac{1}{N} \sum_{q=0}^{N-1} V(qN + k) e^{j\frac{2\pi}{NN}(qN+k)i}}_{v_{k,i}} = a_k C_{k,i} + v_{k,i} ,$$

$$(14)$$

where

$$C_{k,i} = \sum_{q=0}^{N-1} e^{j\frac{2\pi}{N}q(i-k)} H(qN+k) e^{-j\frac{2\pi}{NN}i(k-i)}$$ is an equivalent channel coefficient and

$$v_{k,i} = \frac{1}{N} \sum_{q=0}^{N-1} V(qN+k) e^{j\frac{2\pi}{NN}(qN+k)i}$$ is the independent additive noise.

Now the relationship among each demodulated output R (k), the equivalent channel coefficient , transmitted signal ak , and noise $v\&_i$ is clearly shown in Equation (15).

2.2.5 Channel Compensation

Based on the analysis above, we can observe that the channel introduces distortion on the transmitted signal. The distortion can be compensated by employing frequency doamin equalization methods (W. G. Jeon, 1999) before the transmitted date bits can be recovered.

To compensate for the channel distortion, the frequency domain equalizer is used as shown in Figure 7.
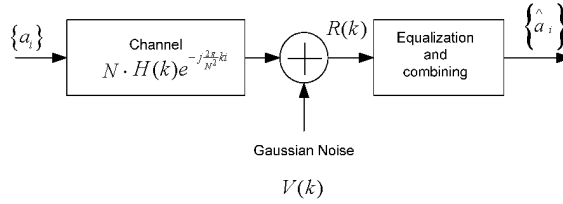

Fig. 7. Frequency domain equalizer.

According to the received signal and channel models, we can derive the algorithms for computing the estimate $\{a_i\}$ at the output of the frequency domain equalizer.

We start the derivation process from the output of the demodulator. The demodulator output signal in the frequency domain with multipath distortions is given as in Equation (12), which is rewritten as follows:

$$R\big(pN+i\big) = Na_i C_{i,p} + V\big(pN+i\big),$$

(15)

Where

$$C_{i,p} = e^{-j\frac{2\pi}{N^2}(pN+i)i} H\big(pN+i\big), p = 0, \ 1,\cdots, \ N-1, i = 0, \ 1,\cdots, \ N-1.$$

(16)

From the equation above, it can be seen that R(pN + i) is composed of N groups, each group forms one subband which consists of N subcarriers, and each subcarrier consists of the corresponding weighted signal and noise. Generally, we assume that the received signal contains M subbands, $M < N$, and use the M subbands to demodulate the transmitted signal. The structure of R(pN + i) can be illustrated in Figure 8.
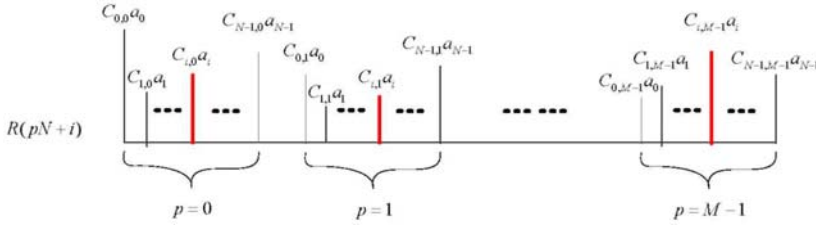
Fig. 8. The structure of $R(k)$.

The transmitted data symbol a_i, $i = 0,1, — , N$ -1, can be recovered from R(pN + i), p = 0,1, ... , M -1, which can be expressed in matrix form as follows:

$$
\begin{pmatrix} R(i) \\ R(N+i) \\ \vdots \\ R\big((M-1)N+i\big) \end{pmatrix} = N \begin{pmatrix} C_{i,0} \\ C_{i,1} \\ \vdots \\ C_{i,M-1} \end{pmatrix} a_i + \begin{pmatrix} V(i) \\ V(N+i) \\ \vdots \\ V\big((M-1)N+i\big) \end{pmatrix}.
$$

According to the above matrix equation, a minimum mean squared error (MMSE) estimate of a_i can be obtained as follows:

$$
\hat{a}_i = \frac{1}{N} \left(\begin{pmatrix} C_{i,0}^* & C_{i,1}^* & C_{i,M-1}^* \end{pmatrix} \begin{pmatrix} C_{i,0} \\ C_{i,1} \\ \vdots \\ C_{i,M-1} \end{pmatrix} + \frac{1}{SNR} \right)^{-1} \cdot \begin{pmatrix} C_{i,0}^* & C_{i,1}^* & \cdots & C_{i,M-1}^* \end{pmatrix} \begin{pmatrix} R(i) \\ R(N+i) \\ \vdots \\ R\big((M-1)N+i\big) \end{pmatrix}
$$

$$
= \frac{1}{N} \sum_{q=0}^{M-1} \frac{C_{i,q}^*}{\sum_{p=0}^{M-1} |C_{i,p}|^2 + \frac{1}{SNR}} R(qN+i)'
$$

where SNR is the signal-to-noise ratio before the MMSE equalization.

To analyze the system BER performance, the normalized MMSE can be expressed as

$$
\varepsilon_{min}^2 = \left(SNR \begin{pmatrix} C_{i,0}^* & C_{i,1}^* & \cdots & C_{i,M-1}^* \end{pmatrix} \begin{pmatrix} C_{i,0} \\ C_{i,1} \\ \vdots \\ C_{i,M-1} \end{pmatrix} + 1 \right)^{-1} = \frac{1}{SNR \sum_{p=0}^{M-1} |C_{i,p}|^2 + 1}.
$$

Therefore, the output signal-to-noise ratio after MMSE equalization is

$$
\gamma_{out} = \frac{1 - \varepsilon_{min}^2}{\varepsilon_{min}^2} = \frac{1}{\varepsilon_{min}^2} - 1 = SNR \sum_{p=0}^{M-1} |C_{i,p}|^2.
$$

Assuming QPSK modulation for data symbols and making a Gaussian distribution approximation for the intersymbol interference after MMSE equalization, we finally obtain the expression for the average BER using the Q function

$$P_e^{(MMSE)} = E\left(Q\left(\sqrt{\gamma_{out}}\right)\right),$$

(17)

where E () denotes ensemble average over channel coefficients.

2.2.6 Maximal Ratio Combining

From Equation (16) we see that a data symbol a_i is transmitted on different subcarriers and experiences independent fading. Thus, another effective way to gain frequency diversity is to directly use the maximal ration combining (MRC) technique to collect signal energy from all information bearing subcariers. In this way the system diversity can be made full use of and the best system performance can be achieved.

Similar to the above MMSE equalization, we assume that only M subbands, $M < N$, are used to recover the transmitted data symbols. The MRC process proceeds as follows.

First, we multiply the conjugated channel coefficient C_i ^ with the received $R(pN + i)$ to compensate the channel phase shift and weight the information bearing subcarrier with a gain proportional to the signal strength. Then, all the weighted subcarriers corresponding to the same data symbol a_i are combined to produce the decision variable U_i for the detection of a_i . That is,

$$U_i = \sum_{p=0}^{M-1} C_{i,p}^* R\left(pN + i\right).$$

To analyze the BER performance of the MRC technique, we substitute $R(pN+i)$ in the above equation with Equation (16) and have

$$U_i = N \sum_{p=0}^{M-1} \left|C_{i,p}\right|^2 a_i + \sum_{p=0}^{M-1} C_{i,p}^* V\left(pN + i\right).$$

The output signal-to-noise ratio after MRC can be obtained as

$$\gamma = \frac{\left(N \sum_{p=0}^{M-1} \left|C_{i,p}\right|^2\right)^2 E\left(\left|a_i\right|^2\right)}{E\left(\left|\sum_{p=0}^{M-1} C_{i,p}^* V\left(pN + i\right)\right|^2\right)} = \sum_{p=0}^{M-1} \left|C_{i,p}\right|^2 SNR$$

where $\quad SNR = \dfrac{N^2 E\left(|a_i|^2\right)}{E\left(\left|V\left(pN+i\right)\right|^2\right)} \quad$ is the signal-to-noise ratio before MRC.

For QPSK modulated data symbols, the average error probability can be therefore written as:

$$P_e^{(MRC)} = E\left(Q\sqrt{\gamma}\right).$$

(18)

From Equation (18) and Equation (19) we see that the MMSE equation and the MRC have the same performance for the serial demodulation.

2.2.7 BER Performance

For diversity performance comparison, the normalized signal-to-noise ratio E_b/No is more often used than SNR. In this section, we briefly discuss the normalization method in our system.

E_b/No is the ratio of bit energy to noise power spectral density, which is used to evaluate the BER performance. Thus, it is desirable to obtain the relation between SNR and Eb/No , so that we can obtain the system BER performance as a function of E_b/No .

In Figure 9, let er². and er² denote the signal variance, or average signal power, and average noise power, respectively. N is the number of subcarriers used to modulate data symbols. E_s, Eb, N_0 denote QPKS symbol energy, bit energy and noise power spectral density, respectively. M ($M < N$) denotes the number of subbands of the received signal after the receiver filter. $M = N$ means that all the subbands of the transmitted signal are used to recover the transmission information. Thus, the relation between SNR and No can be expressed as:

$$SNR = \frac{\sigma_x^2}{\sigma_v^2} = \frac{MN\sigma_x^2}{MN\sigma_v^2} = \frac{MN\sigma_x^2 \cdot T \cdot MN}{MN\sigma_v^2 \cdot T \cdot MN} = \frac{E_s \cdot N}{MN\sigma_v^2 \cdot T \cdot MN}$$

$$= \frac{E_s \cdot N}{N_0 \cdot \dfrac{1}{T} \cdot T \cdot MN} = \frac{E_s}{N_0 \cdot M} \cdot = \frac{2}{M}\frac{E_b}{N_0} \cdot$$

(19)

According to the above relationship between SNR and Eb/No, the system BER performance expression in Equation (19) can be rewritten as:

$$P_e^{(MRC)} = E\left(Q\left(\sqrt{\sum_{p=0}^{M-1}\left|C_{i,p}\right|^2 \frac{2}{M}\frac{E_b}{N_0}} \right) \right).$$
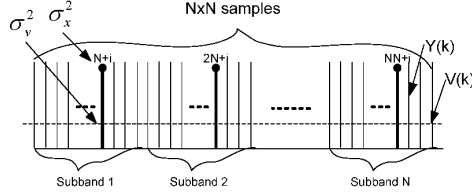

Fig. 9. *SNR* normalization.

3. Adaptive Subband Selection Method

In previous section, the multiple subband ISS-OFDM signal is presented, which consists of *N* subbands, each containing *N* subcarriers. When the signal passes through a multipath fading channel, the received signal experiences fading in different subbands. Meanwhile, interference signals will be imposed on the transmitted subbands.

3.1 Interference avoidance

To avoid interference at the receiver so that system performance can be improved, the interfered subband(s) should be removed from the received signal if the interference power lever is above a predetermined threshold. The mechanism is explained in more detail below. Assume that the value of the threshold is *Y0* (we will discuss how to derive it in the next section). If the interference superimposed on a subband is greater than the threshold value **Y0** , then this interference will cause severe adverse effects on the desired signal. In order to achieve better system BER performance, the interfered subband should be excluded. If the interference is lower than the threshold value **Y0** , this interference can be tolerated and the interfered subbands are kept in the transmission bands. Otherwise, if we remove the interfered subbands with interference lower than the threshold, system performance will be worse than that of the system with the interfered subbands, since when we remove the interfered subbands, the signal energy is removed at the same time. Figure 10 shows the fundamental principle of interference avoidance, in which a 4 subband OFDM signal in the unlicensed 2.4 GHz ISM frequency band is faded and interfered with, but only two of the subbands with interference over the threshold are removed.

3.2 Determination of Interference Thresholds

In the previous subsection, the method of avoiding interference is discussed under the assumption of a known threshold of interference. In this subsection the thresholds of interference under different channel conditions for systems coexistence are investigated.

3.2.1 Thresholds in Gaussian Channels

Assume that M subbands of the N subband signal are transmitted, each having signal power P_S. If the noise power spectral density is N_0, and the total interference signal power is P_J, which is distributed in l subbands of the transmitted M subband OFDM signal, then, the total signal power of the M subbands is $M \cdot P_S$, and the total signal to interference noise ratio ($SINR$) can be expressed as:

$$SINR_M = \frac{M \cdot P_S}{N_0 M \cdot B + P_J} = \frac{SNR}{1 + INR},$$

(20)

Where $\dfrac{MP_S}{MN_0 B} = SNR$, $\dfrac{P_J}{N_0 MB} = INR$, B, is the bandwidth of each subband.

Let G_p denote the processing gain for each subband, the total $SINR$, after considering the processing gain, is

$$SINR^{(1)} = M \cdot G_p \cdot SINR_M = \frac{M \cdot G_p \cdot SNR}{1 + INR}.$$

Similar to $SINR_M$, if the interfered subbands are cut off, the interference energy P_J is correspondingly completely removed. Thus, the total $SINR$, after the interfered subbands are cut off, can be expressed as:

$$SINR_{M-l} = \frac{(M-l) \cdot P_S}{N_0 (M-l) \cdot B} = SNR.$$

(21)

Similar to $SINR^{(1)}$, after the l interfered subbands are removed, the total $SINR$ after considering the processing gain is

$$SINR^{(2)} = (M-l) \cdot G_p \cdot SINR_{M-l} = (M-l) \cdot G_p \cdot SNR.$$

Since the signal energy is distributed in the whole bands, after the l interfered subbands are removed, the signal energy distributed in the l subbands is also removed, and the signal energy is correspondingly reduced. In a Gaussian channel, the BER performance is determined by the normalized signal-to-noise ratio. In order to have higher

SNR after the interfered subband removal, $SINR^{(2)}$ should be higher than $SINR^{(1)}$, and thus we have the following expression:

$$SINR^{(2)} > SINR^{(1)} \implies INR > \frac{l}{M-l} = \gamma_0 .$$

(22)

It can be seen from Equation (25) that the threshold γ_0 of the INR is related to the number of transmitted subbands M and the number of interfered subbands l over Gaussian channels.

Given the interfered subband number l, the INR threshold decreases with the increase of the number M of the transmitted subbands. If the number M of the transmitted subbands is given, the INR threshold increases with the increase of the interfered subband number l.
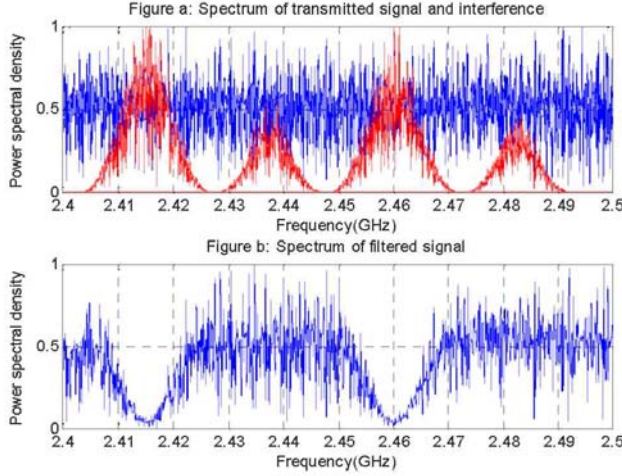

Fig. 10. Subband selection by using adaptive filter.

3.2.2 Thresholds in Multipath Fading Channels

In a multipath fading channel, since channel fading is random, we can not obtain the same thresholds as those in the Gaussian channel. However, it is possible for us to get an estimation of the thresholds by statistically analyzing system BER performance in a multipath fading channel.

Recalling Equation (23) and considering the multipath fading effects, system BER performance before the interfered subbands are removed can be expressed by using the Q function as:

$$P_e^{(1)} = E\left[Q\left(\sqrt{\sum_{p=0}^{M-1}\left|H_p\right|^2 SINR^{(1)}}\right)\right] = E\left[Q\left(\sqrt{\sum_{p=0}^{M-1}\left|H_p\right|^2 \frac{SNR}{1+INR}}\right)\right],$$

where E() denotes the statistical expectation of the conditional error probability over

multipath fading channels, and $\displaystyle\sum_{p=0}^{M-1}\left|H_p\right|^2$ is the sum of the M multipath fading channel

coefficients on the M subcarriers. Due to the QPSK modulation, $SNR = \dfrac{2}{M} E_b / N_0$, which

is derived in Section 2.2.7, where E_b / N_0 is the signal bit energy to noise spectrum density

ratio. The equation above can be rewritten as:

$$P_e^{(1)} = E\left[Q\left(\sqrt{\frac{1}{M} \sum_{p=0}^{M-1} |H_p|^2 \frac{2 \cdot E_b / N_0}{1 + INR}} \right) \right].$$

(23)

Similarly, recalling Equation (24), the BER performance of the system after the interfered subbands are removed in multipath fading channels can be derived as:

$$P_e^{(2)} = E\left[Q\left(\sqrt{\sum_{p=0}^{M-l-1} |H_p|^2 \frac{2 \cdot E_b / N_0}{M}} \right) \right],$$

(24)

where $E(\cdot)$, H_p, E_b / N_0 have the same meanings as defined in Equation (24).

According to Equation (24), we simulate system BER performance with the increase of the

INR when the transmission subband number is 16 and $E_b / N_0 = 10 dB$, which is displayed

in Figure 11.

According to Equation (25) and with the same E_b / N_0, the BER performance of the

system after the interfered subbands (its number is l =1, 2, 4, 8 respectively) are removed is shown as flat lines in Figure 11. The cross points are the interference thresholds **Y0** corresponding to l =1, 2, 4 and l =8. It can be seen from Figur 11 that the interference thresholds are -11dB, -7.5dB, -4.0dB and 1dB. The corresponding thresholds at the same conditions in a Gaussian channel can be calculated from Equation (25).

The interference thresholds in a multipath fading channel with transmission subbands M = 2, 4, 8, 16 and 32 are presented in Table 1. It is observed from Table 1 that the threshold discrepancies between multipath fading channel and Gaussian channel decrease with the increase of the transmission subband number and the discrepancy is approximately zero when the transmission subband number M = 32. That is to say, we can use thresholds in a Gaussian channel to replace the thresholds in a multipath fading channel if the number of transmitted subbands is more than 32.

3.3 Filtering of Multiple Subband Signal

When the multiple subband signal passes through the adaptive filter, the filtering of the multiple subband signal can be viewed as the convolution between the received multiple subband signal and the filter impulse response.

If the bandwidth of the adaptive filter is designed to be equal to the bandwidth of the signal frequency band with N subbands, after filtering, the pass-band signal has N subbands, each of which has the whole transmitted information. If the bandwidth of the filter is designed to be M ($M \leq N$, N is the number of the subcarriers of the base band signal) sub-bands, only M subbands of the signal are passed through the filter, or only M subbands of the signal are received. In the case that the filter pass-band is equal to one subband of the multiple subband signal, the system model is equivalent to a conventional OFDM system.

Assume the value of the threshold is y_0. If the interference superimposed on a subband is greater than the threshold **Y0**, then this interference would cause severe adverse effects on the desired signal. In this case the interfered subband should be excluded. If interference is lower than the threshold **Y0**, the interference can be tolerated by the system and the interfered subbands should be kept in the transmission bands. These transmission signals with different numbers of subbands are very flexible to match different spectrum holes with different bandwidths.

4 System Performance

4.1 PAR Simulation and Analysis
Assume that the data symbol mapping scheme uses QPSK and the number of subcarriers $N = 32$. The generated ISS-OFDM signal has N^2 samples and contains N subbands, each of which contains N subcarriers and carries the same data information. The transmitter is designed so that M subbands are selected, $M = 1, N$, respectively, and thus the transmitted signal contains M subbands and the spectrum spreading factor is M.

We assume that the data symbol has the same energy for different spreading factors. we select the signal bandwidth to contain $M = 1, 2, 4, 8, 16$ and 32 subbands respectively (M is the spreading factor), and the filtered signal is oversampled by a factor of four, which is commonly used to estimate the PAR of an analog signal from its samples. When $M = 1$, there is only one signal subband to be allowed to pass through the filter, resulting in a conventional OFDM signal. When $M = 2, 4, 8, 16$ and 32, respectively, i.e., the spreading factor is 2, 4, 8, 16 and 32, the signal contains 2,4,8,16 and 32 subbands respectively.

Time domain samples of the transmitted signal y(t) in the equivalent complex valued low-pass domain are approximately Gaussian distributed due to the statistical independence of carriers. Resulting PAR of $y(t)$ can be written as:

$$\xi = \frac{\max |y(t)|^2}{E[|y(t)|^2]},$$

(25)

where $\max |y(t)|^2$ is the maximum instantaneous power of the ISS-OFDM signal and $E\left[|y(t)|^2\right]$ is the expected value of $|y(t)|^2$

According to Equation (26) we can compute the PAR performance curves. Figure 12 shows the PAR performance of the ISS-OFDM transmitted signal when PAR exceeds a certain

threshold PAR0 with the increase of the spectrum spreading factor M from 1 to 32. It can be seen from Figure 12 that with the increase of the number of subbands passing through the filter, the PAR performance is improved considerably. The most right-hand-side curve shows the PAR performance of the ISS-OFDM signal when M is equal to 1, which is the same as that of the conventional OFDM signal. As M increases to 2, 4, 8, 16, and 32 respectively, the PAR gains are 0.5dB, 1.5 dB, 3.5 dB, 5 dB and 7 dB respectively.
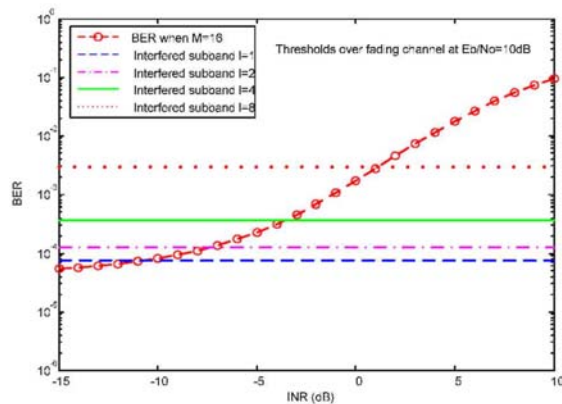

Fig. 11. *INR* thresholds over multipath fading channel.

Interfered Subbands	M=2		M=4		M=8		M=16		M=32	
	Gauss.	Multi.	Gauss.	Multi.	Gauss.	Multi.	Gauss.	Multi.	Gauss.	Multi.
$l=1$	0	4.9	-4.8	-1.7	-8.5	-5.5	-11.8	-11.0	-14.9	-14.5
$l=2$	–	–	0	3.3	-4.8	-2.0	-8.5	-7.5	-11.8	-11.5
$l=4$	–	–	–	–	0	3.0	-4.8	-4.0	-8.5	-8.5
$l=8$	–	–	–	–	–	–	0	1.0	-4.8	-4.5
Discrepancy	≈ 4.9		≈ 3.2		≈ 3		≈ 1.0		≈ 0.3	

Table 1. Threshold comparison between Gaussian and multipath channels at E_b / N_0 = 10 dB.

4.2 BER Performance

Once the system *INR* thresholds are obtained, the interfered subbands with INR over the thresholds can be adaptively cut off to avoid interference. In order to verify the effectiveness of the subband adaptation in a multipath fading channel, the BER of the system without interference, with interference, and with interference removed are simulated. These are shown in Figure 13, Figure 14 and Figure 15, respectively.

Figure 13 indicates the system BER performance without interference with system transmission subband numbers M = 2, 4, 8, 16 and 32. It is observed that with the increase of M, the system BER performance is improved considerably. After M = 16, a further increase of transmission subbands is not obviously beneficial to the improvement of the BER performance, but is helpful for interference avoidance.

Under the same conditions, the system BER with one subband interfered is displayed in Figure 14. The *INR* is 3dB. It is seen that the impact from the interference results in 5 ~6 dB degradation at the BER performance of 10^{-6} compared with the system performance without interference.

As shown in Figure 15 after the interfered subband is removed when M = 4, 8, 16 and 32, the system BER is improved by 0.5~ 5 dB compared with the system performance with interference. In this case performance is very close to that without interference. However, when M = 2 system performance becomes worse, since the subband with INR lower than the threshold ^ = 4.9 dB is removed, which agrees with the previous analysis.
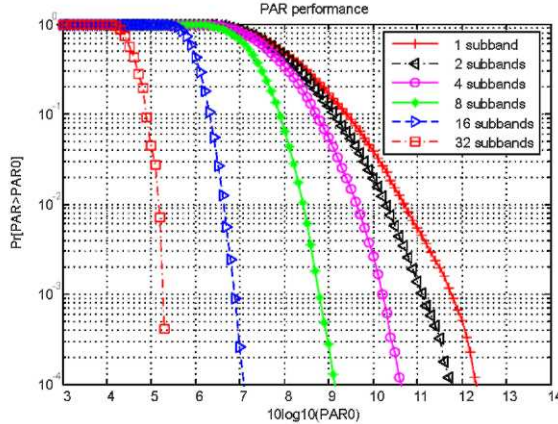

Fig. 12. PAR performance for ISS-OFDM signals with different number of subbands.

5. Conclusions and Future Work

In this chapter, we proposed an adaptive subband selection methods based on OFDM and cognitive radios. Since the transmission signal spectrum is spread to a wide band consisting of many subbands, and each subband contains the whole information needed to be transmitted, any one or more subband can be employed as the system transmission bandwidth to transmit information, which makes the system transmission bandwidth extremely flexible, so that the system bandwidth can be adaptively reconfigured to avoid interference and improve system performance according to the radio scene or interference conditions.

Due to the spectrum spreading of the transmitted ISS-OFDM signal, the signal transmission power is very low, which well prepares the ISS-OFDM system for use in sensor networks and wireless body area networks. It can realize this without interfering with other users.

The simulation results on system BER performance under this assumption are satisfactory and close to the theoretical analysis results. However, to some extent, we idealize the system without considering the impacts of frequency offset and synchronizations. If we add the impacts from system synchronization, system performance may degrade to some extent. Synchronization is a complex question waiting to be solved. Any fault in one of the issues of frame synchronization, symbol synchronization and subcarrier synchronization may cause the system bit error rate to be come worse. Sometimes it may result in the malfunction of the system. Therefore, ignoring system synchronization to simulate the ISS-OFDM system may not be perfect.
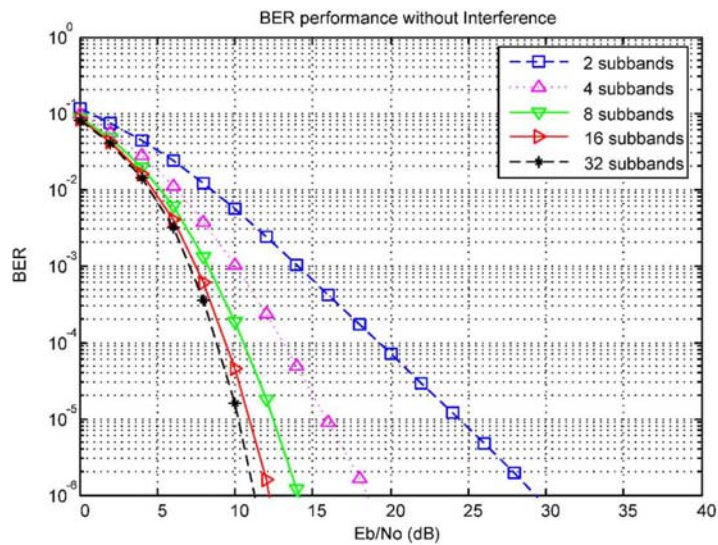
Fig. 13. BER performance without interferences in fading channel.
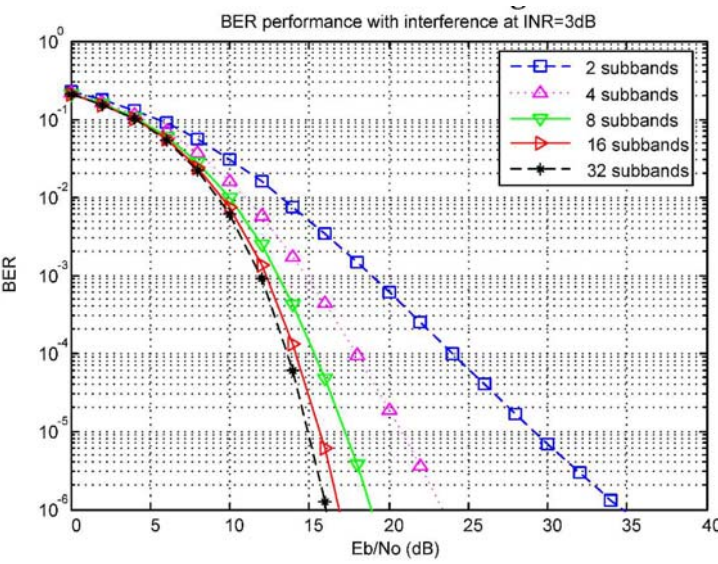

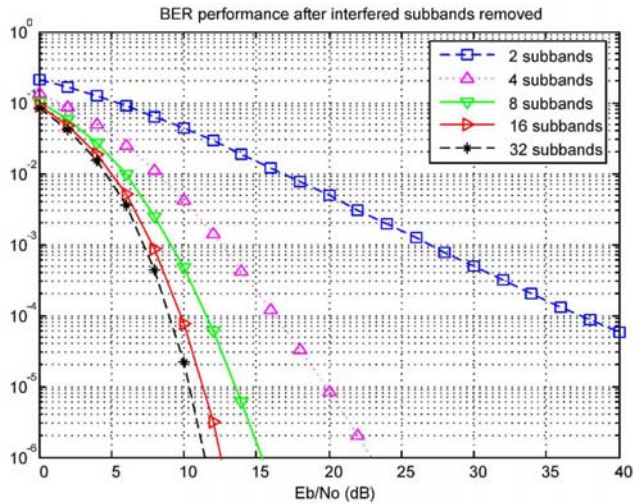Fig. 14. BER performance with interferences in fading channel.

Fig. 15. BER performance after interfered subbands removed adaptively in the fading channels.

6. References

B. C. Jung, Y. J. Hong, D. K. Sung and S. Y. Chung, "Adaptive Sub-band Nulling for OFDM-Based Wireless Communication Systems," in IEEE Wireless Communications and Networking Conference, WCNC, pp. 1491-1495, 2007.

G. Andrea, C. Marco and D. Davide, "Coexistence Issues in Cognitive Radios Based on Ultra-Wide Bandwidth Systems," in 1st International Conference on Cognitive Radio Oriented Wireless Networks and Communications, pp. 1-5, 2006.

J. G. Proakis, *Digital Communications,* 4ed, New York, USA, Mc Graw Hill, 2000.

S. B. Slimane, "Peak-to-Average Power Ratio Reduction of OFDM Signals Using Broadband Pulse Shaping," in 2002 IEEE 56th Proceedings on Vehicular Technology Conference, VTC 2002-Fall, vol. 2, pp. 889-893, 2002.

S. Haykin, "Cognitive Radio: Brain-Empowered Wireless Communications," in *IEEE Journal on Selected Areas in Communications,* vol. 23, pp. 201-219, 2005.

T. Cooklev, *Wireless Communication Standards,* NewYork: Standards Information Network, IEEE Press, 2004.

W. G. Jeon, K. H. Chang and Y. S. Cho, "An Equalization Technique for Orthogonal Frequency Division Multiplexing Systems in Time-Variant Multipath Channels," in *IEEE Transactions on Communications,* vol. 47, pp. 27-32, 1999.

Power Allocation in OFDM-Based Cognitive Radio Systems

Peng Wang, Xiang Chen, Xiaofeng Zhong, Limin Xiao,
Shidong Zhou and Jing Wang
Research Institute of Information Technology, Tsinghua University
Beijing, China

1. Introduction

Spectrum scarcity is becoming a serious problem as the rapid developments of wireless communications. However, recent spectrum measurements show that the licensed spectrum is severely underutilized by the primary users (PUs) in both the time domain and the spatial domain (FCC, 2002; Mchenry, 2005). The secondary users (SUs) are introduced to exploit the spectrum opportunities left by the PUs in order to improve the spectrum utilization. The transmission of SU is required to satisfy the PU's interference limits to protect the PU's performance. Cognitive radio (Mitola, 1999; Haykin, 2005; Zhao & Sadler, 2007) is the key technology enabling flexible, efficient and reliable spectrum utilization by adapting the radio's operating characteristics to the real-time conditions of the environment. The SUs with cognitive radio technology are able to cleverly detect and utilize the spectrum opportunities and thus realize efficient reuse of the licensed spectrum.

It is extremely important to guarantee the transmission of the SU satisfy the interference limits of the PUs for the cognitive radios to be acceptable. Generally, the interference limits can be described by two metrics: the maximum allowable collision probability and the interference power limit (Zhao & Sadler, 2007). The former one is related to the spectrum sensing mechanism of the cognitive radios, which has been widely studied in both the physical (PHY) layer (Sahai et al., 2004; Oner & Jondral, 2007; Cabric et al., 2004) and the medium access control (MAC) layer (Zhao et al., 2007; Kim & Shin, 2007; Ghasemi & Sousa, 2007) in the literature. The latter one introduces power restrictions on the channel output signals in cognitive radio systems, which is different from the conventional wireless systems where the power constraints are only placed on the channel input signals. This feature motivates the studies on how to efficiently utilize the spectrum opportunities under the interference power limits of the PUs (Zhang et al., 2008; Shu et al., 2006; Sudhir & Jafar, 2007; Le & Hossain, 2007).

A cognitive radio system needs to be highly flexible in terms of the spectral shape of the transmit signal. Orthogonal frequency division multiplexing (OFDM) modulation is a promising candidate for such a flexible system because of its reconfigurable subcarrier structure (Weiss & Jondral, 2004; Berthold & Jondral, 2005). With OFDM, the SU has the ability to flexibly fill the spectral gaps left by PUs. Also, the fast Fourier transform (FFT)

components at the OFDM system's receiver may also be used for the SU to execute the channel detection. In

(Weiss & Jondral, 2004), Weiss and Jondral proposed that the band of the SU covers multiple PUs' licensed spectrum, then the SU modulates zero on the subcarriers which belong to the detected PUs's licensed spectrum while utilizing other subcarriers for transmission. In OFDM-based cognitive radio systems, the adaptive subcarrier configuration should consider not only the channel state information (CSI) as in conventional OFDM systems, but also the sensing results of the SU and the interference limits of the PUs. This chapter focuses on investigating the research challenges involved in the power allocation for OFDM-based cognitive radio systems.

The optimal power allocation algorithm for conventional OFDM systems that maximizes the channel capacity is the wellknown *water-filling,* which is derived by solving a convex optimization problem subject to the sum transmit power constraint(TSE & Viswanath, 2005). In OFDM-based cognitive radio systems, the band of the SU can be divided into several subchannels, each of which is corresponding to a licensed band of one PU system. Since the interference limit of each PU introduces the subchannel transmit power constraint for the SU, the power allocation in OFDM-based cognitive radio systems should not only satisfy the sum transmit power constraint but also the subchannel transmit power constraints (Zhao & Sadler, 2007). Therefore, the conventional water-filling algorithm is not applicable in such a scenario.

The transmit power in each subchannel is comprised of the power allocated to the subcarriers inside the subchannel and the sidelodes power of the subcarriers in other subchannels. In this chapter, we first formulate the power allocation problem in the case where the effects of subcarrier sidelobes can be ignored, i.e., there is sufficient guard band between any two neighboring subchannels. Based on the convex optimization theory, an algorithm named *iterative partitioned water-filling* (IPW) is proposed to obtain the optimal power allocation that maximizes the capacity while satisfying both the sum and subchannel transmit power constraints. Then, we extend the results and address the power allocation problem in general OFDM-based cognitive radio systems, where the effects of subcarrier sidelobes are specially considered. In this case, we propose a *recursive power allocation* (RPA) algorithm to obtain the optimal power allocation by decoupling the subchannel power constraints phase-by-phase. The organization of the rest of this chapter is as follows: Sections 2 presents an overview of the cognitive radio and OFDM-based cognitive radio systems. In Section 3, the power allocation problem ignoring the effects of subcarrier sidelobes for OFDM-based cognitive radio systems is formulated and analyzed. The IPW algorithm is proposed and its optimality is proved. Then Section 4 addresses the general OFDM-based cognitive radio systems where the effects of subcarrier sidelobes are considered. The RPA algorithm is proposed and its optimality is proved. Simulation results are presented in Section 5. Section 6 discusses the future research directions and concludes the chapter.

2. System Model

2.1 Cognitive radio system and interference power limit
One of the typical cognitive radio systems is shown in Fig. 1. A certain channel is licensed to the PU system. Since the PU system does not occupy the channel anywhere at any time, the

channel is underutilized in both the spatial domain and the time domain. A channel is said to be a spectrum opportunity if the interference to the PU receivers caused by the SU's transmission is tolerable. The SU is permitted to access the channel if the channel is detected to be a spectrum opportunity. With the participance of the SU, the spectrum
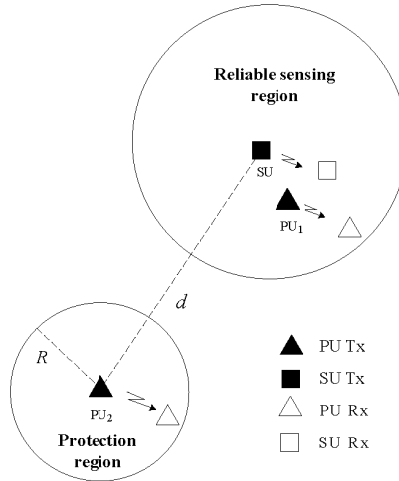

Fig. 1. Cognitive radio system model.

opportunities can be identified and exploited and thus the overall spectral efficiency can be improved.

To guarantee the interference to the PU tolerable is an extremely important problem in the implementation of cognitive radio systems. Generally, the SU has to sense the channel in order to determine whether a spectrum opportunity exists before the transmission. The SU is permitted to transmit on a licensed channel unless the corresponding PU's signal is seen as absent. A detection probability has to be achieved to reduce the collisions between the SU and the PU caused by the detection errors. As in Fig. 1, PU_1's signal can be detected when active and thus the interference to PU_1 receiver caused by the SU's transmission can be avoided. However, due to the signal to noise ratio limit, the SU can only detect the signal of the PU with the required detection probability within a certain region, as the reliable sensing region shown in Fig. 1. Therefore, for the PU_2 transmitter that is outside the SU's reliable sensing region, the SU is unable to detect PU_2's signal with the required detection probability, as depicted in Fig. 1. In this situation, as in (Zhao & Sadler, 2007), PU_2 defines a protection area whose radius is R and requires the interference power at any potential receiver in this area be lower than a certain value, say   . Therefore, when one PU's signal is seen as absent, the SU's transmit power on the PU's licensed channel P_{tx} should be subject to a power constraint, which is given by

$$P_{tx} \leq \eta(d - R)^{\beta}, \tag{1}$$

where d is the distance between the SU transmitter and the nearest undetectable PU transmitter (the PU transmitter outside the reliable sensing region is referred to undetectable PU transmitter in the following), and β is the path attenuation factor. Note that we only consider

the distance-based path loss here for simplicity. Also, the distance between the SU and the nearest undetectable PU transmitter is assumed to be known in advance by the SU
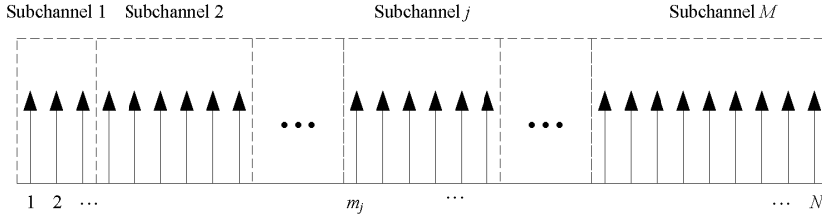

Fig. 2. Spectrum of SU in OFDM-based cognitive radio systems.

in (1). We note that if the SU is unaware of the PU's location, some conservative scheme may be used to estimate P_{tx}. For example, d is set to be the radius of the reliable sensing region, i.e., the PU's transmitter is assumed to be just at the margin of the reliable sensing region.

2.2 OFDM-based cognitive radio systems and subchannel transmit power constraints
In order to efficiently exploit the spectrum opportunities left by different PU systems, a cognitive radio system needs to be highly flexible with respect to the spectral shape of the transmit signal. OFDM modulation is a promising candidate for such a flexible system because of its reconfigurable subcarrier structure. Also, the FFT component at the OFDM receiver can be used for spectrum sensing, which reduces the overhead for the implementation of the cognitive capability.

As in Fig. 2, in an OFDM-based cognitive radio system, the spectrum that can be potentially used by the SU is divided into M subchannels. Each subchannel is corresponding to a PU's licensed band. The total number of the subcarriers is assumed to be N and m_j denotes the index of the first subcarrier in the jth subchannel.

Before the transmission, the SU senses whether each subchannel is occupied firstly, by proper spectrum sensing methods. Then according to the sensing results and the CSI of each subcarrier over SU transmission link, the SU can decide the proper power allocation scheme, modulation type and other parameters for SU's transmission. Therefore, the total transceiver diagram can be shown in Fig. 3. At the transmitter in Fig. 3, the transmission parameters should be decided before serial-to-parallel (S/P) conversion, IFFT operation, parallel-to-serial (P/S) conversion, insertion of a Cyclic Prefix (CP) and filtering. Accordingly, at the receiver, the information on transmission parameters over the link should be obtained through signaling to receive and demodulate OFDM signals. Then the received signal of the *ith* subcarrier within one OFDM symbol can be expressed as:

$$y_i = h_i x_i + n_i \tag{2}$$

where x_i is the signal transmitted on the ith subcarrier by the SU transmitter, h_i is the channel gain and n_i is the additive white Gaussian noise with mean 0 and variance 1. If the corresponding PU transmitter is detected in a subchannel, all the subcarriers in this subchannel will be modulated by zero during the transmission, i.e., the sum power of the subcarriers in this subchannel is set to be zero. Otherwise, the SU can use this subchannel
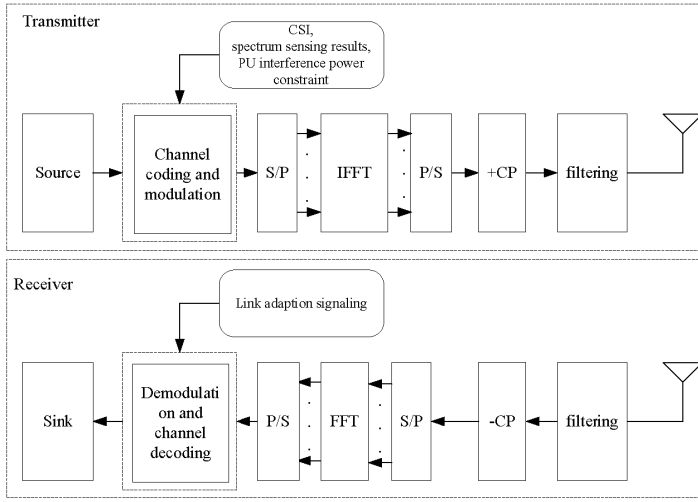
Fig. 3. Transceiver diagram of SU in OFDM-based cognitive radio systems.

but with a certain power constraint which has been described in Section 2.1. Assume G_j is the power constraint on the jth subchannel after the sensing, we have

$$G_j \triangleq \begin{cases} 0 & \text{PU}_j \text{ is detected} \\ \eta_j (d_j - R_j)^{\beta_j} & \text{PU}_j \text{ is not detected} \end{cases} \tag{3}$$

where η_j is the interference power limit of PU_j, R_j is the radius of the protection area of PU_j, d_j is the distance between the SU's transmitter and the nearest undetectable PU_j's transmitter, and β_j is the path attenuation factor.

3. Power Allocation for OFDM-based Cognitive Radio Systems without Considering Subcarrier Sidelobes

In conventional OFDM systems, given the sum transmit power constraint or the target rate, the optimal power allocation that aims at maximizing the sum rate or minimizing the required power can be achieved by the well-known water-filling algorithm. In OFDM-based cognitive radio systems, the power allocation should also satisfy the subchannel transmit power constraints which are introduced by the interference power limits of the PUs. Therefore, the water-filling algorithm needs to be modified. In this section, we first review the power allocation problem in conventional OFDM systems and then formulate that in the cognitive radio scenario.
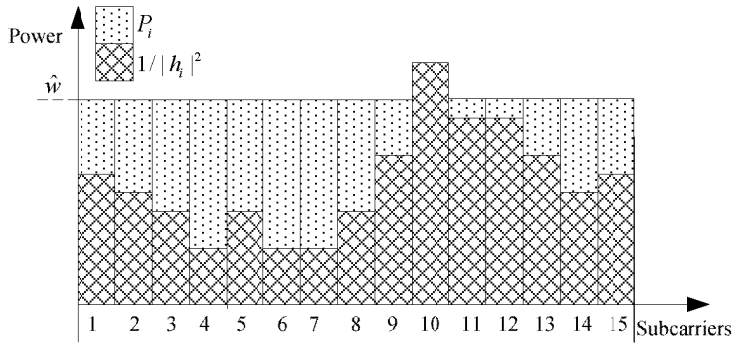
Fig. 4. Optimal power allocation in conventional OFDM systems.

3.1 Power allocation in conventional OFDM systems

The signal model for conventional OFDM systems is as (2) shows. The power allocation problem can be classified into two categories in conventional OFDM-based systems. The first one is to optimize the power allocation across the subcarriers such that the sum rate is maximized with a given sum transmit power constraint. Assume the sum transmit power constraint of an OFDM block is P_t, the allocated signal power on the ith subcarrier is P_i and the number of the subcarriers is L. The optimal power allocation can be given by

$$P_i = \left(\widehat{w} - \frac{1}{|h_i|^2} \right)^+, i = 1, 2, ..., N \tag{4}$$

where $(\cdot)^+$ denotes $\max\{\cdot, 0\}$ and $\widehat{w}$ is determined by

$$\sum_{i=1}^{L} \left(\widehat{w} - \frac{1}{|h_i|^2} \right)^+ = P_t \tag{5}$$

The algorithm to obtain the optimal solution is the well-known water-filling, where the so-called water-level $\widehat{w}$ is calculated based on (5). The other category of power allocation problem can be formulated as optimizing the power allocation across the subcarriers such that the required transmit power is minimized while a given target rate is satisfied. Assume the target rate is R_t. The optimal power allocation has the same expression of (4) while the water-level $\widehat{w}$ is determined by

$$\prod_{i=1}^{L} \left(1 + |h_i|^2 \left(\widehat{w} - \frac{1}{|h_i|^2} \right)^+ \right) = 2^{R_t} \tag{6}$$

The algorithm to obtain the optimal solution is also the water-filling, where $\widehat{w}$ is calculated based on (6). An example of the result of the conventional water-filling is shown in Fig. 4 where $L = 15$.

3.2 Power allocation in OFDM-based cognitive radio systems

The power allocation problem in OFDM-based cognitive radio should consider the subchannel transmit power constraints in addition to the sum transmit power constraint or the target rate. Assume $F_j \triangleq \sum_{i=m_j}^{m_{j+1}-1} P_i$ is the power allocated to the jth subchannel[1]. The power allocation problem with the sum transmit power constraint P_t can be formulated as the following optimization problem:

$$
\begin{aligned}
P^* = \quad &\arg\max \sum_{i=1}^{N} \log(1 + |h_i|^2 P_i) \\
\text{s.t.} \quad &P_i \geq 0 \qquad\qquad i = 1, 2, ..., N \\
&\sum_{i=1}^{N} P_i \leq P_t \\
&F_j \leq G_j \qquad\qquad j = 1, 2, ..., M.
\end{aligned}
\tag{7}
$$

Similarly, the power allocation problem with the target rate R_t can be formulated as

$$
\begin{aligned}
P^* = \quad &\arg\max \sum_{i=1}^{N} P_i \\
\text{s.t.} \quad &P_i \geq 0 \qquad\qquad i = 1, 2, ..., N \\
&\sum_{i=1}^{N} \log(1 + |h_i|^2 P_i) = R_t \\
&F_j \leq G_j \qquad\qquad j = 1, 2, ..., M.
\end{aligned}
\tag{8}
$$

Note that the SU also suffers interference caused by the PU. The interference can be regarded as white noise by the SU, which may lead to different noise levels in different subchannels. However, although the noise power in each subchannel is set to be equal to each other, this fact does not affect our problem formulation since we can rescale the channel gain. Obviously, the above two optimization problems are convex optimization problems with linear constraints, which can be numerically solved by the standard optimization algorithms. However, we intend to develop analytical as well as more efficient algorithms rather than numerical methods in the following two subsections.

3.3 Iterative Partitioned Water-filling: Sum Transmit Power Constraint

We first consider (7), i.e., the power allocation problem with the sum transmit power constraint. The algorithm is developed by three steps. The fist step is to analyze the structure of the optimal solution of (7) based on the convex optimization theory. Then the algorithm is proposed heuristically based on the analysis in the first step. The algorithm is proved to converge to the optimal point at last. We begin with the first step in this subsection.

3.3.1 Analysis of the structure of optimal solution

In (7), if $P_t \geq \sum_{j=1}^{M} G_j$, the sum transmit power constraint is actually meaningless. (7) is degraded to the problem of power allocation with only subchannel transmit power

---

[1] Here, without considering subcarrier sidelobes, the allocated power is exactly the transmit power.

constraints and the solution is simply water-filling on the subcarriers in each subchannel individually with the corresponding subchannel transmit power constraint. Therefore, we only emphasize on the situation that $P_t < \sum_{j=1}^{M} G_j$. A theorem is first derived to describe the structure of the optimal power allocation vector.

Theorem 1 *Under the assumption that* $P_t < \sum_{j=1}^{M} G_j$, *a power allocation vector* **P** *is the solution for (7) if and only if it satisfies:*

$$P_i = \left(w_j - \frac{1}{|h_i|^2} \right)^+ \tag{9}$$

where i = 1, 2, …,N and j is the index of the subchannel which the ith subcarrier belongs to. Assume $A \triangleq \left\{ j \mid F_j < G_j \right\}, B \triangleq \left\{ j \mid F_j < G_j \right\}$. *Then* w_j *is determined by 1. for j ∈ A:*

$$w_j = \hat{w}, \tag{10a}$$

$$\sum_{j \in A} \sum_{i=m_j}^{m_{j+1}-1} \left(\hat{w} - \frac{1}{|h_i|^2} \right)^+ = P_t - \sum_{j \in B} G_j \tag{10b}$$

2. for j ∈ B:

$$\sum_{i=m_j}^{m_{j+1}-1} \left(w_j - \frac{1}{|h_i|^2} \right)^+ = G_j \tag{11a}$$

$$w_j \le \hat{w}. \tag{11b}$$

Proof: The problem (7) can be reformulated into a standard convex optimization form:

$$
\begin{aligned}
\min \quad & \sum_{i=1}^{N} \log\left(1 + |h_i|^2 P_i \right) \\
\text{s.t.} \quad & -P_i \le 0 && i = 1,2,...,N \\
& \sum_{i=1}^{N} P_i - P_t \le 0 \\
& F_j - G_j \le 0 && j = 1,2,...,M.
\end{aligned}
\tag{12}
$$

The constraint conditions obviously satisfy the Slater's conditions, so the Karush-Kuhn-Tucker(KKT) conditions are sufficient and necessary for the optimal vector **P** (Boyd & Vandenberghe, 2004). The first two KKT conditions are the constraint conditions of (12) and the others are given by

$$\lambda_i \ge 0, \tag{13a}$$

$$\alpha_j \ge 0, \tag{13b}$$

$$v \ge 0, \tag{13c}$$

where i =1,2,...,N, j =1,2,...,M.

$$\lambda_i P_i = 0, \tag{14a}$$

$$\alpha_j \left(F_j - G_j \right) = 0, \tag{14b}$$

$$v \left(\sum_{i=1}^{N} P_i - P_t \right) = 0, \tag{14c}$$

where $i = 1,2,...,N$, $j = 1,2,...,M$.

$$-\frac{1}{P_i + 1 / |h_i|^2} - \lambda_i + v + \alpha_j = 0 \tag{15}$$

where $i = 1,2,...,N$ and j is the index of the subchannel which the ith subcarrier belongs to.

*I. Proof of **only if** part*

(15) can be written as

$$\lambda_i = v - \frac{1}{P_i + 1 / |h_i|^2} + \alpha_j \tag{16}$$

Substituting (16) into (13a) and (14a) yields

$$v + \alpha_j \geq \frac{1}{P_i + 1 / |h_i|^2}, \tag{17}$$

$$\left(v - \frac{1}{P_i + 1 / |h_i|^2} + \alpha_i \right) P_i = 0. \tag{18}$$

If $v + \alpha_j < |h_i|^2$, we have $P_i > 0$ by (17). Then (18) leads to

$$P_i = \frac{1}{v + \alpha_j} - \frac{1}{|h_i|^2}. \tag{19}$$

If $v + \alpha_j \geq |h_i|^2$, from (16), we obtain $v - \frac{1}{P_i + 1/|h_i|^2} + \alpha_i > 0$. Then based on (17), it follows that

$P_t = 0$. From (16), we also have

$$v + \alpha_j > 0. \quad \text{where } j = 1,2,...,M \tag{20}$$

Let $w_j \triangleq 1/(v + \alpha_j)$. Then

$$P_i = \left(w_j - \frac{1}{|h_i|^2} \right)^+, \tag{21}$$

where $i = 1, 2, \ldots, N$ and j is the index of the subchannel which the ith subcarrier belongs to.

The subchannels are divided into two sets A and B. Since we assume $P_t < \sum_{j=1}^{M} G_j$, there must exists at least such one subchannel that $Fj < Gj$, i.e., $A \neq \emptyset$.

1) For all $j \in A$, since $Fj < Gj$, we have $\alpha_j = 0$ based on (14b). Let $\widehat{w} = 1/v$, then Wj satisfies the condition (10a) $w_j = \widehat{w}$. From (20), we have $v > 0$. Consequently, (14c) leads to $\sum_{i=1}^{N} P_i = P_t$. Therefore, we come to the equation (10b)

$$
\begin{aligned}
& \sum_{j \in A} \sum_{i=m_j}^{m_{j+1}-1} \left(0, \widehat{w} - \frac{1}{|h_i|^2} \right)^+ \\
=\ & \sum_{j \in A} P_j \\
=\ & \sum_{i=1}^{N} P_i - \sum_{j \in B} F_j \\
=\ & P_t - \sum_{j \in B} G_j .
\end{aligned}
\tag{22}
$$

2) For all $j \in B$, the condition (11a) is obviously satisfied. Since $\alpha_j \geq 0$, we have

$$
w_j = \frac{1}{v + \alpha_j} \leq \frac{1}{v} = \widehat{w}
\tag{23}
$$

Therefore, the inequality (11b) also holds.
So far, we have proved the *only if* part in Theorem 1.
II. Proof of if part
We need to prove that all the KKT conditions can be derived by the power allocation vector defined in Theorem 1. It is easy to see that the first two KKT conditions, i.e., the constraint conditions in (12), hold inherently.

1) Based on (10b) and A≠∅, we have $\widehat{w} > 0$. Define $v \triangleq 1 / \widehat{w}$, then $v > 0$. From (10b), we can also obtain

$$
\begin{aligned}
& \sum_{i=1}^{N} P_i \\
=\ & \sum_{j \in A} F_j + \sum_{j \in B} F_j \\
=\ & \sum_{j \in A} \sum_{i=m_j}^{m_{j+1}-1} \left(\widehat{w} - \frac{1}{|h_i|^2} \right)^+ + \sum_{j \in B} G_j \\
=\ & P_t
\end{aligned}
\tag{24}
$$

Therefore, (13c) and (14c) both hold.
2) Define $\alpha_j \triangleq 1 / w_j - 1 / \widehat{w}$, from (10a) and (11b), we can conclude that

$$
\begin{cases}
\alpha_j = 0, & j \in A \\
\alpha_j \geq 0. & j \in B
\end{cases}
\tag{25}
$$

Therefore, (13b) holds. Since for $j \in B$, we have $Fj = Gj$, (14b) also holds.

3) Define $\lambda_i \triangleq v - \dfrac{1}{P_i + 1/|h_i|^2} + \alpha_j$, where j is the index of the subchannel which the ith

subcarrier belongs to.Then (15) holds inherently. If $w_j > 1/|h_i|^2$, based on (9), we have

$$P_i = w_j - \frac{1}{|h_i|^2}. \tag{26}$$

Since we can derive $w_j = 1/(v + a_j)$ from the definitions of α_j and v, (26) can be written as

$$v - \frac{1}{P_i + 1/|h_i|^2} + \alpha_j = 0. \tag{27}$$

Then, λ=0. Therefore, given $w_j > 1/|h_i|^2$, (13a) and (14a) hold. On the other hand, if

$w_j \leq 1/|h_i|^2$, it follows that $P_i = 0$ from (9). Then, we have

$$\lambda_i = v - |h_i|^2 + \alpha_j = \frac{1}{w_j} - |h_i|^2 \geq 0. \tag{28}$$

Hence, given $w_j \leq 1/|h_i|^2$ (13a) and (14a) also hold.

In conclusion, we have derived all of the KKT conditions and thus the *if* parts also holds. This completes the proof of Theorem 1. □

Comparing (9) with (4), we can find that the power allocation within each subchannel is the same to the conventional water-filling result. However, the water-levels of different sub-channels may be different in (9). From Theorem 1, we know that the subchannels with the optimal power allocation can be divided into two sets: the set A, i.e., the subchannels whose allocated power is strictly smaller than the corresponding subchannel transmit power constraint and the set B, i.e., the subchannels whose allocated power is equal to the cor-responding subchannel transmit power constraint. For the subchannels in *A,* the allocated power is the result of water-filling on all the subcarriers that belong to these subchannels with the power $P_t - \sum_{j \in B} G_j$. Therefore, all the subchannels in *A* have a *common water-level* W. For each subchannel in B, e.g., subchannel j that $j \in B$, the allocated power is the result of water-filling on the subcarriers that belong to subchannel *j* with the power *Gj*. Therefore, each subchannel in *B* has a *unique water-level* w_j and satisfies $w_j \leq \widehat{w}$, i.e., the water-level of the subchannel which has a *unique water-level* is less or equal to the *common water-level.*
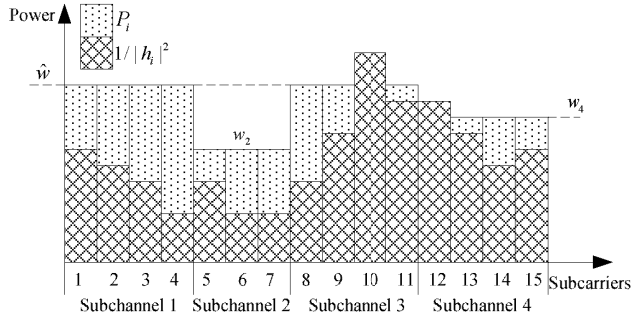
Fig. 5. Optimal power allocation with subchannel transmit power constraints.

A power allocation result satisfying Theorem 1 is shown in Fig. 5. Within each subchannel, it is the same to the conventional water-filling result. However, the water-level of each subchannel is not equal to each other compared to Fig. 4. Subchannel 1 and 3 have the same water-level W, as each allocated power is strictly lower than the corresponding subchannel transmit power constraint. Subchannel 2 and 4 have their unique water-levels w_2 and w_4 respectively, as each allocated power is equal to the corresponding subchannel transmit power constraint. Also, as in Fig. 5, we see that the *unique water-levels w_2, w_4 are lower than the common water-level* w.

3.3.2 Iterative partitioned water-filling

From the above analysis, we can infer that if the partition of the subchannels, i.e., the elements of A and B, is known, the optimal power allocation can be obtained by first water-filling on the subcarriers in each subchannel that belongs to the set B individually with the corresponding subchannel transmit power constraint, and then water-filling on the rest of the subcarriers with the power $P_t - \sum_{j \in B} G_j$. Therefore, the partition of the subchannels plays an important role in the development of the algorithm. An exhaustive search algorithm can be developed intuitively. The basic idea is to consider all the possible partitions and then verify each partition if the conditions in Theorem 1 are satisfied, which is described as follows:

1. Divide the subchannels into two sets, say A and B and there are 2^M kinds of partitions in total.

2. For each partition, perform the conventional water-filling on the subcarriers in each subchannel that belongs to the set B individually with the corresponding subchannel transmit power constraint. Then the water-levels wj where $j \in B$ can be obtained.

3. Remove such partitions that $P_t - \sum_{j \in B} G_j < 0$. Perform the conventional water-filling on the subcarriers in all the subchannels that belong to the set A with the power $P_t - \sum_{j \in B} G_j$. Then the *common water-level w* can be obtained.

4. Verify each partition whether $F_j < G_j$ where $j \in A$ and $w_j \leq \hat{w}$ where $j \in B$ are satisfied. According to Theorem 1, there is only one available partition and the corresponding power allocation vector is the solution.

I. *Initialization:*

$A = \{ j \,|\, j = 1, 2, ..., M \}, B = \varnothing, C = A, \hat{R} = R_t.$

II. *Iterations:*

1. Perform the conventional water-filling on the subcarriers in all the subchannels that belong to the set A with the target rate $\hat{R}$.

2. $F_j = \sum_{i=m_j}^{m_{j+1}-1} P_i$ where $j \in A, C = \{ j \,|\, F_j \geq G_j, j \in A \}$;

3. $A = A \setminus C, B = B \cup C$;

4. Perform the traditional water-filling on the subcarriers in each subchannel j that $j \in C$ individually with the corresponding subchannel transmit power constraint G_j;

5. $R_j = \sum_{i=m_j}^{m_{j+1}-1} \log \left(1 + |h_i|^2 P_i \right)$ where $j \in C$, $\hat{R} = \hat{R} - \sum_{j \in C} R_j$.

III. *End.*

Table 1. The IPW algorithm under Sum Transmit Power Constraint.

However, the complexity of the exhaustive search algorithm is far too high. In the extreme situation, we need to consider 2^M kinds of partitions throughout the algorithm. Therefore, we develop a more efficient algorithm named as *iterative partitioned water-filling* (IPW). The basic idea is to determine the elements of A and B iteratively rather than by exhaustive search. The IPW algorithm is described in Table 1.

In the beginning of each iteration, the conventional water-filling with only sum transmit power constraint P is done on the subcarriers in all the subchannels that belong to the set A. Then those subchannels whose power exceeds its subchannel transmit power constraint are taken out from the set A and then be performed by water-filling with the corresponding subchannel transmit power constraints individually. The iteration stops when each subchannel satisfies its subchannel transmit power constraint. The IPW algorithm is more efficient than the exhaustive search algorithm. In the worst case, the algorithm converges after M iterations.

3.3.3 Proof of optimality

In order to prove the optimality of the IPW, we need to explain that the IPW algorithm converges to the point that satisfies the conditions in Theorem 1. Actually, we only need to prove the inequality (11b) as other conditions are inherently satisfied.

We outline two simple lemmas about the conventional water-filling before the proof. Assume **P** is the power allocation vector of water-filling on N subcarriers with power $\mathbf{1^T P}$ and the water-level is w, where **1** is the column vector of all ones.

Lemma 1 *If we take out n subcarriers from these N subcarriers and the corresponding power allocation vector of these n subcarriers is P_n, P_n is also the power allocation vector of water-filling on the n subcarriers with the power $\mathbf{1^T} P_n$.*

Lemma 2 *If the water-filling is done on these N subcarriers with power P' and the corresponding water-level is w' . We have* $w' \leq w$ *if* $P' \leq 1^T P$ *and vice versa. The two lemmas can be easily obtained from (4) and (5). Based on Lemma 1 and Lemma 2, the following theorem can be proved.*

Theorem 2 *The IPW algorithm converges to the point satisfying wj < w, where j e B, i.e., each unique water level is less or equal to the common water-level.*

Proof: Assume in the fcth iteration, the power $\hat{P}$ used in Step II-1 , the water-level of the subcarriers that belong to the subchannels in the set A after Step II-1 and the temporary set C are denoted as $\hat{P}_k$, w_k and C_k respectively. If we find such j that $Fj \geq Gj$, then the subchannel j needs to be taken out from the set A and put into the set C_k. At Step II-4, the conventional water-filling is done on the subcarriers in each subchannel j that $j \in C_k$ with the corresponding subchannel transmit power constraint Gj individually. The resulting water-levels are just the *unique water-level* w_j when the algorithm converges. From Lemma 1, if the conventional water-filling is done on the subcarriers in each subchannel j that $j \epsilon C_k$ with the power Fj individually, we also get the water-level w_k . From Lemma 2, since $F_j \geq Gj,$ we have

$$w_j \leq \hat{w}_k, \forall j \in C_k \tag{29}$$

after the *k*th iteration.

Based on Lemma 1, for the rest of the subchannels in the set A satisfying $F_j < G_j$, if the conventional water-filling is done on the corresponding subcarriers with the power $P_k - \sum_{j \in C_k} F_j$, we also get the water-level w_h. In the next iteration, we need to perform the conventional water-filling on these subcarriers with the power $\hat{P}_{k+1} = \hat{P}_k - \sum_{j \in C_k} G_j$ and the resulting water-level is assumed to be w_h+_1. Based on Lemma 2, since $\hat{P}_k - \sum_{j \in D} F_j < \hat{P}_{k+1}$ we have

$$\hat{w}_k \leq \hat{w}_{k+1}. \tag{30}$$

When the algorithm converges after $\hat{k}$ iterations, we have $\hat{w} = \hat{w}_{\hat{k}}$. From (30), it follows that $\hat{w}_{\hat{k}} \leq \hat{w}$ where $k = 1, 2,\dots, \hat{k}$. Also, based on (29), since $B = \bigcup_{k=1,2,\dots,\hat{k}} C_k$, we have

$\forall j \in B, \exists k$, so that $w_j \leq \hat{w}_k$. Therefore, we come to the conclusion that $\forall j \in B, w_j \leq \hat{w}$, i.e., each *unique water-level* is less or equal to the *common water-level*. □

So far, we have proved that the IPW algorithm converges to the point satisfying the conditions in Theorem 1 and thus the optimal power allocation can be obtained by the IPW algorithm.

3.4 Iterative Partitioned Water-filling: Target Rate Constraint

In this subsection, we consider (8), i.e., the power allocation problem with the target rate constraint. If the target rate is given, we should first verify if the target rate can be achieved

under the subchannel transmit power constraints. It is easy to see that the maximum rate that can be achieved under the subchannel transmit power constraints is the result of water-filling on the subcarriers in each subchannel with the corresponding subchannel transmit power constraint individually. Therefore, the maximum achievable rate can be expressed as

$$R_{\max} = \sum_{i=1}^{N} \log\left(1 + |h_i|^2 P_i\right), \tag{31}$$

where Pi is determined by

$$P_i = \left(w_j - \frac{1}{|h_i|^2}\right)^+, \tag{32}$$

where $i = 1, 2,\dots, N$ and j is the index of the subchannel which the ith subcarrier belongs to. wj can be determined by

$$\sum_{i=m_j}^{m_{j+1}-1} \left(w_j - \frac{1}{|h_i|^2}\right)^+ = G_j. \, j = 1,2,...,M. \tag{33}$$

Then, if $R_t > R_{max}$, Rt can not be achieved by any power allocation vector under the subchannel transmit power constraints. Also, the optimal power allocation can be determined by (32) and (33) when $R_t = R_{max}$. In the case $R_t < R_{max}$, the algorithm for solving (8) can be developed following the similar derivation as that in Section 3.3. A similar theorem as Theorem 1 can be described as follows.

Theorem 3 *Under the assumption that $R_t < R_{max}$, a power allocation vector $\mathbf{P}$ is the solution for (8) if and only if it satisfies:*

$$P_i = \left(w_j - \frac{1}{|h_i|^2}\right)^+, \tag{34}$$

where $i = 1, 2, \dots ,N$ and j is the index of the subchannel which the ith subcarrier belongs to. Assume $A \triangleq \left\{j \,|\, F_j < G_j\right\}, B \triangleq \left\{j \,|\, F_j = G_j\right\}$. . Then w_j is determined by 1. for $j \in A$:

$$w_j = \widehat{w}, \tag{35a}$$

$$\sum_{j \in A} \sum_{i=m_j}^{m_{j+1}-1} \log\left(1 + |h_i|^2 \left(\widehat{w} - \frac{1}{|h_i|^2}\right)^+\right) \tag{35b}$$

$$= R_t - \sum_{j \in B} R_j,$$

where Rj is determined by

$$R_j = \sum_{i=m_j}^{m_{j+1}-1} \log\left(1 + |h_i|^2 \left(w_j - \frac{1}{|h_i|^2}\right)^+\right).$$

2. *for $j \in B$:*

$$\sum_{i=m_j}^{m_{j+1}-1} \left(w_j - \frac{1}{|h_i|^2}\right)^+ = G_j, \tag{36a}$$

$$w_j \le \widehat{w}. \tag{36b}$$

The proof of Theorem 3 is also similar to that of Theorem 1 and thus is omitted in this paper. Based on Theorem 3, the IPW algorithm for solving (8), which can be developed in a similar way as that in Section 3.3, is described in Table 2.

In each iteration, the conventional water-filling with only target rate constraint $\widehat{R}$ is first done on the subcarriers in all the subchannels that belong to the set *A*. Then those sub-channels whose power exceeds its subchannel transmit power constraint are taken out from the set *A* and put into the set *C*. The conventional water-filling with the corresponding sub-channel transmit power constraints is performed on the subcarriers in each subchannel that belongs to *C* individually. Then the target rate $\widehat{R}$ is updated by subtracting the rates already achieved by the subchannels in the set *C*. The iteration stops when each subchannel satisfies its subchannel power constraint. In the worst case, the algorithm converges after *M* iterations. The proof of the optimality is omitted here as it is also similar to that in Section 3.3.3.

I. *Initialization:*

$A = \{j \mid j = 1, 2, ..., M\}, B = \varnothing, C = A, \widehat{R} = R_t.$

II. *Iterations:*

    1. Perform the conventional water-filling on the subcarriers in all the

       subchannels that belong to the set *A* with the target rate $\widehat{R}$.

    2. $F_j = \sum_{i=m_j}^{m_{j+1}-1} P_i$ where $j \in A, C = \{j \mid F_j \ge G_j, j \in A\}$;

    3. $A = A \setminus C, B = B \cup C$;

    4. Perform the traditional water-filling on the subcarriers in each subchannel *j* that
       $j \in C$ individually with the corresponding subchannel transmit power constraint
       G_j;

    5. $R_j = \sum_{i=m_j}^{m_{j+1}-1} \log\left(1 + |h_i|^2 P_i\right)$ where $j \in C$, $\widehat{R} = \widehat{R} - \sum_{j \in C} R_j$.

III. *End.*

Table 2. The IPW algorithm under Target Rate Constraint.

4. Power Allocation for OFDM-based Cognitive Radio Systems Considering Subcarrier Sidelobes

The *iterative partitioned water-filling* (IPW) algorithm proposed in the above section can obtain the optimal power allocation for OFDM-based cognitive radio systems in the case

where only the interference to the PU caused by the subcarriers inside the corresponding subchannel was considered. However, it is clear that the PU also suffers interference introduced by the subcarrier sidelobes of the neighboring subchannels. Therefore, the assumption in the above section actually implied that there is sufficient guard band between any two subchannels so that the effects of subcarrier sidelobes could be ignored. In this section, we extend the results in the above section and address the power allocation problem in general OFDM-based cognitive radio systems. The effects of subcarrier sidelobes are specially considered in the optimization problem. We propose the power allocation algorithm that maximizes the capacity while satisfying both the sum and subchannel transmit power constraints.

The effects of subcarrier sidelobes in OFDM-based cognitive radio systems were first addressed in (Weiss et al., 2004). The subcarrier sidelobes suppression techniques were further studied in (Cosovic et al., 2005; Pagadarai et al., 2008; Mahmoud & Arslan, 2008). In (Bansal et al., 2007), the authors proposed a power loading algorithm for OFDM-based SU systems whose bands are not overlapped with any PU's. The aim was to maximize the capacity while keeping the interference to the PUs whose licensed bands are in the proximity of the band of SU below a certain threshold.

In the following, we will firstly reformulate the system model and the optimization problem with subcarrier sidelobes in consideration. Then an algorithm will be proposed for a special case with only two non-zero weighted linear inequality constraints. Finally, this algorithm mentioned above will be extended to general cases with multiple non-zero weighted linear inequality constraints, where the extended algorithm is called *recursive power allocation* (RPA) algorithm.

4.1 Power allocation problem reformulation

The basic system model used in this section is the same as that in Fig. 1 and Fig. 2. The SU's transmit power on the PU's licensed channel P_{tx} should still be subject to the power constraint as (1). However, considering the effects of subcarrier sidelobes, it is not proper to require the transmit power in the unavailable subchannel to be zero as (3), since the sidelobes of the subcarriers in other subchannels certainly lead to non-zero transmit power in the unavailable subchannel. So we set a transmit power threshold for the jth subchannel when PUj is detected, say γ_j. The transmit power constraint on the jth subchannel after the sensing G_j is modified from (3) to

$$G_j \begin{cases} \gamma_j & PU_j \text{ is detected} \\ \eta_j (d_j - R_j)^{\beta_j} & PU_j \text{ is not detected} \end{cases} \quad (37)$$

As mentioned in Section 3, in OFDM-based cognitive radio systems, the subchannel transmit power constraints impose further restrictions on the power allocation in addition to the sum transmit power constraint. In fact, the transmit power in one subchannel is comprised of not only the power allocated to the subcarriers inside the subchannel but also the sidelobes power of the subcarriers in other subchannels. We define J_{ij} as the transmit power in the jth subchannel caused by the ith subcarrier with unit power. According to (Weiss et al., 2004), we have

$$J_{i,j} = \frac{1}{\Delta f} \int_{(m_j - \frac{1}{2})\Delta f}^{(m_{j+1} - \frac{1}{2})\Delta f} \left(\frac{\sin\left(\frac{\pi}{\Delta f} \left(f - i\Delta f \right) \right)}{\frac{\pi}{\Delta f} \left(f - i\Delta f \right)} \right)^2 df$$

$$= \int_{m_j - \frac{1}{2}}^{m_{j+1} - \frac{1}{2}} \left(\frac{\sin\left(\pi(x - i) \right)}{\pi(x - i)} \right)^2 dx. \tag{38}$$

Assume P_i be the power allocated to the ith subcarrier. Then the transmit power on the jth subchannel can be expressed as $\sum_{i=1}^{N} J_{i,j} P_i$. Therefore, the subchannel transmit power constraints in (37) introduce M inequality constraints on the power allocation:

$$\sum_{i=1}^{N} J_{i,j} P_i \leq G_j, \quad j = 1, 2, ..., M. \tag{39}$$

Then, the original optimization problem as (7) for optimal power allocation in OFDM-based cognitive radio systems can be modified as the following one:

$$P^* = \arg\max \sum_{i=1}^{N} \log\left(1 + |h_i|^2 P_i \right),$$

$$s.t. \quad P_i \geq 0 \qquad i = 1, 2, ..., N$$

$$\sum_{i=1}^{N} P_i \leq P_t \tag{40}$$

$$\sum_{i=1}^{N} J_{i,j} P_i \leq G_j \quad j = 1, 2, ..., M$$

where $\mathbf{P}^*$ is the optimal power allocation vector, h_i is the ith subcarrier's channel gain, and Pt is the maximum transmit power of the SU transmitter.

Section 3, the subchannel power constraints are approximated by ignoring the subcarrier sidelobes, i.e.,

$$\begin{cases} J_{i,j} = 1 & i \in \left[m_j, m_{j+1} - 1 \right], \\ J_{i,j} = 0 & j \notin \left[m_j, m_{j+1} - 1 \right]. \end{cases} \tag{41}$$

When (41) is satisfied, it can be seen that the M subchannel power constraints in (40) are actually decoupled with each other in terms of the optimization variables. Based on this characteristic, we manage to find an efficient algorithm named iterative partitioned water-filling. In the case where the subcarrier sidelobes are considered, however, the problem turns to be more complicated.

It is clear that (40) is a convex optimization problem with linear inequality constraints. (40) can be expressed in a compact form:

$$P^* = \arg\max \sum_{i=1}^{N} \log\left(1 + |h_i|^2 P_i\right),$$

$$s.t. \quad P_i \geq 0 \qquad i = 1, 2, ..., N \tag{42}$$

$$\sum_{i=1}^{N} J_{i,j} P_i \leq G_j \quad j = 1, 2, ..., M$$

where $J_{i,0} = 1$ for any $i = 1, 2, \dots, N$ and $G_o = P_t$. While the N positivity constraints are easy to handle, the remaining $M + 1$ constraints, which are named *non-zero weighted linear inequality constraints* for ease of description, introduce the main difficulties for the algorithm development.

In order to find the power allocation algorithm for (42), we start with the problem with only two *non-zero weighted linear inequality constraints*. Then the *recursive power allocation* (RPA) algorithm is proposed for the general cases, which is a recursive extension from the above simple problem.

4.2 Power allocation for the case of two non-zero weighted linear inequality constraints

The problem with two *non-zero weighted linear inequality constraints* can be expressed as

$$P^* = \arg\max \sum_{i=1}^{N} \log\left(1 + |h_i|^2 P_i\right).$$

$$s.t. \quad P_i \geq 0 \qquad i = 1, 2, ..., N,$$

$$\sum_{i=1}^{N} J_{i,k} P_i \leq G_k \tag{43}$$

$$\sum_{i=1}^{N} J_{i,l} P_i \leq G_l.$$

The two *non-zero weighted linear inequality constraints* are numbered with C_k and C_l respectively for ease of derivation.

4.2.1 Degraded cases

We begin with solving the problem with C_k excluded. Solving this degraded problem is similar as the derivation of the conventional water-filling algorithm. The solution, which satisfies the constraint C_l with equality, can be expressed as

$$P_i = \left(\frac{1}{J_{i,l}\lambda_l} - \frac{1}{|h_i|^2}\right)^+, \quad i = 1, 2, ..., N \tag{44}$$

where λ_l is determined by

$$\sum_{i=1}^{N} J_{i,l}\left(\frac{1}{J_{i,l}\lambda_l} - \frac{1}{|h_i|^2}\right)^+ = G_l.$$

Substituting (44) into the excluded constraint C_k, if C_k is also satisfied, i.e.,

$$\sum_{i=1}^{N} J_{i,k} \left(\frac{1}{J_{i,l}\lambda_l} - \frac{1}{|h_i|^2} \right)^+ \leq G_k, \tag{45}$$

then (44) is also the solution of the original problem (43).

If (45) is not satisfied, we turn to consider the problem with C_i excluded. Similarly, the solution for this degraded problem can be expressed as

$$P_i = \left(\frac{1}{J_{i,k}\lambda_k} - \frac{1}{|h_i|^2} \right)^+, \quad i = 1, 2, ..., N, \tag{46}$$

where λ_k is determined by

$$\sum_{i=1}^{N} J_{i,k} \left(\frac{1}{J_{i,k}\lambda_k} - \frac{1}{|h_i|^2} \right)^+ = G_k.$$

If the constraint C_l is also satisfied with (46), i.e.,

$$\sum_{i=1}^{N} J_{i,l} \left(\frac{1}{J_{i,k}\lambda_k} - \frac{1}{|h_i|^2} \right)^+ \leq G_l, \tag{47}$$

then (46) is also the solution of the original problem (43).

Therefore, the optimal power allocation for (43) is easy to find in the above two degraded cases, i.e., either (45) or (47) is satisfied.

4.2.2 Algorithm for general cases

In order to obtain the solution in the case that neither (45) nor (47) is satisfied, we first propose the following lemma.

Lemma 3 *Define a convex optimization problem (**OP$_0$**) with differentiate objective function and inequality constraints satisfying Slater's conditions:*

$$\min_{x} \ h_0(x)$$
$$\text{s.t.} \quad h_i(x) \leq 0, i = 1, ..., n. \tag{48}$$

*and let x_0 be any optimal point of **OP$_0$**. Assume x_k is any optimal point of a sub-problem (**OP$_{-k}$**) of **OP$_0$** with $n - 1$ inequality constraints $h_i(x) \leq 0, i = 1, ..., k - 1, k + 1, ..., n$. Then if $h_k(x_k) \geq 0$, we have $h_k(x_0) = 0$.*

Proof: If $h_k(x_0) \neq 0$, then we have $h_k(x_0) < 0$. Since x_0 is an optimal point of the convex optimization problem **OP$_0$**, x_0 satisfies all the KKT conditions of **OP$_0$**, which are given by

$$h_i(x_0) \leq 0, \ \alpha_i \geq 0, \ \alpha_i h_i(x_0) = 0, \ i = 1, ..., n \tag{49a}$$

$$\nabla h_0(x_0) + \sum_{i=1}^{n} \alpha_i \nabla h_i(x_0) = 0, \tag{49b}$$

where α_i is the Lagrange multiplier. Since $h_k(x_0) < 0$ according to $\alpha_k h_k(x_0) = 0$ in (49a), we have $\alpha_k = 0$. Then (49b) can be expressed as

$$\nabla h_0(x_0) + \sum_{i=1, i \neq k}^{n} \alpha_i \nabla h_i(x_0) = 0. \tag{50}$$

Based on (49a) and (50), it is clear that x_0 also satisfies the KKT conditions of $\mathbf{OP_{-k}}$. Therefore, x_0 is also an optimal point of $\mathbf{OP_{-k}}$. Consequently, we know that there is an

Step-1. Calculate the optimal power allocation $\mathbf{P^*}$ with C_k excluded based on (44);
Step-2. Check if $\mathbf{P^*}$ also satisfies C_k. If C_k is satisfied, exit;
Step-3. Calculate the optimal power allocation $\mathbf{P^*}$ with C_l excluded based on (46);
Step-4. Check if $\mathbf{P^*}$ also satisfies C_l. If C_l is satisfied, exit;
Step-5. Calculate the optimal power allocation $\mathbf{P^*}$ based on (53).

Table 3. Power allocation algorithm for the problem including two non-zero weighted linear inequality constraints.

optimal point of $\mathbf{OP_{-k}}$ satisfying $h_k(x_k) < 0$, which contradicts the premise that $h_k(x_k) > 0$. Finally, we come to the conclusion that $h_k(x_0) = 0$.□
Based on Lemma 3, the optimal solution of (43) must satisfy the two constraints C_k and Ci with equalities in the case that neither (45) nor (47) holds. Therefore, solving (43) in this case is equivalent to solving the following problem:

$$P^* = \arg\max \sum_{i=1}^{N} \log\left(1 + |h_i|^2 P_i\right).$$
$$s.t. \quad P_i \geq 0 \qquad i = 1, 2, ..., N,$$
$$\sum_{i=1}^{N} J_{i,k} P_i = G_k \tag{51}$$
$$\sum_{i=1}^{N} J_{i,l} P_i = G_l.$$

To find the solution, we form the Lagrangian:

$$L_i(P_i, \lambda_k, \lambda_l) = \log\left(1 + |h_i|^2 P_i\right) - \lambda_k \left(\sum_{i=1}^{N} J_{i,k} P_i - G_k\right) - \lambda_l \left(\sum_{i=1}^{N} J_{i,l} P_i - G_l\right), \tag{52}$$
$$i = 1, 2, ..., N.$$

The optimal power allocation can be obtained by solving the first order KKT condition $\frac{\partial L_i(P_i, \lambda_k, \lambda_l)}{\partial P_i} = 0$ with the constraint $P_i > 0$, which is given by

$$P_i = \left(\frac{1}{J_{i,k}\lambda_k + J_{i,l}\lambda_l} - \frac{1}{|h_i|^2}\right)^+, \quad i = 1, 2, ..., N \tag{53}$$

where λ_k and λ_l are determined by the following equations:

$$\begin{cases} \sum_{i=1}^{N} J_{i,k} \left(\frac{1}{J_{i,k}\lambda_k + J_{i,l}\lambda_l} - \frac{1}{|h_i|^2}\right)^+ = G_k, \\ \\ \sum_{i=1}^{N} J_{i,l} \left(\frac{1}{J_{i,k}\lambda_k + J_{i,l}\lambda_l} - \frac{1}{|h_i|^2}\right)^+ = G_l. \end{cases}$$

Combining the derivations above, the power allocation algorithm for the problem with two *non-zero weighted linear inequality constraints* is shown in Table 3.

4.3 Recursive power allocation algorithm for general cases

The above algorithm can be further extended to the problem (42) including $M + 1$ *nonzero weighted linear inequality constraints.* Similarly, we first consider $M + 1$ degraded cases each of which excludes one *non-zero weighted linear inequality constraint.* Each of the $M + 1$ degraded optimization problems can be solved by the algorithm for the problem with M *nonzero weighted linear inequality constraints.* If the solution of one of these $M + 1$ degraded

---

Initialization:

    Number the $M+1$ *non-zero weighted linear inequality constraints* as $C_0, C_1, ... C_M$, m=M+1,A={ $C_0, C_1, ... C_M$ }.

Begin

    P*=PAA(m,A).

end

function:PAA (m,A)

begin

    Case m=1:

        Return $f(A)$

    Case m > 1:

        Initialization:

            Set n=1. Define set B.

        Begin:

            Step-1. $B=A/a_n$,**P***=PAA(m-1,B);

            Step-2. if **P*** also satisfies the constraint a$_n$, go to Step-5;

            Step-3. $N=n+1$. if n≤m,go to Step-1;

            Step-4. **P*** = $f(A)$;

            Step-5. return **P***.

        end

end

Table 4. The RPA algorithm for the problem including $M + 1$ non-zero weighted linear inequality constraints

problems also satisfies the corresponding excluded constraint, the optimal power allocation is obtained. Otherwise, the optimal solution of the original problem must satisfy the $M + 1$ constraints with equalities based on Lemma 3. Similar as that in Section 4.2.2, we form the Lagrangian:

$$L_i\left(P_i, \lambda_0, \lambda_1, ..., \lambda_M\right) = \log\left(1 + |h_i|^2 P_i\right)$$
$$- \sum_{j=0}^{M} \lambda_j \left(\sum_{i=1}^{N} J_{i,j} P_i - G_j\right), \quad i = 1, 2, ..., N. \tag{54}$$

The optimal power allocation can be obtained by solving the first order KKT condition $\frac{\partial L_i \left(P_i, \lambda_0, \lambda_1, ..., \lambda_M \right)}{\partial P_i} = 0$ with the constraint $P_i > 0$, which is given by

$$P_i = \left(\frac{1}{\sum_{j=0}^{M} J_{i,j} \lambda_j} - \frac{1}{\left| h_i \right|^2} \right)^+ , \quad i = 1, 2, ..., N \tag{55}$$

where $\lambda_0, \lambda_1, ..., \lambda_M$ are determined by the following $M + 1$ equations:

$$\sum_{i=1}^{N} J_{i,k} \left(\frac{1}{\sum_{j=0}^{M} J_{i,j} \lambda_j} - \frac{1}{\left| h_i \right|^2} \right)^+ = G_k, \quad k = 1, 2, ..., M.$$

Since the algorithm for the problem including two *non-zero weighted linear inequality constraints* has been obtained in Section 4.2.2, the problem for the general case including multiple *non-zero weighted linear inequality constraints* can be solved in a recursive way, i.e., recursive power allocation (RPA) algorithm.
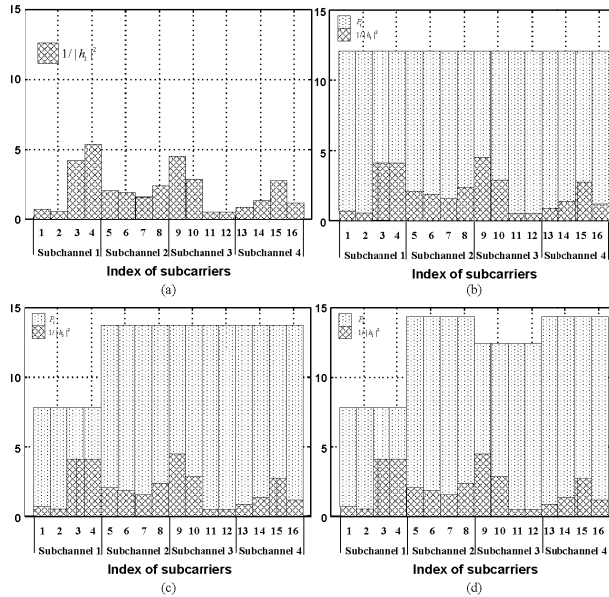

Fig. 6. Process of iterative partitioned water-filling: (a) illustrates the reciprocal of the channel gain of a realization of a frequency selective fading channel; (b), (c) and (d) are the power allocation results after Step II-l (Table 1) of the first, second and third iteration, respectively. The algorithm converges after three iterations.

For ease of derivation, we define the function f (A) where A is a set comprised of all the *non-zero weighted linear inequality constraints* of the optimization problem with the form as (42). f (A) returns the optimal solution that satisfies all the constraints contained in A with equalities. For example, $f(A)$ returns the value determined by (44) when $A = \{C_k\}$ and the value determined by (53) when $A = \{C_k, C_l\}$. We further assume that α_i denotes the ith element of A. The power allocation algorithm for general cases is shown in Table 4.

5. Simulation Results

In this section, two parts of simulation results are presented to verify our proposed algorithms of power allocation for OFDM-based cognitive radio systems. The first part focuses on the

IPW algorithm (as in Table 1)[2] for OFDM-based cognitive radio systems without considering subcarrier sidelobes. The second part focuses on the RPA algorithm (as in Table 4) for OFDM-based cognitive radio systems considering subcarrier sidelobes. ***Part A: IPW algorithm without considering subcarrier sidelobes*** In this part, we set N = 16, M = 4, i.e., each subchannel has 4 subcarriers. The IPW process is shown in Fig. 6. Fig. 6(a) illustrates the reciprocal of the channel gain given a realization of a frequency selective fading channel, i.e., the bottom of the pool for water-filling. The power allocation result after Step in Table 1 in the first iteration is shown in Fig. 6(b), where the power allocation is just the result of the conventional water-filling with only sum transmit power constraint. At Step II-2 in Table 1, we find that the power allocated to Subchannel 1 exceeds its subchannel transmit power constraint G**j..** Therefore, Subchannel 1 should be taken out from the set A and the conventional water-filling is performed on the corresponding subcarriers with the power G_1 at the last step of the first iteration. At Step in Table 1 of the second iteration, the conventional water-filling is performed on the subcarriers in the rest of the subchannels in the set A with the power $P_t - G_1$ and the resulting power allocation is shown in Fig. 6(c). Then the power constraint on Subchannel 3 is found to be destroyed at Step II-2 in Table 1. As a result, we execute Step II-3 and Step II-4 in Table 1 again and then enter the third iteration. Fig. 6(d) illustrates the power allocation after Step II-1 in Table 1 in the third iteration, where only Subchannel 2 and 4 are still remained in the set A. At Step II-2 in Table 1, each subchannel transmit power constraint is found to be satisfied and thus we may end the iterations with C = 0. Finally, the power allocation in Fig. 6(d) is the optimal one that maximizing the sum rate while satisfying both the subchannel transmit power constraints and the sum transmit power constraint. ***Part B: RPA algorithm considering subcarrier sidelobes***

In this part, the frequency selective fading channel is generated according to the Typical Urban (TU) channel model, which is comprised of six paths whose delay parameters are [0.0 0.2 0.5 1.6 2.3 $5.0]\beta s$ and power profile is [0.189 0.379 0.239 0.095 0.061 0.037]. The SU's bandwidth is 5 MHz, which is equally divided into M = 4 subchannels. The number of the subcarriers is supposed to be N = 64 and each subchannel has 16 subcarriers. We use the values 640 for P_t, which implies that the signal to noise ratio is about 10 dB as the noise power on each subcarrier is set to be 1. The four subchannel power constraints are specified as 80, 480, 1.6, 480 respectively. The power constraint on Subchannel 3 is about 25 dB below its neighboring subchannels, which implies that the SU may be very close to or even just

---

[2]Using the IPW algorithm in Table 2 has the similar results as mentioned in Section 3.4.

inside PU_3's protection area. Thus, the constraint on the transmit power in Subchannel 3 should be very rigorous in order to satisfy PU_3's interference power limit. The optimal power allocation is shown in Fig. 7 for a given realization of the TU channel. For comparison, the power allocation result using the IPW algorithm ignoring the effects of subcarrier sidelodes is also given in Fig. 8. Though it can be seen in Fig. 8 that the allocated power on each subchannel is lower than the corresponding subchannel's transmit power constraint, the actual transmit power on Subchannel 3, including the allocated power on Subchannel 3 and the sidelobe power from neighboring subchannels, is about 6 times of its transmit power constraint, which induces to serious destruction to the PU_3's interference constraint. Therefore, the IPW algorithm cannot be applicable for such Subchannel 3 and its neighboring subchannels in this case. On the contrary, the RPA algorithm as shown in Fig. 7 can maximize the capacity while satisfying the subchannel transmit power constraints. Comparing Fig. 7 with Fig. 8, we can see that the power allocated to the subcarriers near Subchannel 3 is suppressed even if the channel states are fine while more power is allocated to the subcarriers that are far away from Subchannel 3. On the other hand, it can also be found that the power allocation in Subchannel 1 is almost the same as that without considering subcarrier sidelobes, which implies that it is reasonable to ignore the effects of subcarrier sidelobes when sufficient guard band between two subchannels is present. Therefore, the IPW algorithm, with higher efficiency compared to the RPA algorithm, can be applied to the subchannels such as Subchannel 1.
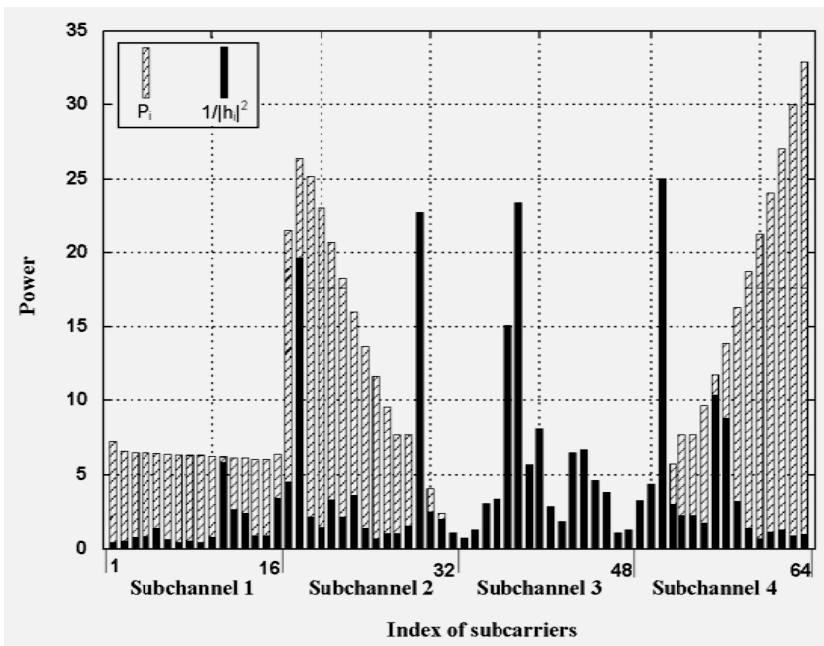

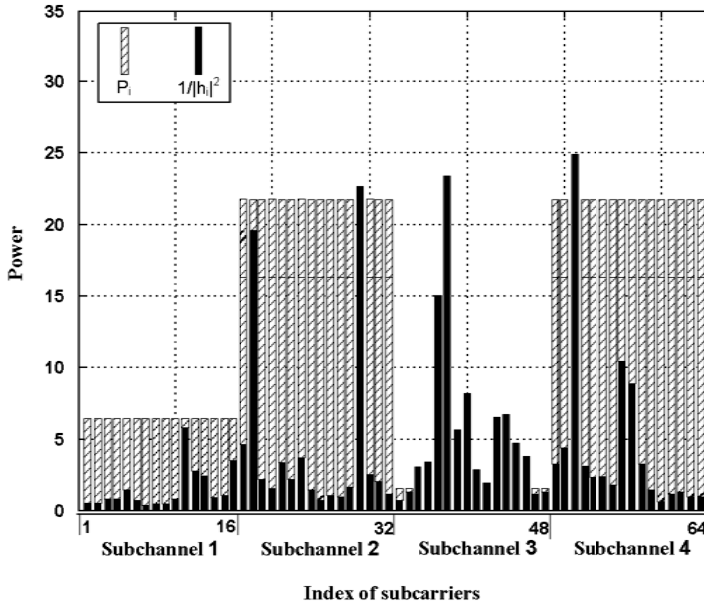Fig. 7. Optimal power allocation by the RPA.

Fig. 8. Optimal power allocation by the IPW without considering subcarrier sidelobes.

6. Conclusions

In this chapter, we study the power allocation problem in OFDM-based cognitive radio systems.

In the cognitive radio scenario, the interference limits of the primary users introduce sub-channel power constraints for the transmission of the SU. Therefore, the power allocation among the subcarriers should satisfy both the sum transmit power constraint, as that in conventional OFDM systems, as well as the subchannel power constraints, which leads to the unavailability of the well-known water-filling algorithm applied in conventional OFDM systems.

In order to find the optimal power allocation, we first propose the IPW algorithms, whose basic computation units is the conventional water-filling. The algorithm is proved to converge to the optimal power allocation, which maximizes the sum rate when the sum transmit power constraint is given or minimizes the required power when the target rate is given, after a finite number of iterations.

Furthermore, we consider the effects of subcarrier sidelobes on the power allocation in practical systems. We note that the actual transmit power in one subchannel is comprised of the allocated power as well as the subcarrier sidelobes power from the neighboring subchannels. The RPA algorithm is proposed for this scenario. The algorithm finds the optimal power allocation recursively by decoupling the subchannel power constraints phase-by-phase. Simulation results show that the IPW algorithm is unavailable when the guard bands between the subchannels are not wide enough, while the RPA algorithm can be

applied in such scenario. In conventional OFDM systems, bit loading is a more practical technique for the system design. Therefore, it is interesting to study bit loading algorithms in OFDM-based cognitive radio systems, where the subchannel transmit power constraints should be considered.[3]

7. References

Federal Communications Commission Spectrum Policy Task Force.(2002). *FCC Report of the Spectrum Eifficiency Working Group,* November 2002.

McHenry, M. (2005). NSF spectrum occupancy measurements project summary, *Shared Spectrum Company Report,* Aug. 2005.

Mitola, J. (1999). Cognitive radio for flexible mobile multimedia communications, *Proc. of IEEE Int. Workshop on Mobile Multimedia Communications,* pp. 3-10, Nov. 1999.

Haykin, S. (2005). Cognitive radio: brain-empowered wireless communications, *IEEE Journal on Selected Areas in Communications,* Vol. 23, pp. 201-220, Feb. 2005.

Q. Zhao & B. Sadler. (2007). A survey of dynamic spectrum access: signal processing, networking, and regulatory policy, *IEEE Signal Processing Magazine,* Vol. 55, No. 5, pp. 2294-2309, May 2007.

Sahai, A.; Hoven, N. & Tandra, R. (2004). Some fundamental limits on cognitive radio, *Proc. of Allerton Conference,* Oct. 2004.

Oner, M. & Jondral, F. (2007). On the extraction of the channel allocation information in spectrum pooling systems, *IEEE Journal on Selected Areas in Communications,* Vol. 25, No. 3, pp. 558-565, Apr. 2007.

Cabric, D.; Mishra, S. & Brodersen, R. (2004). Implementation issues in spectrum sensing for cognitive radios, *Proc. of Asilomar Conference on Signals, Systems and Computers,* pp. 772-776, Nov. 2004.

Zhao, Q.; Tong, L.; Swami, A. & Chen, Y. (2007). Decentralized cognitive MAC for opportunistic spectrum access in Ad Hoc networks: a POMDP framework, *IEEE Journal on Selected Areas in Communications,* Vol. 25, Apr. 2007.

Kim, H. & Shin, K. (2007). Efficient discovery of spectrum opportunities with MAC-Layer sensing in cognitive radio networks, *IEEE Transaction on Mobile Computing,* Vol. 7, No. 5, pp. 533-545, May 2007.

Ghasemi, A. & Sousa, E. (2007). Spectrum sensing in cognitive radio networks: the cooperation-processing tradeoff, *Wiley Wireless Communications and Mobile Computing,* Vol. 7, No. 9, pp. 1049-1060, Nov. 2007.

Zhang, L.; Liang, Y. & Xin, Y. (2008). Joint beamforming and power allocation for multiple access channels in cognitive radio networks, *IEEE Journal on Selected Areas in Communications,* Vol. 26, No. 1, pp. 38-51, Jan. 2008.

Shu, T.; Cui, S. & Krunz, M. (2006). Medium access control for multi-channel parallel transmission in cognitive radio networks, *Proc. IEEE Global Communication Conference,* pp. 1-5, San Francisco, USA, Nov. 2006.

Sudhir, S. & Jafar, S. (2007). Soft sensing and optimal power control for cognitive radio, *Proc. IEEE Global Communication Conference,* pp. 1380 - 1384, Wanshington D.C.,USA, Nov. 2007.

Le, L. & Hossain, E. (2007). QOS-aware spectrum sharing in cognitive radio networks, *Proc. IEEE Global Communication Conference,* pp. 3563 - 3567, Wanshington D.C., USA, Nov. 2007.

Weiss, T. & Jondral, F. (2004). Spectrum pooling: an innovative strategy for the enhancement of spectrum efficiency, *IEEE Communications Magazine,* Vol. 42, pp. S8-S14, March 2004.

Berthold, U. & Jondral, F. (2005). Guidelines for designing OFDM overlay systems, *Proc.of the first IEEE Syposium on New Frontiers in Dynamic Spectrum Access Networks,* November 2005.

TSE, D. & Viswanath, P. (2005). *Fundamentals of Wireless Communication,* Cambridge University Press, May 2005.

Weiss, T.; Hillenbrand, J.; Krohn, A. & Jondral, F. (2004). Mutual interference in OFDM-based spectrum pooling systems, *Proc. IEEE Vehicular Technology Conference Spring,* pp. 1873-1877, Milan, Italy, May. 2004.

Cosovic, I.; Brandes, S. & Schnell, M. (2005). A technique for sidelobe suppression in OFDM systems, *Proc. IEEE Global Communication Conference,* pp. 204-208, St. Louis, USA, Nov. 2005.

Pagadarai, S.; Rajbanshi, R.; Wyglinski, A. & Minden, G. (2008). Sidelobe Suppression for OFDM-based Cognitive Radios Using Constellation Expansion, *Proc. IEEE Wireless Communications and Networking Conference,* pp. 888-893, Las Vegas, USA, Apr. 2008.

Mahmoud, H. & Arslan, H. (2008). Sidelobe suppression in OFDM-based spectrum sharing systems using adaptive symbol transition, *IEEE Communications letters,* Vol. 12, No. 2, pp. 133-135, Feb. 2008.

Bansal, G.; Hossain, J. & Bhargava, V. (2007). Adaptive power loading for OFDM-based cognitive radio systems, *Proc. IEEE International Communication conference,* pp. 5137-5142, Glasgow, Scotland, Jun. 2007.

Boyd, S. & Vandenberghe, L. (2004). *Convex Optimization,* Cambridge University Press

Federal Communications Commission Spectrum Policy Task Force.(2002). *FCC Report of the Spectrum Eifficiency Working Group,* November 2002.

McHenry, M. (2005). NSF spectrum occupancy measurements project summary, *Shared Spectrum Company Report,* Aug. 2005.

Mitola, J. (1999). Cognitive radio for flexible mobile multimedia communications, *Proc. of IEEE Int. Workshop on Mobile Multimedia Communications,* pp. 3-10, Nov. 1999.

Haykin, S. (2005). Cognitive radio: brain-empowered wireless communications, *IEEE Journal on Selected Areas in Communications,* Vol. 23, pp. 201-220, Feb. 2005.

Q. Zhao & B. Sadler. (2007). A survey of dynamic spectrum access: signal processing, networking, and regulatory policy, *IEEE Signal Processing Magazine,* Vol. 55, No. 5, pp. 2294-2309, May 2007.

Sahai, A.; Hoven, N. & Tandra, R. (2004). Some fundamental limits on cognitive radio, *Proc. of Allerton Conference,* Oct. 2004.

Oner, M. & Jondral, F. (2007). On the extraction of the channel allocation information in spectrum pooling systems, *IEEE Journal on Selected Areas in Communications,* Vol. 25, No. 3, pp. 558-565, Apr. 2007.

Boyd, S. & Vandenberghe, L. (2004). *Convex Optimization,* Cambridge University Press.

Cabric, D.; Mishra, S. & Brodersen, R. (2004). Implementation issues in spectrum sensing for cognitive radios, *Proc. of Asilomar Conference on Signals, Systems and Computers,* pp. 772-776, Nov. 2004.

Zhao, Q.; Tong, L.; Swami, A. & Chen, Y. (2007). Decentralized cognitive MAC for opportunistic spectrum access in Ad Hoc networks: a POMDP framework, *IEEE Journal on Selected Areas in Communications,* Vol. 25, Apr. 2007.

Kim, H. & Shin, K. (2007). Efficient discovery of spectrum opportunities with MAC-Layer sensing in cognitive radio networks, *IEEE Transaction on Mobile Computing,* Vol. 7, No. 5, pp. 533-545, May 2007.

Ghasemi, A. & Sousa, E. (2007). Spectrum sensing in cognitive radio networks: the cooperation-processing tradeoff, *Wiley Wireless Communications and Mobile Computing,* Vol. 7, No. 9, pp. 1049-1060, Nov. 2007.

Zhang, L.; Liang, Y. & Xin, Y. (2008). Joint beamforming and power allocation for multiple access channels in cognitive radio networks, *IEEE Journal on Selected Areas in Communications,* Vol. 26, No. 1, pp. 38-51, Jan. 2008.

Shu, T.; Cui, S. & Krunz, M. (2006). Medium access control for multi-channel parallel transmission in cognitive radio networks, *Proc. IEEE Global Communication Conference,* pp. 1-5, San Francisco, USA, Nov. 2006.

Sudhir, S. & Jafar, S. (2007). Soft sensing and optimal power control for cognitive radio, *Proc. IEEE Global Communication Conference,* pp. 1380 - 1384, Wanshington D.C., USA, Nov. 2007.

Le, L. & Hossain, E. (2007). QOS-aware spectrum sharing in cognitive radio networks, *Proc. IEEE Global Communication Conference,* pp. 3563 - 3567, Wanshington D.C., USA, Nov. 2007.

Weiss, T. & Jondral, F. (2004). Spectrum pooling: an innovative strategy for the enhancement of spectrum efficiency, *IEEE Communications Magazine,* Vol. 42, pp. S8-S14, March 2004.

Berthold, U. & Jondral, F. (2005). Guidelines for designing OFDM overlay systems, *Proc.of the first IEEE Syposium on New Frontiers in Dynamic Spectrum Access Networks,* November 2005.

TSE, D. & Viswanath, P. (2005). *Fundamentals of Wireless Communication,* Cambridge University Press, May 2005.

Weiss, T.; Hillenbrand, J.; Krohn, A. & Jondral, F. (2004). Mutual interference in OFDM-based spectrum pooling systems, *Proc. IEEE Vehicular Technology Conference Spring,* pp. 1873-1877, Milan, Italy, May. 2004.

Cosovic, I.; Brandes, S. & Schnell, M. (2005). A technique for sidelobe suppression in OFDM systems, *Proc. IEEE Global Communication Conference,* pp. 204-208, St. Louis, USA, Nov. 2005.

Pagadarai, S.; Rajbanshi, R.; Wyglinski, A. & Minden, G. (2008). Sidelobe Suppression for OFDM-based Cognitive Radios Using Constellation Expansion, *Proc. IEEE Wireless Communications and Networking Conference,* pp. 888-893, Las Vegas, USA, Apr. 2008.

Mahmoud, H. & Arslan, H. (2008). Sidelobe suppression in OFDM-based spectrum sharing systems using adaptive symbol transition, *IEEE Communications letters,* Vol. 12, No. 2, pp. 133-135, Feb. 2008.

Bansal, G.; Hossain, J. & Bhargava, V. (2007). Adaptive power loading for OFDM-based cognitive radio systems, *Proc. IEEE International Communication conference,* pp. 51375142, Glasgow, Scotland, Jun. 2007.

Resource Management of Next Generation Networks using Cognitive Radio Networks

Benon Kagezi Muwonge[1] and H. Anthony Chan[2]
[1]*Coucil of Scientific and Industrial Research*
Johannesburg, South Africa
[2]*Huawei Technologies, Plano, Texas*
USA
Emails: bkagezi@csir.co.za; h.a.chan@ieee.org

1. Introduction

Network resource management is a vastly wide subject that has been studied extensively in many different networks. This chapter studies resource management with respect to cognitive radio networks.

More specifically, this chapter discusses resource management in cognitive radio networks in multimedia enabled Next Generation Networks. At the time of this writing, an emerging multimedia platform being developed is the IP Multimedia Sub-System (IMS). We use the IMS architecture to show how cognitive radio technologies may be incorporated in Next Generation Networks to support multimedia platforms. This is a new area of focus, and we anticipate increase in interest in this area as cognitive radio technologies become a reality. This chapter proposes several resource management scenarios, and highlights open research areas in regards to resource management in cognitive enabled Next Generation Networks.

Delivery of content by cognitive radio technologies will be critical to the successful implementation of cognitive radios in next generation networks. Current trends show that Next Generation Networks will be dominated by multimedia traffic. It is therefore of importance to determine how and where in the architecture of Next Generation Networks cognitive radio technologies will be most effective. Since cognitive radio technologies use white spaces that appear randomly in already assigned spectrum, the manner in which the white space is used is critical to Quality of Service, QoS of the traffic being transported.

The IP Multimedia Subsystem (IMS) is seen as the answer to the much talked-about convergence of data and telecommunication services. The original IMS design was by the 3rd Generation Partnership Project (3GPP) (3GPP TS 23.221, 2007) for delivering IP Multimedia services to end users, using telecommunication infrastructure. 3GPP is a collection of Telecommunication companies. The group's development of the IMS is seen as a way for core network carriers not losing their customers to the fast developing Internet technology.

Cognitive radio technology hopes to make use of unused licensed spectrum, as the secondary user, without interfering with the primary user. While most work has concentrated on predicting of white space in spectrum by a cognitive radio, the concurrent access of multiple cognitive radios still needs more work. Some proposals have proposed cooperation amongst cognitive radios. Should there be hundreds of even thousands of cognitive radios seeking white space conflict resolution should not be a time consuming endeavour. While a solution can be reached in conflict resolution, the appearance of the white space is random, and accessing it should be as soon as it appears.

It should also be expected that primary spectrum users will prefer to have control as to *how*, *when* and *which* cognitive radios access white space in their spectrum. The how: Primary users should be able to determine the rate of transmission in the accessed white space. When: Should the cognitive radio find the white space, how long should it wait before accessing it, and when to exit the white space. Which: Depending on individual behaviours of cognitive radios, primary users should be able to accept or reject a request for access to white space on its spectrum. By addressing the how, when and which concerns of the primary users, management of several cognitive radios in different networks should be more feasible.

The use of cognitive radios in IMS NGN architectures can then be classified as an optimization problem. For optimization, the question would be: At what transmission rate, for how long and which channel/s should a cognitive radio use without interfering with primary bandwidth user/s for best QoS?

The Resource Admission and Control Sub-System, RACS, module in the TISPAN architecture may offer a solution for both fast access of the white space, and managing it as a resource amongst several cognitive radios. We develop our proposal from the TISPAN IMS architecture.

We propose a framework for cognitive radio based resource admission and control based on the RACS architecture that allows for end users to use cognitive radios to search for temporarily unused licensed spectrum. The proposed framework allows for resource admission schemes to operate without need for the primary users to adapt their network usage as a result of spectrum access by secondary users. The framework reduces the need for sensing spectrum usage by primary users, by using capabilities of the RACS to specify a preferable period in which to search for spectrum and on which channels, thereby reducing spectrum sensing time. The major cognitive requirement of the radio will be in predicting when and where the white spectrum will appear within the allotted time period of sensing, and how long it will last.

We consider a scenario in an access network with several wireless communication networks, e.g WLAN, UMTS, GPRS CDMA etc. We assume that each of these technologies has an end-to-end coverage of the communicating users/servers, or connects to an core network technology that can communicate with one of the end users/servers.

The next section reviews some recent work in Next Generation Networks, we then discuss resource management in IMS (Section 3). In Section 4 we discuss traffic engineering issues in IMS enabled networks. In Section 5 we highlight some open research questions in cognitive radio networking in Next Generation Networks with respect to IMS, and propose a framework of resource management in Section 6. Future work on this research and conclusion to this chapter follow in Sections 7 and 8.

2. Recent Developments in Resource Management

This section highlights developments in resource management in cognitive radio networks and next generation wireless networks. Since the area of resource allocation in terms of white space, in cognitive radio networks has not yet been explored by the research community, this section serves as a back ground for resource management in cognitive radio networks. Books on giving comprehensive treatment of resource management in wireless networks can be found in (Cardei et al.,2005) and (Li & Pan, 2006) and citations in this section.

2.1 Resource Management Next Generation Wireless Networks

Resource management is concerned with maintaining a specified QoS in a network, and therefore also concerned with the admission control of new users into a network. The aim of resource management in next generation wireless networks is to allow for the most users into the network without compromising QoS of existing, or already committed resources and predicting future use.

Most work in NGN resource management in wireless networks has been concerned with mobility, and how to maintain connectivity and QoS across networks seamlessly. Over provisioning in wireless networks amounts to wasting of bandwidth. Studies on optimal provisioning of bandwidth have used decision theory (Haas et al., 2000), Game theory (Lin, 2004). Optimal provisioning proposals have had to take into account the rapid change in bandwidth usage of multimedia traffic.

Recent resource provisioning schemes have therefore been adaptive in nature (Nasser & Bejaoui, 2006; Chandramathi, 2008; Lu & Bigham, 2007; Sheu &Wu, 2006; Dharmaraja et al.,2003; Hu &Sharma, 2003). We will not detail the contributions of each of the above references on adaptive schemes since they generally attempt to do the same thing in a similar manner. It is however important to have an understanding of the schemes since resource allocation in cognitive radio networks will need to be as dynamic if not more, to accommodate as many users in a randomly short time period of allocated white space.

In (Haas et al., 2000), a problem formulation of admission control, while maintaining QoS is done using a decision theory approach. Assuming that the network providers provide the utilities and probabilities needed for the decision theory approach, Markov Decision Process is the used to make a decision for admission control. Authors assert that by using the utility and probabilities, it is possible to derive locally optimal policies that fall within the probability measures and utility functions. This would then allow for a trade-off between two contradicting requirements of desired QoS and spectrum utilisation.

Game theory approaches are either cooperative or non-cooperative schemes. Resource management in wireless networks is however modelled as non-cooperative game between the end user and the service provider. This is typically due to their contradicting requirements. For instance, end users want fast access to the network, which requires more bandwidth, while service providers want as many users as possible to maximise of profit. One such proposal is introduced in (Lin, 2004), where a resource management framework that attempts to maximise profits, while taking into account QoS requirements is proposed.

Decision theory and game theory approaches in current network resource management schemes and designs for Next Generation Networks will lay foundation for cognitive network based resource management schemes.

While QoS is important, there have been several publications that have alluded to the importance of the end-user's experience (Mellouk, 2008), (Markaki et al. 2007) while using a service as opposed to traditional QoS. Sometimes referred to as Quality of Experience, QoE, the user's perception of QoS maybe more important than the actual QoS. For example, a user may not mind inter-packet delays while downloading a file, but would rather be concerned with the duration of it takes to download the whole file. It will therefore be important to model user perceptions. In NGN networks, users will be able to choose from several available networks. It will therefore be important to understand user perceptions of a network will be important for network operators, as this will determine how they provision QoS in their networks so as to satisfy their users. In (Pal et al., 2005) a user irritation is modelled using Sigmoid functions. The authors classify users according to how much they pay. Although this method of categorizing is being used in current networks, the modelling of user irritation will be an important aspect for resource management in NGNs.

2.2 Resource Management for Cognitive Radio Networks

Resource management has so far been concerned with spectrum sensing at the physical layer and MAC layer resource allocation on a particular radio. These works have included spectrum sensing, spectrum mobility, spectrum sharing and power control and results from these works have brought about understanding of each of the individual problems. While most work has been directed to solving either of the challenges in cognitive radios separately, it may not be entirely realistic to develop the solutions independent of each other because they are interdependent. For example; a scheme for spectrum sharing may allocate several secondary users to white space on the same channel. The presence of several secondary users will degrade the channel quality, and power control schemes will have to optimise power of the radios to achieve the required QoS of their application, else migrate to another channel using a spectrum mobility scheme. Developments in resource management of white space in cognitive radio networks are now being presented by researchers.

A multi-user resource allocation scheme for delay sensitive application in (Shiang & Schaar, 2009) proposed. The work takes into account time delay restrictions of certain applications, note that it would take it would take too long for global information about the whole network to be learnt. The work therefore employs a multiagent-learning scheme for optimal learning. The approach of designing cognitive radio schemes with applications in mind will prevent the problem of over provisioning of resources.

A distributed scheduling and resource allocation scheme is proposed in (Bazerque , et al.,2008) which allocates power to users so as to maximise the weighted average rate of individual users. The scheme provides a good starting framework for cognitive radio networking. Some of the critical assumptions made in (Bazerque , et al.,2008) such as fairness being defined as guaranteeing minimum requirements for primary users and use of the OFDMA scheme limit the scheme in being used different networks using different modulation schemes as well as asserting limitations to primary users.

3. Resource Management in the IP Multimedia Sub-System

This section details resource management in the RACS. The RACS is responsible for allocating resources. The RACS is made up of Access-Resource and Admission Control Function, ARACF, and Service-based Policy Decision Function, SPDF, shown in Fig. 1. The

two are joined by the Rq interface, which is used for QoS resource reservation information exchange between the SPDF and A-RACF.

In the IMS architecture, an Application Function (AF), that contains multimedia applications being accessed, requests for resource reservations based on the required QoS of the application. The SPDF receives these requests and makes policy decisions, based on the specific access network it is located in. The policy decisions made are then transferred to the ARACF through the Rq interface. The ARACF then implements the resource reservation and admission control.
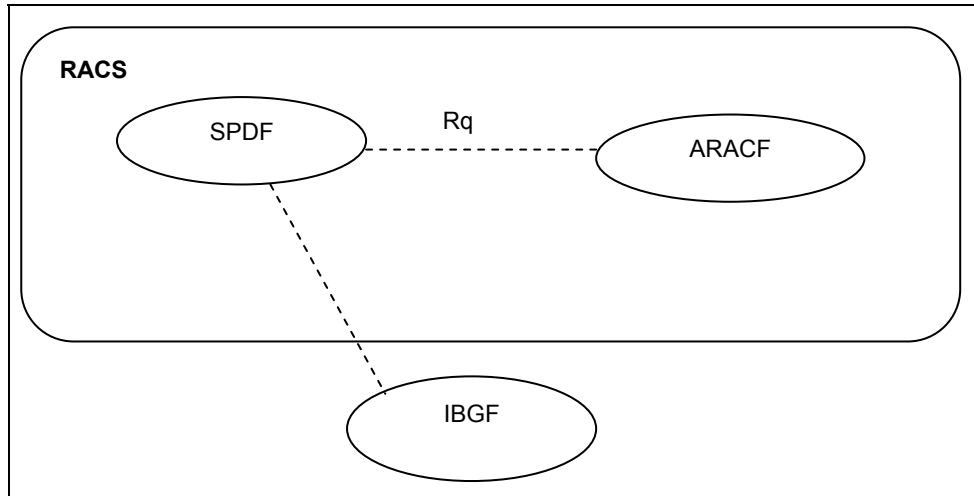
Fig. 1. The RACS architecture based on the TISPAN specification

The communication between the SPDF and ARACF elements is not limited to elements within the same network. One SPDF and ARACF can communicate between several ARACF elements in the same or different networks.

The RACS, a TISPAN NGN specification is meant to provide resource reservation, admission, and policy control to access and core networks. The scope of the RACS is within a given access network to the ingress node of a core network. The RACS therefore has communication interfaces between itself and access nodes, Resource Control and Enforcement Functions (RCEF) and Boarder Gateway Functions (BGF) in the transport layer. The current RACS specification does not take into account the integration of cognitive radio technology at the access nodes and core ingress nodes. The use of cognitive radio technology will allow wireless communication devices to access channels assigned to different access networks and technologies, making it possible for RACS to have access to resources of more than one network. It would also be possible to have several RACS communicate as a result. The RACS can therefore have scope of end-to-end resources across several networks. With available information of the resources available across several networks, better routing strategies would be possible.

4. Traffic engineering issues

While IMS seeks to improve user experience and increase services provided, several issues are still being addressed to make it compatible with existing technologies.

- Increase in Core Traffic

The ease with which services will be deployed in IMS will lead to a tremendous increase in traffic, especially in the core networks. While the strain on current networks due to the growth of the Internet is already a matter of great concern (Willinger, et. Al. 1995) to carrier networks, the implementation of IMS will exacerbate the situation. Increase in core network traffic needs much more efficient signalling schemes than those in current core networks, if QoS for all IMS traffic is to be satisfied.

- Increase in rapidly variable traffic

While to-date, the nature of network traffic has changed significantly from the time voice traffic, traffic characteristics (Willinger, et. al.,1995) will further change due to increased multimedia traffic. The current change of traffic from Poisson modelled traffic to Self-Similar traffic (Willinger, et. Al., 1995) has either led to increased over-provisioning of resources in the network, or reduced QoS. Multimedia traffic is variable in nature, and will thus increase the variance of traffic in the core network. To effectively handle the increased variance of traffic in the core network will require more resource efficient transport protocols, efficient routing schemes and robust Traffic Engineering (TE) solutions.

- Communication across different technologies

For true convergence, where end users can access different services across different networks, probably with different technologies, cognitive radio technologies would have to be used. Cognitive radio technologies would allow for traffic from different technologies to be routed seamlessly. In addition, although different networks with different technologies will co-exist, they will be viewed as one network from the end-user's point view. It our view that communication networks will range from sensor networks, personal area networks to metropolitan networks, all of which have the ability to have IMS applications running on them. Cognitive radio technology will therefore be key to the interoperability of the different technologies.

Specific to the IMS platform, there will be a need for a customised signalling protocol that caters for the following scenarios:

- Several IMS providers

An IMS provider need not be a Telco. Several IMS providers utilising the same IP core can exist on a given network. And an end-user need not be subscribed to a particular network hosting a service in order to access services on its network. Rather, a User Equipment (UE) should be able to specify to access which application service regardless of which network it is on. On a large scale traffic management of the above scenario will pose signalling and routing challenges for core networks. Service Level Agreements (SLA) are needed between the IMS provider and the associated Telco. Appropriate signalling for resource reservation schemes, based on traffic type and length of route to be taken, would have to be used.

- IMS specific application management

Seamless mobility has recently been a key area of research, with some successes and yet some issues still outstanding. With the advent of IMS, an application will (may) have to be accessible in another network should a user move to another network. In addition, should the same application be available in the network moved into, it should be possible to

handover the service from one IMS provider to another. The reason for the application handover would be to maintain the required QoS and/or optimise network resource utilisation. Delay, Jitter and packet loss would change depending on the movement of the end-user. Before the hand-over from one network to another and/or one IMS provider to another, packet routing would have to be re-optimised, based on the new network's status. The re-optimisation of routing would require efficient route discovery schemes across different networks. A generic signalling scheme across heterogeneous networks would be most ideal in such a scenario.

- Different Grades of Service agreements across networks

Across different networks and network types, there will be different Grades of Service (GoS) imposed. Currently internetwork traffic generally experiences poorer end-to-end QoS. Core networks should be able to determine which traffic is more urgent. Traditionally, the network with the lowest capacity or lowest Grades of Service (GoS) would limit the end-to-end QoS, and therefore be a bottleneck the communication. However, NGN IP core networks will be packet based. It will therefore be possible to route a stream of traffic through different routes and even networks, thereby improving end-to-end QoS. Synchronisations of IMS traffic through different routes and networks, and still remain within the strict delay, Jitter and packet loss requirements will be a signalling protocol challenge. To be able to synchronise packets in different networks will require a common signalling protocol between networks, which we propose should be generic in nature.

4. Open Research Areas

The following are open research areas with respect to resource management of cognitive radio networks in Next Generation Networks.

- Signalling protocol:

Current signalling protocols are not appropriate for cognitive radio enabled networks. A signalling protocol in a cognitive radio enabled network should enable the network to:

1. prevent collisions among radios
2. enable cooperation amongst cognitive radios
3. allow for cognitive radios to use available white spaces as efficiently as possible

Traffic behaviour and how access of white spaces by cognitive radios affect traffic modelling

Traffic behaviour has changed significantly from the times of voice communication using the Poisson model. The Poisson model was true for a long while, until telecommunication traffic began including data and multimedia traffic. Recently, traffic has changed from the Poisson model, to bursty traffic, exhibiting self-similar characteristics (Willinger, et. Al 1995). The inclusion of cognitive radios accessing white spectrum will without a doubt have a significant effect on the behaviour of traffic.

While the Poisson model was used for a long while as the traffic model of communication networks long after traffic had changed its behaviour to bursty, we propose that modelling of wireless traffic taking into account access by cognitive radios should be done concurrently with the development of cognitive radio technology. This will go along way in developing adequate resource management methodologies for cognitive radio enabled networks.

- Routing

Cognitive radio networking will not involve end to end resource reservation, due to the dynamic nature of the white spaces. Reservations will most probably only be possible on a link per link basis. Resources for secondary traffic will therefore have to be sought and reserved at every node. This has to be done while maintaining QoS of traffic being transmitted. The choice as to which node traffic should be routed to will depend on how the routing algorithm predicts utilization of a given node or network. Intuitively, secondary traffic should be transmitted into a network with the least utilisation however, delay tolerances of the traffic being transported should not be violated.

- Billing methodology

NGN networks promise single billing for all network access. A billing methodology is still to be developed for cognitive radio enabled networks. Cognitive radio enabled networks will have to develop a billing methodology for the end user so that the user gets billed once. Billing is made difficult by the fact that traffic maybe routed in multiple networks before reaching its destinations. The different SLA in different networks will also complicate accurate billing in cognitive radio networks.

- Classification and priotising of media types and users

Classification of traffic has been done in many traffic engineering solutions. Classification of traffic typically considers parameters such as: delay tolerance, media type, destination, and loss tolerance. In cognitive radio enabled networks, those parameters will have to be used with respect to the characteristics of a particular available white space.

Work needs to be done as to how to classify white spaces that are available randomly, and how to select the most suitable traffic depending on the characteristics of the white space.

5. Proposed Solutions to Some Open Research Areas

In NGNs where end user devices are proposed to have cognitive capabilities, we propose that the RACS controls end user device access to transport networks within its scope. This will be done by giving end user devices time limits within which to make transmissions. Fig. 2 illustrates the proposed architecture. We assume that the two communicating devices are within scope of each of the networks. The control over the use of spectrum by the cognitive radio limits one cognitive radio per time period, thereby reducing conflict in spectrum use by other cognitive radios. This method also lets networks know which cognitive radios are using its spectrum, as well as their individual behaviours. Knowing behaviour of individual radios would allow networks to blacklist any malicious users.
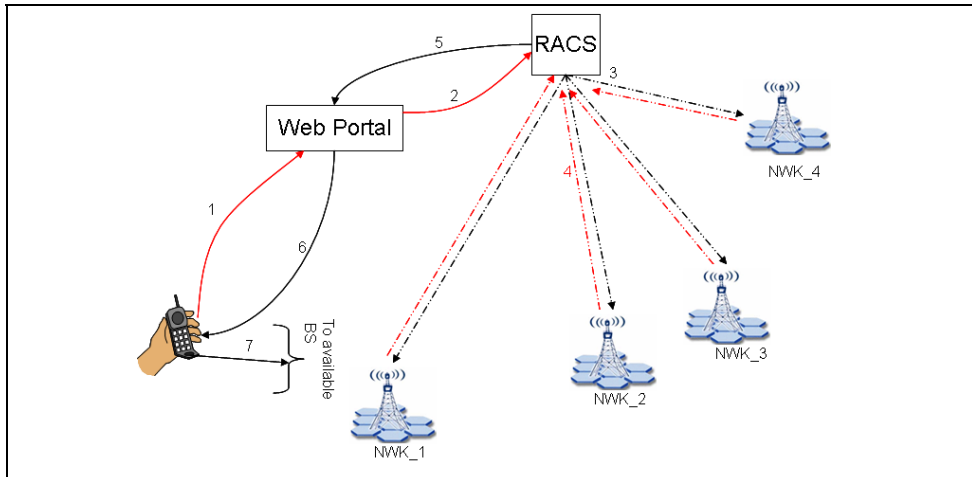
Fig. 2. Proposed use of the RACS for resource allocation in a cognitive radio network

We classify requests from cognitive radio to the Web Portal as Cognitive Radio Requests (CRR) to differentiate them from other bandwidth requests to the RACS. While in normal resource requests, the RACS reserves bandwidth, for CRRs, the RACS instead allocates a time period in which the cognitive radio should have finished or stopped its access to the licensed spectrum.

The sequence of events is as follows:
1. The end user notifies and requests use of licensed spectrum through the web portal.
2. The RACS gets updates from the Web Portal for all requests submitted.
3. For CRRs the RACS sends requests to the base stations within its scope and requests for spectrum access by a cognitive radio.
4. For all networks that accept spectrum access by the cognitive radio, they retain identification of the cognitive radio, and also specify a time period for access to the RACS.
5. The transmission period from each of the networks that accepted CRRs are sent to the Web Portal.
6. The cognitive radio retrieves the information regarding its request.
7. After determining the most optimised transmission solution, the EU transmits in the specified time period in the unused spectrum.

The cognitive radio must send as much data as possible without breaching the time period allocated to it, and without interfering with existing or future spectrum use by primary users. If the time period expires before the cognitive radio finishes transmission, it must request for a new transmission period again.

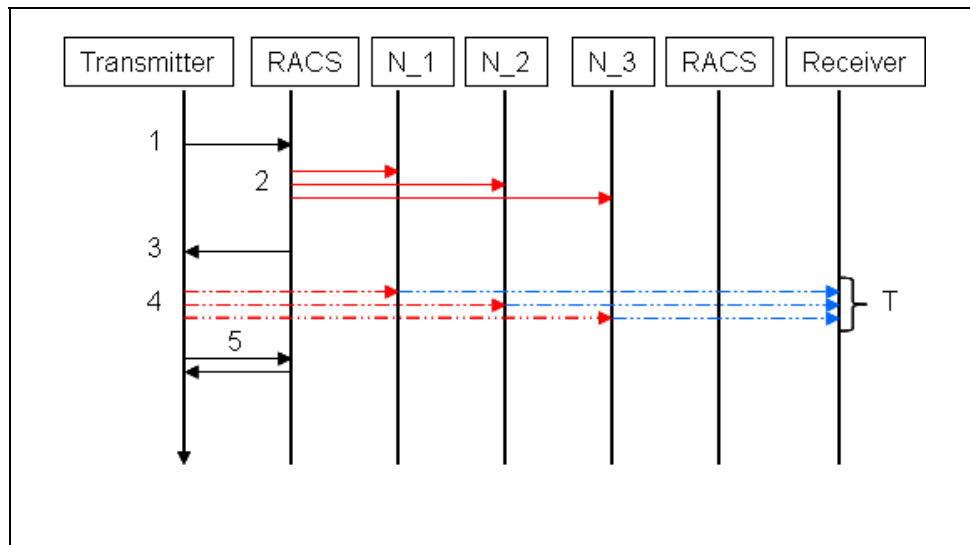We deduce two scenarios from the proposed scheme shown in Fig. 3 and 4.

Fig. 3. A signalling mechanism that extends the scope of RACS in a cognitive radio network

The scenario in Fig. 3 is suitable for real time application, and would take the least time compared to the scenario in Fig. 4. After transmitting, the cognitive radio informs the RACS through the Web Portal.
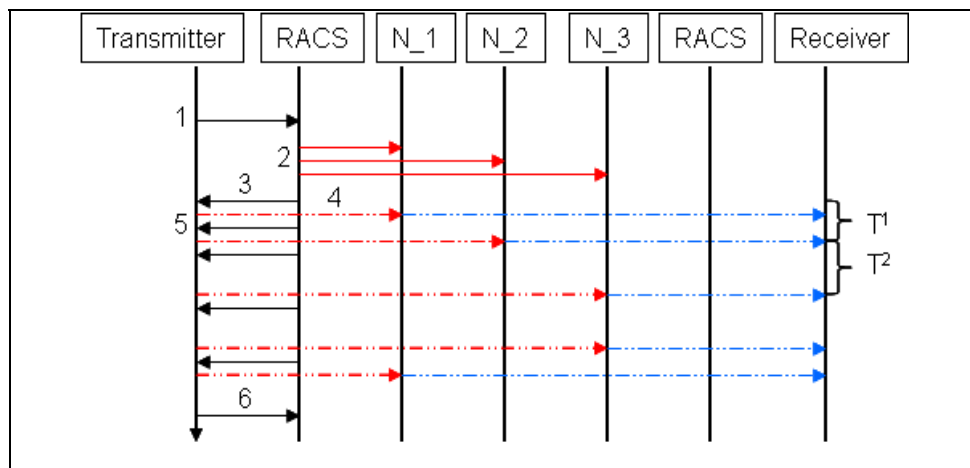

Fig. 4. A more dynamic signalling scheme extending the RACS' scope

The second scenario shown in Fig. 4 requires more signalling and slower than the scenario in Fig. 3. Instead of transmitting to as many networks as accepted, the cognitive radio transmits in the network with the best of two factors: transmitting time period, i.e largest, and least loaded network. However, after each transmission, the cognitive radio must request for new transmission periods. As a result the inter-transmission times vary. The

scheme is therefore not suitable for real time applications. When the cognitive radio finishes transmitting, it must inform the RACS through the Web Portal.

Generally, two main factors to be considered in an NGN cognitive radio based reservation scheme are; number/ types of available networks and capacities of available channels in networks. Since different networks differ in transmission rates temporarily free spectrum gaps in different networks will accommodate different rates of transmission. The greater the number of available networks, the greater the probability of accessing free spectrum. By available networks, we refer to networks that accept cognitive radios to transmit on their spectrum for a specified period.

- Cases for Multiple Cognitive Radio Requests

Partitioning the available transmission times across all the networks to all the available cognitive radios is critical in avoiding conflict during transmission times, and reducing spectrum sensing time. We identify three scenarios in multiple cognitive radio networks:

1. *Scenario* 1: For every cognitive radio request from the Web Portal, give all available transmission time to the first cognitive radio and queue other requests while the first one transmits.
2. *Scenario* 2: Allow the first cognitive radio to take all resources, and reduce transmission times with increase in other cognitive radio requests.
3. *Scenario* 3: For every request, a cognitive radio must identify the RACS through the Web Portal, which predefines a transmission time less than maximum to accommodate future requests.

A. Scenario 1

Scenario 1 would be synonymous with a First in First Out FIFO queuing system for requests. Since each request represents a session from a cognitive radio, priority is then given to the first session over the.

B. Scenario 2

Scenario 2 is a more dynamic type resource reservation scheme. While the first request is given all available transmission times, the allocation of transmission times will be reduced as more requests are allocated, and increased as more resources become available.

C. Scenario 3

In this scenario, the cognitive radio request identifies the type of media and the minimum required transmission rates.
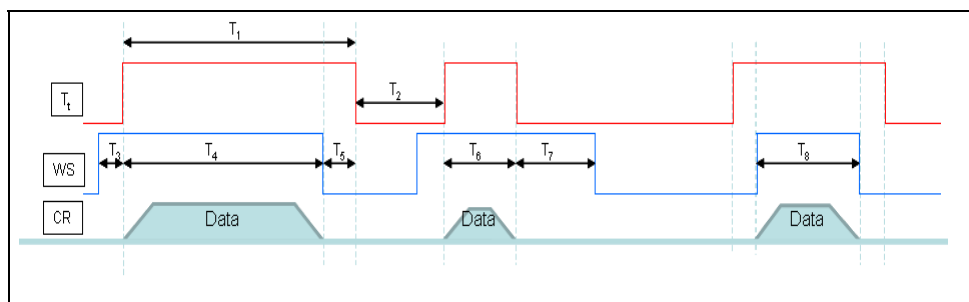

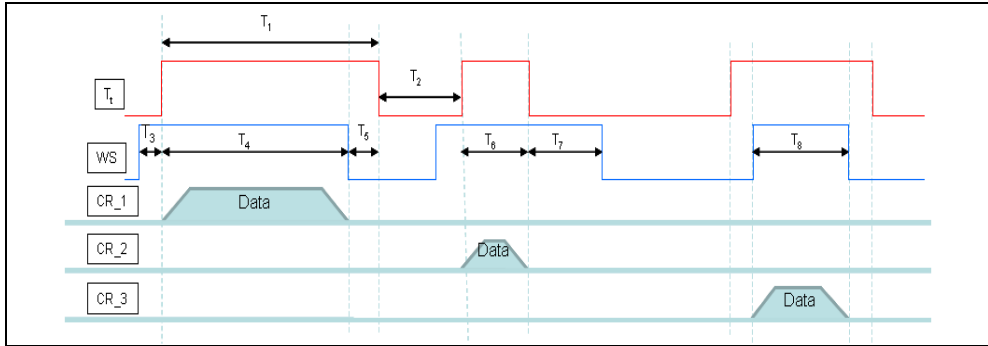Fig. 5. Determining the transmission slot in case of one cognitive radio

Fig. 6. Determining the sharing of time slots in a cognitive radio network

Fig. 5 and 6 show the relationships between the time allowed for by each network for CRs to use their spectrum T_t, the existence of the white spectrum in a given channel WS and transmission times of the CRs. The CRs can only transmit during the allocated transmission times, and when the white spectrum is available. Fig. 5 shows a simple case the first request being served first. And Fig. 6 shows a case where the all requests share available transmission times depending on priority settings in the RACS.

- Determining the transmission time T_t

Consider an access network with N_{nwk} total number of wireless networks, each with n_i channels. For each network and channels, the primary user specifies the time periods Tc for CRs to access the spectrum on given on a channel c. During the Tc the RACS gives permission to specified CRs to transmit data. The CRs then use spectrum sensing to identify white spectrum WS. The CRs then select appropriate transmission times t_{Tt} within both network time periods T_c and white space estimated times WS.

The reservation can be abstracted as an optimisation problem in (1) and (2).The objective functions f(t) and f(x) maximise transmission times and data rates of the cognitive radios. The distinction between transmission times and data rates is due to the different data rates in different network technologies and quality of channels, as such it is taken as a separate function. The RACS receives information of the number of networks and channels per network, Nc that allows access for CRs. Depending on the reservation queuing scheme at the network portal, CRs access the available network channels n_c. A CR in network channel n_i, then determines a transmission time t_{Tt} in the white space and transmission period T_C. Transmission in each white space is restricted to one CR to prevent interference from other CRs. Each request is processed according to a priority, p, given to it by the RACS.

- Maximizing the utility function for transmission time t_{Tt}

The aim is to maximize the transmission periods T_t, in a network of cognitive radios while taking into account the QoS requirements of each of the sessions in each channel n_{ic}. QoS is taken into account by the priority parameter p.

Priority P depends on the following:

     o   Time of request
     o   Position in queue
     o   Type of session, either beginning or session already in session
     o   Delay and jitter tolerance
     o   Packet loss tolerance

The transmission times t_{Tt} must be within the transmission periods T_C for the given channel n_{ic}. The t_{Tt}^{start} and t_{Tt}^{stop} are the starting and end times for Tt.

The selection of a network channel N_{NWKi} depends the quality of the channel, data rates that can be achieved, and network coverage.

$$\max \ f(t) = \sum_{\substack{i_{NT}=1,T_t,\,p=0}}^{\substack{i_{NT}=n_c,\,T_t\leq T_C,\\p=p_{max}}} U_t(n_{ic}, t_{Tt}, P) \tag{1}$$

$$subject \quad to:$$

$$t_{Tt}^{stop} \leq T_C - T_t \qquad t_{Tt} \in T_t$$

$$t_{Tt}^{stop} > t_{Tt}^{start}$$

$$n_c \geq 1,\ N_{NWK} \geq 1,\ n_{ci} \in N_{NWKi},\ i \geq 1$$

- *Maximizing the utility function for data rates in a cognitive radio networks.*

The aim is to maximize the overall data rate in a cognitive radio network with several cognitive radios and several networks with several channels allowing access to their white space. The data rate will depend on the quality of the channel, modulation scheme used on the particular network and loss probability on a given network. At the start of a session, before assignment to a particular channel, the expected amount of data for the session must be estimated, type of data being transmitted, and the delay and jitter tolerance requirements for the session.

$$\max \ f(x) = \sum_{\substack{i_{NT}=1,D_{min},\\r_{CRi}=r_{CR\,min}}}^{\substack{i_{NT}=n_{ic},D_{max},\\r_{CRi}=r_{CR\,max}}} U_D(r_{CRi}, D, n_{ic}) \tag{2}$$

$$subject \quad to:$$

$$D\min \leq D \leq D\max$$

$$D_1 = r_{CRi} * t_{Tt}$$

$$D = \sum_{i=1}^{i=n_c} D_i = \sum_{i=1}^{i=n_c} r_{CRi} * t_{Tt}$$

$$R_{CR} = \sum_{CRi=1}^{CRi=N_{CR}} r_{CRi} = \sum_{CRi=1}^{CRi=N_{CR}} r_{CRi} * n_{ci}$$

The amount of data D to be transmitted for a given session, ranges from a minimum

acceptable data D_{min} for the specific session to meet QoS requirements, to the maximum amount of data D_{max} allowed by the primary user. R_{CR} is the total transmission rate of a cognitive radio, with r_{CRi} being the individual transmission rates per channel.

- Challenges of the Proposed Design

As with most broad proposals, new design challenges surface. We identify the following challenges: Network security, signalling packet overhead and the installation of the proposed cross layer modules.

1. Network security

The proposed signalling scheme shares network resource and traffic information in all connected networks, which maybe a concern to network operators.

2. Signalling packet overhead

Because different networks may use different protocols, signalling packets from one network may need headers that are installed and/or removed at ingress/egress nodes.

6. Future Research

Future work in this area will involve developing a resource management scheme that will take into account QoS requirements of multimedia traffic in the IMS. Due to the location of the RACS in IMS networks, the RACS will be well suited to provide the necessary information required for optimization of resource use by secondary users without interfering with transmission of primary users.

7. Conclusion

Cognitive radio networking should be developed with QoS in mind. By designing schemes that take into account traffic characteristics, better and more dynamic schemes will be developed. This work has highlighted some of the challenges involved in designing cognitive radio enabled networks in Next generation networks. A general framework for developing a resource management scheme for IMS based networks has been introduced.

8. References

3GPP TS 23.221, 2007. 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Architectural requirements (Release 7) http://www.tech-invite.com/

Bazerque, J., Giannakis, G .B. (2008) Distributed Scheduling and Resource Allocation for Cognitive OFDMA Radios, Published online: 23 July 2008, DOI 10.1007/s11036-008-0083-z

Cardei, M., Cardei, I., &Du, D., (2005) Resource management in wireless networking Edition: 2, illustrated Published by Springer, 2005 ISBN 0387238077, 9780387238074 695 pages.

Chandramathi S., Raghuram S.P.P., Srinivas V.S., Satyajit S. H., (2008). Dynamic bandwidth allocation for 3G wireless systems--A fuzzy approach, Applied Soft Computing, Volume 8, Issue 1, January 2008, Pages 274-284, ISSN 1568-4946, DOI: 10.1016/j.asoc.2007.01.006.

Dharmaraja S., Trivedi K. S., Logothetis D. (2003) Performance modeling of wireless networks with generally distributed handoff interarrival times, Computer Communications, Volume 26, Issue 15, Recent Advances in Computer Communications networking, 22 September 2003, Pages 1747-1755, ISSN 0140-3664, DOI: 10.1016/S0140-3664(03)00044-6.

Haas, Z., Halpern, J. Y., Li, L., & Wicker, S. B. (2000). A decision-theoretic approach to resource allocation in wireless multimedia networks. In Proceedings of the 4th international Workshop on Discrete Algorithms and Methods For Mobile Computing and Communications (Boston, Massachusetts, United States, August 11 - 11, 2000). DIALM '00. ACM, New York, NY, 86-95. DOI= http://doi.acm.org/10.1145/345848.345870

Hu, F & Sharma, N. K. (2003). Multimedia call admission control in mobile networks: a dynamical reservation-pool approach, Computer Networks, Volume 43, Issue 3, 22 October 2003, Pages 263-288, ISSN 1389-1286, DOI: 10.1016/S1389-1286(03)00271-8.

Lin, H. (2004). Game Theory Based Resource Management Framework for Revenue Maximization and Differentiated Services in Competitive Wireless Data Networks, PhD dissertation, Dept. of Computer Science and Eng., Univ. of Texas at Arlington, Dec. 2004.

Li, W. & Pan, Y. (2006) Resource Allocation in Next Generation Wireless Networks Edition: illustrated Published by Nova Publishers, 2006 ISBN 1594545839, 9781594545832 346 pages.

Lu, N & Bigham, J. (2007). On utility-fair bandwidth adaptation for multi-class traffic QoS provisioning in wireless networks, Computer Networks, Volume 51, Issue 10, 11 July 2007, Pages 2554-2564, ISSN 1389-1286, DOI: 10.1016/j.comnet.2006.11.012.

Markaki, O.; Charilas, D.; Nikitopoulos, D. (2007) Enhancing Quality of Experience in Next Generation Networks Through Network Selection Mechanisms, Date: 3-7 Sept. 2007, Pages: 1 – 5, Digital Object Identifier 10.1109/PIMRC.2007.4394198

Mellouk, A. (2008) Quality of Service Mechanisms in Next Generation Heterogeneous Networks, ISBN: 978-1-84821-061-5, December 2008, Wiley-ISTE.

Nasser, N. & Bejaoui, T. (2006). Adaptive resource management for cellular-based multimedia wireless networks. In Proceedings of the 2006 international Conference on Wireless Communications and Mobile Computing (Vancouver, British Columbia, Canada, July 03 - 06, 2006). IWCMC '06. ACM, New York, NY, 641-646. DOI= http://doi.acm.org/10.1145/1143549.1143677

Pal, S, Chatterjee, M., & Das, S. K.(2005) A Two-level Resource Management Scheme in Wireless Networks Based on User-Satisfaction. ACM SIGMOBILE Mobile Computing and Communications Review, 9(4):4–14, 2005.

Sheu, T. & Wu, Y. (2006) A preemptive channel allocation scheme for multimedia traffic in mobile wireless networks, Information Sciences, Volume 176, Issue 3, 6 February 2006, Pages 217-236, ISSN 0020-0255, DOI: 10.1016/j.ins.2004.12.002.

Shiang, H. & van der Schaar, M. (2009). Distributed Resource Management in Multihop Cognitive Radio Networks for Delay-Sensitive Transmission, Vehicular Technology, IEEE Transactions on Volume 58, Issue 2, Feb. 2009 Page(s):941 - 953

Shih, D & Lin, Z. (2003) "Bandwidth saturation QoS provisioning for adaptive multimedia in wireless/mobile networks", Published by Elsevier Science B.V. 2003

Wang, J. & Chiang, S. (2004) Adaptive channel assignment scheme for wireless networks, Computers & Electrical Engineering, Volume 30, Issue 6, September 2004, Pages 417-426, ISSN 0045-7906, DOI: 10.1016/j.compeleceng.2004.05.004.

W. Willinger, et. al. (1995), "Self-similarity through high-variability: Statistical analysis of Ethernet LAN traffic at the source level," Proceedings of SIGCOMM '95, pp 100--113, Boston, MA, 1995.

Multi-User Multimedia Transmission overCognitive Radio Networks Using Priority Queuing

Hsien-Po Shiang and Mihaela van der Schaar
University of California Los Angeles
USA

1. Introduction

The emergence of cognitive radio networks have spurred both innovative research and ongoing standards (Mitola et al., 1999; Haykin, 2005; Cordeiro et al., 2006). Cognitive radio networks have the capability of achieving large spectrum efficiencies by enabling *interactive* wireless users to *sense* and *learn* the surrounding environment and correspondingly *adapt* their transmission strategies. In this context, there exist three main challenges for multimedia users to efficiently transmit their delay-sensitive traffic over the cognitive radio networks. The first problem is how these multimedia users should sense the spectrum and timely model the behavior of the primary licensees. The second problem is how these users should manage the available spectrum resources and share the resource to the license-exempt users to satisfy their multimedia traffic requirements while not interfering with the primary licensees. The third problem is how to maintain seamless communication during the transition (hand-off) of selected frequency channels. In this chapter, we focus on the second challenge regarding the resource management problem. For the remaining two challenges, one can find relevant discussions in other existing literatures as in Akyildiz et al., 2006 and Brown, 2005, etc.

Due to the informationally-decentralized nature of cognitive radio networks (Shiang & van der Schaar, 2007b), the complexity of the optimal centralized solutions (Zekavat & Li, 2006; Fu & van der Schaar, 2007; van der Schaar & Fu, 2009) for spectrum allocation is prohibitive for large systems. In addition, the centralized solution might require a large amount of time to process and to collect the required information, which induces delay that can be unacceptable for the delay-sensitive applications, e.g. multimedia streaming. Hence, it is important to implement distributed solutions as in (Shiang & van der Schaar, 2009) for dynamic resource management by relying on the wireless users' capabilities to sense, adapt, and coordinate themselves. Importantly, for the distributed solutions, the coordinated interactions (information exchanges) across the autonomous wireless users are essential, since the decisions of an autonomous wireless user will impact and be impacted by the other users. Without explicit coordinated interactions, the heterogeneous users will consume additional resources and respond slower to the time-varying environment. Such information

exchange can rely on a dedicated control channel for all users (Brik et al., 2005), or using a group-based coordination scheme without a common control channel (Zhao et al, 2005).

In recent years, the research focus regarding multimedia transmission in wireless networks has been to adapt existing multimedia compression (Stockhammer et al., 2003), error protection algorithms (Mohr et al., 2000), and rate-distortion optimized transmission (Chou & Miao 2006) to the rapidly varying resources of wireless networks (see van der Schaar & Chou, 2007 for more references). Significant contributions have been made to enhance the separate performance of the various OSI layers, or jointly for the MAC, PHY, and application layers (van der Schaar et al., 2003; Setton et al. 2005). However, these solutions cannot provide an integrated and realistic cross-layer optimization framework in the cognitive radio networks to support delay-sensitive multimedia streaming applications. Importantly, the cross-layer optimization has been performed in an autonomous, selfish and isolated manner, at each multimedia source/user, and does not consider its impact on the overall wireless infrastructure and the interactions with other information streams. As such, existing solutions do not provide adequate support for multi-user multimedia streaming over spectrum agile network.

In this chapter, we introduce an integrated cross-layer optimization framework for multi-user multimedia transmission over cognitive radio networks. The traffic of the users (including the licensed users and the license-exempt users) and the channel conditions (e.g. Signal-to-Noise Ratio, Bit-Error-Rate) are modeled using stationary stochastic models (Shanker et al., 2005). Based on these models, a novel *priority virtual queue analysis* (Shiang & van der Schaar, 2008) for cognitive radio networks is introduced. This analysis enables a coordination interface to the license-exempt wireless users without requiring to change existing communication protocols, e.g. IEEE 802.22 (Cordeiro et al., 2006). The virtual queues are priority queues for each of the frequency channels. They are emulated in a distributed manner by each autonomous wireless user in order to estimate the delay of selecting a specific frequency channel for transmission. Unlike the majority of prior works assuming the available frequency channels as *spectrum holes* (Haykin, 2005; Akyildiz et al., 2006) that can be accessed using a 2-state on-off channel model (Shanker et al., 2005), we adopt priority queuing to model the users' interactions by assigning the highest priority to the licensed users and adapt the channel access of the license-exempt users (with lower priorities). Importantly, instead of making the primary licensees passively exclude the license-exempt users from using the occupied frequency channels, the introduced approach allows the primary licensees to share the frequency channels and also endows these primary licensees with the preemptive priority to delay the transmissions of the license-exempt users in the same frequency channel. Hence, the introduced approach can further improve the spectrum utilization and reduce the impact of the license-exempt users. The proposed concept can also be applied to the *leased network* as in Akyildiz et al., 2006 and Stine, 2005.

Based on the priority queuing analysis, each wireless user builds an abstraction of the dynamic wireless environment (e.g. wireless condition) and the competing users' behaviors using the same frequency channel (including the primary licensees, to which the highest priority is assigned). Note that the abstraction is important in order to enable intelligent wireless users to learn and adapt their cross-layer transmission strategies (Haykin, 2005; Mitola et al., 1999). Additionally, the necessary multi-agent interactions (information exchanges) are also determined for the priority queuing analysis. This chapter focuses on the delay-sensitive applications such as surveillance, multimedia conferencing, and media

streaming etc., since these applications are most impacted by inefficient spectrum usage. Moreover, this chapter only focuses on the multimedia transmission over a single-hop network infrastructure. Discussion regarding to multimedia transmission over a multi-hop cognitive radio network can be found in Shiang & van der Schaar, 2009.

The organization of the chapter is as follows. In Section 2, we provide the network settings of the cognitive radio network. Section 3 presents the cross-layer problem formulation for multi-user multimedia streaming over such network through a multi-agent interaction. In Section 4, we show that the multi-user multimedia streaming problem over such network can be analyzed using priority queue modeling and hence, facilitate the optimal cross-layer transmission strategies of the multimedia streaming problem through appropriate information exchange. Finally, Section 5 concludes the chapter.

2. Network Settings for Multi-User Multimedia Transmission over Cognitive Radio Networks

2.1 Multimedia traffic characteristics

Assume that there are N multimedia applications $V_1, ..., V_N$ with distinct sources destinations. In this chapter, we define the users as the transmission pairs between predetermined source wireless stations and destination wireless stations. Each encoded multimedia stream is separated into a certain number of classes (quality layers) as in van der Schaar et al., 2006. Assume that the packets within each multimedia class have the same delay deadline. The number of priority classes for a multimedia sequence V_i equals K_i. Assume that the total number of priority classes across all users in the network is K. The priority classes in the network are denoted as $C_1, ..., C_K$. For the purpose of analysis, we reserve the highest priority class C_1 for the primary users in each frequency channel, i.e. $\lambda_1 \gg \lambda_k, 2 \leq k \leq K$. The secondary users can be categorized into the rest of $K - 1$ priority classes ($C_2, ..., C_K$) to access the frequency channels[1]. Hence, the total number of classes across all users in the network equals $K = \sum_{i=1}^{N} K_i + 1$. The users in higher priority classes can preempt the transmission of the lower priority classes to ensure an interference-free environment for the primary users (Kleinrock, 1975). The priority of a user affects its ability of accessing the channel. Each multimedia class C_k is characterized by:

- λ_k, the expected quality impact of receiving the packets in the class C_k. We prioritize the multimedia classes based on this parameter. In the subsequent part of the chapter, we label the K classes (across all users) in descending order of their priorities, i.e. $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_K$.

- L_k, the average packet lengths of the class C_k. The expected quality improvement for receiving a multimedia packet in the class C_k is defined as $\lambda_k \cdot L_k$ (see e.g. Wang & van der Schaar, 2006 for more details).

- N_k, the number of packets in the class C_k in one GOP duration of the corresponding

---

[1] The prioritization of the secondary users can be determined based on their applications, prices paid for spectrum access, or other mechanism design based rules. In this chapter, we will assume that the prioritization was already performed.

multimedia sequence.

- P_k^{succ}, the probabilities of successfully receiving the packets in the class C_k at the destination. Thus, the expected number of the successfully received packets of the class C_k is $N_k \cdot P_k^{succ}$.

- d_k, the delay deadlines of the packets in the class C_k. Due to the hierarchical temporal structure deployed in 3D wavelet multimedia coders (see Wang & van der Schaar, 2006; van der Schaar & Turaga, 2007), for a multimedia sequence, the lower priority packets also have a less stringent delay requirement. This is the reason why we prioritize the multimedia bitstream in terms of the quality impact. However, if the used multimedia coder did not exhibit this property, we need to deploy alternative prioritization techniques $\lambda_k^{video}(\lambda_k, d_k)$ that jointly consider the quality impact and delay constraints (see more sophisticated methods in e.g. Jurca & Frossard, 2007; Chou & Miao, 2006).

At the client side, the expected quality improvement for multimedia V_i in one GOP can be expressed as:

$$u_i = \sum_{C_k \in V_i} \lambda_k L_k N_k P_k^{succ}. \tag{1}$$

Here, we assume that the client implements a simple error concealment scheme, where the lower priority packets are discarded whenever the higher priority packets are lost (van der Schaar & Turaga, 2007). This is because the quality improvement (gain) obtained from decoding the lower priority packets is very limited (in such embedded scalable multimedia coders) whenever the higher priority packets are not received. For example, drift errors can be observed when decoding the lower priority packets without the higher priority packets (Wang & van der Schaar, 2006). Hence, we can write:

$$P_k^{succ} = \begin{cases} 0 & , \text{ if } P_{k'}^{succ} \neq 1 \text{ and } C_{k'} \prec C_k \\ (1 - P_k) = E[I(D_k \leq d_k)], \text{ otherwise,} \end{cases} \tag{2}$$

where we use the notation in (Chou & Miao, 2006) - $C_{k'} \prec C_k$ to indicate that the class C_k depends on $C_{k'}$. Specifically, if C_k and $C_{k'}$ are classes of the same multimedia stream, $C_{k'} \prec C_k$ means $k' < k$ due to the descending priority ($\lambda_{k'} > \lambda_k$). P_k represents the end-to-end packet loss probability for the packets of class C_k. D_k represents the experienced end-to-end delay for the packets of class C_k. $I(\cdot)$ is an indicator function. Note that the end-to-end probability P_k^{succ} depends on the network resource, competing users' priorities as well as the deployed cross-layer transmission strategies. In addition, for the multimedia V_i, let us assume that the multimedia packets are scheduled in a specific order π_i according to the prioritization associated with the multimedia content characteristics.

2.3 Cognitive radio network settings

Assume that there are a total of M frequency channels $\mathbf{F} = \{F_1, ..., F_M\}$ in the cognitive radio network and there are M primary users $\mathbf{PU} = \{PU_1, ..., PU_M\}$, each in a separate frequency channel. These primary users can only occupy their assigned frequency channels. Since the primary users are licensed users, they are provided with an interference-free

environment (Haykin, 2005; Akyildiz et al., 2006). Assume that there are N secondary users $\mathbf{SU} = \{SU_1,...,SU_N\}$ in the system. These secondary users are able to operate their applications across various frequency channels for transmission. Hence, they need to time share the chosen frequency channel. Moreover, these secondary users are usually the license-exempt users, and hence, the resulting impact on the primary users must be eliminated.

Let us further assume that there is a Network Resource Manager (NRM) that coordinates multiple access control scheme for sharing the spectrum resource (by assigning transmission opportunities), while ensuring that the corresponding interference on the primary users is eliminated. The role of the NRM is similar to the coordinator in the current IEEE 802.11e Hybrid Coordination Function (HCF) solutions for multimedia applications (van der Schaar et al., 2006). Note that the NRM will not make decisions for the secondary users, but it will only manage the transmission opportunities of the frequency channels based on the priority classes to avoid interference. In this chapter, we investigate the dynamic resource management problem for the multimedia streaming of the secondary users that are associated with this specific NRM. Primary users in the highest priority class C_1 can always access their corresponding channels at any time. Secondary users, on the other hand, require transmission opportunities from the NRM for transmission based on their priorities.

Multiple users can time share the same frequency channel. Note that even if the same time sharing fractions are assigned to the users choosing the same frequency channel, the experienced channel conditions can be different for the users. A wireless user needs to stream its multimedia over an appropriate frequency channel to minimize the transmission delay D_k and thereby, increase the multimedia quality in equation (1). For a certain frequency channel F_j, the secondary users can experience various channel conditions for the same frequency channel. Besides, the frequency selections of the secondary users also mutually affect each other. Let us denote T_{ij} and p_{ij} as the resulting physical transmission rate and packet error rate for the secondary user SU_i transmitting through a certain frequency channel F_j. Let $R_{ij} = [T_{ij}, p_{ij}] \in \mathcal{R}$ be the channel conditions of the channel F_j for the secondary user SU_i and denote the channel condition matrix as $\mathbf{R} = [R_{ij}] \in \mathcal{R}^{M \times N}$.

The effective transmission rate $T_{ij}^e = T_{ij}(1 - p_{ij})$ depends on 1) the channel conditions, i.e. Signal-to-Noise-Ratio (SNR) x_{ij}, 2) the modulation and coding scheme θ that is adopted by the user, and 3) the multiple access scheme of sharing the channel. For simplicity, in this chapter, let us assume that the NRM adopts the simplest polling mechanism (Bertsekas & Gallager, 1987) similar to IEEE 802.11e that assigns transmission opportunity to the secondary users from the users in higher priority class to the lower priority class. The users in the same priority class will be polled in a round-robin fashion and have the same chance to transmit their packets. More sophisticated MAC protocols are proposed to deal with the spectrum heterogeneity (such as HD-MAC in Zhao et al., 2005). However, the contention-based MAC protocol are not preferable for delay-sensitive applications. Hence, a simple polling-based MAC similar to that used in IEEE 802.11e is considered in this chapter. The NRM only ensures the priority order among the users, but will not decide the secondary

users' actions. The expected physical transmission rate T_{ij}^e and packet error rate p_{ij} can be approximated as sigmoid functions of measured and the adopted modulation and coding scheme as in (Krishnaswamy, 2002):

$$p_{ij}(\theta_i, x_{ij}) = \frac{1}{1 + \exp(\zeta(\theta_i)(x_{ij} - \delta(\theta_i)))}, \tag{3}$$

$$T_{ij}^e(\theta_i, x_{ij}) = \frac{T_{ij}(\theta_i)}{1 + \exp(-\zeta(\theta_i)(x_{ij} - \delta(\theta_i)))}, \tag{4}$$

where $\zeta(\theta_i)$ and $\delta(\theta_i)$ are empirical constants for multimedia user V_i using frequency channel F_j corresponding to the modulation and coding schemes θ_i for a given packet length L_k of class C_k. Thus, even though coordinated by the same NRM, the expected T_{ij} and p_{ij} of the same frequency channel can be different for various secondary users.

3. Cross-Layer Transmission Problem Formulation for Multi-User Multimedia Transmission over Cognitive Radio Networks

3.1 Cross-layer transmission problem formulation

The considered actions of a secondary user V_i as the following cross-layer transmission strategies $\mathbf{a}_i = [\boldsymbol{\pi}_i, \gamma_i, \boldsymbol{\alpha}_i, \theta_i] \in \mathcal{A}_{tot} = \mathcal{A}_\pi \times \mathcal{A}_\gamma \times \mathcal{A}_\alpha \times \mathcal{A}_\theta$ including:

- The physical layer modulation and coding scheme $\theta_i \in \mathcal{A}_\theta$.
- The MAC layer retransmission limit $\gamma_i \in \mathcal{A}_\gamma$.
- The application layer multimedia packet scheduling $\boldsymbol{\pi}_i \in \mathcal{A}_\pi$.
- The selection of the frequency channel for multimedia transmissions $\boldsymbol{\alpha}_i \in \mathcal{A}_\alpha$.

Denote the frequency selection of a secondary user SU_i using $\boldsymbol{\alpha}_i = [a_{i1}, a_{i2}, ..., a_{iM}] \in \mathcal{A}_\alpha = \{0,1\}^M$, where $a_{ij} = 1$ represents the fact that the secondary user SU_i chooses the frequency channel F_j. Otherwise, $a_{ij} = 0$. Let $\boldsymbol{\alpha}_{-i}$ denote the actions of the other secondary users except SU_i. Let $\mathbf{A} = [\mathbf{a}_1, ..., \mathbf{a}_N]$ denote the total action set across all secondary users.

As stated in equation (1), each secondary user has its own multimedia quality function as the utility function to maximize. Conventionally, the utility function of a specific user is often modeled solely based on its own action, i.e. $u_i(\mathbf{a}_i)$ without modeling the other secondary users (Wang & Pottie, 2002; van der Schaar & Shanker, 2005). However, in fact, the utility function in cognitive radio networks depends on the effective delay and throughput that the secondary user can get from the frequency channel it selects, and this is related to the channel sharing, which is coupled with other secondary users. Hence, the utility function u_i is also influenced by the action of other secondary users that select this frequency channel. In other words, the utility function can be regarded as $u_i(\mathbf{a}_i, \boldsymbol{\alpha}_{-i})$. Specifically, each autonomous multimedia user selects its optimal action $\mathbf{a}_i^{opt}$ by solving the optimization problem with the knowledge of $\boldsymbol{\alpha}_{-i}$:

$$\mathbf{a}_i^{opt} = \arg \max_{\mathbf{a}_i \in \mathcal{A}_{ot}} u_i(\mathbf{a}_i, \boldsymbol{\alpha}_{-i}) \tag{5}$$

Importantly, in an informationally-decentralized cognitive wireless network that consists of decentralized secondary users, the secondary user SU_i may not know the exact actions of other secondary users $\boldsymbol{\alpha}_{-i}$. Moreover, even if all the actions are known, it is unrealistic to assume that the exact action information can be collected timely to compute and maximize the actual utility function $u_i(\mathbf{a}_i, \boldsymbol{\alpha}_{-i})$. Hence, a more practical solution is to dynamically model the other secondary users' behavior by updating their probabilistic frequency selection strategy profile based on some available information exchange, and then maximizes the *expected* utility function of SU_i.

Hence, we define a frequency selection strategy profile of a secondary user SU_i as a vector of probabilities $\mathbf{s}_i = [s_{i1}, s_{i2}, ..., s_{iM}] \in \mathcal{S}_\alpha = \mathcal{S}^M$, where $s_{ij} \in \mathcal{S}$ ($\mathcal{S} \in [0,1]$) represents the probability of the secondary user SU_i choosing the frequency channel F_j ($a_{ij} = 1$). Hence, the summation over all the frequency channels, $\sum_{j=1}^{M} s_{ij} = 1$. Note that s_{ij} can also be viewed as the fraction of data from SU_i transmitted on frequency channel F_j, and hence, multiple frequency channels are selected for a secondary users with $s_{ij} > 0$. Let $\mathbf{S} = [\mathbf{s}_1^T, ..., \mathbf{s}_N^T] \in \mathcal{S}^{M \times N}$ denote the total strategy profile across all the secondary users. The expected utility function, given a fixed strategy profile $\mathbf{S} = (\mathbf{s}_i, \mathbf{s}_{-i})$ is

$$U_i(\tilde{\mathbf{a}}_i(\mathbf{s}_i), \mathbf{s}_{-i}) = E_{(\mathbf{s}_i, \mathbf{s}_{-i})}[u_i(\mathbf{a}_i, \boldsymbol{\alpha}_{-i})], \tag{6}$$

where $\mathbf{s}_{-i}$ denotes the collected frequency selection profile. $\tilde{\mathbf{a}}_i(\mathbf{s}_i) = [\boldsymbol{\pi}_i, \gamma_i, \mathbf{s}_i, \theta_i] \in \tilde{\mathcal{A}}_{ot} = \mathcal{A}_\pi \times \mathcal{A}_\gamma \times \mathcal{S}_\alpha \times \mathcal{A}_\theta$ represents the transmission strategy using $\mathbf{s}_i$ instead of $\boldsymbol{\alpha}_i$. Then, the optimization problem in equation (5) becomes:

$$\tilde{\mathbf{a}}_i^{opt} = \arg \max_{\tilde{\mathbf{a}}_i \in \tilde{\mathcal{A}}_{ot}} U_i(\tilde{\mathbf{a}}_i(\mathbf{s}_i), \mathbf{s}_{-i}) = \arg \max_{\tilde{\mathbf{a}}_i \in \tilde{\mathcal{A}}_{ot}} \sum_{C_k \in V_i} \lambda_k L_k N_k P_k^{succ}(\tilde{\mathbf{a}}_i(\mathbf{s}_i), \mathbf{s}_{-i}) \tag{7}$$

Figure 1 provides the system architecture of the secondary users. In Section 4, we will discuss how to model the strategy (behavior) $\mathbf{s}_{-i}$ and the impact of the other secondary users using priority queuing analysis.
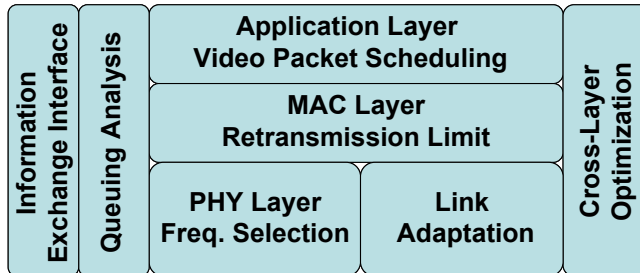

Fig. 1. The system architecture of the cross-layer optimization for video streaming over cognitive radio network

3.2 Cross-layer optimization examples

We first look at the case with 6 secondary users with multimedia streaming applications ("Coastguard", frame rate of 30Hz, CIF format, delay deadline 500ms) sharing 10 frequency channels ($N = 6, M = 10$). We compare the discussed cross-layer optimization using priority queuing analysis with other two resource allocation algorithms – the "Static Assignment" (Tekinay & Jabbari, 1991) and the "Dynamic Least Interference" (Wang & Pottie, 2002). In the "Static Assignment" algorithm, a secondary user will statically select a frequency channel with the best effective transmission rate without interacting with other secondary users. In the "Dynamic Least Interference" algorithm, a secondary user will dynamically select a single frequency channel that has the least interference from the other users (both secondary users and primary users). We consider 100 random frequency channel conditions as well as the traffic loadings and then compute the average the video PSNR and the standard deviation of the PSNR over the one hundred cases in Table 1 for the 6 video applications. There are primary users randomly appears in each frequench channel (occupying different frequency channels with a fixed loading). The results show that the cross-layer optimization outperforms the other two algorithms for delay-sensitive multimedia applications in terms of video quality (PSNR). The standard deviations of the cross-layer optimization are also smaller than the other two solutions. Unlike the "Dynamic Least Interference" that only considers how a single user adapts to the experienced environment, the multi-agent cross-layer optimization allows the secondary users to track the frequency selection strategies of the other users and adequately optimize the cross-layer transmission strategies. Hence, the users will be able to self-organize into various cognitive radio channels while adapting to the new environmental conditions. In Table 2, we see that the multi-agent cross-layer optimization approach still outperforms the other two approaches with different loadings of the primary users in each frequency channel.

Average $T_{ij}(\theta_i^{opt}) = 1$ Mbps	"Static Assignment - Largest-Bandwidth"		"Dynamic Least Interference"		"Cross-layer Optimization"	
	Average Y-PSNR (dB)	Y-PSNR Standard Deviation	Average Y-PSNR (dB)	Y-PSNR Standard Deviation	Average Y-PSNR (dB)	Y-PSNR Standard Deviation
SU_1	29.48	4.94	29.89	4.32	32.42	1.97
SU_2	29.90	4.89	30.35	4.29	32.62	2.42
SU_3	29.69	5.02	30.37	4.41	32.36	2.26
SU_4	30.59	4.98	30.87	4.37	32.75	2.31
SU_5	29.48	4.98	29.87	4.41	32.40	2.33
SU_6	30.01	5.04	30.65	4.46	32.26	2.67

Table 1. Comparisons of the channel selection algorithms for video streaming with $N = 6, M = 10$.

Average $T_{ij}(\theta_i^{opt})$ = 1 Mbps	No primary users	Primary users randomly appear ($\rho_i = 0.25$)	
	Average Y-PSNR (dB)	Average Y-PSNR (dB)	Y-PSNR Standard Deviation
"Static Assignment"	33.84	29.69	5.02
"Dynamic Least Interference"	33.90	30.37	4.41
"Cross-layer Optimization"	35.61	32.36	2.26

Table 2. Comparisons of the various resource management schemes for video streaming with $N = 6, M = 10$.

Next, let us take a look at the impact of different numbers of secondary users with video streaming applications. Figure 2 shows the average packet loss rate and the average PSNR over the N video streams (instead of over 100 different channel conditions in the previous simulation). The empirical average $T_{ij}(\theta_i^{opt})$ of the frequency channels is shown to be 3 Mbps, instead of 1 Mbps in the previous example. Larger N reduces the available resources that can be shared by the video streams, and hence, increases the application layer packet loss rate (due to the expiration of the delay deadline) and hence, decreases the received video quality. The result shows that the cross-layer optimization using the priority queuing analysis outperforms the other two algorithms for multi-user video streaming applications.
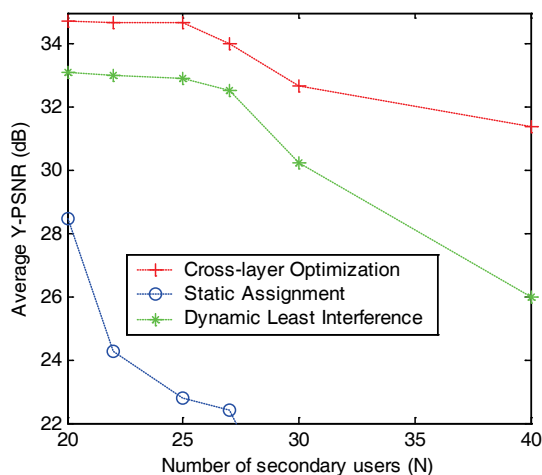
Fig. 2. Average Y-PSNR versus number of secondary users using different resource management schemes

4. Priority Queuing Analysis for Multimedia Transmission in Cognitive Radio Networks

In this section, we discuss the priority queuing analysis for the multimedia streaming problem over cognitive radio networks. The goal is to provide an abstraction of the dynamic

wireless environment and the competing wireless users' behaviors that impact the secondary user's utility. It is important to note that the packets of the competing wireless users are physically waiting at different locations. Figure 3 gives an example of the physical queues for the case of M frequency channels and N secondary users. Each secondary user maintains M physical queues for the various frequency channels, which allows users to avoid the well-known head-of-line blocking effect (Wang et al., 2004).
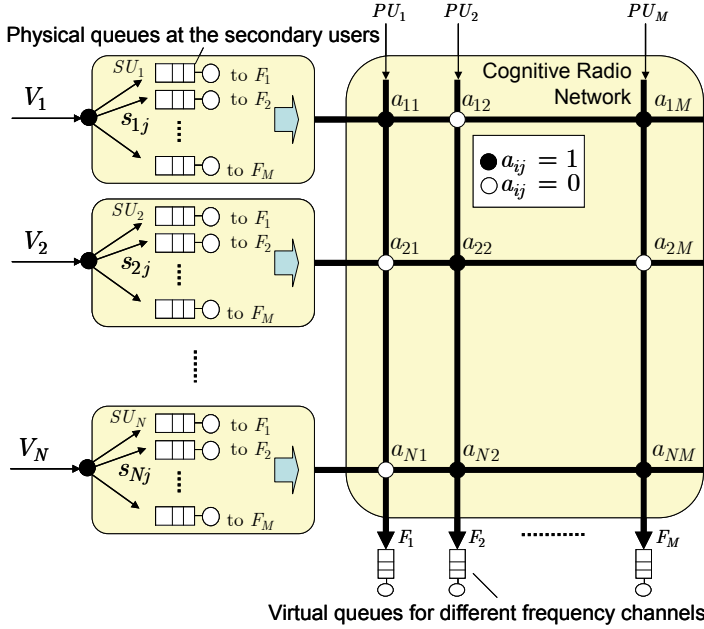
Fig. 3. Frequency selection of the secondary users a_{ij} and physical queues and virtual queues for each frequency channel.

4.1 Traffic models of the users

- **Traffic model for primary users –** Assume that the stationary statistics of the traffic patterns of primary users can be modeled by all secondary users. The packet arrival process of a primary user is modeled as a Poisson process with average packet arrival rate r_j^{PU} for the primary user PU_j using the frequency channel F_j. We denote the m_{th} moments of the service time distribution of the primary user PU_j in frequency channel F_j as $E[(X_j^{PU})^m]$. Note that this traffic model description is more general than a Markov on-off model (Shanker et al., 2005), which is a sub-set of the introduced queuing model with an exponential idle period and an exponential busy period. As in Shanker et al., 2005, an M/G/1 model is adopted in this chapter for the traffic descriptions.
- **Traffic model for secondary users –** Assume that the average rate requirement for the secondary user SU_i is B_i (bit/s). Let r_{ij} denote the average packet arrival rate of the

secondary user SU_i using the frequency channel F_j. Since the strategy s_{ij} represents the probability of the secondary user SU_i taking action a_{ij} (transmitting using the frequency channel F_j), we have $r_{ij} = s_{ij}B_i / L_i(\boldsymbol{\pi}_i)$, where $L_i(\boldsymbol{\pi}_i)$ denotes the average packet length of the secondary user SU_i depending on which priority class of multimedia packets is chosen in $\boldsymbol{\pi}_i$. If a certain secondary user SU_i can never use the frequency channel F_j, we fix its strategy to $s_{ij} = 0$, and hence, $r_{ij} = 0$. For simplicity, we also model the packet arrival process of the secondary users using a Poisson process.

Since packet errors are unavoidable in a wireless channel, let us assume that packets will be retransmitted, if they are not correctly received. This can be regarded as a protection scheme similar to the Automatic Repeat Request protocol in IEEE 802.11 networks. Hence, the service time of the users can be modeled as a geometric distribution. Let $E[X_{ij}(\boldsymbol{\pi}_i, \gamma_i, \theta_i)]$ and $E[X_{ij}^2(\boldsymbol{\pi}_i, \gamma_i, \theta_i)]$ denote the first two moments of the service time of the secondary user SU_i using the frequency channel F_j. For simplicity, we denote $E[X_{ij}(\boldsymbol{\pi}_i, \gamma_i, \theta_i)]$ and $E[X_{ij}^2(\boldsymbol{\pi}_i, \gamma_i, \theta_i)]$ using $E[X_{ij}]$ and $E[X_{ij}^2]$ hereafter in this chapter. Based on the physical layer link adaptation (Krishnaswamy, 2002), $R_{ij} = [T_{ij}, p_{ij}]$ in equation (3) and (4), we have:

$$E[X_{ij}] = \frac{(L_i(\boldsymbol{\pi}_i) + L_o)(1 - p_{ij}(\theta_i)^{\gamma_i+1})}{T_{ij}(\theta_i)(1 - p_{ij}(\theta_i))} \approx \frac{(L_i(\boldsymbol{\pi}_i) + L_o)}{T_{ij}(\theta_i)(1 - p_{ij}(\theta_i))}, \tag{8}$$

$$E[X_{ij}^2] \approx \frac{(L_i(\boldsymbol{\pi}_i) + L_o)^2(1 + p_{ij}(\theta_i))}{T_{ij}(\theta_i)^2(1 - p_{ij}(\theta_i))^2}, \tag{9}$$

where L_i is the average packet length of the secondary user SU_i and L_o represents the effective control overhead including the time for protocol acknowledgement, information exchange, and NRM polling delay, etc.

To describe the traffic model, we define the traffic specification[2] for the secondary user SU_i as $\mathbf{TS}_i = [C_k, B_i, L_i, \mathbf{X}_i, \mathbf{X}_i^2]$, if $SU_i \in C_k$, where $\mathbf{X}_i = [E[X_{ij}] \mid j = 1, ..., M]$ and $\mathbf{X}_i^2 = [E[X_{ij}^2] \mid j = 1, ..., M]$ are the vectors of the first two moments of service time when user SU_i transmitting in each frequency channel. This information needs to be exchanged among the secondary users, which will be discussed in more details in Section 4.4.

4.2 Queuing analysis for the secondary users with the same priority

In this subsection, we first consider the case that all packets have the same priority by ignoring the impact of the primary users. In the next subsection, we will generalize these results by considering the impact of the primary users using priority queuing.

In order to solve the dynamic resource management problem, we need to calculate the distribution of the end-to-end delay $D_i(\mathbf{a}_i, \boldsymbol{\alpha}_{-i})$ for the secondary user SU_i to transmit its

---

[2] The traffic specification is similar to the TSPEC in current IEEE 802.11e for multimedia transmission.

packets. The expected end-to-end delay $E[D_i]$ of the secondary user SU_i can be expressed as:

$$E[D_i(\mathbf{a}_i, \boldsymbol{\alpha}_{-i})] = \sum_{j=1}^{M} s_{ij} E[D_{ij}(R_{ij}(\mathbf{a}_i, \boldsymbol{\alpha}_{-i}))], \tag{10}$$

where $E[D_{ij}(R_{ij}(\mathbf{a}_i, \boldsymbol{\alpha}_{-i}))]$ is the average end-to-end delay if the secondary user SU_i chooses the frequency channel F_j. Recall that $R_{ij} = [T_{ij}, p_{ij}] \in \mathcal{R}$ is the channel condition of the channel F_j for the secondary user SU_i. Note that s_{ij} is the strategy of the action a_{ij}. We use the queuing model in Figure 3 and apply queuing theory to calculate the end-to-end delay.

In Figure 3, for a specific secondary user SU_i, each arriving packet will select a physical queue to join (action a_{ij}) according to the strategy s_{ij}. Assume that once a packet selects a physical queue, it cannot jockey to another queue (change position to the other queues). Thus, a queued packet waits in line to be served on the selected frequency channel.

Note that there are N physical queues from N secondary users for a frequency channel F_j. Only one of them can transmit its packets at any time. Hence, we form a "virtual queue" for the same frequency channel (see Figure 4). In a virtual queue, the packets of the different secondary users wait to be transmitted. Importantly, the total sojourn time (queue waiting time plus the transmission service time) of this virtual queue now becomes the actual service time at each of the physical queues. The concept is similar to the "service on vacation" (Bertsekas & Gallager, 1987) in queuing theory, and the waiting time of the virtual queue can be regarded as the "vacation time".
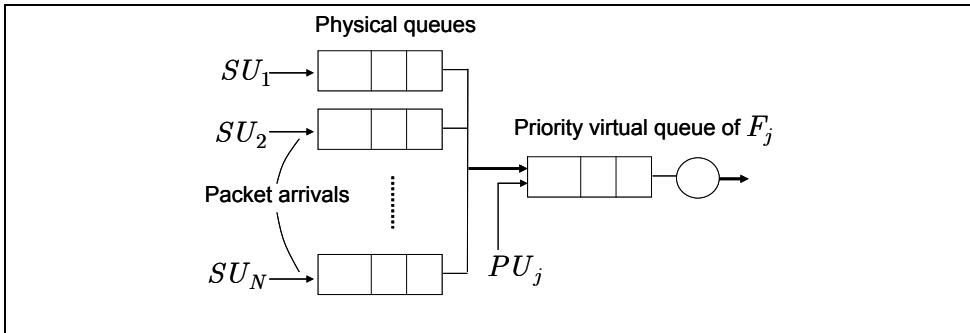
Fig. 4. Priority virtual queue for a specific frequency channel.

Since the number of the secondary users in a regular cognitive radio network is usually large, we can approximate the virtual queue using the FIFO (First-In-First-Out) M/G/1 queuing model (i.e. when $N \to \infty$, the input traffic of the virtual queue can be modeled as a Poisson process). The average arrival rate of the virtual queue of the frequency channel F_j is $\sum_{i=1}^{N} r_{ij}$. Let us denote the first two moments of the service time for the virtual queue of the frequency channel F_j as $E[\tilde{X}_j]$ and $E[\tilde{X}_j^2]$. For a packet in the virtual queue of

frequency channel F_j, we determine the probability of the packet coming from the secondary user SU_i as:

$$f_{ij} = \frac{s_{ij}}{\sum_{k=1}^{N} s_{kj}}. \tag{11}$$

Hence,

$$E[\tilde{X}_j] = \sum_{i=1}^{N} f_{ij} E[X_{ij}], \;\; E[\tilde{X}_j^2] = \sum_{i=1}^{N} f_{ij} E[X_{ij}^2]. \tag{12}$$

Let $\tilde{W}_j$ and $\tilde{D}_j$ represent the queue waiting time and sojourn time of the virtual queue of the frequency channel F_j, respectively. $E[\tilde{W}_j]$ can be obtained from the Pollaczek-Khinchin formula (Bertsekas & Gallager, 1987). Then, $E[\tilde{D}_j]$ can be obtained as:

$$E[\tilde{D}_j] = E[\tilde{W}_j] + E[\tilde{X}_j] = \frac{\sum_{i=1}^{N} r_{ij} E[\tilde{X}_j^2]}{2\left[1 - \sum_{i=1}^{N} r_{ij} E[\tilde{X}_j]\right]} + E[\tilde{X}_j]. \tag{13}$$

Then, we apply G/G/1 approximation based on the work of (Abate et al., 1995; Jiang et al., 2001) for the virtual queuing delay distribution:

$$\text{Prob}\left(\tilde{D}_j > t\right) = \sum_{i=1}^{N} r_{ij} E[\tilde{X}_j] \exp\left(-\frac{t\sum_{i=1}^{N} r_{ij} E[\tilde{X}_j]}{E[\tilde{D}_j]}\right). \tag{14}$$

This virtual queuing delay distribution is the service time distribution of the physical queues at the secondary users. Since the service time of the physical queue is an exponential distribution (see equation (14)), the average end-to-end delay of the secondary user SU_i sending packets through frequency channel F_j is approximately:

$$E[D_{ij}] = \frac{E[\tilde{D}_j]}{1 - r_{ij} E[\tilde{D}_j]}, \text{ for } r_{ij} E[\tilde{D}_j] < 1. \tag{15}$$

Strategies $(\mathbf{s}_i, \mathbf{s}_{-i})$ such that $r_{ij} E[\tilde{D}_j] \geq 1$ will result in an unbounded delay $E[D_{ij}]$, which is undesirable for multimedia streaming. The advantage of this approximation is that once the average delay of the virtual queue $E[\tilde{D}_j]$ for a certain frequency channel F_j is known by the secondary user SU_i, the secondary user can immediately calculate the expected end-to-end delay $E[D_{ij}]$ of a packet transmitting using the frequency channel F_j.

4.3 Queuing analysis with the impact of higher priority users

Based on the derivations in the previous subsection, we now consider the impact of primary users. First, let us consider the case that there are only two priority classes (i.e. $K = 2$, $\mathbf{PU} \in C_1$, $\mathbf{SU} \in C_2$). Note that in the introduced queuing model in Figure 4, the packets from the primary users will not be seen at the physical queues of the secondary users, but

only have impact on the virtual queues of the frequency channels. Since the primary users are the first priority in each of the frequency channels, we modeled the virtual queues for a particular frequency channel using a priority M/G/1 queue instead of a FIFO M/G/1 queue. Recall that the average input rate of the primary user PU_j is λ_j^{PU}, and the first two moments of the service time is $E[X_j^{PU}]$ and $E[X_j^{PU2}]$. By applying the Mean Value Analysis (MVA) in queuing theory (Kleinrock, 1975), we modify equation (13) into a priority M/G/1 queuing case:

$$E[\tilde{D}_j] = E[\tilde{W}_j] + E[\tilde{X}_j]$$

$$= \frac{r_j^{PU} E[(X_j^{PU})^2] + \sum_{i=1}^{N} r_{ij} E[\tilde{X}_j^2]}{2\,(1 - r_j^{PU} E[X_j^{PU}])(1 - \sum_{i=1}^{N} r_{ij} E[\tilde{X}_j] - r_j^{PU} E[X_j^{PU}])} + E[\tilde{X}_j], \quad (16)$$

$$= \frac{\rho_j^2 + \mu_j^2}{2\,(1 - \rho_j)(1 - \rho_j - \mu_j)} + E[\tilde{X}_j]$$

ρ_j represents the normalized loading of the primary user PU_j for the frequency channel F_j, and

$$\rho_j = r_j^{PU} E[X_j^{PU}], \quad \rho_j^2 = r_j^{PU} E[(X_j^{PU})^2]. \qquad (17)$$

μ_j represents the summation of the normalized traffic loading of all the secondary users using the frequency channel F_j, and

$$\mu_j = \sum_{i=1}^{N} r_{ij} E[\tilde{X}_j], \ \ \mu_j^2 = \sum_{i=1}^{N} r_{ij} E[\tilde{X}_j^2]. \qquad (18)$$

Hence, we substitute the $E[\tilde{D}_j]$ of equation (16) into equation (15), and determine the average end-to-end delay $E[D_{ij}]$ of the secondary user SU_i sending packets using frequency channel F_j while considering the impact of the primary user PU_j.

The derivation can be generalized to K priority classes among users ($K > 2$, $\mathbf{PU} \in C_1$, $\mathbf{SU} \in \{C_2, ..., C_K\}$). Similar to the two priority classes' case, the priority queuing model only affects the virtual queues for different frequency channels. Since the secondary users now have different priorities, the secondary users in different priority classes will experience different virtual queuing delay. Let $E[\tilde{D}_{jk}]$ represent the virtual queuing delay experienced by the secondary user in class C_k in the virtual queue for the frequency channel F_j. Let μ_{jk} represent the normalized traffic loading of all the class C_k secondary users using the frequency channel F_j. Based on the definition in the two priority users' case, we have:

$$\mu_{jk} = \sum_{\forall SU_i \in C_k} r_{ij} E[\tilde{X}_j], \text{ and } \mu_{jk}^2 = \sum_{\forall SU_i \in C_k} r_{ij} E[\tilde{X}_j^2]. \qquad (19)$$

By applying the Mean Value Analysis (MVA) (Kleinrock, 1975), we have:

$$E[\tilde{D}_{jk}] = E[\tilde{W}_{jk}] + E[\tilde{X}_j] = \frac{\rho_j^2 + \sum\limits_{l=2}^{k} \mu_{jl}^2}{2\,(1 - \rho_j - \sum\limits_{l=2}^{k-1} \mu_{jl})(1 - \rho_j - \sum\limits_{l=2}^{k} \mu_{jl})} + E[\tilde{X}_j]. \qquad (20)$$

Hence, for a secondary user $SU_i \in C_k$ using the frequency channel F_j, its end-to-end delay $E[D_{ij}]$ and probability of packet loss $P_{ij}(\tilde{\mathbf{a}}_i(\mathbf{s}_i), \mathbf{s}_{-i})$ become:

$$E[D_{ij}] = \frac{E[\tilde{D}_{jk}]}{1 - r_{ij} E[\tilde{D}_{jk}]}, \ \text{for } r_{ij} E[\tilde{D}_{jk}] < 1, \ SU_i \in C_k, \qquad (21)$$

$$P_{ij}(\tilde{\mathbf{a}}_i(\mathbf{s}_i), \mathbf{s}_{-i}) = r_{ij} \exp(-\frac{r_{ij}(1 - r_{ij}E[\tilde{D}_{jk}])d_i}{E[\tilde{D}_{jk}]}), \ \text{for } SU_i \in C_k. \qquad (22)$$

Therefore, we can approximate the objective function in equation (7) for the multimedia streaming of the secondary user SU_i as (note that $\sum_{j=1}^{M} s_{ij} = 1$):

$$\operatorname*{maximize}_{\tilde{\mathbf{a}}_i \in \mathcal{A}_{ot}} \sum_{C_k \in V_i} \lambda_k L_k N_k \cdot P_k^{succ}(\tilde{\mathbf{a}}_i(\mathbf{s}_i), \mathbf{s}_{-i})$$

$$\Rightarrow \operatorname*{maximize}_{\tilde{\mathbf{a}}_i \in \mathcal{A}_{ot}} \sum_{C_k \in V_i} \lambda_k L_k N_k \cdot \sum_{j=1}^{M} s_{ij}(1 - P_{ij}(\tilde{\mathbf{a}}_i(\mathbf{s}_i), \mathbf{s}_{-i})) \qquad (23)$$

$$\Rightarrow \operatorname*{maximize}_{\tilde{\mathbf{a}}_i \in \mathcal{A}_{ot}} \sum_{C_k \in V_i} \lambda_k L_k N_k \sum_{j=1}^{M} s_{ij}(1 - r_{ij} \exp(-\frac{r_{ij}(1 - r_{ij}E[\tilde{D}_{jk}(\tilde{\mathbf{a}}_i(\mathbf{s}_i), \mathbf{s}_{-i})])d_i}{E[\tilde{D}_{jk}(\tilde{\mathbf{a}}_i(\mathbf{s}_i), \mathbf{s}_{-i})]}))$$

Note that only $E[\tilde{D}_{jk}]$ depends on the strategies $\mathbf{s}_{-i}$ of other secondary users.

We provide here an example that considers a simple network with two secondary users and three frequency channels (i.e. $N = 2$, $M = 3$). In the simple example, the behavior of the proposed cognitive radio model can be clearly understood. Assume that each secondary user can choose all three frequency channels. The two secondary users are in the same priority class. The simulation parameters of the secondary users are presented in Table 3 including the channel conditions $R_{ij} = [T_{ij}, p_{ij}]$, and initial strategies $\mathbf{s}_i(0)$, etc. The normalized traffic statistics of the primary users are given in Table 4.

Secondary users	Physical transmission rate $T_{ij}(\theta_i^{opt})$ (Mbps)			Physical packet error rate $p_{ij}(\theta_i^{opt})$			Rate requirement B_i (Mbps)
	F_1	F_2	F_3	F_1	F_2	F_3	
SU_1	1.88	1.27	1.07	0.11	0.17	0.17	0.92
SU_2	1.32	1.68	1.20	0.03	0.05	0.11	0.74

Table 3. Considered parameters of the secondary users in the example.

Primary users	Normalized loading ρ_j	Second moment normalized loading ρ_j^2
PU_1	0.2	1×10^{-4}
PU_2	0.1	1×10^{-4}
PU_3	0.3	1×10^{-4}

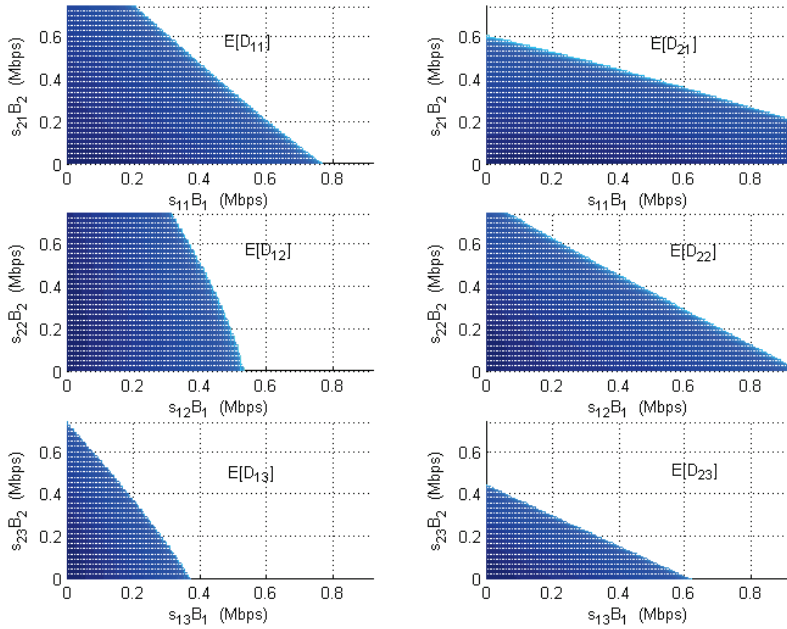Table 4. Considered parameters of the primary users in the example.

Fig. 5. Analytical expected delay of the secondary users with various strategies in different frequency channels, shadow part represents a bounded delay below the delay deadline (stable region).

Given the statistics, Figure 5 provides the different strategy pairs (s_{1j}, s_{2j}) in the three frequency channels that keep the analytical experienced delays $E[D_{ij}]$ (using equation (21)) bounded by the delay deadlines for the two secondary users. Importantly, a strategy pair (s_{1j}, s_{2j}) that results in an unbounded $E[D_{ij}]$ will make the multimedia quality drop abruptly for the delay-sensitive applications, which is undesirable for these secondary users. Figure 5 clearly shows when the channel conditions become worse (from F_1 to F_3), the selectable frequency strategy pairs becomes less. Hence, equation (21) provides the analytical operation points for the frequency selection strategy pairs.

4.4 Realistic framework for multimedia transmission over cognitive radio networks using queuing analysis

The priority virtual queue analysis requires the following information to compute μ_{jl} and μ_{jl}^2 in (20):

- Priority: the secondary users' priorities.
- Normalized loading: the secondary users' normalized loading parameters $r_{ij}E[\tilde{X}_j]$, which not only include the information of $\mathbf{s}_i$, but also reflects the input traffic loading and the expected transmission time using a specific frequency channel.
- Variance statistics: the secondary users' variance statistics with the normalized parameter $r_{ij}E[\tilde{X}_j^2]$.

Hence, two kinds of information exchange are defined for the priority virtual queue analysis:

- Other secondary users' traffic specification $\mathbf{TS}_{-i}$.
- The frequency selection information of the other secondary users to model the strategies $\mathbf{s}_{-i}$.

Figure 6 shows the block diagram of the introduced priority virtual queue interface (Shiang & van der Schaar, 2008) together with the cross-layer optimization approach in Section 3.1. Since the traffic specification $\mathbf{TS}_i$ only varies when the frequency channels change dramatically, the traffic specification can be exchanged only when a secondary user joins the network. On the other hand, the frequency selection information can be exchanged more frequently (e.g. once per service interval in van der Schaar et al., 2006). Note that since the users in the higher priority classes will not be affected by the users in the lower priority classes, they do not need the information from the users in a lower priority class. Hence, the information exchanges (overheads) and computational complexity will be scalable and will increase as the traffic priority decreases, thereby benefiting the high priority and low-delay traffic.
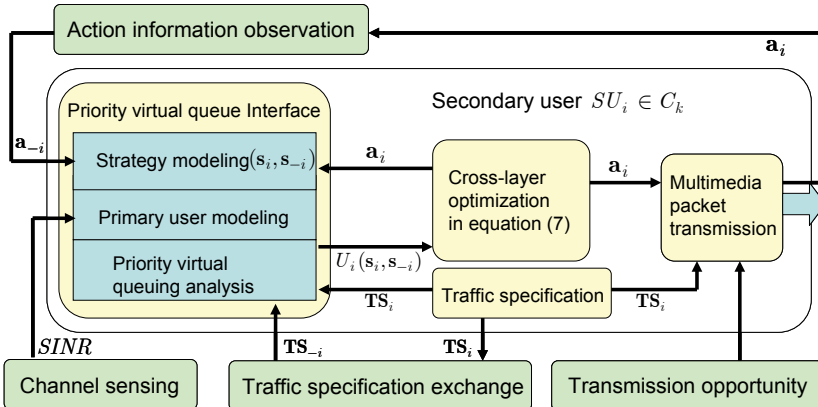
Fig. 6. Block diagram of the priority virtual queue interface and the cross-layer optimization for multimedia streaming over the cognitive radio networks.

5. Conclusions

In this chapter, we discussed the priority virtual queuing architecture for heterogeneous and autonomous secondary users in cognitive radio networks, based on which they can time share the various frequency channels in a distributed fashion. With the information exchange defined by the proposed interface, the secondary users can build an abstraction of the dynamic wireless environment as well as the competing wireless users' behaviors and learn how to efficiently adapt their transmission strategies for multimedia streaming. Importantly, unlike conventional channel allocation schemes that select the least interfered channel merely based on the channel estimation, the introduced multi-agent priority queue modeling allows the secondary users to track the other users and adequately adapt their own transmission strategies to the changing multi-user environment. It can be shown that the introduced cross-layer optimization that applies priority queuing analysis significantly outperforms the fixed channel allocation and the current dynamic channel allocation that selects the least interfered channel, in terms of multimedia quality. Finally, we discuss the required information exchange that is required for the queuing analysis and present a realistic framework for the secondary users to transmit multimedia traffic over cognitive radio networks.

6. References

Abate, J.; Choudhury, G. L. & Whitt, W. (1995). "Exponential approximations for tail probabilities in queues I: Waiting times", in *Operations Research*, vol. 43, no. 5, pp 885-901, 1995.

Akyildiz, I. F.; Lee, W. –Y.; Vuran, M. C. & Mohanty, S. (2006). "NeXt generation/dynamic spectrum access/cognitive radio wireless networks: a survey," *Computer Networks: The International Journal of Computer and Telecommunication Networking*, vol. 50, no. 13, Sep 2006.

Bertsekas, D. & Gallager, R. (1987). *Data Networks*, Prentice Hall, Inc. Upper Saddle River, NJ, 1987.

Brik, V. ; Rozner, E.; Banarjee, S. & Bahl, P. (2005). "DSAP: a protocol for coordinated spectrum access," in *Proc. IEEE DySPAN 2005, Nov. 2005*, pp. 611-614.

Brown, T. X. (2005). "An analysis of unlicensed device operation in licensed broadcast service bands," in *Proc. IEEE DySPAN 2005*, Nov 2005, pp. 11-29.

Chou, P. A. & Miao, Z. (2006). "Rate-Distortion Optimized Streaming of Packetized Media," in *IEEE Transactions on Multimedia*, vol. 8, no. 2, Apr. 2006, pp. 390-404.

Cordeiro, C.; Challapali, K.; Birru D. & Shankar S. (2006). "IEEE 802.22: An Introduction to the First Wireless Standard based on Cognitive Radios," in *Journal of Communications, Academy Publishers*, vol. 1, no. 1, Apr 2006.

Fu, F. & van der Schaar, M. (2007). "Noncollaborative Resource Management for Wireless Multimedia Applications Using Mechanism Design," in *IEEE Trans. Multimedia*, vol. 9, no. 4, pp. 851-868, Jun. 2007.

Haykin, S. (2005). "Cognitive Radio: Brain-Empowered Wireless Communications," in *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, Feb 2005.

Jiang, T.; Tham, C. K. & Ko, C. C. (2001). "An approximation for waiting time tail probabilities in multiclass systems", *IEEE Communications Letters*, vol. 5, no. 4, pp

175-177. April 2001.

Jurca, D. & Frossard P. (2007). "Packet Selection and Scheduling for Multipath video streaming," *IEEE Transactions on Multimedia*, vol. 9, no. 2, Apr. 2007.

Kleinrock, L. (1975). *Queuing Systems Volume I: Theory*, NY: Wiley-Interscience Publication, 1975.

Krishnaswamy, D. (2002). "Network-assisted link adaptation with power control and channel reassignment in wireless networks," *3G Wireless Conference*, pp. 165-170, 2002.

Mitola, J. III & Maguire, G. Q. Jr. (1999). "Cognitive radio: Making software radios more personal," in *IEEE Pers. Commun.*, vol. 6, no. 4, pp. 13-18, Aug. 1999.

Mohr, A. E.; Riskin, E. A. & Ladner, R. E. (2000). "Unequal loss protection: Graceful degradation of image quality over packet erasure channels through forward error correction," in *IEEE J. Sel. Areas Commun.*, vol. 18, no. 6, pp. 819–829, Jun. 2000.

Setton, E.; Yoo, T.; Zhu, X.; Goldsmith A. & Girod, B. (2005). "Cross-layer design of Ad hoc Networks for real-time video streaming," *IEEE Wireless Communications Mag.*, pp. 59-65, Aug 2005.

Shankar, S.; Chou, C. T.; Challapali, K. & Mangold, S. (2005). "Spectrum agile radio: capacity and QoS implementations of dynamic spectrum assignment," *Global Telecommunications Conference*, Nov. 2005.

Shiang, H. –P. & van der Schaar, M. (2007a). "Multi-user video streaming over multi-hop wireless networks: A distributed, cross-layer approach based on priority queuing," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 4, pp. 770-785, May 2007.

Shiang, H. -P. & van der Schaar, M. (2007b). "Informationally Decentralized Video Streaming over Multi-hop Wireless Networkss," in *IEEE Trans. Multimedia*, vol. 9, no. 6, pp. 1299-1313, Sep 2007.

Shiang, H. -P. & van der Schaar, M. (2008). "Queuing-Based Dynamic Channel Selection for Heterogeneous Multimedia Applications over Cognitive Radio Networks," in *IEEE Trans. Multimedia*, vol. 10, no.5, pp. 896-909, Aug. 2008.

Shiang, H. -P. & van der Schaar, M. (2009), "Distributed Resource Management in Multi-hop Cognitive Radio Networks for Delay Sensitive Transmission," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 2, pp. 941-953, Feb 2009.

Stine, J. A. (2005). "Spectrum management: the killer application of ad hoc and mesh networking," in *Proc. IEEE DySPAN 2005,* Nov. 2005, pp. 184-193.

Stockhammer, T.; Hannuksela, M. M. & Wiegand, T. (2003). "H. 264/AVC in Wireless Environments," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, July 2003, pp. 657-673.

Tekinay S. & Jabbari, B. (1991). "Handover and Channel Assignment in Mobile Cellular Networks," *IEEE Communication Magazine*, vol. 29, pp. 42-46, Nov 1991.

van der Schaar, M.; Andreopoulos, Y. & Hu, Z. (2006). "Optimized Scalable Video Streaming over IEEE 802.11a/e HCCA Wireless Networks under Delay Constraints," *IEEE Trans. On Mobile Computing*, vol. 5, no. 6, pp. 755 – 768, June 2006.

van der Schaar, M. & Chou, P. (2007). *Multimedia over IP and Wireless Networks*, Academic Press, 2007.

van der Schaar, M. & Fu, F. (2009). "Spectrum Access Games and Strategic Learning in Cognitive Radio Networks for Delay-Critical Applications," accepted to *Proc. of IEEE, Special issue on Cognitive Radio*, to appear Early 2009.

van der Schaar, M. & Shankar, S. (2005). "Cross-layer wireless multimedia transmission: challenges, principles, and new paradigms," *IEEE Wireless Commun. Mag.*, vol. 12, no. 4, pp. 50-58, Aug. 2005.

van der Schaar, M. & Turaga, D. S. (2007). "Cross-layer Packetization and Retransmission Strategies for Delay-sensitive wireless Multimedia Transmission," *IEEE Transactions on Multimedia*, vol. 9, no. 1, pp. 185-197, Jan 2007.

Wang, C. C. & Pottie, G. J. (2002). "Variable Bit Allocation for FH-CDMA Wireless Communication Systems," in *IEEE Transactions on Communications*, vol.50, no. 6, Oct 2002.

Wang, M. & van der Schaar, M. (2006). "Operational Rate-Distortion Modeling for Wavelet Video Coders," in *IEEE Transactions on Signal Processing*, vol. 54, no. 9, pp. 3505-3517, Sep. 2006.

Wang, J.; Zhai, H. & Fang, Y. (2004). "Opportunistic packet scheduling and media access control for wireless LANs and multi-hop ad hoc networks," in *IEEE Wireless Communications and Networking Conference*, vol. 2, pp. 1234-1239, Mar 2004.

Zekavat, S. A. & Li, X. (2006). "Ultimate Dynamic Spectrum Allocation via User-central Wireless Systems," *Journal of Communications, Academy Publishers*, vol. 1, issue 1, pp. 60-67, Apr 2006.

Zhao, J.; Zheng, H. & Yang, G.-H. (2005). "Distributed coordination in dynamic spectrum allocation networks," in *Proc. IEEE DySPAN 2005*, Nov 2005, pp. 259-268.

Platform for Inter-Radio System Switching with Cognitive Radio

Seishi Hanaoka, Masashi Yano and Shinji Nishimura
Hitachi, Ltd., Central Research Laboratory
Japan

1. Introduction and background

Frequencies below the 6 GHz band are well known to be suitable for mobile communication systems; hence, many wireless communication systems such as 3rd generation cellular systems and wireless LANs are assigned within this frequency band in Japan. As a consequence, there are not enough spare bands for future wireless broadband systems. In this situation, there is a need to use these frequencies, which are "finite resources," in a more efficient manner, including the utilization of multiple wireless communication systems with intelligence. Many technologies related to efficient utilization of multiple wireless communication systems have been researched. Cognitive radio is also one of the most effective technologies to resolve this issue..

There are two trends in cognitive radio especially in Japan. One trend is the so-called "Multiple Systems," which switches wireless communication systems according to the radio conditions. The other trend is the so-called "Dynamic Spectrum Access," which recognizes spare frequencies of a primary system and allocates them to be used for communication of a secondary system to such an extent that the primary system would not be affected.

Japanese regulations assign a unique frequency band to a particular wireless system both for licensed and unlicensed bands; therefore era of "Multiple Systems" comes earlier than "Dynamic Spectrum Access" system that needs more changes in the regulations. From time-to-market point of view, we focus on "Multiple Systems" approach here.

Based on the "Multiple systems" concept, the MIRAI architecture had already been proposed as one of the network architectures to support multiple systems. In MIRAI architecture, all access points of wireless communication systems are connected to a CCN (Common Core Network) and switching between systems is executed via mobile IP. The mobile IP protocol supports mobility transparently above the IP level and allows nodes to change their location. Mobile IP is generally adopted as a macro-mobility solution. In general, a few seconds is taken for system handover by using the mobile IP protocol, so that is less well suited for fast system handover in which an environment of mobile terminals changes dynamically.

From a terminal point of view, SDR (Software Defined Radio) terminals that support multiple wireless systems have been proposed. Mobile terminals can measure radio information and report that to the base station, and the base station decides whether to

switch to other systems according to this report from the mobile terminal. However, that is not sufficient for maximizing system capacity and satisfying requirements for user communication quality because system load and information that can be acquired from the network (e.g., the number of terminals that connect to an access point) are not taken in account.

We have studied a cognitive radio system that covers multiple wireless communication systems from the network point of view. The main difference from conventional system switching technologies like MIRAI architecture is using radio information that is acquired in the physical layer and system load that is acquired in the MAC and higher layers for system switching. As a result, system throughput is enhanced with efficient frequency utilization.

In our system, a control node is newly set inside a cognitive base station to support fast system switching and multiple transmissions, and one local IP address is assigned to the terminal regardless of the number of wireless communication systems that the terminal communicates with. The radio environment, system load, and information that can be acquired from network side are taken into account to maximize system capacity.

In this chapter, we describe the architecture of the cognitive radio system and then, the simulator and the testbed system built based on the proposed architecture. In section 2, we describe the approach and system concept of our cognitive radio system. In section 3, we describe the system architecture. In section 4, we show some simulation results. In section 5, we show the experimental results obtained from the testbed system.

2. Concept of cognitive radio

2.1 Overview of cognitive radio

Japanese regulations assign a unique frequency band to a particular wireless system both for licensed and unlicensed bands. However, the time ratio of frequency utilization varies widely according to location, time, day of week, wireless communication system, and communication carrier company, for example. By using these spare radio resources adaptively, the time ratio of frequency utilization can be increased.

The concept of cognitive radio based on "Multiple Systems" approach is shown in Fig. 1. In this figure, the concept is expressed using both the frequency domain and time domain.
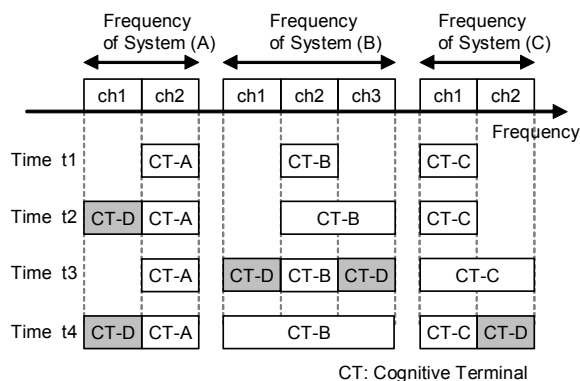
Fig. 1. Concept of Cognitive Radio.

For example, we assume we can communicate with three systems: A, B and C.

In this figure, when the current time is t2, cognitive terminal (CT) D communicates using the frequency of System A, and when the current time becomes t3, terminal D changes the wireless communication system from System A to System B to communicate using the frequency of System B. As shown in Fig. 1, the number of wireless communication systems used simultaneously is not limited to one, and the cognitive system can transmit and receive data with multiple wireless communication systems simultaneously.

Furthermore, terminals of the cognitive radio system (cognitive terminal) switch the wireless communication system frequently according to the radio conditions as stated. Therefore, in cognitive radio, the corresponding node need not know which wireless system is being used.

Based on this concept, we provide two requirements to achieve cognitive radio below:

1. system architecture for fast system handover, which can reflect radio environments that change dynamically, and system load and information that can be acquired from the network.
2. assignment of one local IP address to the terminal regardless of the number of wireless communication systems that the terminal communicates with.

2.2 System concept of cognitive radio

Provided that EV-DO (cdma2000 1x <u>E</u>volution <u>D</u>ata <u>O</u>ptimized) is system A, WiMAX (<u>W</u>orldwide <u>I</u>nteroperability for <u>M</u>icrowave <u>Ac</u>cess) is system B and wireless LAN is system C in Fig. 1, terminal D can communicate with EV-DO, wireless LAN, and WiMAX adaptively according to the radio conditions. However, terminal D can use different radio systems, as shown in Fig. 1, only when terminal D is located in the area where EV-DO, WiMAX, and wireless LAN are in service. The service areas of each system differ from each other due to the difference in frequency performance and difference in service (carrier, bit rate, and charge, for example), so we need to consider the architecture of the cognitive base station.

When we set up a cognitive BS (Base Station) that supports multiple wireless systems like that shown in Fig. 2, only the center area of the base station, which is covered by all kinds of wireless communication systems, can satisfy the conditions shown in Fig. 1. This architecture is simple and easy to construct; however, an area where cognitive radio can be adopted is narrow and limited. Actually, WiMAX access points are not always located in the same place where EV-DO access points are located; therefore, this architecture is not suitable and it would seem that the realization probability of the architecture is low.
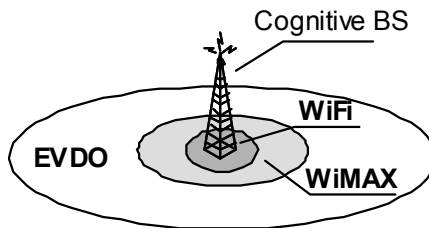
Fig. 2. Cognitive Base Station with Multiple Radio Systems.

To expand the area that satisfies the conditions shown in Fig. 1, we newly define the cognitive base station as described below:

1. The area of the cognitive base station is equivalent to the area in which access points of a cellular system are covered, which is the widest area among the access points of other wireless systems.
2. A cognitive base station has the function of access points of a cellular system, the function of access points of WiMAX and wireless LAN in the cognitive base station area, and a control node to integrate these functions.

The concept of a cognitive BS (base station) based on this definition is shown in Fig. 3.
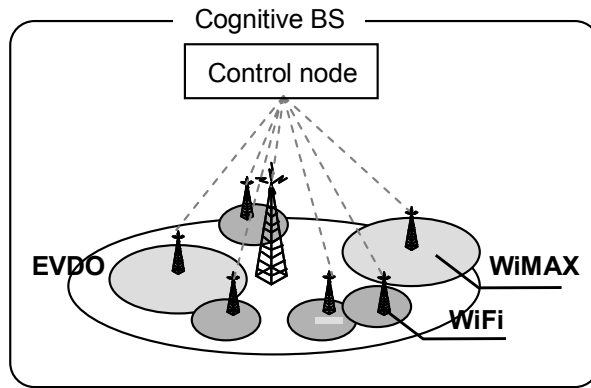
Fig. 3. Concept of Cognitive Base Station.

From this definition, the area that satisfies the conditions shown in Fig. 1 can be expanded. Actually, WiMAX access points are not always located in the same place where the EV-DO access point is located; therefore, this architecture is more realistic. Moreover, placing a control node inside the cognitive base station is one characteristic. The control node controls these systems below the IP layer. Thus, converging multiple access points of multiple systems inside the cognitive base station enables us to treat the radio resources spread throughout the cognitive base station area as an "internal" radio resource of the cognitive base station. Consequently, that is expected to achieve fast system handover.

3. System architecture

3.1 System architecture

Based on the concept described in section 2, we propose a system architecture of our cognitive radio system as shown in Fig. 4. Cognitive base station consists of PDSN (Packet Data Serving Node) to integrate an EV-DO access point to the cognitive base station, ASN-GW (Access Service Network Gateway) to integrate multiple WiMAX access points, PDIF (Packet Data Interworking Function) to integrate multiple wireless LAN access points, control node, monitoring node in addition to multiple access points of multiple radio systems. The monitoring node collects radio information and system load from each access

point and information that can be acquired from network side and recognizes radio condition. Based on the information from the monitoring node, the control node switches the communication system in a packet-by-packet basis. Moreover, location of control node is not above each access point, but above PDSN, ASN-GW and PDIF. PDSN terminates PPP (Point-to-Point Protocol) session to the EV-DO terminal function, ASN-GW controls multiple WiMAX access points, and IPSec (Security Architecture for Internet Protocol) tunnel is established between PDIF and the terminal, and PDIF controls multiple access points of wireless LAN. Therefore, it is reasonable for future system migration to locate control node above PDSN/ASN-GW/PDIF.

To place control node to converge wireless systems inside the cognitive base station and its location above PDSN, ASN-GW and PDIF are major characteristics of our system and these are main difference from MIRAI architecture.

Cognitive terminal consists of EV-DO terminal module, WiMAX CPE (Customer Premises Equipment) module, wireless LAN access terminal module and control node (application) to integrate the data received from these modules.

PDSN: Packet Data Serving Node
ASN-GW: Access Service Network Gateway
PDIF: Packet Data Interworking Function
HA: Home Agent
AAA: Authentication, Authorization and Accounting
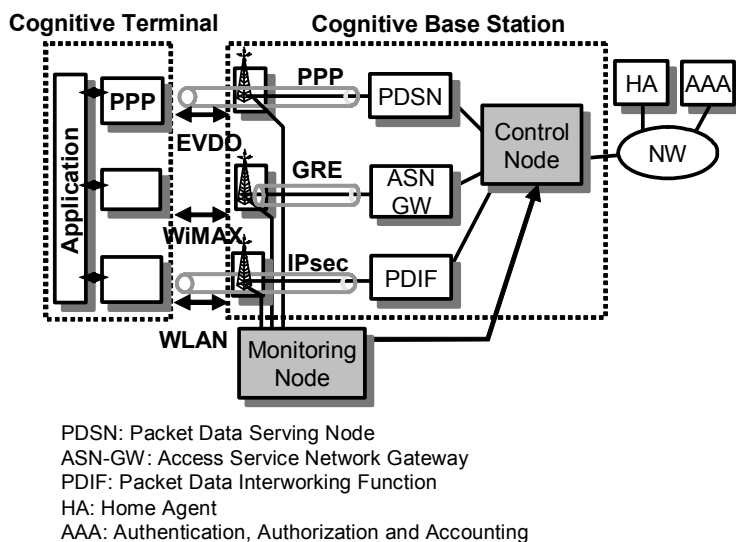
Fig. 4. System Architecture of Cognitive Radio

3.2 IP address assignments

As described previously, wireless communication system are switched dynamically according to radio environment in cognitive radio system, therefore it is desirable the corresponding node does not need to know which wireless system is being used. Unconscious switching nodes to node are achieved with mobile IP, however mobile IP is generally adopted to macro-mobility solution; in general, it takes a few seconds for system handover by using mobile IP protocol, therefore it will be not enough for fast system handover, in which an environment of mobile terminal changes dynamically.

To realize the single local IP address on a cognitive terminal, we propose IP configuration and control sequence, as depicted for the case of switching wireless system from EV-DO to wireless LAN as shown in Fig. 5.

When the terminal communicates with EV-DO for the first step, PPP (Point-to-Point Protocol) is established between the terminal and PDSN, and then it is authenticated with PAP/CHAP (PAP: Password Authentication Protocol, CHAP: Challenge Handshake Authentication Protocol). After finishing the authentication process, PDSN transmits an access request message to AAA (Authentication, Authorization and Accounting). AAA assigns an IP address and sends it back to PDSN along with the information of HA (Home Agent), and PDSN relays these information to the terminal.

PDSN: Packet Data Serving Node
ASN-GW: Access Service Network Gateway
PDIF: Packet Data Interworking Function
HA: Home Agent
AAA: Authentication, Authorization and Accounting
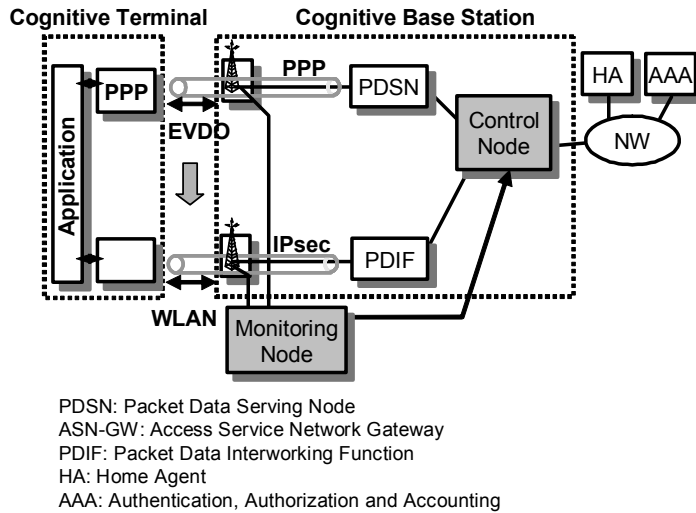
Fig. 5. System Architecture of Cognitive Radio.

When the terminal moves to the wireless LAN area, terminal authentication is established between the terminal and PDIF with IKEv2 (Internet Key Exchange version 2). After finishing authentication, PDIF transmits access request message to AAA (Authentication, Authorization and Accounting).

In the present system, AAA of wireless LAN system is independent of that of cellular system, but in our proposal, AAAs of wireless LAN and cellular system would be unified or have cooperation with each other. When one communication operator (carrier) operates multiple systems, unified AAA is easy to construct. When plural communication operators cooperate with each other to achieve cognitive system shown in Fig. 5, cooperation of AAAs is needed to identify the terminals and to assign one local IP address to them. This is also one characteristic of our system. Before assigning a IP address to the terminal, AAA identifies whether access request message is sent from the terminal that has same product number or USIM number. If the access request message is sent from the terminal that has same product number or USIM number, AAA assigns the same IP address that is already assigned using EV-DO system (Fig. 6). In this sense, AAAs are needed to be unified or have cooperation with each other.
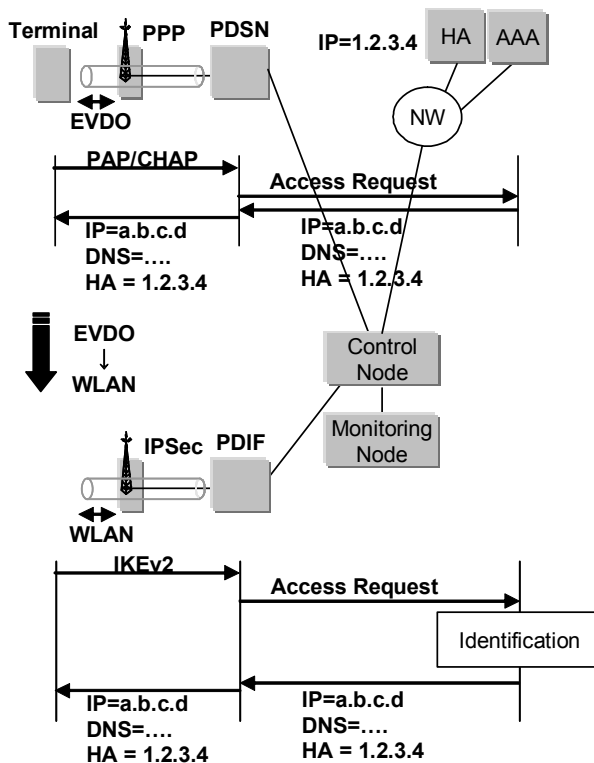
Fig. 6. Control Sequence for IP Address Assignments

3.3 Controls for system switching

To achieve the single local IP address on a terminal, control node acts as a Foreign Agent to the Home Agent, and acts as Home Agent to the PDSN/PDIF at the same time. These relationships are shown in Fig. 7. Control node has the table that relates between IP address of cognitive terminal and IP address of Foreign Agent. To relate multiple IP addresses of Foreign Agents with one cognitive terminal is one major characteristic of our proposed system and cognitive terminal keeps multiple sessions during data transmission. Due to this characteristic, system switching can be done by packet-by-packet basis and expected system switching delay can be a few milliseconds.

Regarding IP packet format, there are many approaches to realize transfer of IP packets to the nodes (PDSN/PDIF). We adopt IP capsulation, because the process of header replacement can be done by the hardware implementation and high speed switching could be expected.
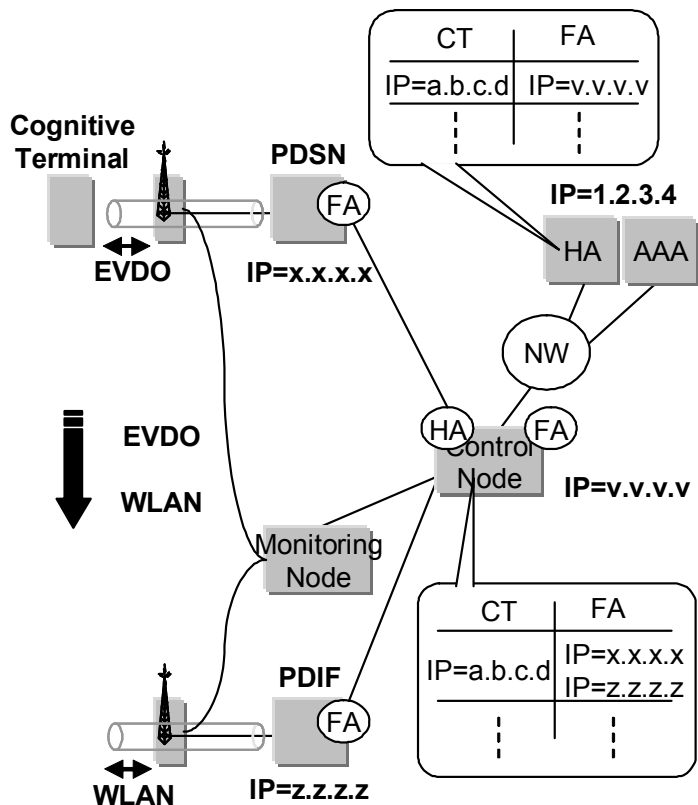
Fig. 7. Function of Control Node

4. Simulations

4.1 Preliminary simulator overview

Based on the architecture described in section 3, we have developed a preliminary simulator that supports both WiMAX and wireless LAN.

As described in section 3, cognitive node switches the wireless system to communicate with according to the radio condition and system load. In scenario 1, monitoring node monitors RSSI (Received Signal Strength Indicator) value of wireless LAN and based on this value, control node switches the system, because WiMAX service is provided in all area of the simulator. In scenario 2, adding to scenario 1, system load is taken into account.

Simulator overview is shown in Fig. 8, and system parameters of each wireless system for this simulator are shown in Table 1.
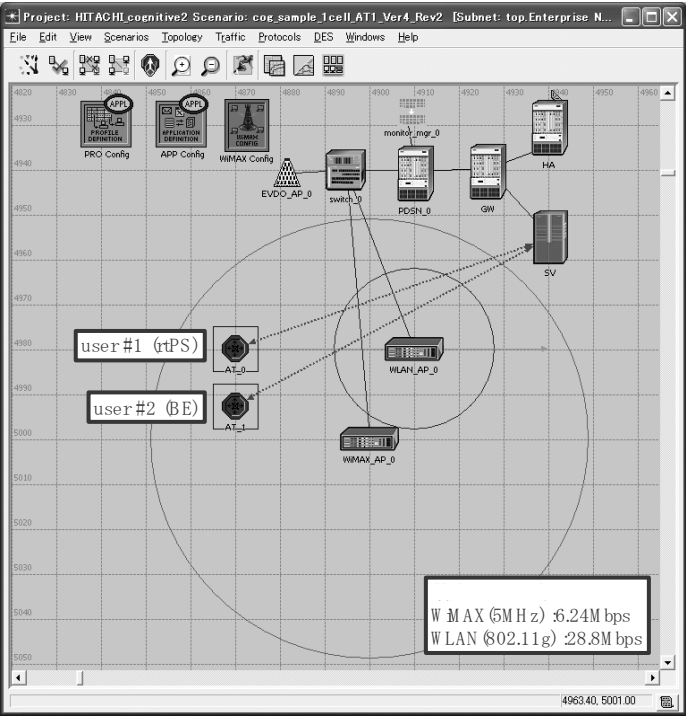
Fig. 8. Overview of Preliminary Simulator.

Item	Specifications	
WiMAX	Based on	IEEE802.16e
	Bandwidth	5 MHz
	QoS Mode	User #1: rtPS
		User#2: BE
Wireless LAN	Based on	IEEE802.11g
	Bandwidth	20 MHz
	Radio information	RSSI Level
Data	UDP	User #1: 4.8Mb/s
		User #2: 28.8Mb/s

Table 1. Preliminary Simulator Specifications.

4.2 Scenario 1: Switching based on RSSI level

In scenario 1, we assume two terminals. The terminals move along a line in cognitive base station area. For the first step, these terminals connect with WiMAX, and then, the terminals can use either WiMAX or wireless LAN area in the overlapped area as shown in Fig. 8.
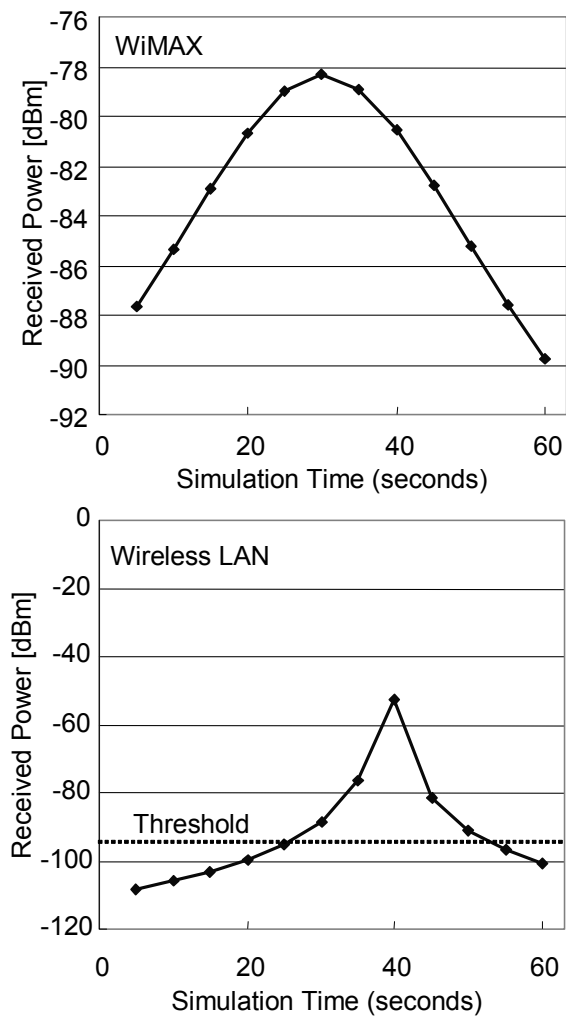
Fig. 9. RSSI Level of Each System.

To realize the system switching between two systems, we set the threshold to switch the wireless system according to RSSI value of wireless LAN. RSSI of each system is shown in Fig. 9. The terminal can connect with wireless LAN when the RSSI level exceeds -95dBm, therefore the threshold level is set to -95dBm.

Fig. 10 shows the history of user throughputs with using WiMAX and wireless LAN. User #1 and user #2 move same way with same speed.
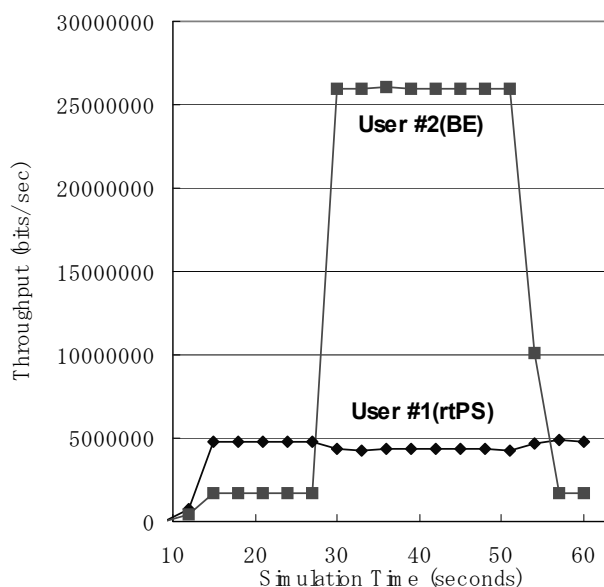
Fig. 10. Throughputs of Scenario 1.

When both users are in WiMAX area, due to QoS function of WiMAX, user #1, that has higher priority than user #2, can communicate with required bit rate, and user #2 communicates with vacant bandwidth.

When both users come near an access point of wireless LAN, RSSI level of wireless LAN becomes higher, therefore based on RSSI level, both user #1 and user #2 switch the system from WiMAX to wireless LAN.

However, wireless LAN cannot support 33.6Mb/s (that equals a sum of the data rates of user #1 and user #2), throughput performance degradation occurred.

In this scenario, we have found that not only radio condition, but also system load is taken into account to switch radio systems.

4.3 Scenario 2: Switching based on RSSI level and load balancing

In scenario 2, the terminal communicates with WiMAX area first, and then moves to wireless LAN area as shown in Fig. 8, and switches wireless system according to not only the RSSI of wireless LAN and but also system load or users' QoS, etc.

In this case, QoS of user #1 is set to rtPS (Real Time Packet Service), and QoS of user #2 is set to BE (Best Effort), so user #1 has more priority.

rtPS class ensures the real time transmission and user's required bandwidth. On the contrary, BE class uses the rest bandwidth, so there are no guarantee regarding transmission rate and throughput.

The result is shown in Fig. 11. When both users are in WiMAX area, due to QoS function of WiMAX, user #1 that has higher priority than user #2, can communicate with required bit

rate, and user #2 communicates with vacant bandwidth. Performance of this area is same as that of scenario 1.
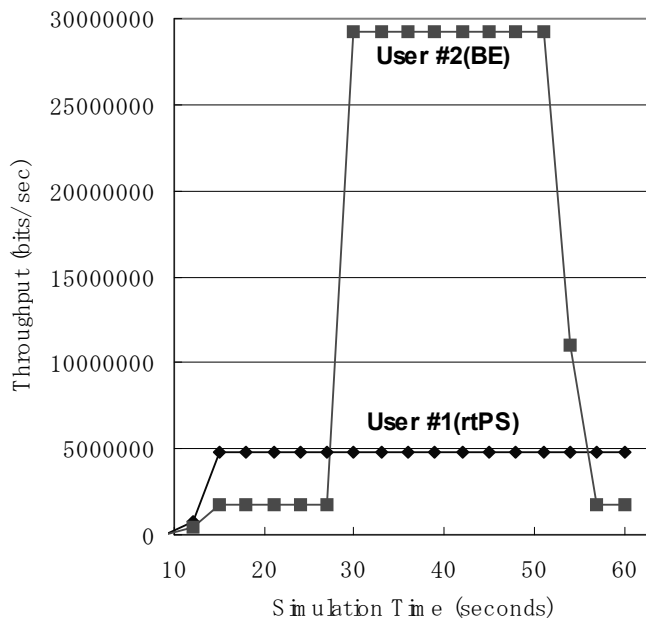
Fig. 11. Throughputs of Scenario 2.

When both users come near an access point of wireless LAN, RSSI level of wireless LAN becomes higher. User #1 has a priority, therefore control node decide that user #1 continues communication with WiMAX (not switching to wireless LAN). On the contrary, user #2 does not have priority, therefore based on RSSI level and control nodes' decision of user #1, user #2 switches the system from WiMAX to wireless LAN.

As a result, any performance degradation can not be seen in both user #1 and user #2 transmission.

4.4 Simulator that supports EV-DO, WiMAX and wireless LAN

Based on the architecture described in section 3, we have also developed a simulator that supports EV-DO, WiMAX and wireless LAN. Simulator overview is shown in Fig. 12, and system parameters of each wireless system for this simulator are shown in Table 2.

In this simulator, cognitive node switches the wireless system to communicate with according to the radio condition and system load.
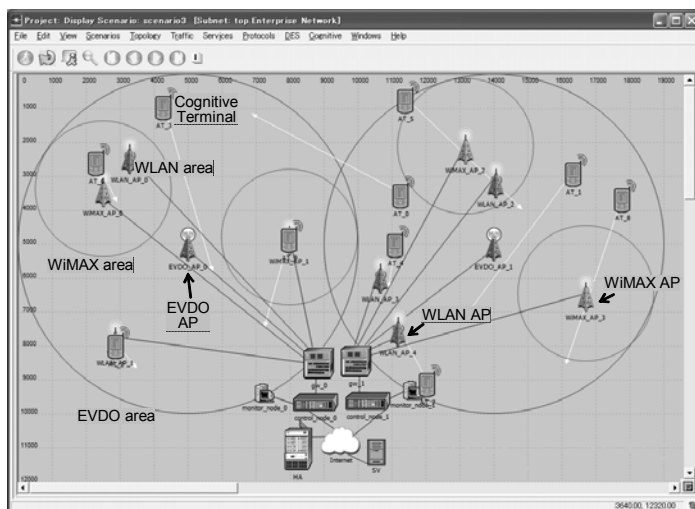
Fig. 12. Simulator Overview.

System	Specifications	
EV-DO	Based on	cdma2000 1x EV-DO Rev.0 (simplified model)
	Frequency	2.0GHz
	Maximum Rate	2.4Mb/s
	Maximum Tx Power	10W
WiMAX	Based on	IEEE802.16e based (OFDMA/TDD)
	Frequency	2.5 GHz
	Maximum Rate	6.2Mb/s (downlink)
	Maximum Tx Power	27dBm
WirelessLAN	Based on	IEEE802.11g
	Frequency	2.4 GHz
	Maximum Rate	28.8Mb/s
	Maximum Tx Power	17 dBm

Table 2. Simulator Specifications.

4.5 Scenario 3: Average throughput evaluation

In this simulator, ten terminals moves randomly in the simulation area and required rate of each user is set to 400kb/s. We assumed web browsing and small size streaming as an application.

For example, we picked up the performance of user #5 out of ten users.

In this simulation, user #5 started from EV-DO area for the first time, and then moved into WiMAX area and moved into wireless area, and finally moved into EV-DO area again.

Simulation result of user #5 is shown in left side of Fig. 13. From Fig. 13, performance of user #5 is improved by switching to other systems according to RSSI level of each system.
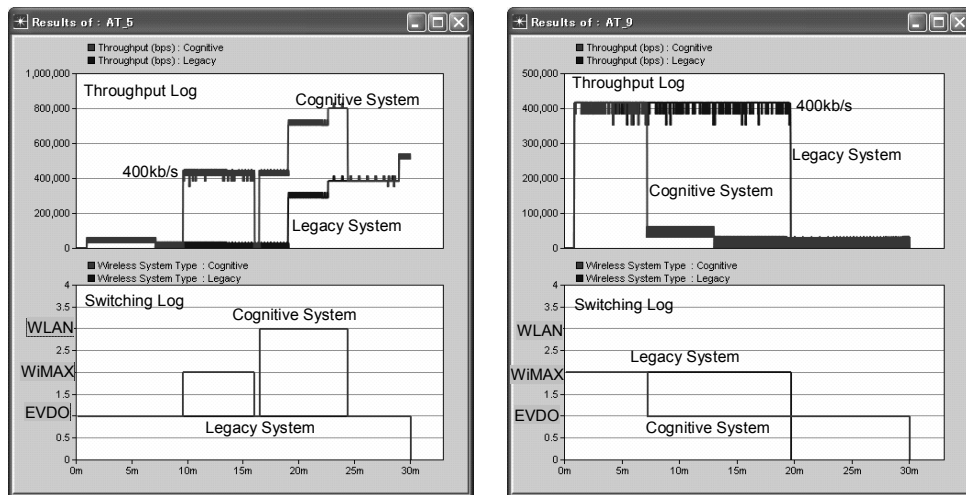
Fig. 13. Simulation results of user #5 (left), user #9 (right).

On the contrary, we picked up user #9 as another case. User #9 started from WiMAX area for the first time, and then moved into EV-DO area. Simulation Result of user #9 is shown in right side of Fig. 13. In this case, legacy system keeps connection with WiMAX until simulation time = 20minutes and then is disconnected due to long distance between WiMAX access point and user terminal, however proposed cognitive system decides to switch system from WiMAX to EV-DO when simulation time = 7minutes and then keep connection with EV-DO until user #9 is in EV-DO area. From Fig. 13, user throughput improvement is NOT seen by switching from WiMAX to EVDO.

The reason is WiMAX has wider bandwidth than EVDO, and expected throughput of WiMAX is faster than that of EV-DO. When user's required rate becomes higher, using WiMAX as long as possible gives better throughput performance as a total. Right side of Fig. 13 is one example that proposed system cannot achieve better performance than legacy system.

As a total, average throughput of ten terminals of both legacy system and proposed cognitive system is shown in Table 3. Regardless the example such as right side of Fig. 13, we have proved that throughput enhancement can be achieved by using proposed architecture.

System	Average Throughput
Legacy System	135.4kb/s
Cognitive System	230.9kb/s

Table 3. Average Throughput Comparison.

5. Experiments with testbed system

5.1 Preliminary testbed system

Based on the system architecture described in section 3, we have developed a preliminary testbed system that supports both WiMAX and wireless LAN. An overview of the testbed system is shown in Fig 14. In the preliminary testbed system, we connect the base station and the terminal via RF cables with fading simulators and variable attenuators inserted in the middle to emulate wireless radio propagation. The attenuation level of each system can be changed independently, manually and continuously, simulating the fluctuation of radio conditions.

In our experiments, we prepare two terminals just like as preliminary simulation scenario, that is, an application of user #1 is set to data streaming, and an application of user #2 is set to file downloading. .

System parameters of each wireless system for this experiment are shown in Table 4.
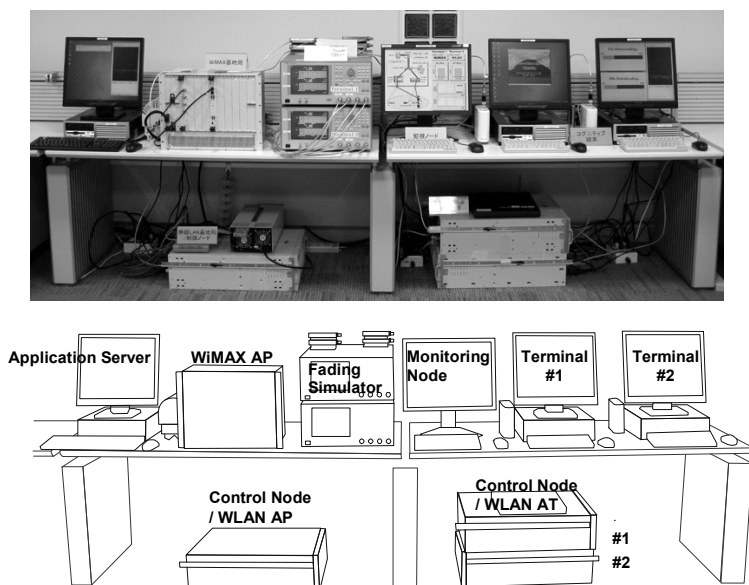
Fig. 14. Overview of Preliminary Testbed System.

RSSI level of wireless LAN can be adjusted by using fading simulators. Streaming data (equivalent to rtPS priority) is transmitted from the streaming server shown on the left of Fig. 14, to the cognitive terminal #1 shown in the right of Fig. 14, and at the same time, file download data (equivalent to Best Effort priority) is transmitted from the server shown on the left of Fig.14, to the cognitive terminal #2 shown in the right of Fig.14.

Moreover, overview of monitoring node is shown in Fig. 15. Lines in the left side shows the links that the terminals connect with, and RSSI value and its history are shown in the right side of the screen. Monitoring node has GUI interface to change the algorithm for system switching. Concretely, we can choose whether RSSI level is used or not and whether load

balancing is taken into account or not. These GUI interface is located on the center top of the screen.

Item		Specifications
WiMAX	Based on	IEEE802.16e (OFDM/TDD)
	Bandwidth	5 MHz
	Freq. Band	2.5 GHz
	Max Tx Power	30 dBm
	QoS Mode	rtPS / BE
Wireless LAN	Based on	IEEE802.11g
	Bandwidth	20 MHz
	Max. rate	54 Mb/s
	Freq. Band	2.4 GHz
	Max Tx Power	16 dBm
Data		Streaming / File Download

Table 4. Preliminary Testbed System Specification.
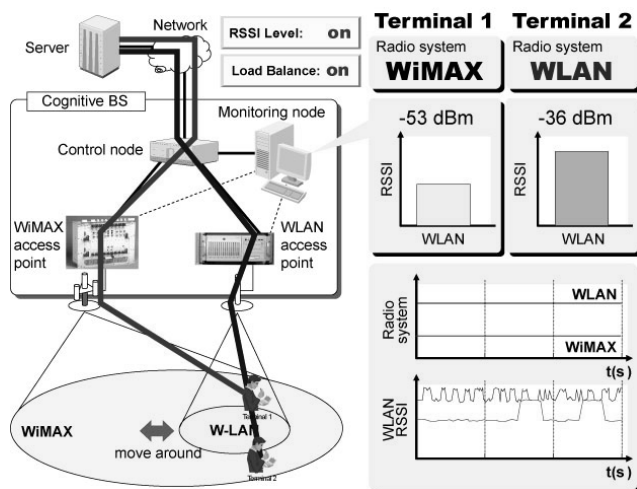
Fig. 15. Overview of Monitoring Node.

And also, the monitoring node has a function to set propagation loss of the fading simulators according to the terminal location located at the bottom of the screen (terminal location is same as the picture of standing men).

When both users are in WiMAX area, due to QoS function of WiMAX, user #1 that has higher priority than user #2, can communicate with required bit rate, and user #2 communicates with vacant bandwidth. Performance of this area is same as that of scenario 1 and 2 of the simulation (see the period (1) in Fig.16).

When both users come near an access point of wireless LAN, RSSI level of wireless LAN becomes higher. Provided that we switch the system based on both RSSI level of wireless LAN, both user #1 and user #2 switch to wireless LAN (see the period (2) in Fig.16). Provided that we switch the system based on both RSSI level of wireless LAN and also system load, control node decide that user #1 continues communication with WiMAX (not

switching to wireless LAN). On the contrary, user #2 does not have priority, therefore based on RSSI level and control nodes' decision of user #1, user #2 switches the system from WiMAX to wireless LAN.

As a result, any performance degradation can not be seen in both user #1 and user #2 transmission with the testbed system. User #1 can enjoy streaming service without any block noise or delay, and User #2 can download the files faster (see the period (3) in Fig.16). These performance enhancements are same as that of scenario 2 of the simulation.

Screenshot of the experiment is shown in Fig.16.
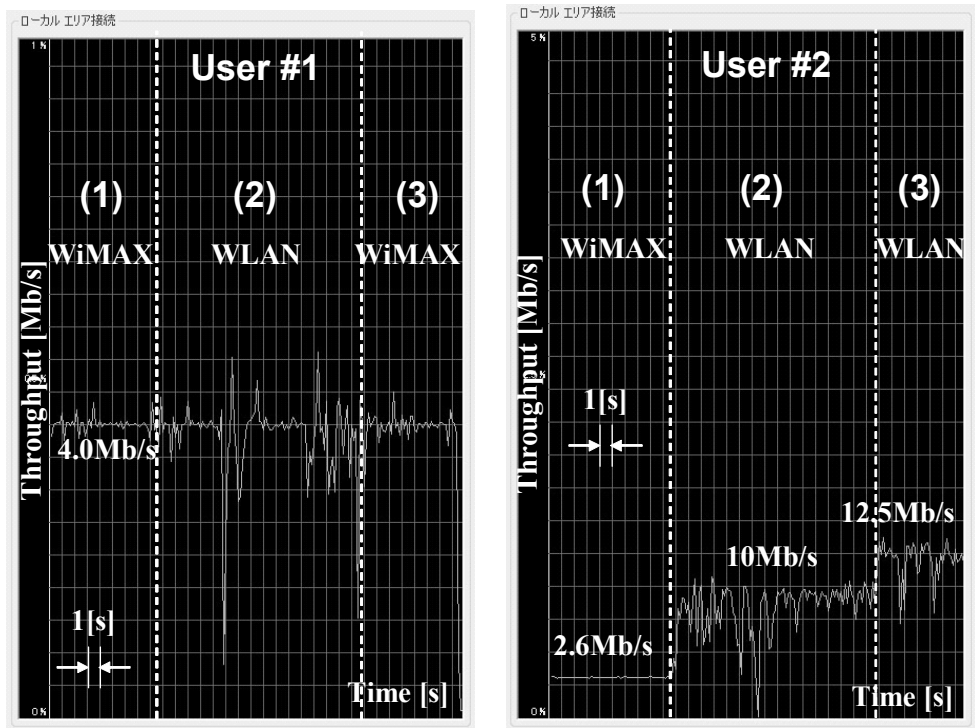
Fig. 16. Screenshot of the experiment.

5.2 Testbed system

In previous section, we described the results of experiments conducted on the preliminary testbed system that supports both WiMAX and wireless LAN. In this section, we describe a testbed system that supports EV-DO (Cellular), WiMAX, and wireless LAN. Specifications of the testbed system are shown in Table 5 and testbed overview is shown in Fig. 17. We had received experimental license for the testbed, and evaluate the testbed under outdoor environment.

System	Specifications	
EV-DO	Based on	cdma2000 1x EV-DO Rev.A
	Frequency	2.0GHz
	Maximum Rate	3.1Mb/s
	Maximum Tx Power	AP: 5dBm, AT: 24dBm
WiMAX	Based on	IEEE802.16e based (OFDM/TDD)
	Frequency	2.5 GHz
	Maximum Rate	6.2Mb/s (downlink)
	Maximum Tx Power	AP: 26dBm, AT: 14dBm
WirelessLAN	Based on	IEEE802.11g
	Frequency	2.4 GHz
	Maximum Rate	28.8Mb/s
	Maximum Tx Power	18 dBm
Data Transmission	UDP Packets	

Table 5. Specifications of testbed system
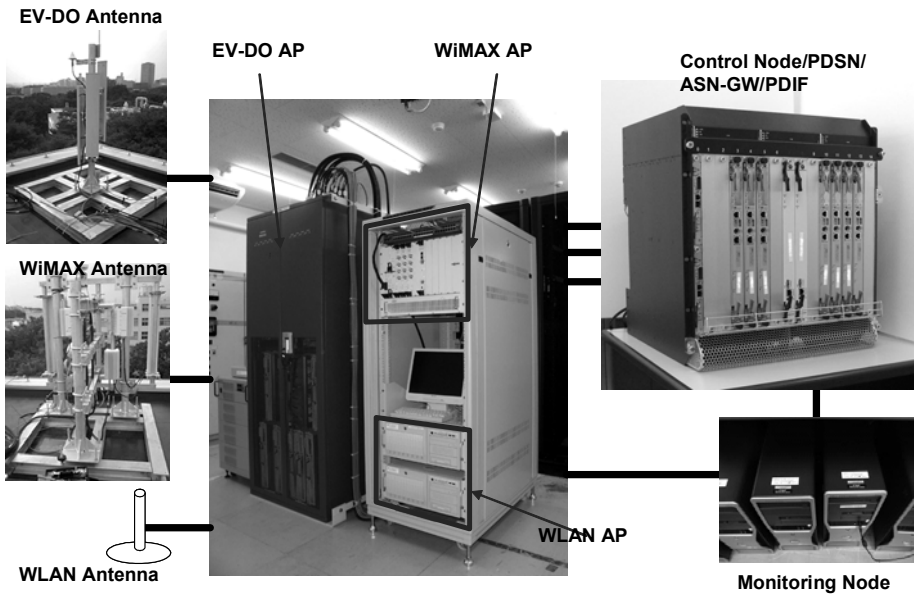
Fig. 17. Structure of Testbed System.

Outdoor experiment examples with the testbed system are shown in Fig. 18 and Fig. 19.
As shown in Fig. 18, we confirmed when received power of wireless LAN became lower, system decided to switches to EV-DO system automatically, and during system switching, connection with the terminal was continued.

Moreover, as shown in Fig. 19, we also confirmed when another terminal came into wireless LAN area, other terminal that connects with wireless LAN switches to EVDO to avoid conjestion and to achieve maximize system capacity.

Fig. 18. System switching example under outdoor environment.

Terminal #1 – WLAN   Terminal #2 starts   Terminal #1 – EVDO
Terimnal #2 - sleep   communication with WLAN   Terimnal #2 - WLAN
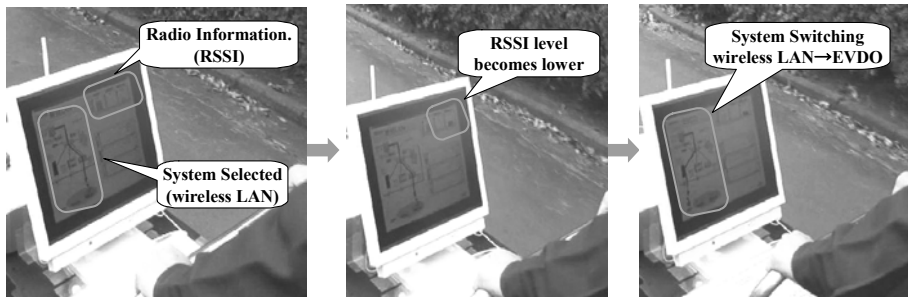
Fig. 19. System switching example under outdoor environment.

Through these experiments, we confirmed that system switching works correctly according to radio condition and system load under outdoor environment and that the system architecture described here is a reasonable architecture to achieve a convergence with plural wireless systems.

6. Conclusion

We described the architecture to integrate multiple-radio system with cognitive radio. Furthermore, we described the simulator and testbed system based on the architecture.

Through simulation and experiment with testbed system, we have proved that total system capacity was increased with using proposed architecture. These results suggest that the system architecture described here is one of reasonable platform to achieve a convergence with plural wireless systems.

Part of this project is funded by Ministry of Internal Affairs and Communications of the Japanese Government.

7. References

D.Bourse & K.El-Khazen. (2005). End-to-End Reconfigurability (E2R) research perspectives, *IEICE Technical Report SR2005-3*, pp. 21-28, May 2005, IEICE, Japan

S.Hanaoka; M.Yano. & T.Hirata. (2008). Testbed System of Inter-Radio System Switching for Cognitive Radio, *IEICE Trans. Communications, Vol. E91-B, No. 1*, pp. 14-21, Jan 2008, IEICE, Japan

S.Hanaoka; J.Yamamoto & M.Yano. (2008). Platform for Load Balancing and Throughput Enhancement with Cognitive Radio, *IEICE Trans. Communications, Vol. E91-B, No. 8*, pp. 2501-2508, Aug 2008, IEICE, Japan

H.Harada. (2005). A Study on Cognitive Radio and its applications, *IEICE, SR2005-18*, pp. 117-124, May 2005, IEICE, Japan

J.Mitola. (1999). Cognitive radio for flexible mobile multimedia communications, *Proceedings of 1999 IEEE Int Workshop on Mobile Multimedia Communications Digest*, pp. 3-10, Nov. 1999, IEEE

J.Mitola & G.Q.Maguire, J. (1999). Cognitive Radio: Making Software Radios More Personal, *1999 IEEE Personal Communication, Vol. 6, No. 4*, pp. 13-18, Aug. 1999, IEEE

T.Shono; K.Uehara. & S.Kubota. (2001). Proposal for System Diversity on Software Defined Radio, *IEICE Trans. Fundamentals, Vol. E84-A, No. 9*, pp. 2346-2358, Sep 2001, IEICE, Japan

G.Wu.; M.Mizuno & P.J.M.Havinga. (2002). MIRAI architecture for Heterogeneous Network, *IEEE Communications Magazine*, pp. 126-134, Feb 2002, IEEE

Power Amplification issues related to Dynamic Spectrum Access in the Cognitive Radio Systems

J. Palicot, Y. Louët and S. Hussain
SUPELEC/IETR
France

1. Introduction

For all sorts of telecommunication systems, in order to properly transmit the information, the signal has to be amplified to cope with the channel impairments and attenuations. Hence Power Amplifier (PA) is a vital element in a telecommunication system transmission chain. However, PA is still hard to design because of the key factors related to its performance which include the emitted mask power, battery consumption (linked to power amplifier efficiency), the channel propagation attenuation (linked to power amplifier gain), and the carrier-to-intermodulation ratio requirements (linked to power amplifier linearity). The design problem of PA is exacerbated due to the conflicting parameters of linearity and efficiency: the efficiency is the maximum near the saturation point, the point where non-linear effects are most severe.

In parallel to the PA non-linearity problem, the ever-increasing demands of high data rate transmission, bandwidth improvement, and the greater spectral efficiency have resulted in the migration of modulation schemes from single carrier to multi carrier. This trend has increased the signal power fluctuations by a great extent, a feature generally characterized by the term Peak-to-Average Power Ratio (PAPR). Consequently the efficient power amplification of multi carrier signals in the absence of any special processing induces severe impairments, the well known non-linear distortions. Hence turns up the problem of efficient power amplification of large envelope signals.

This chapter will discuss this problem in the context of Cognitive Radio (CR) systems. Software Radio (SWR), being the enabling technology of CR systems, inherits very high PAPR because of the multi-carrier nature of its signals. Moreover, as a consequence of Dynamic Spectrum Access (DSA), this high PAPR will be modified in real time. We will highlight the importance of this phenomenon and propose how to deal with this new situation. We explain how the PAPR could be considered as a sensor in the Cognitive Radio context during DSA process. Then we will propose and explain a frequency view of this sensor metric which could be of great interest to visualize the problem of PAPR modifications over DSA.

2. A brief historical note on the problem

PA has been considered as a very important transmitter component since long. So as the problem of PA non linear distortion is not new and has been extensively studied in past for different communication systems. For example,

- For satellite transmission, Traveling Wave Tube (TWT) amplifier is used and its non linear problems are discussed during 80's and 90's (Bremenson et al., 1980, Steck & Pham, 1987).
- For multi carriers systems the problem is studied extensively during 90's and 2000's (Han & Lee, 2005).
- Recently, for SWR system a European project, TRUST, highlights this issue (Macleod et al., 2000).
- Also, for Multiple-Input Multiple-Output (MIMO) systems the problem of PA non linearity is discussed (Lee et al., 2005, Tam et al., 2005).

This chapter deals with the PA non linearity problem in the context of CR systems where DSA can enhance this non linearity issue.

3. High Power Amplifier Characteristics

There exist several non-linear devices in a telecommunication transmission chain like Digital to Analog (DAC) converters, High Power Amplifiers (HPA). Transmitters employ HPA in order to send signals in the free space with sufficient power to counter channel fading. HPA being a non linear device causes in and out of band distortions, inter modulations and consequently Bit Error Rate (BER) degradation when operated in the non linear region. The obvious solution would be to "back off" the signal to get it operated in the linear amplifier region but this is accompanied by considerable PA efficiency penalty. Low PA efficiency means increased battery consumption which is an important issue especially for handset terminals. The other solution would be the linearization of the amplifier which faces amplifier design constraints problem (Kenington, 2000). Below we present HPA characteristics to highlight the non linearity issue.

3.1 Non linear gain characteristics
Mainly two types of HPAs are used in communications systems, Travelling Waves Tube mainly for satellite transmission and Solid State Power Amplifier (SSPA) commonly used in terrestrial transmissions and particularly in cellular systems for the handsets. Both the types of HPA present input and output amplitude and phase non-linear characteristics. The classical gain characteristics of SSPA and TWTA (TWT Amplifier) amplifiers are illustrated in the following curves:
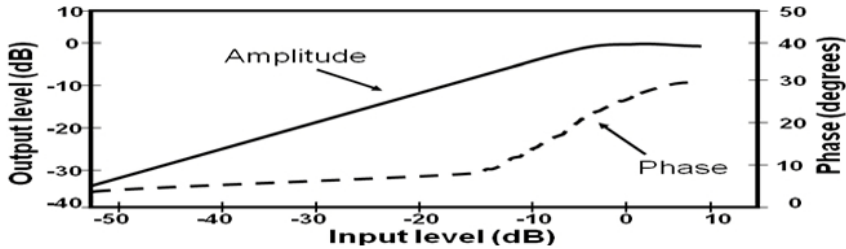
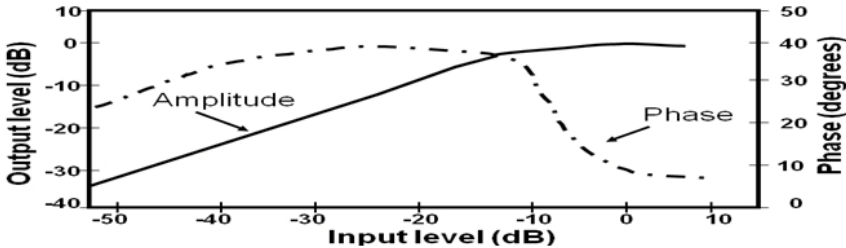Fig. 1. Amplitude and Phase characteristics of an SSPA power amplifier

Fig. 2. Amplitude and Phase characteristics of a TWTA power amplifier

The relation between amplifier input and output amplitude is represented by Amplitude /Amplitude (AM/AM) curve while Amplitude/Phase (AM/PM) curve describes the phase shift with the input variation. The AM/AM curve can be divided into three regions.

- **Linear Zone**: In this zone, HPA has approximately a linear behaviour i-e the output power is proportional to the input power. This proportionality is the gain of the amplifier without any degradation of the signal at output.

- **Compression Zone**: In this zone the proportionality between the input power and the output power no longer exists and signals perturbations appear. One very important point in this region is 1 dB compression point which is the point at which there is 1 dB difference between ideal gain curve and practical gain curve.

- **Saturation Zone**: In this zone the output power is quasi constant i-e output power is saturated denoted by P_{sat}.

It could be seen in Fig. 1 and Fig. 2 that AM/AM curves for the two types of HPAs are almost similar. But the AM/PM curves are quite different as in the case of SSPA, the phase curve is almost constant during the linear region but varies a lot for TWTA in the same region.

3.2 Efficiency characteristics

Power amplifier needs a dc source to amplify the signal. This dc value is associated to the efficiency of the amplifier. Figure.3 shows a simplified budget of the different powers associated to power amplifier.
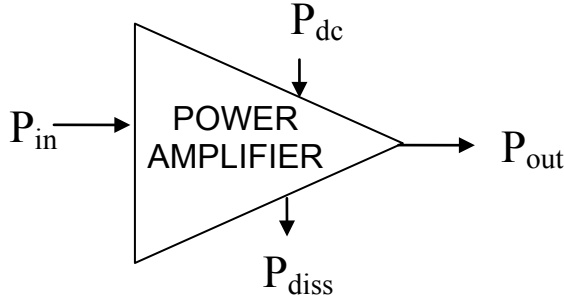
Fig. 3. Simplified power budget diagram of a power amplifier.

The power consumption budget is given by:

$$P_{in} + P_{dc} = P_{out} + P_{diss}$$

Power amplifier efficiency is the ratio between the output power and the source (dc) power. This is an important parameter as it is related to amplifier consumption which itself would be a great challenge for future communication systems. PA efficiency is given by,

$$\eta = \frac{P_{out}}{P_{dc}}.$$

There is another efficiency definition, power-added efficiency, which takes into account the input power and is given by,

$$\eta_{added} = \frac{P_{out} - P_{in}}{P_{dc}}.$$

3.3 Input and Output back-off

To nullify or reduce the non-linear effects, the operating point of power amplifier is backed-off. The amount of input and output back-off from 1 dB compression point or saturation point is called Input Back Off (IBO) and Output Back Off (OBO) respectively.

Let P_{in} be the signal mean input power, P_{out} mean output power, $P_{sat,out}$ saturation output power and $P_{sat,in}$ the corresponding input power as shown in Fig. 4. The IBO and OBO would then defined as,

$$IBO = \frac{P_{sat,in}}{P_{in}} \quad or \quad IBO(dB) = P_{sat,in}(dBm) - P_{in}(dBm) \quad .$$

$$OBO = \frac{P_{sat,out}}{P_{out}} \quad or \quad OBO(dB) = P_{sat,ou}(dBm) - P_{out}(dBm) \quad .$$

IBO and OBO can also be defined in terms of $P_{out,1dB}$ and $P_{in,1dB}$ (Kenington, 2000) where $P_{out,1dB}$ is the output power at 1 dB compression point and $P_{in,1dB}$ is the corresponding input power.

$$IBO = \frac{P_{in,1dB}}{P_{in}} \quad or \quad IBO(dB) = P_{in,1dB}(dBm) - P_{in}(dBm) \quad .$$

$$OBO = \frac{P_{out,1dB}}{P_{out}} \quad or \quad OBO(dB) = P_{out,1dB}(dBm) - P_{out}(dBm) \quad .$$

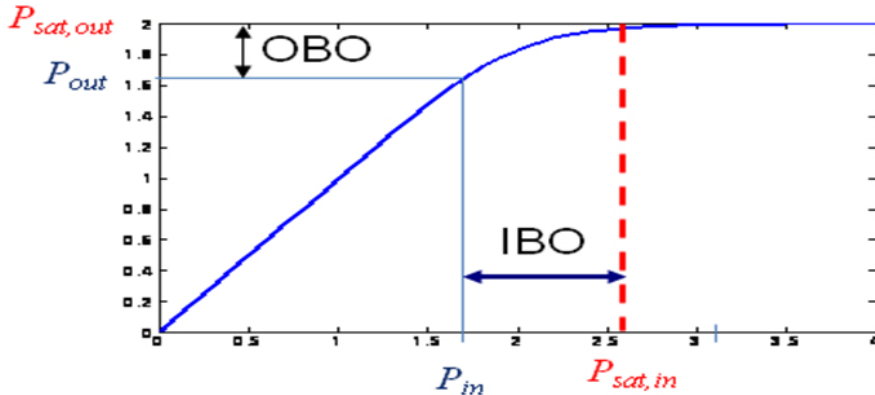Fig. 4. Different power levels on the PA gain curve.

3.4 Carrier to Intermodulation (C/I) ratio

It is the ratio of useful carrier power level to intermodulation products power level. Normally this term is used to describe intermodulation effects in narrow band systems. C/I_n is the C/I ratio of useful power to the intermodulation product of order n.

3.5 Interception point

The n-order intercept point relates nonlinear products caused by the nth order term in the nonlinearity to the linearly amplified signal. Generally 3rd order intercept point is used to relate intermodulation products to input signal. Fig. 5 depicts the Interception point concept.
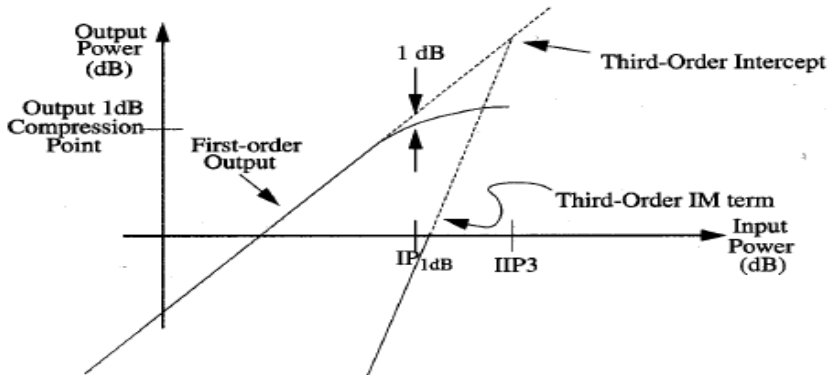
Fig. 5. Third order interception point

3.6 Memory effect

With the rapid growth of communication systems, the future systems are supposed to operate on a very large bandwidth. And HPA characteristics should be frequency

independent otherwise the performance of the amplifier shall be affected by the frequency of the signals being amplified (Soury & Ngoya, 2005). When HPA performance is dependent on frequency it is said to have 'memory'. Memory effects of an amplifier have a very important impact on the overall performances of the HPA and should not be neglected for future systems.

3.7 HPA modeling

There are several HPA models presented in literature with and without memory effect considerations. Below we shall very briefly discuss some HPA models.

3.7.1 HPA modeling without memory effect

For an input signal $S(t) = |S(t)|e^{j(\omega_c t + \varphi_c(t))}$ with $s(t) = |S(t)|\cos(\omega_c t + \varphi_c(t)) = \mathrm{Re}\{S(t)\}$ the transmitted signal, the output signal is given by:

$$y(t) = A(|S(t)|)\cos(\omega_c t + \varphi_c(t) + \Phi(|S(t)|)) ,$$

where $A(|S(t)|)$ is relative to the AM/AM transfer function and $\Phi(|S(t)|)$ is relative to the AM/PM transfer function. For an ideal linear amplifier $A(|S(t)|)$ is constant as well as phase $\Phi(|S(t)|)$ independent of the input power $|S(t)|$.

The most commonly used model, mainly for TWTA was the **Saleh model** (Kenington, 2000) with,

$$A(|S(t)|) = \frac{v|S(t)|}{1 + \beta_a |S(t)|^2} ,$$

$$\Phi(|S(t)|) = \frac{\alpha_\phi |S(t)|}{1 + \beta_\phi |S(t)|^2} ,$$

and

$$A_s = \frac{1}{\sqrt{\beta_a}}$$

where v is the small signal gain and $\sqrt{\beta_a}$ is the input saturation power . The maximum phase shift is $\Phi_\infty = \frac{\alpha_\phi}{\beta_\phi}$.

Rapp model is commonly used for SSPA in transmission systems (Kenington, 2000) with

$$A(|S(t)|) = \frac{v|S(t)|}{\left\{1 + \left[\frac{v|S(t)|}{A_0}\right]^{2p}\right\}^{\frac{1}{2p}}} ,$$

$$\Phi(|S(t)|) \approx 0 ,$$

where $A_0 = vA_S$ is the output saturation value, p is an integer. When p becomes large the AM/AM curve approaches the ideal transformation known as 'soft limiter', i-e $A(|S(t)|) = v|S(t)|$ for $0 \le |S(t)| \le A_S$ and $A(|S(t)|) = vA_S$ for $A_S \le |S(t)|$.

3.7.2 HPA modeling with memory effect

In practice, AM/AM and AM/PM curves are different as function of the frequency. This phenomenon should be taken into account in wideband application such as SWR. Models which take into account this frequency parameter are called models with memory. There are several models like Saleh, Bösch, Blum and Meghdadi (Kenington, 2000) which consider memory effect. Only Saleh model is presented here to give an idea.

Saleh model is given by the following 4 coefficients $\alpha_p, \beta_p, \alpha_q, \beta_q$, with the rules proposed by Saleh and illustrated by the Fig. 6.

$$H_p(f) = \sqrt{\beta_p(f)} \quad and \quad H_q(f) = \sqrt{\beta_q(f)}$$

$$P(A) = \frac{\alpha_p(f)A}{1 + \beta_p(f)A^2} \quad and \quad Q(A) = \frac{\alpha_q(f)A^2}{1 + \beta_q(f)A^2}$$

$$G_p(f) = \frac{\alpha_p(f)}{\sqrt{\beta_p(f)}} \quad and \quad G_q(f) = \frac{\alpha_q(f)}{\beta_q(f)\sqrt{\beta_q(f)}}$$
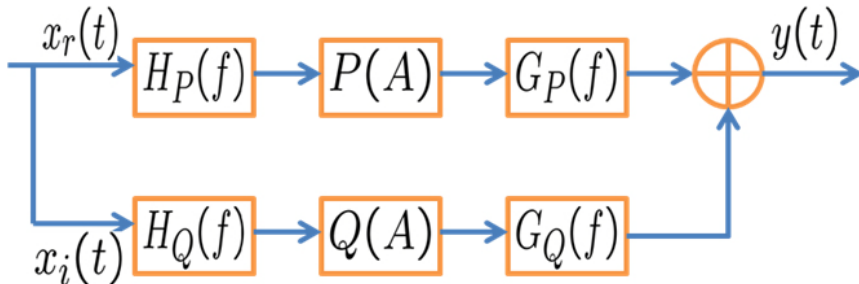
Fig. 6. Saleh model with memory effect.

4. Envelope characteristics of multi-carrier signals

There are two types of signals with respect to signal carrier frequency. First type is that of single carrier signals where the data is transferred on a single carrier frequency like in Global System for Mobile communications (GSM) where each user's data is transmitted on a particular frequency. Second type is that of multi-carrier signals where the data is transmitted over many carrier frequencies like in Orthogonal Frequency Division Multiplexing (OFDM) where the information is modulated on several carriers. The envelope characteristics of single carrier and multi-carrier signals are different from each other and in this part we shall highlight these differences.

4.1 Single carrier signals

In single carrier communication like the transmission of a Quadrature Phase Shift Keying (QPSK) or Gaussian Minimum Shift Keying (GMSK) modulated data, the information is transmitted over a single carrier frequency. Thus the power fluctuations of this signal depend upon the constellation scheme and shaping filter characteristics. For example in case of amplitude modulations like Quadrature Amplitude Modulation (QAM), more power fluctuations should be observed compared to those of phase modulations like QPSK as maximum and mean power of the symbols is the same in case of phase modulation. Also the characteristics of the shaping filter affect the signal envelope, as in case of Root Raised Cosine (RRC) filter, the envelope varies with the change in filter's roll-off factor. Fig. 7 shows the power fluctuations of QPSK modulated single carrier signal with roll-off factor of 0.22.
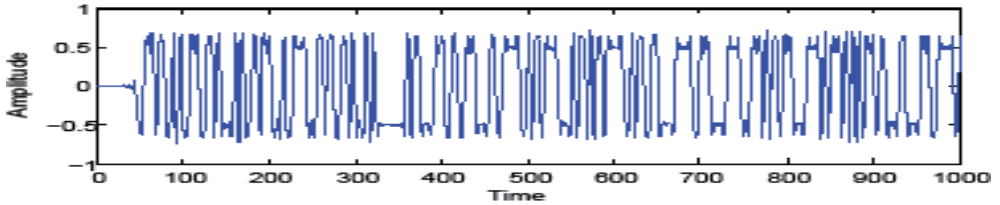
Fig. 7. Power fluctuations in single carrier signals (QPSK modulation, roll-off=0.22).

4.2 Multi-carrier signals

4.2.1 OFDM signals

In case of multi carrier communication like OFDM, a large number of sub carriers are added simultaneously to make a multi carrier signal. For example an OFDM symbol containing N sub carriers would be formed as,

$$S(t) = \sum_{j=-\infty}^{+\infty} \sum_{k=0}^{N-1} C_{j,k} g(t - jT_s) e^{2i\pi f_k t}.$$

(1)

where $g(.)$ is the rectangular pulse of duration T_S, $C_{j,k}$ represents the complex information symbol of carrier k with $f_k = \frac{k}{T_s}$ of jth OFDM symbol . For any l^{th} OFDM symbol, the baseband signal $S_l(t)$ is given by,

$$S_l(t) = \sum_{k=0}^{N-1} g(t - lT_S) C_{l,k} e^{2i\pi f_k t},$$

(2)

or,

$$S_l(t) = \sum_{k=0}^{N-1} C_{l,k} e^{2i\pi f_k t}, t \in [lT_S, (l+1)T_S].$$

(3)

According to the central limit theorem, when a large number of independent and identically distributed (i.i.d) random variables are added simultaneously, their distribution becomes Gaussian which means that there is a big gap between mean and maximum values. As OFDM signal can be treated like a summation of many i.i.d random variables, its distribution tends to be Gaussian following the central limit theorem which results in high power fluctuations as shown in Fig. 8.
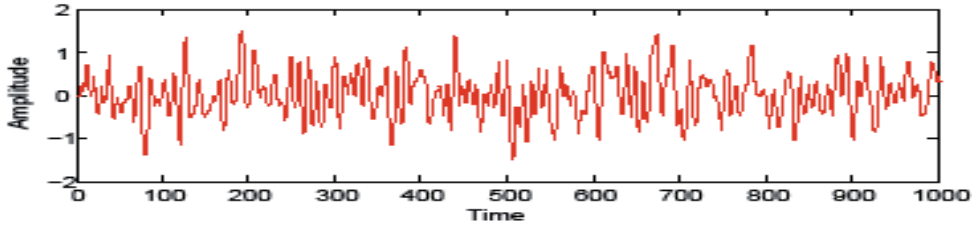
Fig. 8. Power fluctuations in multi carrier signals (N=64, QPSK Mapping).

4.2.2 Software Radio signals

SWR can be defined as a system able of modulating and demodulating any kind of signal, any time anywhere, on any network (Fasbender et al., 1999). Consequently any SWR signal $x(t)$ is a composite signal given by,

$$x(t) = \sum_{i=1}^{S} S_i(t),$$

(4)

where S is the number of standards contained in the composite SWR signal and S_i represents the ith standard. Each signal $S_i(t)$ associated to a given standard of P_i carriers can be expressed as,

$$S_i(t) = \sum_{p=1}^{P_i} r_{i,p}(t) e^{2i\pi f_{i,p} t},$$

where $r_{i,p}(t)$ represents the modulated and filtered signal associated to *carrier* p of the *standard* i. In this case, $r_{i,p}(t) = fem_i(t) * m_{i,p}(c(t))$, where $m_{i,p}(c(t))$ represents the modulation relative to the *carrier* p and standard i and $fem_i(t)$ is the shaping filter function of standard i. Using the two aforementioned equations, a multi-standard signal can be written as,

$$x(t) = \sum_{i=1}^{S} \sum_{p=1}^{P_i} (fem_i(t) * m_{i,p}(c(t)) e^{2i\pi f_{i,p} t}.$$

(5)

As it could be seen that SWR signal is a multi-carrier signal and due to this property it demonstrates high signal fluctuations.

4.3 Peak to Average Power Ratio (PAPR)

As discussed previously that HPA is a non-linear device and to operate it in the high efficiency zone, the power fluctuations of the signal should be minimized otherwise non linear distortions would occur. These power fluctuations are described by various terms in literature but most common of all is PAPR (Peak to Average Power Ratio). It is defined as the ratio between maximum instantaneous power and mean instantaneous power. For a complex base band signal, $s(t) = s_I(t) + s_Q(t)$, PAPR would be defined as,

$$\text{PAPR}\{s(t)\} = \frac{\max_t |s(t)|^2}{\lim_{T \to \infty} \frac{1}{T} \int_0^T |s(t)|^2}.$$

(6)

For Radio Frequency (RF) signal with $f_0 = \frac{\omega_0}{2\pi}$ being the carrier frequency. The PAPR is defined as,

$$\text{PAPR}\{s(t)\} = \frac{\max_t |Re(s(t)e^{j\omega_0 t})|^2}{\lim_{T \to \infty} \frac{1}{T} \int_0^T |Re(s(t)e^{j\omega_0 t})|^2 dt}.$$

(7)

The relation between base band signal PAPR, $PAPR_{BB}$, and RF signal PAPR, $PAPR_{RF}$, is given by (Tellado-Mourelo, 1999),

$$\text{PAPR}_{RF} \; dB \approx \text{PAPR}_{BB} \; dB + 3 \; dB.$$

(8)

All in all, PAPR is the term which demonstrates the variations in the peak of the signal power with respect to its mean power.

4.4 Equivalence between OFDM and SWR signals

SWR and OFDM both are multi-carrier signals and we shall try to make analogies between these two types of signals. First analytical equivalence is demonstrated and then PAPR distribution is discussed.

4.4.1 Analytical equivalence

If the carrier interspacing is constant for all standards, SWR signal will become a multicarrier signal, given by,

$$x(t) = \sum_{i=1}^{S} \sum_{p=1}^{P_i} r_{i,p}(t) e^{2i\pi((p-1)\triangle p)t}.$$

(9)

where $\triangle p$ is the carrier interspacing which is supposed to be constant for all standards. Generally speaking, this condition cannot be verified because the distance between carriers is different from one standard to the other (Roland & Palicot, 2002). It varies from 25 kHz for Personal Digital Cellular (PDC) to 50 MHz for Hiperlan. When SWR signal contains more than one standard, it cannot be a 'typical' multicarrier signal. Therefore we cannot make any connection between an OFDM signal and a multi-standard SWR signal.

We suppose now a mono standard SWR signal. By definition then,

$$S_i(t) = \sum_{p=1}^{P_i} r_{i,p}(t) e^{2i\pi f_{i,p} t}, i = 1.$$

(10)

which is equivalent to

$$S_i(t) = \sum_{p=1}^{P_i} r_{i,p}(t) e^{2i\pi((p-1)\triangle p)t}, i = 1.$$

(11)

From an analytical point of view, Eq. 11 is formulated in the same way as equation Eq. 2 models an OFDM signal. In all other cases, an OFDM signal remains a particular case of a SWR signal.

4.4.2 Gaussian Equivalence between OFDM and SWR Signals

OFDM PAPR distribution has been discussed a lot in literature. (Ochiai & Imai, 2001) deals with PAPR distribution for OFDM signals and gives some Complementary Cumulative

Distribution Function (CCDF) upper bounds. With large oversampling factors (continuous signals) and knowing the fact that there is approximately 3 dB PAPR increase due to Radio Frequency (RF) transposition, PAPR CCDF for RF OFDM continuous signals is approximated by,

$$Pr[PAPR \geq PAPR_0] \approx 1 - e^{-\sqrt{\frac{\pi}{6}}N\sqrt{PAPR_0 e^{-PAPR_0}}}.$$

(12)

Below we prove that OFDM PAPR CCDF is similar to PAPR CCDF for multi standard SWR signals.

PAPR analysis of a multi standard SWR signal containing OFDM, MC-GMSK and MC-QPSK signal, is considered. General specifications used for each standard are same i.e. $N = 64$ carriers/subcarriers and 10^4 symbols. This analysis is done in RF conditions and is depicted in Fig. 9. First, it can be shown that all standards have almost similar PAPR CCDF, confirming Gaussian equivalence. Secondly, from a statistical point of view, as SWR signal is generated by the combination of these standards, PAPR CCDF of a SWR signal also demonstrates Gaussian distribution. In conclusion, any multiplex of carriers or standards, exhibits Gaussian distribution, independent of the modulation techniques and the carrier frequencies. From a PAPR point of view, SWR and OFDM show similar behaviors.
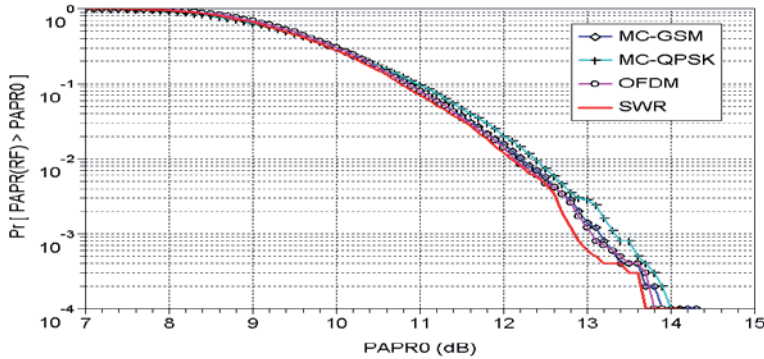

Fig. 9. PAPR distribution of a SWR signal with MC-GMSK, MC-QPSK and OFDM modulated standards in RF.

5. High Power Amplification of multi-carrier signals

As discussed previously that multi-carrier signals have high power fluctuations and these fluctuations can cause interference when the signal is amplified by HPA. We shall see some of the interference types produced on non-linear amplification.

5.1 Non linear characteristics

We can approximate any non linear amplifier f(.) by a polynomial function :

$$f(.) = a_0 + a_1 x + a_2 x^2 + ... + a_n x^n$$

For an OFDM signal s(t) defined in Eq. 2, the output of power amplifier would be,

$$y = f(s(t)) = a_0 + a_1 s(t) + a_2(s(t))^2 + ... + a_n(s(t))^n$$

All the aforementioned terms generate components whose frequencies F_s are of the form,

$$F_s = \sum_{i=0}^{N-1} m_i f_i$$

(13)

F_S is called intermodulation product of order $\sum_{i=1}^{N} |m_i|$ where m_i are arbitrary integer values.

Some of the F_S frequencies fall in the useful band resulting in BER degradation while the rest fall out of the useful band causing spectrum re-growth. Some terms are defined below to better understand the intermodulation concept.

5.1.1 Adjacent Channel Power Ratio (ACPR)
In wideband systems, ACPR is used instead of C/I. ACPR is the ratio of signal power inside the useful band to the adjacent channel power.

$$ACPR_{dB} = 10 \cdot \log \frac{\int_{f_1}^{f_2} P(f)df}{\int_{f_3}^{f_4} P(f)df + \int_{f_5}^{6} P(f)df}$$

The different frequency positions are described in Fig.10. In (Ragusa et al., 2005), ACPR definition for multi carrier OFDM signal is given as,

$$ACPR = \frac{a_3^2 . A^4 . (3/8)^2 \sum_{i=1}^{N-1} u_i^2}{a_1^2 . N/4}$$

where a_3 and a_1 are third and first intermodulation product coefficients, A is the normalized amplitude of the multi carrier signal, N is the number of OFDM subcarriers and $u = \frac{(N-i)(N-i+1)}{2}$.
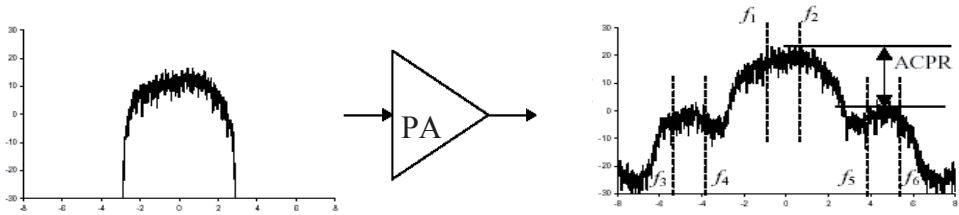

Fig. 10. Spectrum distortion after power amplification

5.1.2 EVM and MER ratios
In OFDM systems, Error Vector Magnitude (EVM) and Modulation Error Rate (MER) are the terms used to define the distortion caused by HPA in frequency and time domain respectively.

Power Amplification issues related to
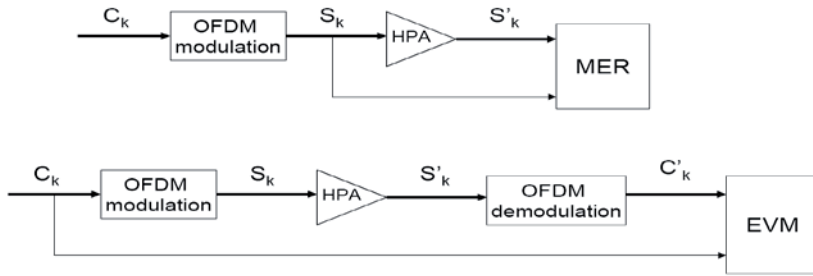Dynamic Spectrum Access in the Cognitive Radio Systems
329

Fig. 11. MER and EVM graphical illustrations
EVM represents the influence of HPA on constellation symbols while MER describes the influence of HPA on time domain signal.

5.2 Solution to avoid HPA non-linearity

To amplify the signal at high efficiency, it needs to be driven near the saturation point. Consequently the signal should have low PAPR in order to not enter the saturation zone. Thus there was need of PAPR reduction methods and a huge number of methods have been developed since 15 years to reduce multi-carrier signal PAPR. The main reason is the success of multi-carrier modulations like OFDM as modulation and multiple access techniques in many communication standards like Digital Audio Broadcasting (DAB), Digital Video Broadcasting (DVB), Asymmetric Digital Subscriber Line (ADSL), and Wireless MAX (WiMAX) etc. A high PAPR being the drawback of OFDM, extensive research has been carried out to resolve this issue. Historically, the signal used to be backed-off and transmitted at low efficiency without PAPR reduction but with the development of technology and emerging of the new standards, battery consumption has gained a lot of attention and in turns PAPR reduction.
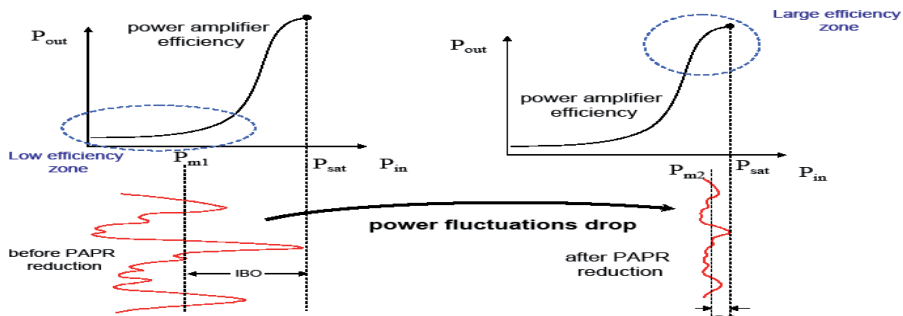

Fig. 12. High efficiency PA operation demands reduction in PAPR.

5.3 Classification of PAPR reduction methods

A classification of PAPR reduction methods has been proposed in (Louët & Palicot, 2008). It mainly concerns OFDM PAPR reduction but some methods can be applied to any modulation scheme. The idea behind this classification is to gather the methods with similar characteristics into clusters and facilitate the selection of PAPR reduction method depending

upon system requirements. There are different criteria to differentiate between existing PAPR reduction methods. Few of them are outlined below:

- Downward compatibility (DC): A method is called downward compatible if it does not imply any receiver change. When downward compatible, the PAPR reduction is implemented only at the transmitter meaning that the input data is manipulated to generate low PAPR. In that case BER needs to be investigated. Thus we have two subclasses when the method is downward compatible.

- o   Cluster 1 : DC without BER degradation (e-g Active Constellation Extension)
- o   Cluster 2 : DC with BER degradation (e-g Clipping, Peak windowing)

-If there is no DC, PAPR reduction method is implemented on both transmitter and receiver (the function and its "inverse") and theoretically BER remains unchanged. In this case data rate loss has to be investigated.

- o   Cluster 3 : no DC without date rate loss (e-g Pulse shaping, Tone Injection)
- o   Cluster 4 : no DC with date rate loss degradation (e-g Coding, Tone Reservation, Partial Transit Sequences)

These are the few examples given to illustrate the classification. The complete tree diagram of method classification is given in Fig. 13. The classification presented also refers to linearization methods which do not take into account the PAPR reduction techniques.

As already described that SWR signal is a multiplex of many modulated carriers and demonstrates a complex Gaussian distribution as OFDM thus its PAPR can be reduced using the methods developed for OFDM PAPR reduction with appropriate changes.
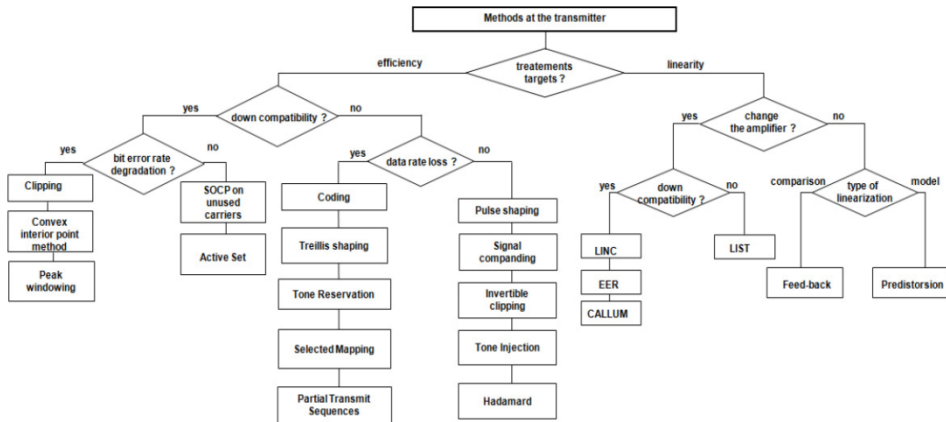

Fig. 13. Classification of PAPR reduction methods.

6. Dynamic Spectrum Access (DSA) in the Cognitive Radio context

The term "Cognitive Radio" was first coined by Mitola in 1999 who gave the following definition of CR (Mitola & Maguire, 1999),

"A radio or system that senses, and is aware of, its operational environment and can dynamically and autonomously adjust its radio operating parameters accordingly."

In other words, CR refers to self-aided and self-established reconfigurable communication system where decisions of reconfiguration are taken with the help of a set of sensors. One of these sensors is the spectrum sensor which using an opportunistic approach senses any available bandwidth and thus assists in establishing communication on suitable frequencies as suggested by Federal Communication Commission (FCC) in (FCC, 2003),

"A radio frequency transmitter/receiver that is designed to intelligently detect whether a particular segment of the radio spectrum is currently in use, and to jump into (and out of, as necessary) the temporarily-unused spectrum dynamically, without interfering with the transmissions of other authorized users."

This FCC definition is given in the context of spectrum allocation and utilization. The reason of this dynamic spectrum access is the under utilization of radio spectrum. In fact around 90% of spectrum is quiet on average and there is a great need to 'recover' the wasted spectrum resources. There are a lot of ways discussed in literature to 'recover' the spectrum (Zhao & Swami, 2007).

6.1 The inconveniences of spectrum access

In the spectrum access scenario, a CR equipment should be able to access any available bandwidth in order to fulfill its data needs. Anyhow, this spectrum access in a CR context makes sense only if the bandwidths proposed by spectrum sensor do not imply any modification to the CR equipment's Analog Front End (AFE). AFE consists of elements like filters, antennas, power amplifiers etc. The design and performance of these AFE components depend upon temporal and spectral properties of the signal, for example, antennas are designed for a specific frequency range or amplifier design is input amplitude dependent etc.

Also spectral characteristics affect temporal signal properties like signal bandwidth and constellation size relation for a specific data rate demand. Moreover, spectrum access can result in out of band distortions. This results in decrease of ACPR which degrades system performance as bit error rate increases.

Concluding, spectrum access can affect CR equipment's AFE design and also temporal signal characteristics. As CR should not modify AFE design thus spectrum should be accessed as to keep signal's temporal and spectral properties within AFE design limits. One of the AFE's elements, HPA, is of particular interest in this chapter. As HPA is susceptible to high PAPR signals, spectrum sensor should take into account the PAPR value of the emitted signal on the available bandwidth to keep the PAPR value under an acceptable level which does not imply any changes to power amplifier design. Thus the selection of suitable bandwidth should be made according to some PAPR metrics.

6.2 CR signal and its PAPR

CR uses SWR as its implementation technology. In SWR based-systems, PAPR may be quite large due to the fact that the transmitted signal is a sum of a large number of modulated carriers as discussed before. Moreover, a reconfigurable transmitter, Software Defined Radio (SDR) base station for instance, should be able to amplify single carrier and multi carrier modulation signals as well as many non-constant envelope modulation signals. Thus CR signals inherit high PAPR and this factor asks for PAPR information based spectrum access

decisions. This reason emphasizes the necessity of a frequency domain PAPR approach in a spectrum access scenario.

6.3 Spectrum access and PAPR problem

In a CR context spectrum access is about finding the free bandwidth within a standard or in between the standards for communication using an opportunistic approach as shown in Fig. 14. There are certain parameters which should be kept in mind during spectrum access for example Quality of Service (QoS), out of band interference etc. Our emphasis in this chapter is on the influence of spectrum access on PA efficiency.
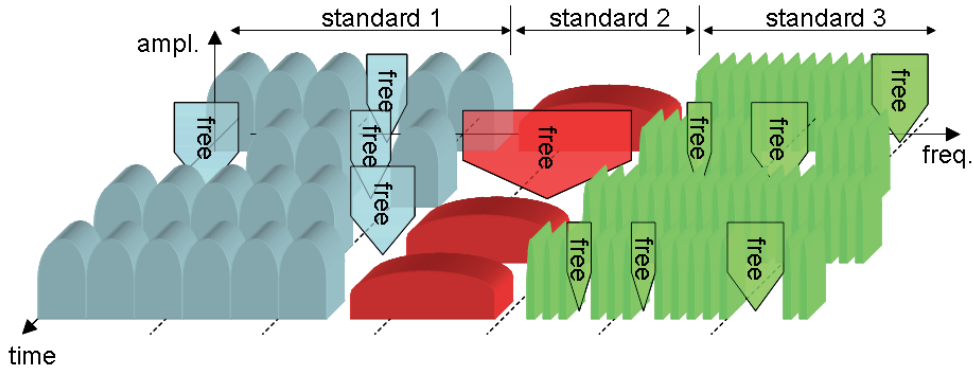
Fig. 14. Spectrum access in a multi-standard system scenario.

HPA efficiency curve is shown in Fig. 15 following the context of the discussion here. Generally it is preferred to operate PA in the maximum efficiency region near saturation point. But for that signal fluctuations should be kept as small as possible otherwise non-linearity would occur because of HPA operation in saturation region. Now spectrum access would definitely modify signal properties namely mean power, bandwidth, amplitude etc. As the signal fluctuations are modified on spectrum access, the operating point of HPA needs to be modified also so as to avoid non-linearity. Consider that the signal fluctuations increase on spectrum access which is generally the case after signal addition to useful band, one solution can be large IBO labeled as `Solution 1: Large IBO' in Fig. 15. This large IBO results in the PA operation in less efficient zone which is obviously not recommended. Also, one needs to add some electronic components for real time operating point modifications according to signal fluctuations. This hardware implementation not only comes at the cost of more energy consumption but also it is difficult to realize on the transmitter circuitry.

The solution which we propose, labeled as `Solution 2: Insertion with PAPR constraint' in Fig. 15, is to insert the added tones respecting the signal fluctuation constraints. If the added signal does not increase signal fluctuations, high efficiency PA operation is easily possible. Thus a frequency view of PAPR is highly essential to see the effect of the added signal on PAPR and to decide whether to insert the signal on some available band or not and if yes then what should be the characteristics of the added signal like amplitude or constellation size, bandwidth and position of the added carriers.
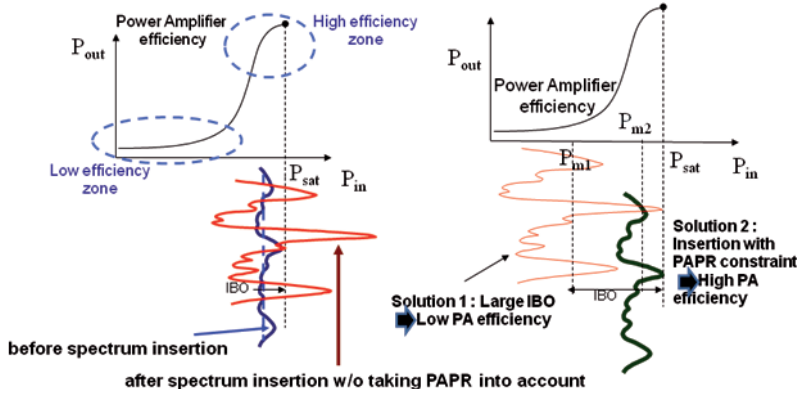
Fig. 15. Spectrum access effect on PAPR and in turns on PA efficiency.

7. Frequency view of PAPR metric

7.1 Frequency Domain Interpretation of PAPR Metric

Since PAPR is varied on spectrum access, its frequency domain interpretation would help in better understanding of these variations corresponding to the frequency components of signal (Hussain et al., 2008). To get the frequency vision of the signal, temporal SWR signal is first sliced into N_s symbols of N samples each. This process of slicing the temporal signal is illustrated in Fig. 16. These symbols fill the N_s rows of table $'S'$ in Fig. 17. Then, each row is processed with N point FFT to get the spectral components of the signal and thus table $'C'$ is filled in Fig.17.
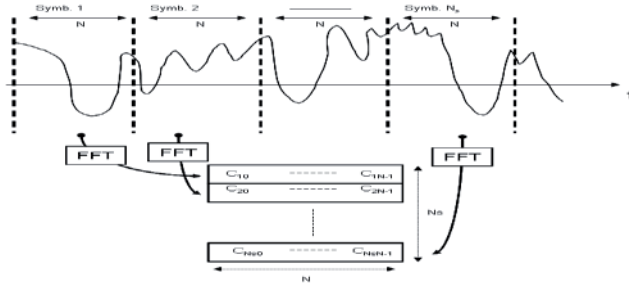

Fig. 16. Temporal signal slicing for PAPR upper bound calculation.

Note that, N should be large enough to contain all the frequency components in SWR signal. In Fig. 17, each row in table $'C'$ refers to frequency components $X_i(n)[i \in [J]]$ present in temporal SWR symbol $S_i(n)$ given by,

$$X_j(n) = \sum_{k=0}^{N-1} s_{j,k} e^{-2i\pi \frac{kn}{N}}, n \in [(j-1)N, jN-1], j \in [1, N_S].$$

(14)

Here $J = [1, N_S]$ and each column is associated to a signal $\nu_k(n)$ given by,

$$\nu_k(j) = \sum_{m=0}^{N_S-1} C_{m,k} e^{2i\pi \frac{jm}{N_s}}, k \in K, j \in J.$$

$$(15)$$

Here $K = [0, N-1]$. A column k is associated to a carrier f_k and the idea is to calculate PAPR column wise and to show that it is similar to PAPR calculated row wise.
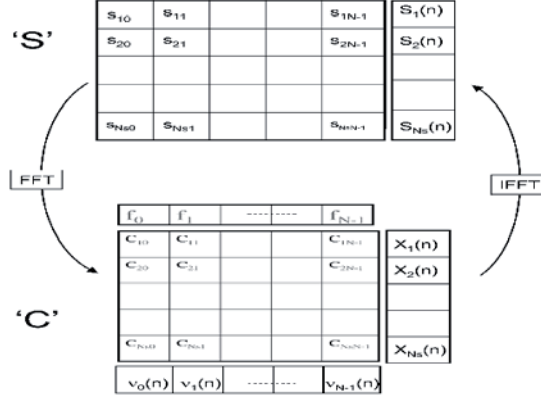

Fig. 17. Time and frequency vision of an SWR signal.

This new vision paves the way to a carrier wise analysis of PAPR whose objective is to find a relation between $PAPR_{N_S}$ (PAPR calculated over time domain table 'S') and power ratios $PAPR_{f_k}$ calculated on each carrier f_k given by,

$$PAPR_{f_k} = \frac{\max_{j \in J} |C_{j,k}|^2}{\frac{1}{N_S} \sum_{j=1}^{N_S} |C_{j,k}|^2} = \frac{\max_{j \in J} |C_{j,k}|^2}{P_m(k)},$$

$$(16)$$

where $P_m(k)$ refers to mean power calculated on carrier f_k. By using temporal expression of the SWR symbols, it can be easily shown that,

$$PAPR_{N_s}(S(n)) = \frac{\max_{k \in K}(\max_{j \in J}(|s_{j,k}|^2))}{\frac{1}{N_s} \sum_{j=1}^{N_s} E[|s_{j,k}|^2]}.$$

$$(17)$$

Using Fourier transform relation,

$$|s_{j,k}|^2 = \sum_{p=1}^{N} |C_{j,p}|^2 + \lambda(j,k), j \in J, k \in K.$$

$$(18)$$

with

$$\lambda(j,k) = \sum_{p=1}^{N} C_{j,p} \sum_{p' \neq p} \overline{C_{j,p'}} e^{2i\pi \frac{k(p-p')}{N}}, j \in J, k \in K.$$

As

$$\max_{j \in J} |C_{j,k}|^2 = P_m(k) \times PAPR_{f_k},$$

$$(19)$$

the final result would be,

$$PAPR_{N_s}(S(n)) \leq \frac{1}{\sum_{k=1}^{N} P_m(k)} [(\sum_{k=1}^{N} P_m(k) \times PAPR_{f_k})$$
$$+ \max_{k \in K}(\max_{j \in J}(\lambda(j,k)))].$$
(20)

In this equation, $PAPR_{N_s}$ is upper bounded by addition of the weighted average of all $PAPR_{f_k}$ metrics and a constant depending on the frequency component correlation. From now onwards $PAPR_{N_s}$ will be termed as $PAPR_{temp}$ to underline temporal PAPR.

7.2 Application of Frequency PAPR definition to SWR signal

A composite SWR signal of multi carrier modulation schemes of two OFDM based standards was simulated. Standard A was composed of 64 OFDM carriers while Standard B was composed of 256 OFDM carriers. Standard A transmits with 25 dB less power than Standard B. This SWR signal was mapped to the mentioned table format where each row of the table $'S'$ representing time domain SWR symbols. 10^3 SWR signals of length 4096 were simulated. The table was transformed to frequency domain using 1024 point FFT to fill $'C'$ of size 4×1024. Fig. 18 demonstrates the application of Eq. 20 to SWR signal.
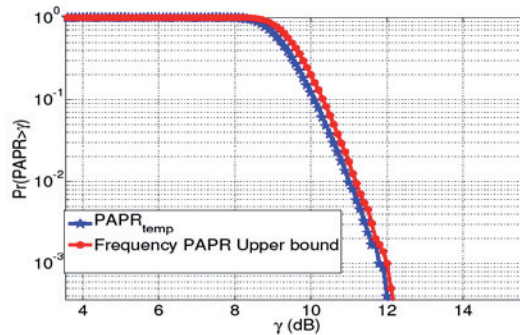

Fig. 18. Frequency PAPR calculation for SWR signal.

7.3 PAPR variations on dynamic spectrum access: mathematical developments

A study on the PAPR variation due to spectrum access is carried out and mathematically formulated. Let $PAPR_{new}$ be the PAPR after spectrum access. It can be upper bounded by the sum of initial PAPR of the primary users' signal before spectrum access $PAPR_p$ given by Eq. 20 and the influence of the secondary users' signal on PAPR, to be given after mathematical developments.

Consider the case of a CR signal, where different standards occupy their specific bandwidths while the free bandwidth is used by the secondary users. Transforming the temporal CR signal into the frequency domain following the carrier per carrier vision presented above. Let the signal be divided in to N_S symbols and N be frequencies of the whole spectrum. Out of these N frequencies, N_p are the primary user frequencies and N_s be the frequencies that are used by secondary users. Here the hypothesis is made that the spectrum occupied by secondary users is very small compared to primary users' spectral

occupancy i-e $N_s << N_p$ and also $P_m(p) + P_m(s) \approx P_m(p)$ due to the fact that the secondary signal mean power $P_m(s)$ is negligible compared to primary user mean power $P_m(p)$ due to spectral mask constraints.

A signal sample $s_{j,k}$ can be written as,

$$s_{j,k} = \sum_{c=0}^{N-1} C_{j,c} e^{i2\pi kc/N},$$

(21)

where $j \in \lceil 1, N_S \rceil$, $k \in \lceil 0, N-1 \rceil$. Also $c \in \lceil 0, N-1 \rceil$ is the number of carriers, i-e size of c is the sum of N_p and N_s. Now following the above equation

$$|s_{j,k}|^2 = |\sum_{c=0}^{N-1} C_{j,c} e^{i2\pi kc/N}|^2.$$

(22)

Dividing the carriers in to primary and secondary user carriers,

$$|s_{j,k}|^2 = |\sum_{p \in P} C_{j,p} e^{i2\pi kp/N} + \sum_{s \in S} C_{j,s} e^{i2\pi ks/N}|^2.$$

(23)

Here P and S are the index sets of primary and secondary user frequencies respectively. Now following the inequality which stats that for complex vectors a and b,

$$|a + b| \leq |a| + |b|.$$

and consequently

$$(|a + b|)^2 \leq (|a| + |b|)^2.$$

(24)

Thus Eq. 23 becomes,

$$|s_{j,k}|^2 \leq (|\sum_{p \in P} C_{j,p} e^{i2\pi kp/N}| + |\sum_{s \in S} C_{j,s} e^{i2\pi ks/N}|)^2,$$

(25)

and

$$|s_{j,k}|^2 \leq |\sum_{p \in P} C_{j,p} e^{i2\pi kp/N}|^2 + |\sum_{s \in S} C_{j,s} e^{i2\pi ks/N}|^2$$
$$+ 2|\sum_{p \in P} C_{j,p} e^{i2\pi kp/N}||\sum_{s \in S} C_{j,s} e^{i2\pi ks/N}|.$$

(26)

Let

$$\zeta_{j,k} = |\sum_{p \in P} C_{j,p} e^{i2\pi kp/N}|^2 + |\sum_{s \in S} C_{j,s} e^{i2\pi ks/N}|^2$$
$$+ 2|\sum_{p \in P} C_{j,p} e^{i2\pi kp/N}||\sum_{s \in S} C_{j,s} e^{i2\pi ks/N}|.$$

(27)

Now following Eq. 20,

$$PAPR_{N_s}(S(n)) \leq \frac{\max_{k \in K}(\max_{j \in J}(\zeta_{j,k}))}{\sum_{k=1}^{N} P_m(k)}.$$

(28)

And thus the above equation can be written as,

$$PAPR_{N_s}(S(n)) \leq PAPR_p + \delta.$$

(29)

Here $PAPR_p$ is the PAPR contribution because of primary user's carriers i-e,

$$PAPR_p \approx \frac{\max_{k \in K}(\max_{j \in J}(|\sum_{p \in P} C_{j,p}e^{i2\pi kp/N}|^2))}{\sum_{k=1}^{N} P_m(k)}.$$
(30)

And δ is the amount of variation in the initial PAPR, $PAPR_p$, because of dynamic spectrum access. The factor δ can be written as,

$$\delta = \delta_s + \delta_m.$$
(31)

where,

$$\delta_s \approx \frac{\max_{k \in K}(\max_{j \in J}(|\sum_{s \in S} C_{j,s}e^{i2\pi ks/N}|^2))}{\sum_{k=1}^{N} P_m(k)} = \frac{\sum_{s \in S} P_m(s)}{\sum_{p \in P} P_m(p)}.PAPR_s.$$
(32)

In Eq. 32, the factor $\frac{\sum_{s \in S} P_m(s)}{\sum_{p \in P} P_m(p)}$ is the ratio of secondary user's mean power to the primary user's mean power which is quite small and that's why the factor δ_s does not contribute much to the PAPR variations. Also $PAPR_s$ is the secondary user PAPR in frequency domain, i-e,

$$PAPR_s = \frac{\max_{k \in K}(\max_{j \in J}(|\sum_{s \in S} C_{j,s}e^{i2\pi ks/N}|^2))}{\sum_{s \in S} P_m(s)}.$$
(33)

And

$$\delta_m \approx \frac{1}{\sum_{k=1}^{N} P_m(k)}[\max_{k \in K}(\max_{j \in J}(2|\sum_{p \in P} C_{j,p}e^{i2\pi kp/N}|.|\sum_{s \in S} C_{j,s}e^{i2\pi ks/N}|))].$$
(34)

Here δ_m is the contribution because of the mutual correlation of the primary and secondary user carriers. It should be noted that the influence on PAPR due to spectrum access can be calculated with the help of spectral information of the primary and secondary signal. This vision facilitates the spectrum access phenomenon in the context of CR by directly relating the effect on PAPR with the spectral information of signal.

This formulation is applied on aforementioned bi-standard SWR signal. QPSK modulated [10] carriers are added on free bands in between these two standards and their effect on PAPR is demonstrated in Fig. 19.

7.4 Frequency PAPR view benefit

Let us now explain the advantages of using frequency view for PAPR computation over a temporal PAPR calculation approach in a CR spectrum access scenario. As mentioned previously that spectrum access influences the transmitted signal properties and subsequently its PAPR. PAPR of the signal to be amplified i-e signal between f_{min} and f_{max} in Fig. 20 is increased after spectrum access where W is the available free bandwidth and B is the allocated bandwidth for secondary user during spectrum access. This increase in PAPR can cause signal distortion. Therefore, in order to see the effects on PAPR due to the allocated bandwidth, a classical temporal vision would need constant exchanges between frequency domain (for bandwidth allocation) and time domain (for PAPR computation) in order to select a bandwidth which does not increase a lot the PAPR of the signal to be amplified after spectrum access. This implies a tremendous increase in computational

complexity. In the PAPR frequency view, the effects on PAPR due to any available band allocation can be seen in the frequency domain without going back to time domain for PAPR calculations. Moreover, this spectrum knowledge can be used by other CR sensors like (Hachemani et al., 2007) uses power spectral density of the signal for blind standard recognition. Fig. 20 explains that this approach facilitates bandwidth allocation with respect to PAPR metric in CR context. A two standard SWR signal is depicted with an available bandwidth W. After slicing of the temporal signal and FFT operation, a table is filled with signal's spectral components. Then a suitable bandwidth B is allocated by the CR system while $PAPR_{temp}$ computations are performed staying in the frequency domain during bandwidth selection procedure.

This frequency vision also helps in PAPR mitigation e-g PAPR reduction by 'Tone Reservation' (Tellado-Mourelo, 1999) is about reserving tones in frequency domain to reduce PAPR for OFDM signals. This process is same as bandwidth allocation. Knowing the fact that OFDM and SWR signals show similar characteristics (Hussain & Louët, 2008) and carrier per carrier vision facilitates bandwidth allocation, OFDM PAPR mitigation methods can easily be applied to SWR signals.
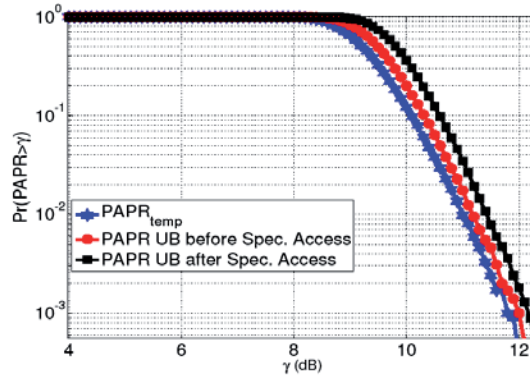

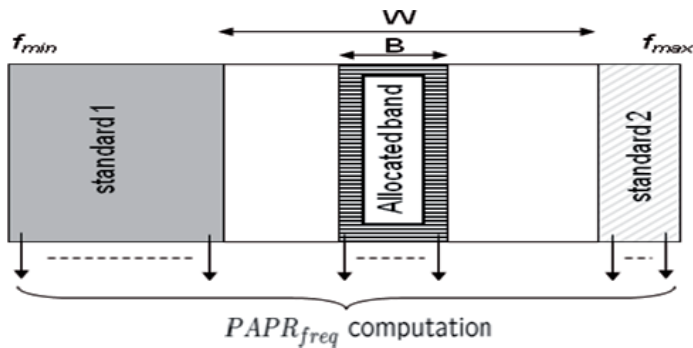Fig. 19. Upper bound is modified after spectrum access

Fig. 20. Illustration of PAPR computation with a two standards SWR signal and an allocated bandwidth B.

8. Conclusion

This chapter explores the non linearity issues in SWR based CR systems in the spectrum access scenario by exploiting the frequency vision of PAPR. The importance of this issue can be estimated by a simple fact that reduced PAPR would result in higher PA efficiency and in turns reduced energy consumption which is closely related to 'Green communication' concept (ICC, 2009). Also the problem is addressed at the receiver end on Low Noise Amplifier (LNA) (Marshall, 2009). High PAPR being a major problem in multi carrier communications and CR signal being a multi carrier signal suffers the same problem. Spectrum access in the CR context is about using the free spectrum under certain constraints like QoS maintenance of the primary users. Another constraint for spectrum access is discussed in this chapter which is PAPR as high PAPR affects transmitter specifications. A carrier per carrier vision approach is presented which associates PAPR with spectrum. This spectral knowledge based PAPR definition helps in spectrum access by providing PAPR information about available bandwidths. An upper bound is given for PAPR after spectrum access which is based on the spectral knowledge of primary and secondary users. The process of spectrum access under PAPR metric constraint is facilitated with this frequency PAPR vision.

9. References

Bremenson, C. et al. (1980), Linearizing TWT Amplifiers in Satellite Transponders, Systems Aspects and Practical Implementation, *Proc. AIAA 8th Communications Satellite System Conference*, April 1980, pp. 80-89.

Fasbender, A.; Reichert, F.; Geulen, E.; Hjelm, J. & Wierlemann, T. (1999). Any Network, Any Terminal, Anywhere, *IEEE Personal Communication*, Vol. 6, No. 2, April 1999, pp. 22-30.

FCC. (2003). http://www.fcc.gov/oet/cognitiveradio/

Hachemani, R., Palicot, J. & Moy, C. (2007). A New Standard Recognition Sensor for Cognitive Radio Terminals, *EUSIPCO*, September 2007.

Han, S.H., Lee, J.H. (2005) An overview of Peak to Average Power Ratio Reduction techniques for Multicarrier Transmission, *IEEE Wireless Communications*, April 2005, pp. 56–65.

Hussain, S., Palicot, J., Louët, Y., Zabré, S. (2008). *Frequency Domain Interpretation of Power Ratio Metric for Cognitive Radio Systems*, Proceedings of IET Communications Journal, Vol. 2, Issue 6, July 2008, pp.783–793.

Hussain, S. & Louët, Y. (2008). Peak to Average Power Ratio Analysis of Multi-carrier and Multi-standard signals in software radio context, *IEEE ICTTA 08*, Damas, April 08.

ICC. (2009). http://www.ieee-icc.org/workshops.html#GreenComm

Kenington, P. B. (2000). *High-linearity RF amplifier design*, Artech House Inc. Boston-London. ISBN: 1580531431.

Lee, H., Liu, D.N., Zhu, W., Fitz, M.P. (2005). Peak power reduction using a unitary rotation in multiple transmit antennas, *IEEE International Conference on Communication,* Vol. 4, May 2005, pp. 2407-2411.

Louët, Y. & Palicot, J. (2008). A classification of methods for efficient power amplification of signals, *Annals of Telecommunications*, Vol. 63, Issue 7-8, July/August 2008, pp. 351-368.

Macleod, J., Beach, M., and al. (2000). IST-TRUST, 1999-12070, D3.1.1, *European Project*, Septembre 2000.

Marshall, P. F. (2009). Cognitive Radio as a mechanism to manage front end linearity and dynamic range, *IEEE Communications Magazine*, March 2009, pp. 81-87.

Mitola, J. & Maguire, G. Q. Jr. (1999). Cognitive Radio: Making Software Radios More Personal, *IEEE PCS Magazine*, Vol. 6, No. 4, August 1999, pp. 13-18.

Ochiai, H. & Imai, H. (2001). On the distribution of the Peak to Average Power Ratio in OFDM Signals, *IEEE transactions on communications*, Vol. 49, No. 2, February 2001, pp. 282-289.

Ragusa, S., Palicot, J., Roland, C., Lereau, C. (2005). Adjacent channel power ratio analysis for an OFDM signal, *IEEE ISSPIT 2005*, pp. 495-500.

Roland, C. & Palicot, J. (2002). A Self Adaptive Universal Receiver, *Annals of Telecom*, 57, No. 5-6, May- June 2002.

Soury, A., Ngoya, E. (2005). Modeling long term memory effects in microwave power amplifiers for system level simulations, *Annals of Telecom*, Vol. 60, No. 11-12, 2005, pp. 1488-1506.

Steck, J. & Pham, D. (1987). A New TWT Linearizer for Satellite-Communications Transmitters, *Microwave System News and Communications Technology*, Vol. 17, No. 9, June 1987, pp. 28-42.

Tan, M., Latinovic, Z., and Bar-Ness, Y. (2005). STBC MIMO-OFDM Peak-to-Average Power Ratio Reduction by Cross-Antenna Rotation and Inversion, *IEEE Communication Letters*, Vol. 9, No. 7, July 2005.

Tellado-Mourelo, J. (1999). *Peak to Average Power Ratio Reduction for multicarrier modulation*, Ph.D thesis, Stanford University, September 1999.

Zhao, Q., Swami, A. (2007). A survey of Dynamic spectrum access Signal Processing and networking perspectives, *ICASSP*, Vol. 4 April 2007, pp 1349-1352.