Mohammad A. Matin   *Editor*

Spectrum Access and Management for Cognitive Radio Networks

Springer

Signals and Communication Technology

Mohammad A. Matin
Editor

Spectrum Access and Management for Cognitive Radio Networks

Springer

Editor
Mohammad A. Matin
Electrical and Electronic Engineering
Universiti Teknologi Brunei
Bandar Seri Begawan
Brunei Darussalam

Preface

The radio spectrum is one of the most precious resources which must be managed to ensure efficient access for the wireless communication services which use it. The allocation and management of spectrum are administered by the regulatory authorities. Traditionally, spectrum allocation is carried out exclusively of its use in large geographic areas and assigning frequency bands to specific users or service providers is proved to be inefficient. Recently, substantial knowledge about dynamic spectrum access scheme has been accumulated to enable efficient spectrum sharing. The cognitive radio based on dynamic spectrum access has emerged as a new technology to enhance the efficiency of spectrum usage. The key characteristics of cognitive radio system is to sense their environment and dynamically adapt their transmission parameters (e.g., modulation waveform, channel access method, spectrum use, and networking protocol) as needed for reliable communication and efficient utilization of radio spectrum. The main objective of this book is to treat some of the core functionalities related to cognitive radios and provide an in-depth and wide range of knowledge in spectrum access and management for enabling the technology.

Chapter "Introduction" is the introductory chapter which is devoted to kick-start with the topic. It provides the preliminaries for understanding the rest of the chapters on spectrum access and management.

Chapter "Spectrum Sensing for Half and Full-Duplex Cognitive Radio" explores the spectrum sensing approaches in cognitive radio networks for half duplex (HD) and full duplex (FD) mode of transmission. FD techniques can achieve the simultaneous spectrum sensing and data transmission in cognitive radio networks. Two spectrum sensing algorithms are developed based on the Cumulative Power Spectral Density (CPSD) of the received signal, dealing with the HD mode. These algorithms outperform the traditional Energy Detector (ED), the well-known Cyclostationary Detector (CSD) based on the Generalized Likelihood Ratio Test (GLRT) and the Autocorrelation Detector (ACD). Furthermore, these algorithms are robust against the noise variance, so that the dependence on Noise Uncertainty (NU) presented in ED is avoided. In addition, the proposed algorithms are blind as

they do not require any prior information on the PU's signal, contrary to Cyclo-stationary or Waveform detectors. The material in this chapter covers the main trends on the state of the art.

Chapter "Dynamic Spectrum Access Algorithms Without Common Control Channels" presents the most recent Dynamic Spectrum Access (DSA) algorithms, the metrics of the algorithms, and the mathematical background. These algorithms are classified into Homogeneous Role Algorithms, Heterogeneous Role Algorithms, Symmetric Available Channel Model, and Asymmetric Available Channel Model. The algorithms have bounded Maximum Time-to-Rendezvous (MTTR) to avoid long-time delay in CRNs. State of the art have nearly optimal MTTR. This kind of dynamic spectrum access algorithms is one of the key MAC layer functions in the cognitive radio networks (CRNs).

Chapter "Quantified Dynamic Spectrum Access Paradigm" introduces a quantified approach of QDSA which enables easier understanding and interpretation of the outcomes to spectrum sharing. The quantified approach transforms spectrum into a commodity that can be exchanged with service providers, and a resource that can be precisely controlled for making an efficient use. It facilitates easier interpretation and requires less skill of its users. Furthermore, quantified dynamic spectrum sharing enables spatial overlap of multiple RF systems while protecting their spectrum rights.

Chapter "A Case Study of Cognitive Radio Networks: Secure Spectrum Management for Positive Train Control Operations" presents a multi-tiered cognitive radio network architecture that is being developed to address the spectrum allocation for Positive Train Control (PTC) operations. In addition to efficient spectrum sharing, this cognitive radio will also detect spectrum abuse, any potential cyber threats to PTC operations and improve security of radio communication between the PTC nodes.

Chapter "Centralized and Distributed Algorithms for Stable Communication Topologies in Cognitive Radio Ad hoc Networks" presents two algorithms for stable communication topologies in CRAHN—the first one is generic benchmarking algorithm that determines a sequence of stable communication topologies spanning all of the SU nodes such that the number of transitions from one instance of the topology to another is the global minimum. The second one is distributed local spectrum knowledge-based stable path routing protocol called the Maximum Common Primary User channel-based Routing (MCPUR) protocol that prefers to choose an SU-SU source-destination (s-d) path with the largest value for the sum of the number of common PU channels available for use across each of its constituent SU-SU edges.

Chapter "Cognitive Radio Testbeds: State of the Art and an Implementation" summarizes various software defined radio (SDR) implementations of CR systems in the literature. Most popular SDR tools are GNU Radio, LabVIEW and USRP hardware. The authors present a prototype implementation of an energy-based wideband spectrum sensing testbed by using SDR tool, LabVIEW, and NI USRP-2921 hardware in real time.

Chapter "Visible Light Based Throughput Downlink Connectivity for the Cognitive Radio Networks" presents a visible light (VL)-based throughput downlink connectivity scheme for the CR network, by exploiting both CR and VL technologies in IoT communication scenario.

Chapter "Reconfigurable MIMO Antennas for Cognitive Radios" will shine some light on the most recent CR front ends focusing on MIMO-based integrated architecture. The front end of a CR consists of a sensing antenna for spectrum scanning and a reconfigurable antenna or reconfigurable multiple-input multiple-output (MIMO) antenna system for communication over several frequency bands.

Chapter "Cognitive Radio and the New Spectrum Paradigm for 5G" focuses on the introduction and access management of new spectral resources to support the substantial traffic increase in 5G. An extensive review on the current state of cognitive radio techniques as well as possible applications of cognitive radio in 5G wireless communications is presented in this chapter.

This book attempts to present current and emerging trends in research and development of spectrum access techniques and management of cognitive radio networks featuring a structured approach. Features include the following:

- Reviews the current state of the art as well as offering an in-depth treatment of spectrum sensing and management of cognitive radio networks.
- Provides up-to-date materials for cognitive radio networks and practical design information and extensive discussion through case studies and descriptions of cognitive radio platforms and testbeds.
- The chapters are written by several experts; highlight current design and engineering practices, emphasizing challenging issues of cognitive radio networks together with implementation details.
- Supplements its content with extensive references to enable researchers for further investigation of spectrum and access and management of cognitive radio networks.

It is hoped that this book will serve as a reference book for graduate level course in believing that it would be a useful resource for readers who want to understand this emerging technology, as well as those who wish to enhance their knowledge conducting research and development in this field.

Bandar Seri Begawan, Brunei Darussalam   Mohammad A. Matin

Contents

Introduction

Mohammad A. Matin

Introduction

The ever increasing demand for new wireless services and applications for ubiquitous computing in modern lives and businesses creates a shortage of available wireless spectrum. In the traditional approach, the spectrum allocation is done via assigning the spectrum entirely to a licensed user and the system has to operate within that frequency band. This leads to inefficient utilization of spectrum because most of the time, the large portion of the bands remain underutilized. In 2002, FCC has reported that there are huge temporal and spatial variations in the allocated spectrum usage ranging from 15 to 85 %. The underutilization of spectrum has stimulated engineers and regulation communities in hunting for better spectrum management policies. The cognitive radio concept provides an innovative technique for such spectrum management. It offers new techniques of exploiting the available spectrum. The main aspect of cognitive radio is to take advantage of unused spectrum to provide a new route to spectrum access. However, the two main issues hinder the realization of cognitive radio. The first and most fundamental issue is primary users (PUs') concern about potential interference from secondary users (SUs') communications. If licensed users allow unlicensed users (or SUs) to access their licensed spectrum bands in an opportunistic manner, there will be a risk of interference due to SUs' communications. Such interference can degrade the QoS of PUs', which can lead to the loss of PUs' business. Therefore, SUs must provide efficient mechanisms to protect PUs' communications in detecting spectrum hole through spectrum sensing.

The second issue is the lack of economic motivations to PUs for spectrum sharing. PUs will not perhaps be convinced to open up their spectrum resources to

M.A. Matin (✉)
Electrical and Electronic Engineering, Universiti Teknologi Brunei,
Bandar Seri Begawan, Brunei Darussalam
e-mail: matin.mnt@gmail.com

unlicensed users even if SUs can almost avoid interference with PUs' transmission since there is no comprehensible incentive for them to share their spectrum. Therefore, it is needed to provide efficient economic mechanisms, such as spectrum pricing strategies, that can provide incentives for facilitating SUs transmission. Without resolving the above two main issues, cognitive radio might not be accepted extensively. Therefore, to make it acceptable, there is a lot of progress on cognitive radio network happen in the last decade. This chapter provides an overview of spectrum allocation policies, CR technology, and CR network architecture. Then describe spectrum sensing, allocation and sharing techniques and spectrum mobility concepts. This chapter is ending up with the applications of CR systems.

Spectrum Allocation Policies

Spectrum allocation policies can be categorized into two-static spectrum allocation and dynamic spectrum allocation. These two polices have been discussed in the following subsections.

Static Spectrum Allocation

Traditional wireless networks feature static spectrum allocation policies, according to which licensees are approved to the primary users exclusively for the use of spectrum bands on a long-term basis over huge geographical regions. This policy has generated inefficient utilization of spectrum because it is very rare that all licenses make full use of the frequencies assigned to them, most of the time the large portion of the bands remain underutilized. Moreover, for the greater success of wireless applications, the unlicensed band usage has demanded, thus causing possible wireless spectrum shortage. This spectrum crisis has motivated and pushed forward to the concept of dynamic spectrum-allocation policies (shown in Fig. 1).
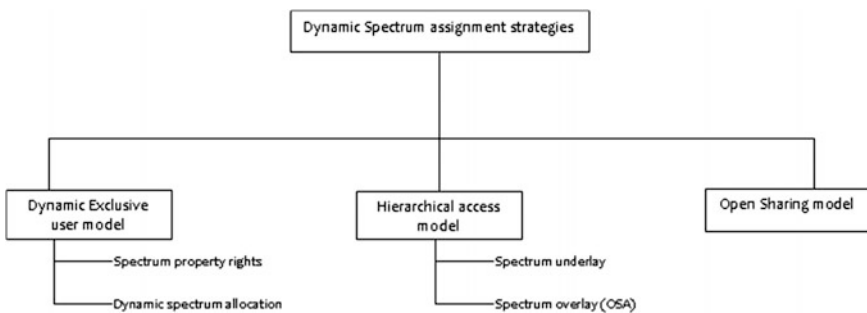

Fig. 1 Dynamic spectrum allocation policies

Dynamic Spectrum Allocation

Under dynamic spectrum allocation policy, three different models have been presented in this section.

A. Dynamic Exclusive Use Model

In this model, spectrum bands are licensed for exclusive use to PU services, however, flexible to improve spectrum utilization. Two approaches have been stated under this model for efficient utilization, the first one is spectrum property rights [1, 2] and the second one is dynamic spectrum allocation [3]. The former approach allows licensees to sell and deal spectrum. Economy and the market will, therefore, drive the most cost-effective use of this limited resource. Remember that although licensees have the authority to rent or share the spectrum for mutual benefit, such sharing is not directive by the regulatory authorities. The latter approach, dynamic spectrum allocation, was got forwarded by the European DRiVE project [3]. It seeks to improve efficient spectrum utilization through dynamic spectrum allocation by exploiting the spatial and temporal traffic statistics of diverse applications.

B. Open Sharing Model

This is also termed to as spectrum commons [4, 5], which opens sharing among users as the basis for dealing spectral region. This model is operated in the unlicensed ISM band (e.g., WiFi). Centralized [6, 7] and distributed [8–10] spectrum sharing strategies have been investigated to deal the technological demands under this spectrum management model.

C. Hierarchical Access Model

The basic idea of this model is to open licensed spectrum to secondary users without or limiting interference with the licensed user. Two methods for spectrum sharing between PU and SU have been considered which are: spectrum underlay and spectrum overlay.

The underlay method entails severe constraints on SUs transmission power and SU operates below the noise floor of PUs. The transmitted signals of SUs spread out over a wide frequency band, which can realize high data rate at low transmission power in a shorter distance. In the worst-case scenario where the PUs transmit all the time, does not exploit spectrum white space.

The other method is spectrum overlay which was first visualized by Mitola [11] under the term "spectrum pooling". It is then explored by the DARPA XG program [12] under the term "opportunistic spectrum access (OSA)". This approach differs from spectrum underlay which does not provide severe limitations on the transmitted power of SUs, but rather focus on when and where they can transmit. It exploits spatial and temporal spectrum by allowing SUs to recognize and utilize local and instantaneous spectrum availability in a non-intrusive manner.

The hierarchical model is possibly well-suited with the current spectrum management policy and is inheritance wireless systems in compare to the dynamic exclusive use and open sharing models. Furthermore, the underlay and overlay methods can be utilized simultaneously for further improvement of spectrum utilization.

Cognitive Radio Technology

The main ideas of CR technologies are the cognitive radio techniques that are capable of sharing the spectrum opportunistically. The CR mostly consists of radio, sensor, knowledge database, learning engine, optimization tools, and a reasoning engine. CR systems involve PU and SU of the spectrum; primary users are license owners while secondary users hunt for the spectrum to use opportunistically through CR. A CR-enabled device adapts its operational parameters as a function of its environment [13–16].The two main features of CR technology are cognitive capability and reconfigurability. The cognitive capability allows CR to sense its radio environment and collect information (e.g., different signals and their modulation types, noise, transmission power, etc.) through real-time interaction and, select the best available spectrum and exploit spectrum holes without or minimum interference with the licensed user. The reconfigurability features of CR allow it to optimally adapt the best spectrum band and the operational parameters as a function of the sensed information. Based on the reconfigurability features, CR networks can access both unlicensed and licensed spectrum. This can be done by two main processes: licensed band operation and unlicensed band operation.

a. Licensed band operation: Since, PUs are the owner to use the licensed bands primarily, the CR users need to focus on the detection of the reappearance of PUs. Ideally, CR communication takes place in the event of idle primary users. Once the reappearance of PU is detected, CR must evacuate this spectrum band and leave straight away. This process is called channel mobility.
b. Unlicensed band operation: All CRs have the same right with the PU to access unlicensed band. However, effective spectrum sharing algorithms are important for CR networks to develop.

Cognitive Network Architecture

Network Architecture

The cognitive networks appear as feasible architectural solutions to the current wireless network problems resulting from scarcity of spectrum availability and inefficient use of spectrum. The network can be in different forms for example-cognitive wireless sensor network, cognitive cellular network, cognitive wireless
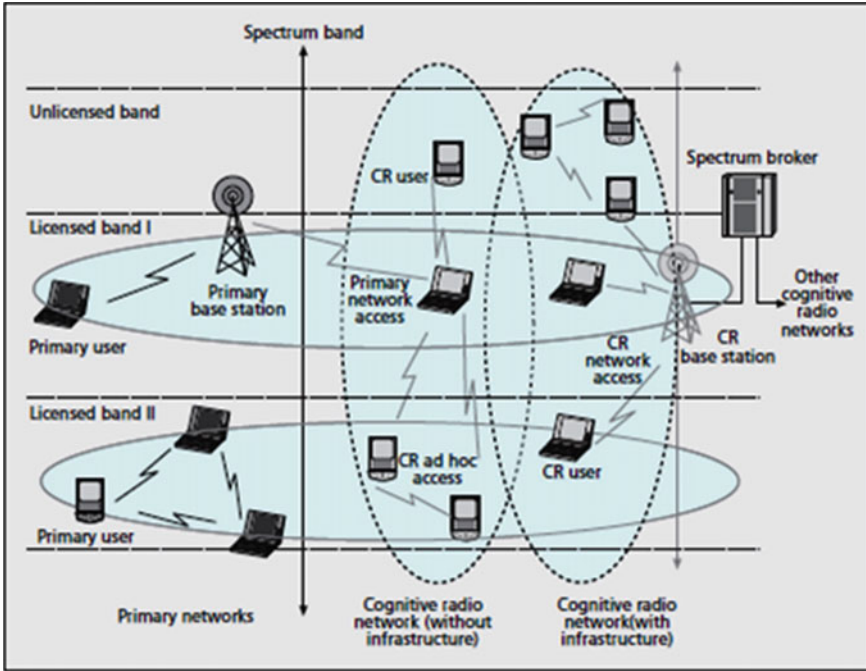
Fig. 2 Cognitive radio network architecture [17]

mesh network, cognitive device-to-device network, cognitive wireless local area network, cognitive wireless body network. Two approaches can be implemented for the spectrum access of cognitive users according to the used radio transmission and/or the network requirement which are spectrum overlay and spectrum underlay.

In cognitive radio networks (CRNs), cognitive nodes are equipped with cognitive transceivers and the networks can be either an infrastructure-based or an infrastructure-less communication architecture shown in Fig. 2 [17]. In the infrastructure-based architecture, a CR base station coordinates data communication among CR users. The CR network does not have the right to operate in the license band. Hence, additional functionalities are required for CR users to share the license spectrum band. Again, the network can be either centralized or distributed dynamic spectrum access architecture [18]. CR network can include spectrum brokers in distributing the spectrum resources among different CR networks [19].

Network Heterogeneity

Heterogeneous radio access technologies (RATs) provide reliable spectrum sensing and access services for efficient utilization of spectrum in cognitive radio

(CR) technology. This offers the opportunity of switching between different frequency channels, which can significantly increase the mobility, flexibility and energy efficiency of the system. The throughput of the user can be considerably increased by executing a "spectrum handoff" operation, by which the user switches to a "sensed-as-idle" channel and keeps on data transmission if the current channel is detected as busy in a heterogeneous network (HetNet). However, the channel switching is not always expected due to the extra energy consumption, delay, and transmission collision that could happen. In [20], the authors consider joint optimization of the trade-off between the user's sensing period and throughput, as well as the tradeoff between switching and staying in a given channel during the spectrum access in a multi-channel scenario.

Spectrum Management in CRNs

The tree diagram of spectrum management process in cognitive radio networks is shown in Fig. 3. The unique challenges of CRNs are interference avoidance, QoS awareness, and seamless communication. To address these challenges, different functionalities such as spectrum sensing, spectrum allocation and sharing, and finally spectrum mobility are necessitated for spectrum management in CR networks.
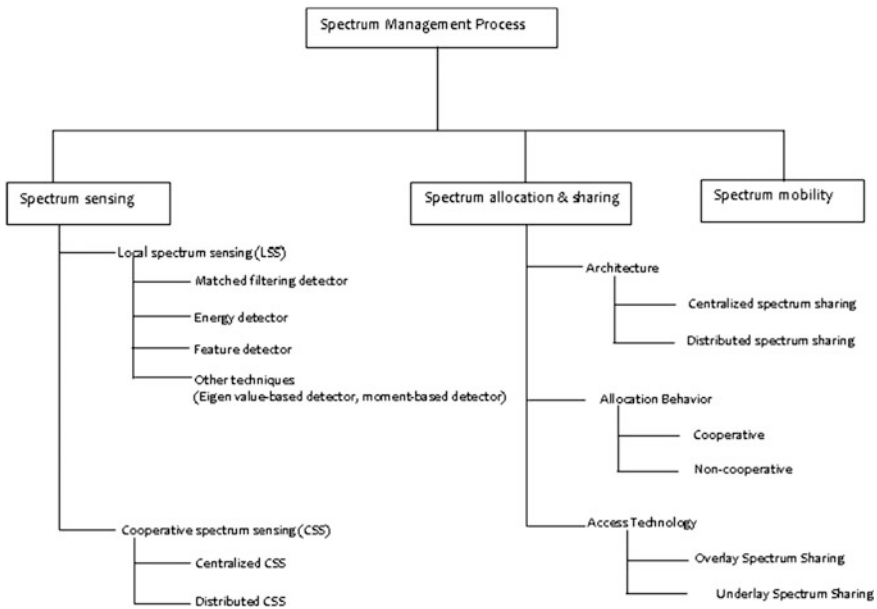

Fig. 3 Tree diagram for spectrum management

Spectrum Sensing

Spectrum sensing is one of the main functions in detecting the presence of PU for the establishment of cognitive radio. It provides the ability to the radio to sense and learns the parameters related to channel characteristics to detect temporarily spectrum holes or white spaces without causing harmful interference to the licensed or primary user activities [21]. Once available, spectrum sharing and allocation functionalities are in place to use the spectrum bands. The performance of spectrum sensing is important since all their functionalities will be based on the information of spectrum availability but in practical application scenarios, it is often concession by many influential factors such as multipath fading, shadowing, and the receiver uncertainty problem [22]. Collaborative spectrum sensing (CSS) [23–25] is a successful approach to tackling these challenges and get better sensing performance.

Spectrum sensing in cognitive radio is a more important term that involves getting the spectrum usage characteristics across different dimensions such as time, space, frequency, and code. To make better utilization of available spectrum, cognitive radio enables spectrum access which aspires to effective spectrum sharing between licensed and unlicensed users. It seeks to improve spectrum efficiency through dynamic spectrum allocation via exploiting the spatial and temporal spectrum holes (SHs). The temporal spectrum hole appears during a certain time when there is no PU transmission whereas spatial Spectrum hole appears if the x transmission of PU is inside the area. In the latter case, the SU has to use the spectrum outside that area.

Challenges in Spectrum Sensing

In this section, we will discuss some challenges associated with the spectrum sensing for cognitive radio.

A. Hardware Requirements

High sampling rate, high-resolution analog to digital converters (ADCs), and high-speed signal processors are required for spectrum sensing in cognitive radio applications. For optimal receiver designs as well as for improved handoff, power control, and channel allocation techniques, noise variance estimation techniques can be used [26].

B. Hidden PU Problem

The hidden PU problem is alike to the hidden node problem in carrier sense multiple access (CSMA). This is due to several factors including severe multipath fading or shadowing effects while scanning for PU transmissions.

C. Sensing Time Duration

Primary users can claim their licensed bands during the transmission of CR users. Cognitive radio users should be able to perceive the presence of PUs as fast as possible to avoid interference to and from license holders, and to vacate the band without delay. Therefore, sensing should be done within certain duration to identify the presence of primary users. This requirement poses a performance limit of sensing algorithm and creates a challenge for cognitive radio design.

D. Decision in Cooperative Sensing

In the case of cooperative sensing, the cognitive radios share information themselves and results from various measurements are combined. This is a challenging task. The shared information can be soft or hard decisions made by each cognitive device [27].

E. Security

In cognitive radio, a selfish or malicious user can adjust its air interface to mimic a primary user signal's characteristics in a hostile environment. Thereby, it can delude the spectrum sensing performed by genuine users. Such a behavior or attack is investigated in [28] and it is termed as primary user emulation (PUE) attack. Most of the research presumes that the cognitive users are assumed to be honest, which is not true forever. In fact, cognitive radio networks are vulnerable to various security attacks like all other wireless networks. Malicious cognitive users may report false spectrum sensing information purposely for disrupting the final decision.

Security has received growing interest in cognitive radio technology [29–31]. The conventional security mechanisms such as cryptography, intrusion detection system, and authentication are not enough to prevent spectrum sensing against attacks and intrusions [29] as attacks are smarter, easily bypassing common security mechanisms [29, 32]. In light of these constraints, new designs in spectrum sensing need to be provided for resilience and security and it should be more flexible to cope with attacks mutations [33].

Spectrum Allocation and Sharing

To ensure the advantage brought by CR technology, efficient spectrum sensing along with spectrum sharing and allocation are necessary. In the earlier section, the spectrum sensing methods for CR technology has been discussed. The sensing results provide information to the SUs about the status of channels even though the channel status may alter quickly based on the behavior of the PUs. In this section, the spectrum allocation and sharing schemes have been discussed to address these issues. So as to gain better performance, SUs should choose two things at a time-the first one is which channel and the second one is when and how to get admission to the channel. The PU system should be protected from the harmful interference

caused by the SUs. In addition, the SU should take into account the activities of other co-existing SUs in the case of cooperative spectrum sharing. But in the case of non-cooperative spectrum sharing, the selfish CR user does not consider the interference in other CR nodes which could result in reduced spectrum utilization.

Based on the architecture, the spectrum sharing and allocation can be classified into two: centralized and distributed. In centralized spectrum sharing, the central entity controls the spectrum allocation and access procedures. A distributed sensing procedure can be used such that measurements of the spectrum allocation are forwarded to the central entity, and a spectrum allocation map is constructed. Furthermore, the central entity can lease spectrum for a specific amount of time in a limited geographical region. On the other hand, distributed spectrum allocation and access are based on local (or possibly global) policies that are carried by each node distributively [34].

Depending on the uses of spectrum bands by SUs, the spectrum sharing techniques can be partitioned into two- open spectrum sharing and licensed spectrum sharing [35, 36]. During the open spectrum sharing system, all the users in the system have the equal opportunities to facilitate spectrum sharing. The SUs share the unlicensed spectrum among themselves under the open spectrum sharing system. The licensed spectrum sharing may also be termed as hierarchical spectrum access model. In such systems, the PUs have superior priority as license holders than the unlicensed SUs. Generally, PUs do not conflict among themselves in a CR network as all of them have their own licensed bands. The SUs necessitate in adjusting their parameters, such as transmitting power and transmission strategy, to evade the disturbance of the PUs activities. The hierarchical spectrum access model can be further partitioned into two: spectrum underlay and spectrum overlay [36] conforming to the access strategies of the SUs. The spectrum underlay technique is a spectrum management principle where the SUs can coexist with PUs and are permitted to send signals at the same time as being used by the PUs. But the interference caused by the SUs at PUs remains acceptable. Exceeding the predefined tolerable level could degrade dramatically the primary signal. On the other hand, in the spectrum overlay systems, the secondary user uses a channel from a primary user and can only do transmission when it is not being occupied. Spectrum overlay techniques are based on detection and interference avoid mechanism. Spectrum overlay is also termed as opportunistic spectrum access (OSA). Spectrum sharing techniques can also be classified into two: spectrum sharing inside a CR network (intranetwork spectrum sharing) and among multiple coexisting CR networks (internetwork spectrum sharing).

Based on the spectrum availability, the CR network is able to choose the best available band according to the QoS requirements for the applications of each SU. This spectrum decision [37, 38] is very much related to the channel behaviors and operations of PUs. Spectrum decision considers long-term channel characteristics in compare to traditional spectrum sharing. Usually, the spectrum decision follows two main steps: channel characterizing and spectrum assigning [39]. To quantify or characterizing the channel, the following parameters such as path loss, link errors, and link delay can be estimated based on both the current situation of SUs and

statistical information of PUs. The best spectrum in its vicinity is then allocated to each SU to meet QoS requirements. In [38], the author presents a spectrum decision framework for CR networks.

Spectrum Mobility

The PUs can claim their own channels while SUs continue to transmit and if this happens, SUs need to cease transmission, and find other alternative channels to keep on transmission, which is called spectrum handoff [40]. Each time, a CR user changes its spectrum; the network protocols may require modifications to the operation parameters. The purpose of the spectrum mobility management in CR networks is to ensure smooth and quick spectrum handoff leading to minimum performance degradation. An important requirement of mobility management protocols is information about the duration of a spectrum handoff. When a SU changes its frequency of operation, its transmission hangs up, this always leads to latency increase. Therefore, a good spectrum handoff mechanism is required to ensure imperceptible and quick transition with minimum performance degradation, [41–45]. To improve the performance or reduce latency, a certain number of spectrum bands can be reserved for spectrum handoff [41]. SUs straight away use the exclusively reserved spectrum bands for ongoing transmission. However, we should be careful in selecting the number of the reserved bands to balance the spectrum efficiency and handoff performance. The other approach is proactive spectrum handoff [42], where SUs discharge channels before PUs attempt to use them to avoid conflict between Pus and SUs. This is clearly too idealistic situation as it is difficult for SU to predict the behavior of PU in most of the case.

CR Applications

The spectrum sensing and management techniques for spectrum sharing enable CR applications in diverse areas. Here, some of the areas are mentioned.

A. TV White Spaces

The main regulators FCC (US), OFCOM (UK), ECC (Europe) have committed to the unlicensed use of TV white spaces. FCC released the final regulation for using the TV white space in September 2010 [46] after putting the long effort that led to the conclusion of this field. In the meantime, other agencies have also been receiving progress [47]. Both the FCC and Ofcom have considered three methods to ensure that SU devices do not make harmful interference: beacons, geolocation combined to access to a database, and sensing. The accessible database (centralize-fashion) of free TV bands for SUs communication seems to be the best solution for interference avoidance.

B. Cellular Networks

The CR applications in cellular networks are promising in recent time due to facing high demand for spectrum. To mitigate the indoor coverage problem and acclimatize the traffic growth, small cells concept, such as femtocells, has been proposed in 3GPP LTE-Advanced (LTE-A) [48] and IEEE 802.16m [49], and companies like PicoChip driving femtocell revolution.

C. Military Applications

CR technique in military applications is emerging. The enemy's communication can be recognized using CR techniques and protect their own communication. The US department of defense (DoD) has already established programs such as SPEAKeasy radio system (US Military project) and next Generation (XG) to exploit the benefits of CR techniques [40].

D. Emergency Networks

Under emergency conditions, e.g., natural disasters such as hurricanes, earthquakes, wild fires and accidents, the infrastructure of the current wireless communication can be collapsed. Therefore, there is an urgent need for a means of communications to help the recue team to facilitate and locate the disaster survivors. The CRN techniques can be used for such efficient and reliable emergency network transmission.

E. Smart Grid Network

Smart grid network consists of home/building area networks, the advanced metering infrastructure (AMI) or field area networks (FAN) and the wide area networks. Like other unlicensed devices, CR enabled advanced metering infrastructure or field area network devices are not immune from interference or congestion. CR enabled AMI/FAN should go beyond just dynamic spectrum access and develop self-coexistence mechanisms to coordinate spectrum usage and can even prioritize spectrum use according to the class of smart grid traffic, e.g., real time versus non-real time, emergency report versus demand response.

F. Public Safety Network

The cognitive radio is capable of adapting an available communications channel which is important for public safety radio devices to carry the operational requirements of ubiquitous communications in case of critical situations.

G. Wireless Medical Networks

Cognitive radio technology can be implemented in wireless medical applications. The transmission parameters of SUs can be adjusted based on electromagnetic interference constraints to overcome the interference with PUs. CR technology can also improve the QoS of wireless communication between medical devices by defining the priority levels of each device [50].

Conclusion

In this chapter, cognitive radio technology has been discussed. It provides efficient spectrum usage to deal with the increasing demand for wireless spectrum. Since, the licensed users have the legal rights to use the spectrum bandwidth, while SU is an opportunistic user that can transmit on that bandwidth whenever it is vacant through spectrum sensing technique. The spectrum sensing technique determines which portions of the spectrum are available to access without interference with the signal of PU. Different spectrum allocation and sharing schemes and spectrum handoff functionalities are being investigated for spectrum management. Moreover, the potential applications of CR networks are also explored in this chapter.

References

1. Coase, R.: The federal communications commission. J. Law Econ., 1–40 (1959)
2. Hatfield, D., Weiser, P.: Property rights in spectrum: taking the next step. In: Proceedings of the first IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks (2005)
3. Xu, L., Tonjes, R., Paila, T., Hansmann, W., Frank, M., Albrecht, M.: DRiVE-ing to the internet: dynamic radio for IP services in vehicular environments. In: Proceedings of 25th Annual IEEE Conference on Local Computer Networks, pp. 281–289 (2000)
4. Benkler, Y.: Overcoming agoraphobia: building the commons of the digitally networked environment. Harvard J. Law Technol (1997–1998, Winter)
5. Lehr, W., Crowcroft, J.: Managing shared access to a spectrum commons. In: Proceedings of the first IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks (2005)
6. Raman, C., Yates, R., Mandayam, N.: Scheduling variable rate links via a spectrum server. In: Proceedings of the first IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks, pp. 110–118 (2005)
7. Ileri, O., Samardzija, D., Mandayam, N.: Demand responsive pricing and competitive spectrum allocation via a spectrum server. In: Proceedings of the first IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks (2005)
8. Chung, S., Kim, S., Lee, J., Cioffi, J.: A game-theoretic approach to power allocation in frequency-selective Gaussian interference channels. In: Proceedings of the IEEE International Symposium on Information Theory, pp. 316–316 (2003) (Submitted to IEEE Signal Processing Magazine, 29 Sept 2006)
9. Etkin, R., Parekh, A., Tse, D.: Spectrum sharing for unlicensed bands. In: Proceedings of the First IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks (2005)
10. Huang, J., Berry, R., Honig, M.: Spectrum sharing with distributed interference compensation. In: Proceedings of the first IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks (2005)
11. Mitola, J.: Cognitive radio for flexible mobile multimedia communications. In: Proceedings of the IEEE International Workshop on Mobile Multimedia Communications, pp. 3–10 (1999)
12. DARPA: The Next Generation (XG) Program. http://www.darpa.mil/ato/programs/xg/index.htm
13. Federal Communications Commission, Spectrum Policy Task Force report. ET Docket (02–135), 215 (2002)
14. Haykin, S.: Cognitive radio: brain-empowered wireless communications. IEEE J. Sel. Areas Commun. 23(2), 201–220 (2005)

15. Akyildiz, I.F., Lee, W.-Y., Vuran, M.C., Mohanty, S.: Next generation/dynamic spectrum access/cognitive radio wireless networks: a survey. Comput. Netw. **50**(13), 2127–2159 (2006)
16. Wang, B., Liu, K.R.: Advances in cognitive radio networks: a survey. IEEE J. Sel. Topics Signal Process. **5**(1), 5–23 (2011)
17. Akyildiz, I.F., et al.: A survey on spectrum management in cognitive radio networks. IEEE Commun. Mag. **46**(4), 40–48 (2008)
18. Hossain, E., Niyato, D., Han, Z.: Dynamic Spectrum Access and Management in Cognitive Radio Networks. Cambridge University Press, Cambridge (2009)
19. Ileri, O., Samardzija, D., Mandayam, N.B.: Demand responsive pricing and competitive spectrum allocation via spectrum server. Proc. IEEE DySPAN **2005**, 194–202 (2005)
20. Wang, S., et al.: Energy-efficient spectrum sensing and access for cognitive radio networks. IEEE Trans. Veh. Technol. **61**(2), 906–912 (2012)
21. Yucek, T., Arslan, H.: A survey of spectrum sensing algorithms for cognitive radio applications. IEEE Commun. Surv. Tutorials **11**, 116–130 (2009)
22. Akyildiz, I., Lee, W., Vuran, M., Mohanty, S.: Next generation/dynamic spectrum access/cognitive radio wireless networks: a survey. Comput. Netw. **50**(13), 2127–2159 (2006)
23. Mishra, S., Sahai, A., Brodersen, R.: 'Cooperative sensing among cognitive radios'. Proc. IEEE ICC **4**, 1658–1663 (2006)
24. Sun, C., Zhang, W., Letaief, K.: Cooperative spectrum sensing for cognitive radios under bandwidth constraints. In: Proceedings of the IEEE WCNC, pp. 1–5 (2007)
25. Sun, C., Zhang, W., Ben, K.: 'Cluster-based cooperative spectrum sensing in cognitive radio systems. In: Proceedings of the IEEE ICC, pp. 2511–2515 (2007)
26. Yucek, T., Arslan, H.: MMSE noise plus interference power estimation in adaptive OFDM systems. IEEE Trans. Veh. Technol. (2007)
27. Visotsky, E., Kuffner, S., Peterson, R.: On collaborative detection of TV transmissions in support of dynamic spectrum sharing. In: Proceedings of the IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, Baltimore, Maryland, USA, Nov, pp. 338–345 (2005)
28. Chen, R., Park, J.-M.: Ensuring trustworthy spectrum sensing in cognitive radio networks. In: Proceedings of the IEEE Workshop on Networking Technologies for Software Defined Radio Networks (held in conjunction with IEEE SECON 2006) (2006)
29. Burbank, J.: Security in cognitive radio networks: the required evolution in approaches to wireless network security. In: Proceedings of the International Conference on Cognitive Radio Oriented Wireless Networks and Communications, pp. 1–7 (2008)
30. Soto, J., Queiroz, S., Nogueira, M.: Taxonomy, flexibility, and open issues on pue attack defenses in cognitive radio networks. IEEE Wirel. Commun. **20**, 59–65 (2013)
31. Stevenson, C., Chouinard, G., Lei, Z., Hu, W., Shellhammer, S., Caldwell, W.: IEEE 802.22: the first cognitive radio wireless regional area network standard. IEEE Commun. Mag. **47**, 130–138 (2009)
32. Zhang, Y., Xu, G., Geng, X.: Security threats in cognitive radio networks. In: Proceedings of the IEEE International Conference on High Performance Computing and Communications, pp. 1036–1041 (2008)
33. Soto, J., Nogueira, M.: A framework for resilient and secure spectrum sensing on cognitive radio networks. Comput. Netw. **79**, 313–322 (2015)
34. Zhao, Q., et al.: Decentralized cognitive MAC for opportunistic spectrum access in Ad Hoc networks: a POMDP framework. IEEE JSAC **25**(3), 589–599 (2007)
35. Akyildiz, I.F., Lee, W.-Y., Mohanty, S.: Next generation/dynamic spectrum access/ cognitive radio wireless networks: a survey. Comput. Netw. **50**, 2127–2159 (2006)
36. Zhao, Q., Sadler, B.: A survey of dynamic spectrum access. IEEE Signal Process. Mag. **24**(3), 79–89 (2007)
37. Akyildiz, I.F., Lee, W.-Y., Vuran, M.C., Mohanty, S.: A survey on spectrum management in cognitive radio networks. IEEE Commun. Mag. **46**(4), 40–48 (2008)
38. Lee, W.-Y., Akyildiz, I.F.: A spectrum decision framework for cognitive radio networks. IEEE Trans. Mobile Comput. **10**(2), 161–174 (2011)

39. Lu, L., Zhou, X, Onunkwo, U., Li, G.Y.: Ten years of research in spectrum sensing and sharing in cognitive radio. EURASIP J. Wirel. Commun. Netw., 1–16
40. Akyildiz, I.F., Lee, W.-Y., Mohanty, S.: Next generation/dynamic spectrum access/ cognitive radio wireless networks: a survey. Comput. Netw. **50**, 2127–2159 (2006)
41. Zhu, X., Shen, L., Yum, T.: Analysis of cognitive radio spectrum access with optimal channel reservation. IEEE Commun. Lett. **11**(4), 304–306 (2007)
42. Song, Y, Xie, J.L.: Prospect: a proactive spectrum handoff framework for cognitive radio ad hoc networks without common control channel. IEEE Trans Mobile Comput. (to appear, 2011)
43. Wang, L., Wang, C., Chang, C.: Modeling and analysis for spectrum handoffs in cognitive radio networks. IEEE Trans Mobile Comput. (to appear, 2011)
44. Kushwaha, H., Xing, Y., Chandramouli, R., Heffes, H: Reliable multimedia transmission over cogntiive radio networks using fountain codes. Proc IEEE **96**(1), 155–165 (2008)
45. Willkomm, D., Gross, J., Wolisz, A.: Reliable link maintenance in cognitive radio systems. In: Proceedings of the IEEE International Symposium on New Frontier in Dynamic Spectrum Access Networks (DySPAN) (2005)
46. Unlicensed Operations in the TV Broadcast Bands: Second Memorandum Opinion and Order, FCC 10-174 (2010)
47. Wang, J., Shosh, M., Challapali, K.: Emerging cognitive radio applications: a survey. IEEE Commun. Mag. **49**(3), 74–81 (2011)
48. Technical Specification Group Radio Access Network: Evolved Universal Terrestrial Radio Access; Medium Access Control Protocol Specification, (Release 8), 3GPP Standard TS 36.321, (2010)
49. System Description Document: IEEE Standard 802.16 m, (2009)
50. Phunchongharn, P., Hossain, E., Niyato, D., Camorlinga, S.: A cognitive radio system for e-health applications in a hospital environment. IEEE Wirel. Commun. **17**, 20–28 (2010)

Spectrum Sensing for Half and Full-Duplex Cognitive Radio

A. Nasser, A. Mansour, K.C. Yao and H. Abdallah

Introduction

According to the Federal Communications Commission (FCC) [1–3], the typical radio channel occupancy is less than 15 %. The Dynamic Spectral Access has been proposed as a solution to increase the spectrum usage, and to solve the problem of the scarcity in the frequency resources due to the increasing in the demand on the wireless technologies. The Cognitive Radio (CR) technology is presented as a good candidate for applying the Dynamic Spectral Access. According to FCC, CR is an aware system, that can detect its own electromagnetic environment, so it can intelligently detect which communication channels are occupied and which are in use, and instantly move into vacant channels while avoiding occupied ones. This optimizes the use of available radio-frequency (RF) spectrum while minimizing interference. The CR idea is based on the sharing of the spectrum between users: licensed users are called Primary Users (PU) and unlicensed users are called Secondary Users (SU). SU can use the channel only when PU is idle. Generally, this activity is done in a non-cooperative context. In other words, PU is not obligated to notify SU on its activity.

A. Nasser (✉) · A. Mansour
LABSTICC, ENSTA Bretagne, 2 Rue François Verny, 29806 Brest, France
e-mail: abbass.nasser@ensta-bretagne.fr; abs.nasser@gmail.com

A. Mansour
e-mail: mansour@ieee.org

A. Nasser · H. Abdallah
Faculty of Science, AUCE, Beirut, Lebanon
e-mail: haniabdallah@auce.edu.lb

K.C. Yao
LABSTICC, UBO, 6 Av. Le Gorgeu, 29238 Brest, France
e-mail: koffi-clement.yao@univ-brest.fr

In such non-cooperative situation, SU must monitor the PU activities by sensing continuously the channel. When SU wants to transmit, he must sense the channel to check the PU status. If PU is active, then SU cannot access the channel and he should remain idle, or search for another available bands.

The spectrum monitoring is allocated to the Spectrum Sensing part in CR. Spectrum Sensing has to provide the CR with the PU status, in order to protect PU against the interference and in the other hand, to ensure an efficient use of vacant bands. Any error in the decision about the PU status can lead to either, harmful interference to PU or inefficient use of the spectrum.

Many methods have been proposed in order to perform the Spectrum Sensing [2]. Those methods can be categorized into blind, such as Energy Detector (ED) [4] and Autocorrelation Detector (ACD) [5] and non-blind, such as Cyclostationary Detector (CSD) [6] and Waveform Detector (WFD) [7]. WFD requires a perfect knowledge on the PU's signal [2, 7, 8]. Therefore it cannot be implemented in real non-cooperative scenarios, since CR deals with a great variety of signals. The classification into blind and non-blind is done according to the pre-information about the PU signal required to establish the Spectrum Sensing.

Energy Detector (ED) is the most used one in the blind category due to its low complexity. The main idea of ED is to compare the energy of the received signal to a predefined threshold depending on the noise variance which should be estimated before performing the Spectrum Sensing. Therefore, ED becomes vulnerable to the Noise Uncertainty (NU) problem incapable to diagnose the real PU status independently of detection time [9].

ACD requires an oversampling of the baseband received signal (i.e. number of samples per symbols: $N_s \geq 2$). The autocorrelation of a non-zero lag vanishes for a white noise. Otherwise, the autocorrelation will have a non zero value and indicates the presence of a communication signal (The PU's signal). The corresponding test statistic combines linearly the autocorrelation measures for different non-zero lags before making the decision on the PU status. The performance of this algorithm increases with N_s.

CSD measures the cyclostationary features of the received signal. The majority of telecommunication signals are cyclostationary thanks to the modulation process, the carrier frequency, the periodic pilot, *etc*. This fact makes CSD a good candidate for distinguishing between PU's signal and noise since the noise does not have cyclostationary features. In order to be applied, CSD needs to know (or estimate) in advance the cyclic frequency of the PU's signal. CSD tests the cyclostationary at a given cyclic frequency in order to examine the channel status.

Throughout this chapter, T_X denotes the test statistic corresponding to the detector X evaluated to make a decision on the channel status.

Performance Analysis

The Spectrum Sensing consists of making a decision on the presence of PU in the bandwidth of interest (BoI). The problem formulation on the presence/absence of the PU can be presented in a classic Bayesian detection problem: Under H_0, the PU is absent, whereas under H_1 PU exists.

$$\begin{cases} H_0 : y(n) = w(n) \\ H_1 : y(n) = hs(n) + w(n) \end{cases} \tag{1}$$

where h is the complex channel gain. $w(n) = w_r(n) + iw_i(n)$, is assumed to be an i.i.d zero mean circular symmetric complex Gaussian noise, i.e. $E[w(n)] = E[w^2(n)] = 0$, where $E[.]$ stands for the mathematical expectation. The real part, $w_r(n)$, and the imaginary part, $w_i(n)$, of $w(n)$ are independent and have equal variance.

$$E[w_r^2(n)] = E[w_i^2(n)] = \frac{\sigma_w^2}{2} = \sigma_w^2 = E[|w(n)|^2] \tag{2}$$

Without loss of generality, we can assume that $s(n)$ is normalized. In this case, the Signal to Noise Ratio (SNR), γ, is defined as follows:

$$\gamma = \frac{|h|^2}{\sigma_w^2} \tag{3}$$

A wrong decision about the channel status can affect either the PU transmission or the efficient use of the channel. In fact, a missed detection can cause a harmful interference which is caused by the transmission of the SU in the same band of the PU. A false alarm, however, decreases the profit of the channel. Therefore, the probability of detection (p_d) should be increased as much as possible, by keeping the probability of false alarm (p_{fa}) low. This can be considered as a Neyman-Person's detection method.

Energy Detector

ED consists in evaluating T_{ED} on N received samples, and compares it to a predefined threshold λ [4, 10, 11].

$$T_{ED} = \frac{1}{N} \sum_{n=1}^{N} |y(n)|^2 \tag{4}$$

The distribution of T_{ED} under H_0 tends toward a central χ^2 distribution with $2N$ degree of freedom. When N becomes large, the distribution of T_{ED} tends toward a Normal Distribution according to Central Limit Theorem. Since we are interested by low SNR, where a very large number of samples must be used during the detection process. In this case, the distribution of T_{ED} can be considered as a Gaussian with a

mean μ_0^{ED} and a variance V_0^{ED}: $\mathcal{N}(\mu_0^{ED}, V_0^{ED})$.

$$T_{ED} \overset{H_0}{\sim} \mathcal{N}(\mu_0^{ED}, V_0^{ED}) \tag{5}$$

where $\overset{H_i}{\sim}$ stands for the tendance of distribution under H_i, $i \in \{0, 1\}$, and $\mathcal{N}(a, b)$ represents a Normal distribution of mean a and variance b.

Under H_1, the distribution of T_{ED} depends on that of $s(n)$ and $w(n)$. Some researchers suggest to use the Gaussian distribution as generic model when $s(n)$'s distribution is totally unknown [10]. Such case leads to obtain a Normal Distribution of T_{ED} under H_1 for a large number of samples according to Central Limit Theorem.

$$T_{ED} \overset{H_1}{\sim} \mathcal{N}(\mu_1^{ED}, V_1^{ED}) \tag{6}$$

PU's signal is assumed in some cases to be deterministic with unknown value, in order to derive the analytic distribution of the test statistic [10, 12, 13]. When no assumption is taken on the distribution of the PU signal, the distribution of the test statistic becomes non-derivable [10].

The probability of False Alarm and Detection of ED for a large N can be derived as follows [10]:

$$p_{fa}^{ED} = Q\left(\frac{\lambda - \sigma_w^2}{\frac{1}{\sqrt{N}}\sigma_w^2}\right) \tag{7}$$

$$p_{d}^{ED} = Q\left(\frac{\lambda - (\sigma_w^2 + \sigma_s^2)}{\frac{1}{\sqrt{N}}(\sigma_w^2 + \sigma_w^2)}\right) \tag{8}$$

where $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{+\infty} \exp\left(\frac{-t^2}{2}\right)dt$ is the Q-function

Autocorrelation Detector

For ACD, the PU signal is assumed to be oversampled. The test statistic to be derived, T_{ACD}, is presented as follows [5]:

$$T_{ACD} = \frac{1}{\hat{r}_{yy}(0)} \sum_{m=1}^{N_s-1} Re\{\hat{r}_{yy}(m)\} \tag{9}$$

where $\hat{r}_{yy}(m) = \frac{1}{N}\sum_{n=1}^{N} y(n)y^*(n-m)$ is an estimator of the autocorrelation function $r_{yy}(m) = E[y(n)y^*(n-m)]$, and N_s is the number of samples per symbols.

The division operation by $\hat{r}_{yy}(0)$ avoids the vulnerability of ACD to the noise uncertainty, since ACD becomes independent of the noise variance. This operation makes ACD a Constant False Alarm Rate (CFAR) detector, so that the threshold of comparison is only depending to the number on samples used in the detection process.

According to [5], the False Alarm and Detection probabilities of ACD are expressed as follows:

$$p_{fa}^{ACD} = Q\left(\lambda\sqrt{\frac{N}{\lambda^2 + \Sigma}} \right) \tag{10}$$

$$p_{d}^{ACD} = Q\left(\frac{\lambda - \frac{\beta\gamma}{\gamma+1}}{\sqrt{V_{ACD}}} \right) \tag{11}$$

where $\Sigma = \frac{1}{2}\sum_{l=1}^{N_s-1} e_i^2, \beta = \sum_{l=1}^{N_s-1} e_l \frac{r_{ss}(l)}{r_{ss}(0)}, V_{ACD} = \frac{(1+\gamma)\lambda^2 - 4\beta\gamma\lambda + \Sigma + \eta\gamma}{N(\gamma+1)^2}$ and $\eta = \sum_{l=1}^{N_s-1}\sum_{k=1;l\neq k}^{N_s-1} e_l e_k \frac{r_{ss}(l-k) + r_{ss}(l+k)}{r_{ss}(0)}$, With $\{e_l\}_{l=1}^{N_s-1}$ is a set of weighting coefficients.

Cyclostationary Detector

In [6], the authors propose the Generalized Likelihood Ratio Test (GLRT) detector based on the cyclostationary features of the PU signal:

$$T_{CSD} = N\hat{r}_y \mathbf{K}\hat{r}_y^T \tag{12}$$

where the upper script T stands for the transpose of a matrix, and r_y is defined as follows:

$$r_y = \left[Re\{\hat{r}_{yy}(\tau_1, \alpha)\}, Re\{\hat{r}_{yy}(\tau_2, \alpha)\} ..., Re\{\hat{r}_{yy}(\tau_M, \alpha)\}, \right.$$

$$\left. Im\{\hat{r}_{yy}(\tau_1, \alpha)\}, Im\{\hat{r}_{yy}(\tau_2, \alpha)\} ..., Im\{\hat{r}_{yy}(\tau_M, \alpha)\} \right] \tag{13}$$

$\hat{r}_{yy}(\tau_i, \alpha)$ is the estimated cyclic autocorrelation function (CAF) at a cyclic frequency α and a time lag τ_i, and M is the number of the time lags used in the detection process:

$$\hat{r}_{yy}(\tau_i, \alpha) = \sum_{n=1}^{N} y(n)y^*(n - \tau_i)e^{-j2\pi\alpha n} \tag{14}$$

K stands for the covariance:

$$\mathbf{K} = \frac{1}{2} \begin{bmatrix} Re\{\mathbf{P} + \mathbf{Q}\} & Im\{\mathbf{Q} - \mathbf{P}\} \\ Im\{\mathbf{P} + \mathbf{Q}\} & Re\{\mathbf{P} - \mathbf{Q}\} \end{bmatrix} \tag{15}$$

where $\mathbf{P} = \left(P_{pq} \right)$ and $\mathbf{Q} = \left(Q_{pq} \right)$ are found as follows:

$$P_{p,q} = \sum_{m=\frac{1-L}{2}}^{\frac{L-1}{2}} f(m)\hat{r}_{yy}\left(\tau_p, \alpha + \frac{2\pi m}{N} \right)\hat{r}_{yy}^*\left(\tau_q, \alpha + \frac{2\pi m}{N} \right)$$

$$P_{p,q} = \sum_{m=\frac{1-L}{2}}^{\frac{L-1}{2}} f(m)\hat{r}_{yy}\left(\tau_p, \alpha + \frac{2\pi m}{N} \right)\hat{r}_{yy}\left(\tau_q, \alpha - \frac{2\pi m}{N} \right) \tag{16}$$

$f(m)$ is a normalized window, so that $\sum_{m=1}^{L} f(m) = 1$. The distribution of T_{CSD} is provided [6] under H_0 by a central χ^2 of $2M$ degrees of freedom, and H_1 by a non-central χ^2 distribution with $2M$ degree of freedom and a non-centrality parameter of $N\hat{r}_x \mathbf{K}\hat{r}_x^T$.

$$\begin{cases} T_{CSD} \overset{H_1}{\sim} \chi_{2M}^2 \\ T_{CSD} \overset{H_1}{\sim} \chi_{2M}^2\left(N\hat{r}_x \mathbf{K}\hat{r}_x^T \right) \end{cases} \tag{17}$$

The probability of false alarm and detection are given based on the distributions under H_0 and H_1 respectively:

$$p_{fa}^{CSD} = \gamma(\lambda/2, M) \tag{18}$$

$$p_d^{CSD} = Q_M\left(\sqrt{\lambda}, \sqrt{N\hat{r}_x \mathbf{K}\hat{r}_x^T} \right) \tag{19}$$

where $\gamma(a, b)$ is the upper incomplete gamma function and $Q_M(a, b)$ is the is the generalized Marcum Q-function.

Half and Full Duplex Cognitive Radio

Recently, the Full-Duplex transmission has gained a lot of attention thanks to recent advancements in the Self-Interference Cancellation (SIC), which makes the appli-

cation of FD communication possible. FD has been proposed in order to double the channel efficiency, where the same bandwidth is used simultaneously for both transmission and reception. However, FD transceiver should be able to eliminate the self interference. Due to many imperfections, a perfect elimination of the self-interference cannot be reached in real world applications [14, 15]. In wireless systems, the FD is considered as achieved if the Residual Self Interference (RSI) power becomes lower than the noise level. For that, an important SIC gain is required (around 110 dB for a typical WiFi system [14]). This gain can be achieved using the passive suppression and the active cancellation. The passive suppression is related to many factors that reduce the Self Interference (SI) such as the transmission direction, the absorption of the metals and the distance between the transmitting antenna, T_x, and the receiver, R_x. The active cancellation reduces the Self-Interference (SI) by using a copy of the transmitted signal at T_x. This copy is used in addition to the estimated channel between T_x and R_x to eliminate the self interference. The estimation of channel coefficients becomes an essential factor in the active cancellation process. Any error in the channel estimation leads to decreasing the SIC gain.

In the context of CR, FD means that SU is able to transmit and sense the channel as the same time. The SU makes a decision on the PU status using a Test Statistic depending on the PU signal and the noise. Any residual interference from the SU can affect the test statistic norm and leads to a wrong decision on the presence of PU.

Traditionally, Spectrum Sensing algorithms are deigned to deal with a Half-Duplex (HD). HD-CR divides the active period into two successive slots: the sensing slot and the transmission slot. During the sensing slot, SU must stop transmission in order to do not affect the decision reliability.

In FD-CR, SU remains active during the Sensing slot. This can be done without affecting the channel decision reliability by doing the SIC on the SU signal. After performing SIC, SU applies a Spectrum Sensing techniques to examine the PU status.

IN HD-CR, the detection problem to diagnose between H_0 and H_1 can be presented as in (17), whereas under FD-CR, the detection problem becomes as follows:

$$\begin{cases} H_0 : y(n) = gz(n) + w(n) \\ H_1 : y(n) = hs(n) + gz(n) + w(n) \end{cases} \tag{20}$$

where $z(n)$ is the SU signal and g represents the channel effect between T_x and R_x. By applying SIC, CR can eliminate the reflected SU signal from the received mixed signal. In ideal situation, where $gz(n)$ is totally eliminated, Eq. (20) refers to that of HD hypothesis (17).

Figure 1 shows the sensing-transmitting period of CR under FD and HD modes. It is clear that under HD, SU must stop the transmission each time the spectrum sensing is established. This issue leads to decrease the Data-Rate of SU. This problem is solved with FD-CR, but with a cost of degradation of Spectrum Sensing performance if the SIC is not efficient enough. For the two modes, SU can continue using the same bandwidth only if the Spectrum Sensing decides that PU is absent.
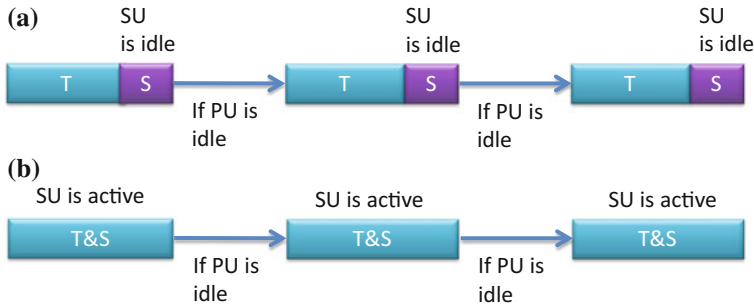
(a)

(b)

Fig. 1 CR Activity Structure for FD and HD transmission. Under HD, the activity period is divided into two slots: transmission slot and sensing slots, whereas under FD, CR performs the sensing while it is transmitting. **a** Half-Duplex Cognitive Radio activity. **b** Full-Duplex Cognitive Radio activity

Work Objectives

In our work, we are aiming at developing a new detector under HD. This detector should be blind that is make the sensing process more flexible. In addition, this fact avoid any pre-requisite information required by the non-blind detectors such as CSD. Moreover, the developed detector takes into account the problems that are frequently existing with the blind detectors. For example, ED suffers a performance degradation due to the noise uncertainty problem. Over more, ACD requires a high oversampling rate in order to present an efficient performance, so that at the limiting case, when the number of samples per symbol is two, ACD is poor comparing to ED. Also, this limitation will be avoided by our developed detectors. Besides that, to overcoming the problems of the state of the art detectors, the developed detectors present an efficient performance, require a short observation time and have a moderate complexity.

For FD mode, our main goal is to eliminate the distortion of the low noise amplifier (LNA), which is a main limiting factor that prevent the FD to be established. The residual power of this distortion affects the test statistic examining the channel status. To ensure a reliable evaluated test statistic, we are aiming at eliminating the LNA's distortion by using a simple and efficient algorithm. The proposed algorithm shows its superiority relatively to another state of the art one recently proposed.

Approaches Based on Cumulative Power Spectral Density

In this section, we present our new Spectrum Sensing algorithms based on the cumulative sum of the Power Spectral Density [16]. First, the baseband PU signal, $s(n)$, that might exist in the BoI can be modeled as follows:

$$s(n) = \sum_k b_k g(n - kN_s) = s_p(n) + js_q(n) \tag{21}$$

b_k are the symbols to be modulated, $g(n)$ is the shaping window, N_s stands for the number of samples per symbol (sa/sy), and satisfies the Nyquist criterion, $N_s = \frac{F_s}{B} \geq 2$, where F_s is the sampling frequency and $s_r(n)$ and $s_i(n)$ are the real and imaginary parts of $s(n)$ respectively. $s(n)$ has a bandwidth B and a real and even autocorrelation function. The same model of $s(n)$ has been used in [12] and [13], where $s(n)$, is assumed to be complex-valued zero mean unknown deterministic signal. $s_r(n)$ and $s_i(n)$ are assumed to be independent and have the same autocorrelation function.

Our new proposed Spectrum Sensing detectors are mainly based on the cumulative sum of the Power Spectral Density (PSD) of the received signal. When assuming the whiteness of the noise samples, the noise's PSD becomes flat. This aspect does not exist for the PU signal, which is assumed to be oversampled. Therefore its PSD is no longer flat. If the PU is absent, the cumulative sum of the received signal PSD has a close shape to a straight line. Whereas a curved shape is obtained when PU exists.

To enhance the robustness of our contribution, hard and soft cooperative schemes are introduced. In those two schemes, the spectrum is divided into two parts. The first part corresponds to the negative frequency points, while the second part deals with the positive frequency points. Hence, two test statistics are calculated, based on the Cumulative PSD of each of those two parts, and they are then combined according to the considered scheme.

The performance of our detectors is better than ones of the ED, ACD, and CSD under Gaussian and Rayleigh fading channels. Besides that, our detectors are less sensitive to the Noise Uncertainty than ED and they can be made independent from the noise variance.

Introduction to the Power Spectral Density

The power spectral density (PSD), $P_x(\nu)$, of a signal $x(n)$ can be estimated as the Fourier Transform of $\hat{r}_{xx}(m)$ as follows [17]:

$$P_x(\nu) = \lim_{N \to +\infty} \sum_{m=N/2-1}^{N/2} \hat{r}_{xx}(m) \exp\{-j2\pi\nu\frac{m}{N}\} \tag{22}$$

Since $w(n)$ is i.i.d., its autocorrelation becomes a dirac function:

$$r_{ww}(m) = \sigma_w^2 \delta(m) \tag{23}$$

In this case, $P_w(\nu)$ becomes flat on the band $[-B, B]$. On the other hand, $P_w(\nu)$ is real and even since $\hat{r}_{ww}(m)$ is real and even. It is obvious to deduce that the cumulative sum of the PSD becomes a straight line, with a slope σ_w^2. The over-sampling aspect of $s(n)$ makes it non i.i.d., then its PSD, $P_s(\nu)$, becomes not constant on $[-B; B]$.

Consequently, $P_y(\nu)$, becomes real and even, since $w(n)$ and $s(n)$ are assumed to be independent. $P_y(\nu)$ is flat under H_0, whereas it is non flat under H_1.

Based on this fact, the differentiation between the two hypotheses: H_0 and H_1 can be realized using the shape of the cumulative sum of PSD of random signals.

The PSD of a random signal $y(n)$ can be estimated as follows [17]:

$$\hat{P}_x(\nu) = \frac{1}{N}|Y(\nu)|^2 \tag{24}$$

where $Y(\nu)$ is the Discrete Fourier Transform (DFT) of the signal $y(n)$ with N samples:

$$Y(\nu) = \sum_{m=N/2-1}^{N/2} y(n) \exp\{-j2\pi\nu\frac{m}{N}\} \tag{25}$$

Cumulative Power Spectral Density-Based Detector

To diagnose the channel status, the test statistic to be estimated in should take into account the relative position of the obtained CPSD shape with respect to the reference straight line, which stands for the noise-only case (H_0). This test statistic is then compared to a threshold, λ, to obtain a decision on the presence of the PU signal. The value of λ must be set according to a target probability of false alarm, p_{fa}.

Let us define the CPSD, $C_y(\nu)$, of the received signal $y(n)$, over a discrete frequency interval $I = [k; l]$:

$$C_y(\nu) = \sum_{u=k}^{\nu} \hat{P}_y(u); \quad k \leq \nu \leq l, \tag{26}$$

where $\hat{P}_y(\nu)$ is the estimated PSD of $y(n)$ and can be found similarly to Eq. (24). In this manuscript, the test statistic is related to the Normalized CPSD, $\Gamma(\nu)$, of $y(n)$. $\Gamma(\nu)$ is defined as follows:

$$\Gamma(\nu) = \frac{C_y(\nu)}{(l-k+1)\sigma_w^2} \tag{27}$$

The term $(l-k+1)\sigma_w^2$ corresponds to the mean value of the last term of $C_w(\nu)$: $C_w(l)$. $C_w(\nu)$ stands for the CPSD of $w(n)$.

$$E[C_w(l)] = \frac{1}{N} \sum_{v=k}^{l} E[|W(\nu)|^2]$$
$$= (l-k+1)\sigma_w^2 \tag{28}$$

Thanks to the flat PSD of $w(n)$, the shape of $C_y(\nu)$ is closed to the straight line $R(\nu; k, l)$ under H_0.

$$R(v; k, l) = \frac{v - k + 1}{l - k + 1} \tag{29}$$

Under H_1 this property is not satisfied, since the PSD of $s(n)$ is not constant, and $\Gamma(v)$ has higher values than the ones obtained under H_0 due to the power of $s(n)$.

Figure 2 shows $\Gamma(v)$ under H_0 and H_1 with a SNR of 0 dB, $N = 10,000$ samples and $N_s = 2$ sa/sy. The $\Gamma(v)$ shape presented in this figure, can be considered as a detection criterion, if this shape is closed to $R(v; k, l)$, then H_0 is selected by the detector, else, H_1 is decided.

Motivated by the discussion above, the test statistic T can be obtained as the difference between $\Gamma(v)$ and the reference straight line $R(v; k, l)$. Accordingly, we introduce the two following detectors:

1. T_p: The normalized CPSD of $y(n)$ over the positive frequency points: $1 \leq v \leq N/2$), $\Gamma_p(v)$, can be defined as follows:

$$\Gamma_p(v) = \frac{1}{M\sigma_w^2} \sum_{u=1}^{v} P_y(u)$$

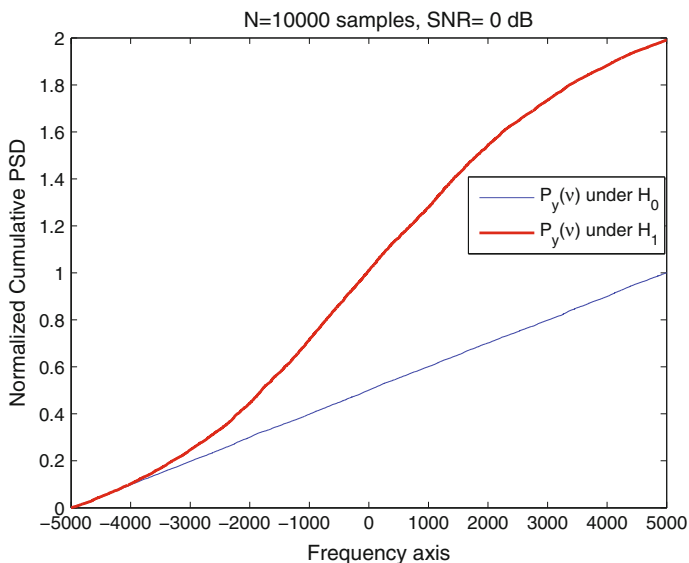$$= \frac{2}{N^2 \sigma_w^2} \sum_{u=1}^{v} |Y(u)|^2 \tag{30}$$

Fig. 2 Normalized Cumulative PSD of the received signal under H_0 and H_1

The T_p detector stands for the difference between $\Gamma_p(v)$ and $R(v; 1, M) = \frac{v}{M}$ [16]

$$
\begin{aligned}
T_p &= \sum_{v=1}^{M} \left(\Gamma_p(v) - \frac{v}{M} \right) \\
&= \sum_{v=1}^{M} \Gamma_p(v) - \frac{M+1}{2} \\
&= \frac{2}{N^2 \sigma_w^2} \sum_{\mu=1}^{M} (M - v + 1)|Y(v)|^2 - \frac{N/2 + 1}{2}
\end{aligned}
\tag{31}
$$

2. T_a: The detection process based on the CPSD of all frequencies of $y(n)$ (i.e. $-M + 1 \leq v \leq M$), T_a, can be defined similarly to (31) as follows:

$$
\begin{aligned}
T_a &= \sum_{v=-M+1}^{M} \left(\Gamma_a(v) - \frac{v+M}{N} \right) \\
&= \sum_{v=-M+1}^{M} \Gamma_a(v) - \frac{N+1}{2}
\end{aligned}
\tag{32}
$$

T_a is an extended version of T_p to cover the positive and negative frequencies.

Proposed Cooperative Detectors

In order to exploit the even parity of the PSD of the PU signal, and the i.i.d property of $W(v)$ [16], two cooperative detectors are proposed. The first proposed detector, T_{or}, aims to exploit all the bandwidth of the signal $y(n)$, by applying two test statistics: T_p tests the shape of the CPSD of $\hat{P}_y^p(v)$ (which is T_p) and T_n tests the CPSD shape of $\hat{P}_y^n(v)$ defined by:

$$
\hat{P}_y^n(v) = \hat{P}_y(-v + 1), \ 1 \leq v \leq M
\tag{33}
$$

$\hat{P}_y^n(v)$ corresponds to the negative frequency points. The detector, T_{or}, takes a final decision by performing a logical OR operation on the decisions of T_p and T_n.

The cooperative detector, T_{av}, performs the average between $P_y^p(v)$ and $P_y^n(v)$, before the computation of the CPSD. The averaging process makes the PSD more smooth.

OR Detector

T_{or} acts as hard cooperative detector of T_p and T_n by using an OR-rule.

$$T_{or} = OR(D_{T_p}, D_{C_n}) \tag{34}$$

where D_{T_p} and D_{C_n} are the detection results of T_p and C_n respectively.

Averaging Detector

T_{av} is introduced by averaging $\hat{P}_y^p(v)$ and $\hat{P}_y^n(-v+1)$, $1 \leq v \leq \frac{N}{2}$, to obtain $P_y^{av}(v)$. The average operation leads to evaluate a more accurate test statistic since the estimator $P_s(v)$ is deterministic and even while $\hat{P}_w(v)$ is i.i.d.

$$
\begin{aligned}
P_{av}(v) &= \frac{\hat{P}_y^p(v) + \hat{P}_y^n(v)}{2} \\
&= \frac{\hat{P}_y(v) + \hat{P}_y(-v+1)}{2}; \quad 1 \leq v \leq \frac{N}{2}
\end{aligned}
\tag{35}
$$

T_{av}, can be considered as a soft combining detectors of T_p and T_n.

$$
\begin{aligned}
T_{av} &= \sum_{v=1}^{M} \left(\Gamma_{av}(v) - \frac{v}{M} \right) \\
&= \sum_{v=1}^{M} \Gamma_{av}(v) - \frac{M+1}{2}
\end{aligned}
\tag{36}
$$

where Γ_{av} is the normalized CPSD corresponding to $P_{av}(v)$ (see Eq. (30)), and it is defined similarly to (30). The average PSD, $P_{y,av}(v)$, can illustrate a more smooth PSD of the received signal, especially for the noise component. This smoothing can help the detector to well distinguish between the two hypotheses, H_0 and H_1 (Fig. 3).

Performance Evaluation

In this section, we compare the performance of our detectors with that of ED [4], ACD [5] and CSD [6], under Gaussian and Rayleigh channels. A 16-QAM baseband modulated PU signal is considered with a rectangular shaping window.
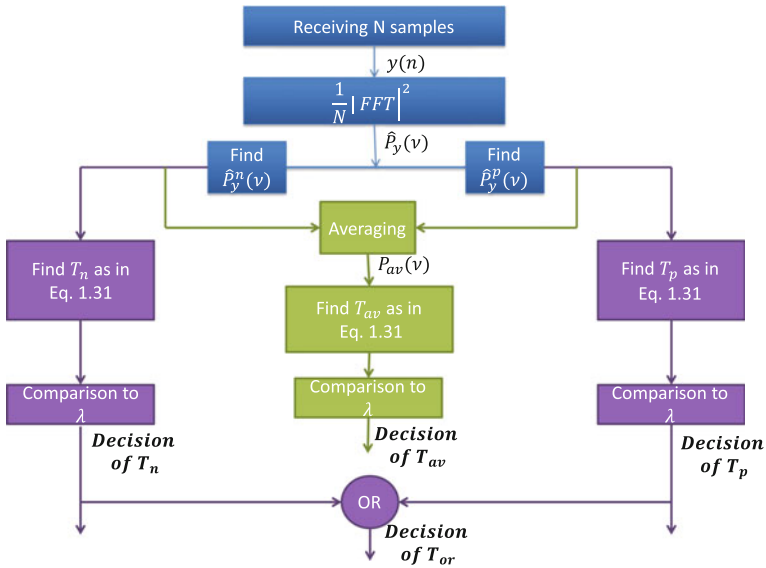
Fig. 3 The architecture of the proposed detectors

Gaussian Channel

First, Let us present the analytic False Alarm and Detection Probabilities of our proposed detectors, derived under Gaussian channel [16].[1]

1. The False Alarm probability of T_p, $p_{fa,p}$, is found as follows:

$$p_{fa,p} = Q\left(\frac{\lambda}{\sqrt{V_0}}\right) \tag{37}$$

where V_0 is the variance of T_p under H_0, and it is defined as follows:

$$\begin{aligned} V_0 &= E[T_p^2] - E^2[T_p] \\ &= \frac{N}{6} + \frac{1}{2} + \frac{1}{12N} \end{aligned} \tag{38}$$

On the other hand, the probability of detection of T_p, $p_{d,p}$, is:

$$p_{d,p} = Q\left(\frac{\lambda - \mu_1}{\sqrt{V_1}}\right) \tag{39}$$

---

[1]The analytic study of the false alarm alarm and probabilities of the proposed detectors is presented in details in [16].

where μ_1 and V_1 are the mean and the variance of T_p under H_1 respectively:

$$\mu_1 = \frac{2|h|^2}{N^2 \sigma_w^2} \sum_{v=1}^{M} (M - v + 1)|S(v)|^2 \tag{40}$$

$$V_1 = E[T_p^2] - E^2[T_p]$$

$$= V_0 + \frac{8\gamma}{N^3} \sum_{v=1}^{M} (M - 1 + v)^2 |S(v)|^2 \tag{41}$$

T_a is an extension of T_p, which covers all the N frequency point instead of only the positive frequency points (N/2 points). In order to obtain the corresponding False Alarm, $p_{fa,a}$, and detection, $p_{d,a}$ probabilities, we can simply replace $M = N/2$ by N in the expressions of $p_{fa,p}$ and $p_{d,p}$ respectively.

2. T_{or}:

As T_p and T_n are independent and have the same statistics and T_{or} applies the OR-rule, the probability of false alarm $p_{fa,or}$, and the probability of detection $p_{d,or}$, of T_{or} can be found as follows [18]:

$$p_{fa,or} = 1 - \left(1 - p_{fa,p}\right)^2 \tag{42}$$

$$p_{d,or} = 1 - \left(1 - p_{d,p}\right)^2 \tag{43}$$

3. T_{av}:

The probability of false alarm, $p_{fa,av}$, and detection, $p_{d,av}$, of T_{av} are expressed as follows:

$$p_{fa,av} = Q\left(\frac{\lambda - \mu_0^{av}}{\sqrt{V_0^{av}}}\right) \tag{44}$$

$$p_{d,av} = Q\left(\frac{\lambda - \mu_1^{av}}{\sqrt{V_1^{av}}}\right) \tag{45}$$

where: $\mu_0^{av} = 0$ and $V_0^{av} = \frac{V_0}{2}$ are the mean value and the variance of T_{av} under H_0 respectively, and $\mu_1^{av} = \mu_1$ and $V_1^{av} = \frac{V_1}{2}$ are the mean value and the variance under H_1 respectively.

The analytic and the simulated ROC curves of proposed detectors, are with good agreement as shown in Fig. 4. The simulation was done under the following conditions: 16-AM modulation, $\gamma = -12$ dB, $N = 1000$ samples and $N_s = 3$ sa/sy.

As shown in the same figure, T_{av} is the most efficient. T_{av} and T_{or} outperform significantly T_a and T_p. For that reason, we will only compare T_{av} and T_{or} to other known detectors.
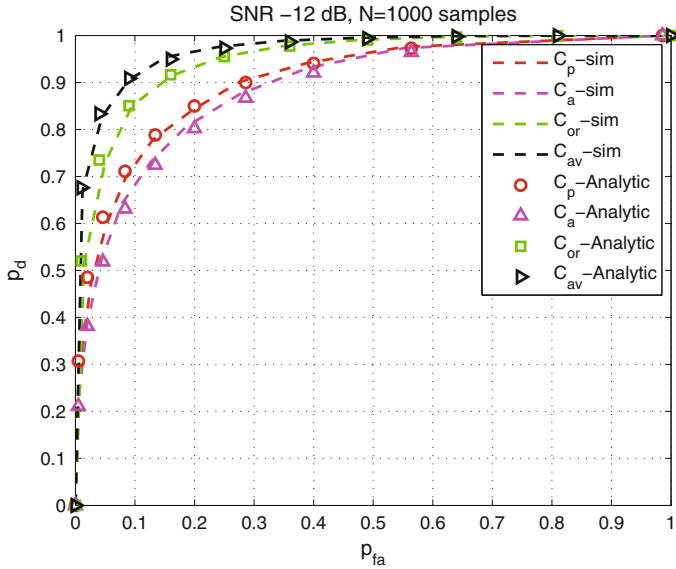
Fig. 4 Analytic results versus Simulation results for the proposed detectors under Gaussian channel.
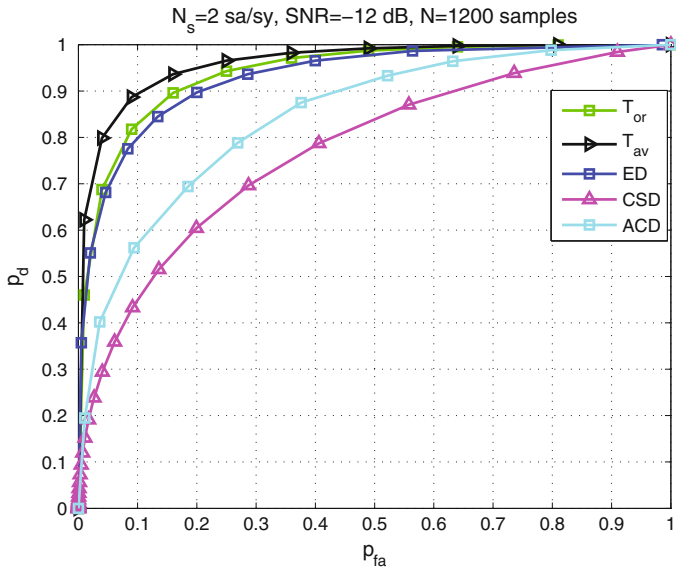

Fig. 5 ROC curves of our proposed detectors with T_{ed} and T_{ac} for $N_s = 2\,sa/sy$

Figure 5 presents the ROC curves under Gaussian channel of various detectors for $N_s = 2$. The simulation is done under $N = 1200$ samples and $\gamma = -12$ dB. CSD and ACD shows a poor performance relatively to T_{av}, T_{or} and ED, while T_{av} outperforms
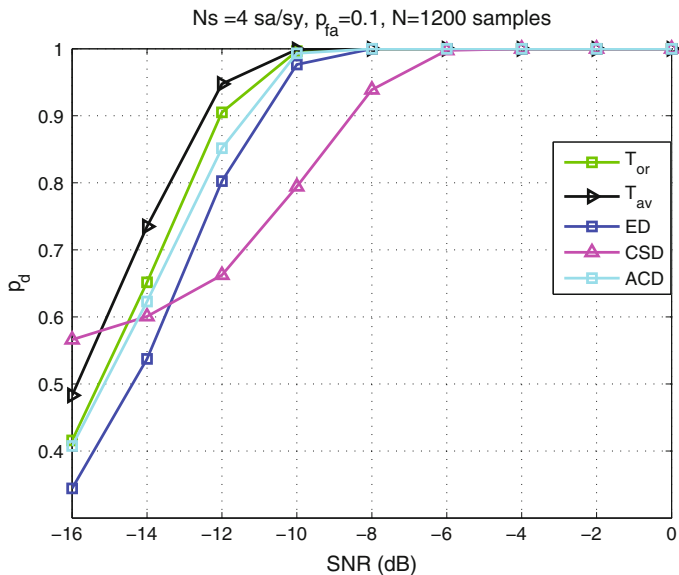
Fig. 6 Performance of our proposed detectors with ED and CSD for $p_{fa} = 0.1$ and $N_s = 4\,sa/sy$

all other detectors. Figure 6 shows the variation of the probability of detection in terms of SNR for a $p_{fa} = 0.1$, N = 1200 samples and $N_s = 4\,sa/sy$. Our detectors, T_{av} and T_{or} reach a higher probability of detection than the other detectors N_s. As shown in this figure, ACD outperforms ED for $N_s = 4\,sa/sym$, this is because ACD's performance increases with N_s.

Sensing Time

According to IEEE 802.2 regulation [19], the time sensing should be no longer than 2 s, with a maximal $p_{fa} = 0.1$ and a minimal $p_d = 0.9$. For that reason, the observation time required to ensure a reliable sensing process becomes a challenge for any Spectrum Sensing algorithm. In this section, we compare the sensing time of proposed algorithm to these of ED, ACD and CSD by evaluating the required number of samples to reach $(p_{fa}; p_d = 0.1; 0.9)$ under different SNR.

The simulation is performed under $N_s = 3\,sa/sy$. Figure 7 shows that when SNR decreases the number of samples increases, this is because a high number of samples can make the detection process more reliable. As shown in this figure, T_{av} and T_{or} can reach this target at a given SNR with lower number of samples than ED, ACD and
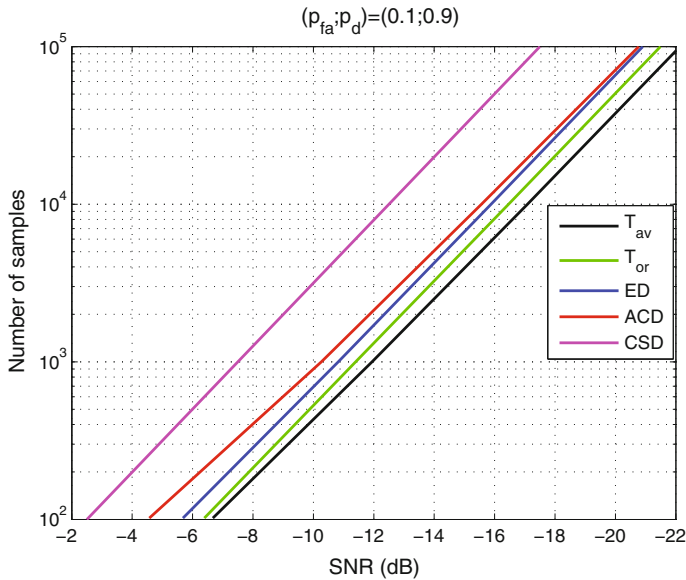
Fig. 7 The number of samples required to reach $(p_{fa}; p_d) = (0.1; 0.9)$ in terms of SNR

CSD. For example, with SNR of -10 dB, T_{av} and T_{or} reach this target with $N \simeq 400$ and $N \simeq 500$ samples respectively, whereas ED, ACD and CSD need $N \simeq 700$, 900 and $N \simeq 3500$ respectively to reach $(p_{fa}; p_d) = (0.1; 0.9)$.

Rayleigh Channel

In this section, we evaluate the performance of our proposed detectors under Rayleigh fading channel. The analytic probability of false alarm remains the same since it is independent of the channel gain h, which is assumed to be a circular symmetric Gaussian variable, but it keeps a constant value during a spectrum sensing period. On the contrary, the probability of detection depends on h, therefore it will have a new form depending on the distribution of the SNR, γ, $f_\gamma(\gamma)$, in the Rayleigh channel.

$$f_\gamma(\gamma) = \frac{1}{\bar{\gamma}} \exp\left(-\frac{\gamma}{\bar{\gamma}} \right) \tag{46}$$

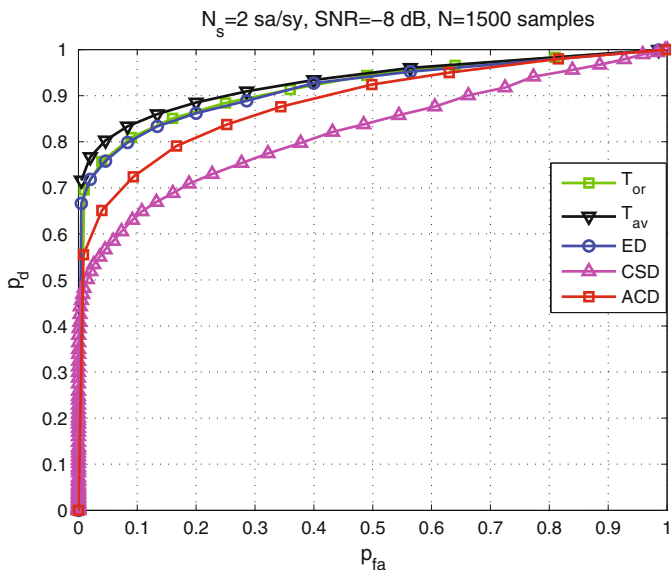where $\bar{\gamma}$ is the average SNR.

Fig. 8 ROC curve under Rayleigh fading channel for $N_s = 2\,sa/sy$

To find the probability of detection, P_{dray}, under Rayleigh channel, the probability of detection, p_{dgauss} under Gaussian channel should be averaged with respect to $f_\gamma(\gamma)$.[2]

$$p_{dray} = \int_0^{+\infty} P_{dgauss} f_\gamma(\gamma) d\gamma \qquad (47)$$

Figure 8 shows the ROC curves under Raleigh Fading channel for $N_s = 2\,sa/sy$ with $N = 1500$ samples and average SNR of -8 dB.

Under Rayleigh fading channel, T_{av} and T_{or} are still outperforming ED, ACD and CSD. The impact of the same fading affected by $P_y^p(v)$ and $P_y^n(v)$ on the cooperation established by T_{av} and T_{or} can be shown in Fig. 8, where the gap of performance between our detectors on the one side and ED, ACD, and CSD on the other becomes smaller when comparing it to that under Gaussian channel.

Complexity Analysis

Our simulations show that our proposed detectors outperforms the Energy, the Autocorrelation and the Cyclostationary detectors. Here, we will compare the complexity of our detectors to the complexity to the other detectors. According to Eq. (31), the

---

[2] A detailed derivation of the probability of detection of the proposed detectors is presented is [16].
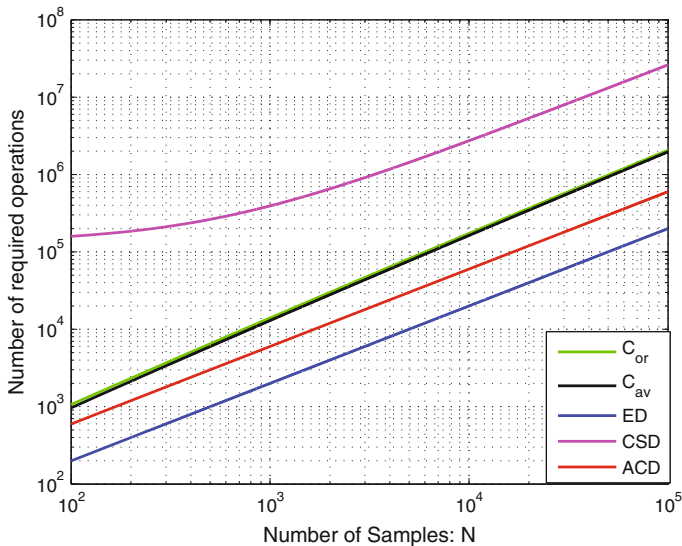
Fig. 9 The complexity of our proposed detector comparing to ED and CSD

Table 1 Comparison of different Spectrum Sensing Algorithms

Algorithm	PU signal requirement	Limitations	Complexity
ED	Blind	Sensitive to noise uncertainty	Low
ACD	Blind	Poor performance for low oversampling rate	Low
CSD	Cyclic frequency	Long sensing time	High
T_{or}	Blind	Applicable on oversampled signals	Moderate
T_{av}	Blind	Applicable on oversampled signals	Moderate

complexity of our proposed detector can be found. The complexity, Q, is linearly proportional to the number samples used in the detection process:

$$Q = L_q N \tag{48}$$

where $L_q > 2$ depends on the used detectors.[3]

Figure 9 shows that T_{av} is of less complexity than T_{or}. Our proposed detectors are slightly more complicated than ED and ACD. On the other hand, T_{av} and T_{or} have

---

[3]The derivation of the complexity of the proposed detectors can be found in [16].

less complexity and CSD. Consequently, our proposed detectors are of moderate complexity relative to CSD and ED.

Table 1 shows a briefed comparison among the proposed detectors and the state of the art detectors. This comparison is done with respect to the pre-requirement, the limitation and the complexity.

Robustness of Our Proposed Detectors Under Noise Uncertainty

In this section, we analyze the effect of the noise uncertainty (NU) on our proposed detectors. Due to the normalization (see Eq. (27)), the noise variance should be pre-estimated. That means, the performance of our proposed detectors is sensitive to the estimation of the noise variance. In this section, the impact of the NU on the robustness of our proposed detectors is evaluated.

The estimated noise variance $\hat{\sigma}_w^2$ can be modeled as follows:

$$\hat{\sigma}_w^2 \in \left[\frac{1}{r}\sigma_w^2; r\sigma_w^2\right] \tag{49}$$

σ_w^2 is the nominal value of the noise variance and r stands for the noise uncertainty factor with $r \geq 1$. In this manuscript, the distribution of $\hat{\sigma}_w^2$, $f_{\hat{\sigma}_w^2}(\hat{\sigma}^2)$, is assumed to be uniform in logarithmic scale.

$$f_{\hat{s}}(\hat{s}) = \begin{cases} \dfrac{1}{2\rho}, & s - \rho \leq \hat{s} \leq s + \rho \\ 0, & elsewhere \end{cases} \tag{50}$$

where $s = 10\log_{10}(\sigma_w^2)$, $\hat{s} = 10\log_{10}(\hat{\sigma}_w^2)$ and $\rho = 10\log_{10}(r)$.

The performance of the proposed detectors is numerically found, based on the model of Eq. (50). Figure 10 shows that the ROC curves of our proposed detectors and ED under an uncertainty factors $\rho = 0.6$ dB. It can be shown that T_{av}, T_{or} and T_p outperform significantly ED that has a similar performance to T_a. In addition, the performance of T_{av} and T_{or} tends toward that of T_p, then the NU affects the gain of the cooperation established by T_{av} and T_{or} to enhance the performance.

Figure 11 presents the performance loss of our detectors and ED for $p_{fa} = 0.1$, this figure shows $\Delta p_d = p_d(0) - p_d(\rho)$, where $p_d(0)$ and $p_d(\rho)$ stand for the probability of detection for the case where there is no NU and where NU is equal to ρ dB respectively. Figure 11 shows that our detectors are less sensitive to the noise uncertainty than the energy detector for $p_{fa} = 0.1$.
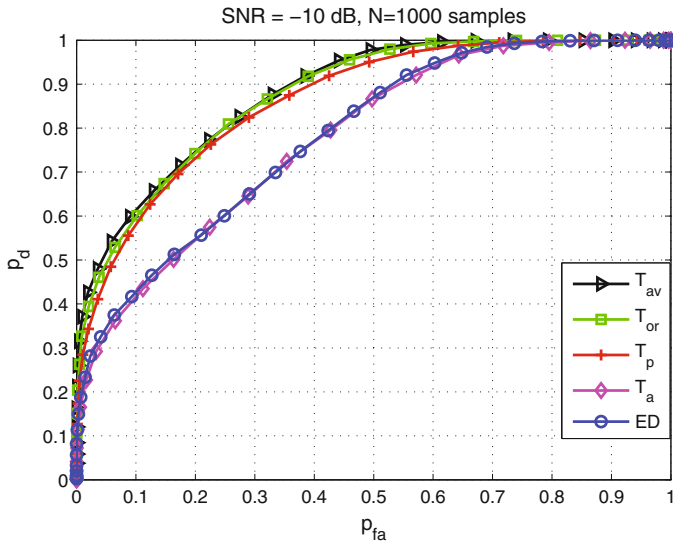
Fig. 10   ROC curves of the proposed schemes under noise uncertainty (NU)
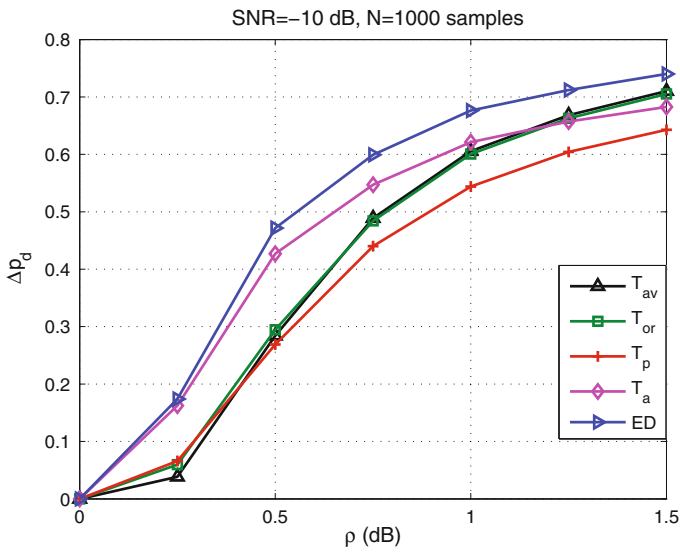

Fig. 11   Performance loss of p_d in terms of noise uncertainty for $p_{fa} = 0.1$

Spectrum Sensing Based on Self Normalized CPSD

According to Eq. (27), the CPSD should be normalized by a factor containing the noise variance, that necessitates a pre-estimation of the noise variance. This fact leads to a possible noise uncertainty. In this section, we introduce a new model for the proposed detectors, which is based on a self normalization of the CPSD. Instead of normalizing by the mean value of the last term of $CP_w(v)$, the CPSD will be normalized using its last term. The self normalized CPSD of the received signal is defined as follows:

$$sc_y(v) = \frac{CP_y(v)}{CP_y(l)}; \quad k \leq v \leq l \tag{51}$$

For $[k; l] = [-M + 1; M]$, $CP_y(M)$ is the estimated energy of the received signal. Using Parseval's theorem, $CP_y(M)$ becomes:

$$CP_y(M) = \frac{1}{N} \sum_{v=-M+1}^{M} |Y(v)|^2$$

$$= \sum_{n=1}^{N} |y(n)|^2 \tag{52}$$

For $[k; l] = [1; M]$, $CP_y(M)$ is the half of the energy of the received signal, this is because of the symmetric property of $P_w(v)$ and $P_s(v)$.

This normalization is equivalent to a scaling and does not change the shape form of the CPSD.

Without taking into account the relative position of $sc_y(v)$ with respect to the reference straight line, the decision about the presence of the PU signal is made by comparing the $sc_y(v)$ shape to the reference line. Figure 12 shows $sc_y(v)$ under H_0 and H_1 for a SNR $ss = 0$ dB. We have a straight line under H_0 and a curved shape under H_1.

Similarly to the detectors T_p, T_a, T_{or} and T_{av}, we define the test statistics of the Self-Normalized detectors: T_p^s, C_a^s, T_{or}^s and T_{av}^s as follows respectively:

$$T_p^s = \sum_{v=1}^{M} |sc_y(v) - d(v)| \tag{53}$$

$$T_a^s = \sum_{v=-M+1}^{M} |sc_y(v) - D(v)| \tag{54}$$
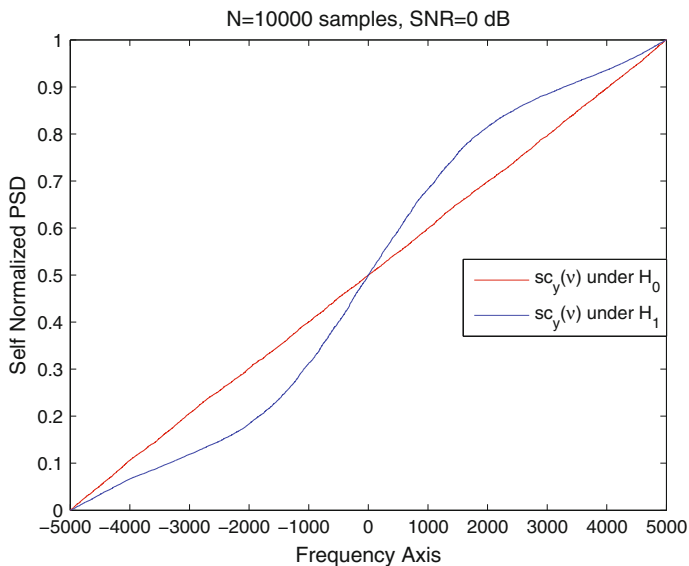
Fig. 12 Self normalized CPSD under H_0 and H_1

$$T_{or}^s = OR(D_{T_p^s}, D_{T_n^s}) \tag{55}$$

$$T_{av}^s = \sum_{v=1}^{M} |sc_y^{av}(v) - d(v)| \tag{56}$$

where $D_{T_p^s}$ and $D_{T_n^s}$ are the made decisions based on T_p^s and T_n^s respectively. T_n^s is defined similar to T_p^s, but it corresponds to the negative frequency points, and $sc_y^{av}(v)$ is given by:

$$sc_y^{av}(v) = \frac{\sum_{u=1}^{v} P_y^{av}(u)}{\sum_{u=1}^{M} P_y^{av}(u)} \tag{57}$$

Figure 13 shows the performance of those proposed detectors under NU of 0 and 0.5 dB with $N_s = 3\,sa/sy$. For NU = 0 dB, T_{av}, T_{or}, T_{or} and T_a outperform T_{av}^s, T_t^s, T_{or}^s and T_a^s respectively. When the NU 0.5 dB, the performance of T_{av}^s, T_t^s, T_{or}^s and T_a^s was not affected, whereas the detectors T_{av}, T_{or}, T_{or} and T_a suffer a performance degradation and become less robust than the self normalization detectors, as shown in Fig. 13a, b.

(a)

SNR = −10 dB, N=1500 samples

(b)

SNR = −10 dB, N=1500 samples

Fig. 13 ROC curves of the proposed schemes under NU = 0 and 0.5 dB. **a** $\rho = 0$ dB. **b** $\rho = 0.5$ dB

Spectrum Sensing for Full-Duplex Cognitive Radio

Applying FD in CR requires efficient techniques to achieve an important SIC, in order to eliminate the SU signal existing in the received mixture and avoid any impact on the decision of the Spectrum Sensing process.
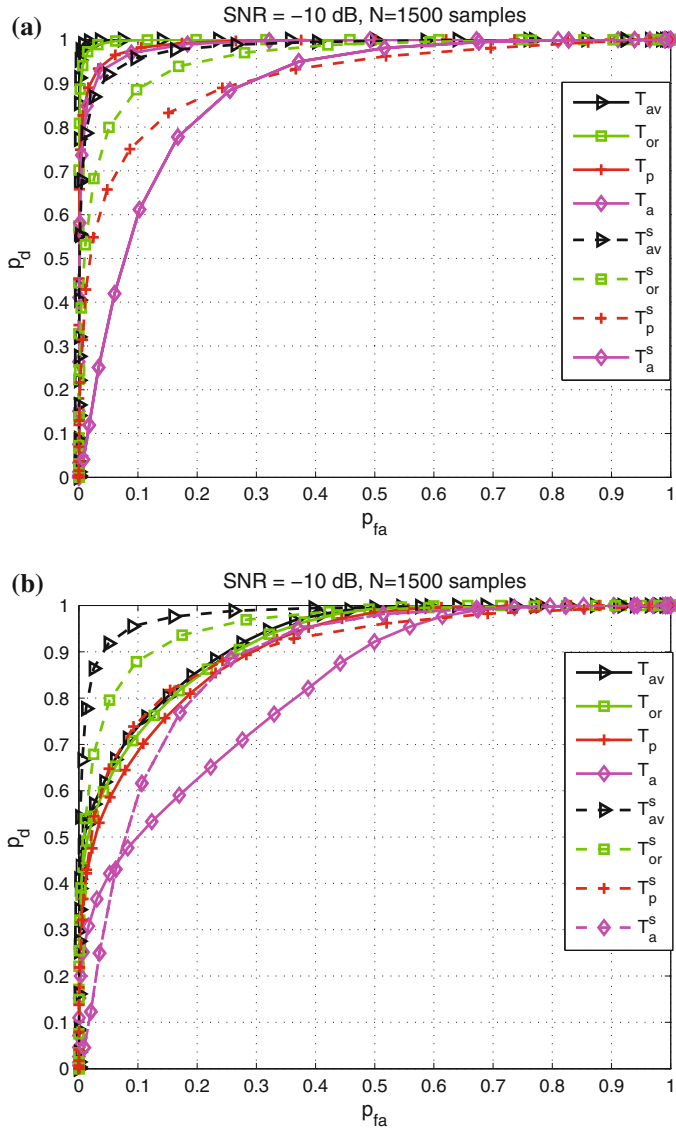
SIC should consider the receiver impairments such as such as the non-linearity of amplifiers and the oscillator noise. In fact, experimental results show the effect of the hardware imperfections that considered as the main limiting performance factors [14, 20–22]. These imperfections can affect the SIC process by preventing a good estimation of the channel between T_x and R_x and introducing residual power to the received mixture.

The authors of [14] modify their method previously proposed in [23] to estimate the channel and the Non-Linearity Distortion (LND) of the receiver Low-Noise Amplifier (LNA). Their method requires two training symbol periods. During the first period, the channel coefficients are estimated in the presence of the NLD. The non-linearity of the amplifier is estimated in the second period using the already estimated channel coefficients. It is worth mentioned that the estimation of the NLD parameters in the second phase depends on the one of the channel coefficients done in the first phase. However the estimation of the channel coefficients in the first phase can be depending on unknown NLD parameters. To solve the previous dilemma, we propose hereinafter an estimation method of the NLD in such way that the estimation of the channel cannot be affected by the NLD.

Many works have deal with the application of FD in CR [11, 24–27], In [24–27], the RSI is modeled as a linear combination of the SU signal without considering hardware imperfections. In [11, 25] the Energy Detection (ED) is studied in FD mode and the probability of detection and false alarm are found analytically. According to our best knowledge, there was no analytic relationship between the RSI, p_d and p_{fa} for both HD and FD mode.

This work deals with the Spectrum Sensing in real world applications. At first we analytically address the impact of the RSI power on the detection process. For that objective, we derive a relation between the RSI power, the probabilities of detection and false alarm under HD and FD modes. Secondly, we analyze the NLD impact on the channel estimation and the Spectrum Sensing performance. Hereinafter, a novel method is proposed to suppress the NLD of LNA without affecting the channel estimation process. Further, our proposed method outperforms significantly the method proposed in [14]. In addition, using our method, the receiver requires only one training symbol period to perform the estimation of the channel and the NLD estimation.

System Model

To deal with a real scenario, we admit, in our work, the signal waveform of [28], where we assume that PU signal and SU signals are wideband OFDM signal. Throughout the following, uppercase letters represent frequency-domain signals and

lower-case letters represent signals in time-domain. In this work, we are interesting to the additive receiver distortion which is dominated by the NLD of the LNA [14], the received signal can be modeled as follows:

$$Y_a(m) = GZ(m) + W(m) + D(m) + \eta HS(m) \tag{58}$$

G is the channel between the SU transmitter antenna T_x and the SU receive antenna R_x, $Z(m)$ is the SU signal, $W(m)$ is an Additive White Gaussian Noise (AWGN), $D(m)$ represents the NLD of the LNA, $S(m)$ is image of the the PU signal on R_x and $\eta \in \{0, 1\}$ is the PU indicator ($\eta = 1$ under H_1 and $\eta = 0$ under H_0).

After applying the SIC and the receiver imperfections mitigation, the obtained signal, $\hat{Y}(m)$ becomes as follows;

$$\hat{Y}(m) = \xi(m) + W(m) + \eta HS(m) \tag{59}$$

where $\xi(m)$ is the RSI and is defined as:

$$\xi(m) = (G - \hat{G})Z(m) + D(m) - \hat{D}(m) \tag{60}$$

$\hat{G}$ and $\hat{D}(m)$ are the estimated channel and the NLD respectively. Ideally $\hat{G} = G$ and $\hat{D}(m) = D(m)$, therefore Eq. (59) becomes: $\hat{Y}(m) = W(m) + \eta HS(m)$, which corresponds to a HD mode. Any mistake in the cancellation process may lead to a wrong decision about the PU presence.

Residual Self Interference Effect

In order to decide the existence of the PU, we use commonly used Energy Detector (ED) by comparing the energy, T, of the received signal to a predefined threshold, λ.

$$T = \frac{1}{N} \sum_{m=1}^{N} |\hat{Y}(m)|^2 \underset{H_0}{\overset{H_1}{\gtrless}} \lambda \tag{61}$$

By assuming the i.i.d property of $\epsilon(n)$, $w(n)$ and $s(n)$ ($\epsilon(n)$ is the time domain version of $\xi(m)$), then $\xi(m)$, $W(m)$ and $S(m)$ become i.i.d. [29]. In this case, the distribution of T should asymptotically follow a normal distribution for a large number of samples, N, according to Central Limit Theorem. Consequently, the probabilities of False Alarm, p_{fa}^F, and the Detection, p_d^F, under the FD mode can be obtained as follows [29] (Fig. 14):

$$p_{fa}^F = Q(\frac{\lambda - \mu_0}{\sqrt{V_0}}) = Q\left(\frac{\lambda - (\sigma_w^2 + \sigma_d^2)}{\frac{1}{\sqrt{N}}(\sigma_w^2 + \sigma_d^2)}\right) \tag{62}$$

$$p_d^F = Q(\frac{\lambda - \mu_1}{\sqrt{V_1}}) = Q\left(\frac{\lambda - (\sigma_w^2 + \sigma_d^2 + \sigma_x^2)}{\frac{1}{\sqrt{N}}(\sigma_w^2 + \sigma_d^2 + \sigma_x^2)}\right) \tag{63}$$

$$y = gz(t) + \eta s(t)$$
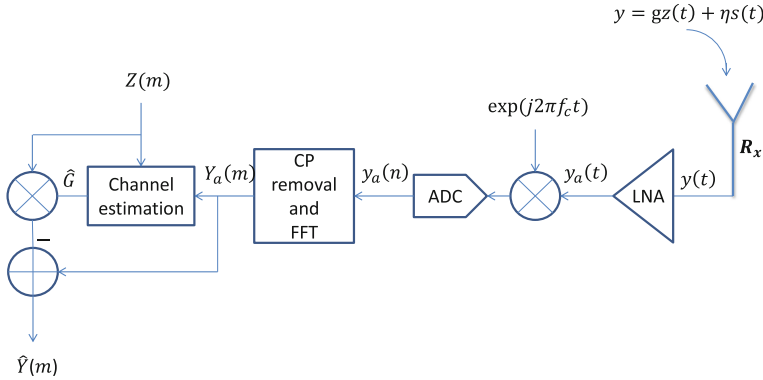
Fig. 14 Classical SIC circuit for OFDM receiver [29]
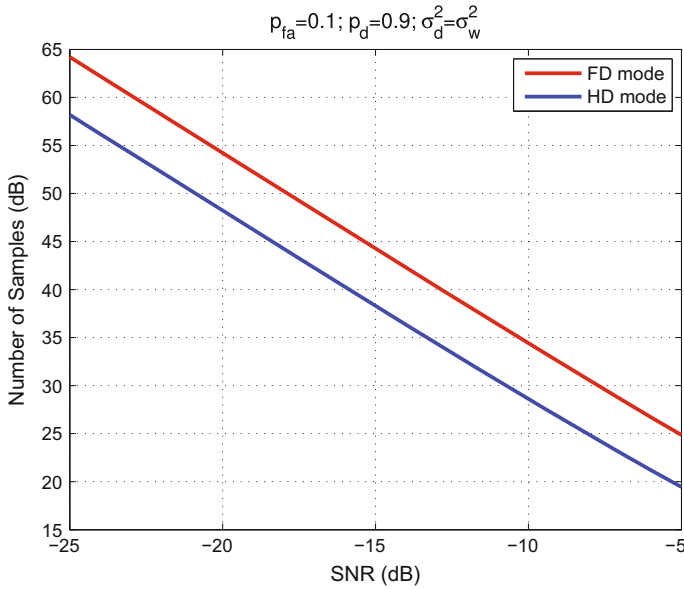
Fig. 15 The number of samples required to reach $P_d = 0.9$ and $P_{fa} = 0.1$ [29]

where μ_i and V_i are the mean and the variance of T under H_i respectively, $i \in \{0; 1\}$, $\sigma_d^2 = E[|\xi(m)|^2]$ represents the RSI power, $\sigma_w^2 = E[|W(m)|^2]$ and $\sigma_s^2 = E[|HS(m)|^2]$. The SNR, γ_s, is defined as: $\gamma_s = \frac{\sigma_s^2}{\sigma_w^2}$. If the SIC is perfectly achieved, i.e. $\sigma_d^2 = 0$, p_{fa}^F and p_d^F take their expressions under the HD mode.

Figure 15 shows the required number of samples to reach $p_d = 0.9$ and $p_{fa} = 0.1$ under the HD and FD modes with respect to the SNR. In FD mode, we set $\sigma_d^2 = \sigma_w^2$ as the target values of σ_d^2 in digital communication.

Figure 15 shows that the number of required samples slightly increases under the FD modes. For example if $\gamma_x = -5$ dB, then 85 samples are enough to reach the target $(P_d; P_{fa})$ under the HD mode while under FD mode, around 300 samples are needed. This means that in CR, the RSI power should be minimized as much as possible in order to do not affect the Spectrum Sensing performance $(\sigma_d^2 \ll \sigma_w^2)$.

Let us define the Probability of Detection Ratio (PDR), δ, for the same probability of false alarm under FD and HD modes, as follows:

$$\delta = \frac{p_d^F}{p_d^H} \; with \; p_{fa}^F = p_{fa}^H = \alpha \tag{64}$$

where p_d^H and p_{fa}^H are the probabilities of detection and false alarm under HD respectively, $0 \leq \alpha \leq 1$ and $0 \leq \delta \leq 1$. As with an excellent SIC, the probability of detection reached in FD can have mostly the same value as that in HF for the same probability of false alarm. In order to show the effect of RSI on δ, let us define the RSI to noise ratio γ_d as follows:

$$\gamma_d = \frac{\sigma_d^2}{\sigma_w^2} \tag{65}$$

Using (62) and (64), the threshold, λ, can be expressed as follows:

$$\lambda = \left(\frac{1}{\sqrt{N}} Q^{-1}(\alpha) + 1 \right) (\sigma_w^2 + \sigma_d^2) \tag{66}$$

By replacing (66) in (63), γ_d can be expressed as follows:

$$\gamma_d = \frac{(1 + \gamma_x)Q^{-1}(\delta P_d^H) - Q^{-1}(\alpha) + \sqrt{N}\gamma_s}{Q^{-1}(\alpha) - Q^{-1}(\delta P_d^H)} \tag{67}$$

If $\delta = 1$, then we can prove that γ_d becomes zero, which means that the SIC is perfectly achieved. Figure 16 shows the curves of γ_d for various values of PDR, δ, with respect to the SNR, γ_x, for $P_d^H = 0.9$ and $\alpha = 0.1$. This figure shows that as δ increases γ_d decreases. To enhance the PDR, the selected SIC technique should mitigate at most the SI. In fact, for $\gamma_x = -5$ and a permitted loss of 1 % (i.e. $\delta = 0.99$), γ_d is about -15 dB, which means the RSI power becomes 15 dB under the noise level.

Effect of the Amplifier Distortion

In real world applications, many imperfections prevents the application of the full duplex transceiver. One of the most important performance limiting factor is the NLD of LNA [14, 20–23]. According to NI 5791 datasheet [28], the NLD power is of 45 dB below the power of the linear amplified component. We present a new efficient algorithm, which shows a reliable performance comparing to that proposed in [14] and make the channel estimation performed without the influence of NLD. In addition, the Spectrum Sensing keeps a good performance under FD using this algorithm contrary to that proposed in [14].
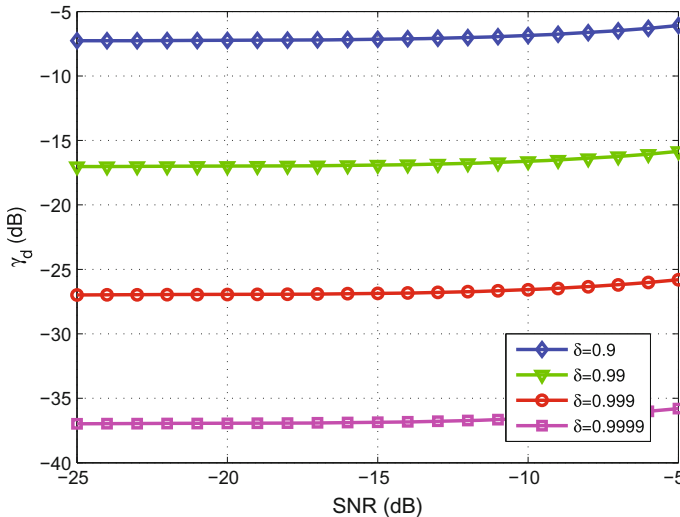
Fig. 16 Evolution of γ_d with respect to γ_x for various values of δ, $(P_{fa}^H ; P_d^H) = (0.1 ; 0.9)$ [29]

Estimation of the Non-linearity Distortion of LNA

The NLD stands for the odd degrees greater than one [30]. By limiting to the third degree since the component of higher degrees are of neglected power [14], the NLD component can be presented as follows:

$$d(t) = \beta y^3(t) \tag{68}$$

where β is the NLD coefficient. The estimation of β can be helpful to suppress the LNA output. In this case, the channel estimation is no longer affected by the NLD. The overall output signal of the LNA, $y_a(t)$, can be expressed as follows:

$$y_a(t) = \theta y(t) + \beta y^3(t) \tag{69}$$

θ and $y(t)$ are the power gain and the input signal of the LNA respectively. To estimate θ and β, by a and b respectively, one can minimize the following cost function based on the Least Square Error:

$$\mathscr{J} = E\left[\left(y_a(t) - (ay(t) + by^3(t))\right)^2\right] \tag{70}$$

By deriving $\mathscr{J}$ with respect to a and b we obtain:

$$\frac{\partial \mathscr{J}}{\partial a} = 0 \Rightarrow aE[y^2(t)] + bE[y^4(t)] = E[y_a(t)y(t)] \tag{71}$$

$$\frac{\partial \mathscr{J}}{\partial b} = 0 \Rightarrow aE[y^4(t)] + bE[y^6(t)] = E[y_a(t)y^3(t)] \tag{72}$$

Using Eqs. (71) and (72), a linear system of equations can be obtained:

$$\begin{bmatrix} a \\ b \end{bmatrix} = A^{-1}B \tag{73}$$

where:

$$A = \begin{bmatrix} E[y^2(t)] & E[y^4(t)] \\ E[y^4(t)] & E[y^6(t)] \end{bmatrix} ; \ B = \begin{bmatrix} E[y_a(t)y(t)] \\ E[y_a(t)y^3(t)] \end{bmatrix} \tag{74}$$

Once the non-linearity coefficient, β, is estimated, the non-linearity component can be subtracted from the output signal of the amplifier.

Numerical Results

Figure 17 shows the residual power of the NLD cancellation. The NLD power is fixed to 45 dB under the linear component [28]. This power is reduced to less than -300 dB after the application of our method. The method of [14] reduces the NLD power by around 50 dB. β is estimated using various numbers of training symbols, N_e. In this simulation, OFDM modulations are used with 64 sub-carriers and a CP length equal to 16. The received power is fixed to -5 dBm and the noise power to -72 dBm [28]. As shown in Fig. 17, the residual power of NLD decreases with an increasing of N_e when the method of [14] is applied. However, the proposed method
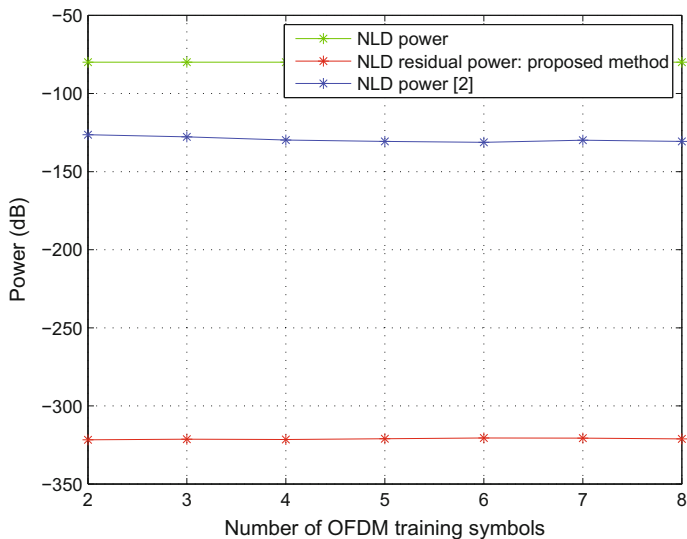

Fig. 17 The effect of the number of training symbols on the NLD residual power [29]

keeps a constant value of this power. Our technique outperforms significantly the method proposed in [14]. To show the impact of NLD on the channel estimation and the RSI power, Fig. 18 shows the power of $\hat{Y}(m)$ obtained in FD under H_0. The channel is estimated according to the method previously proposed by [31] as follows:

$$\hat{g} = IDFT\left\{ \frac{1}{N_e} \sum_{k=0}^{N_e} \frac{Y_a^k(m)}{Y^k(m)} \right\}$$

$$with\ \hat{G} = DFT\{\hat{g}(1,\dots,n_{tap})\} \tag{75}$$

where *IDFT* stands for the inverse discrete Fourier transform and n_{tap} is the channel order. The number of training symbols, N_e, is fixed to 4 symbols. To deal with a practical scenario, the number of sub-carrier is 64, the transmitted signal is of −10 dB (i.e. 20 dBm), and the noise floor is −102 dB (i.e. −72 dBm) [28]. The transceiver antenna is assumed to be omni-directional with 35 cm separation between T_x and R_x, so that a passive suppression of 25 dB is achieved [15]. According to the experimental results of [15], in a low reflection environment, 2 channel taps are enough to perform the SIC when the passive suppression is bellow 45 dB. Furthermore, the line of sight channel is modeled as Rician channel with K-factor about 20 dB. The non-line of sight component is modeled by a Rayleigh fading channel.

Figure 18 shows that our method leads to mitigate almost all the self interference, so that the power of $\hat{Y}(m)$ becomes very closed to the noise power. However, with the method of [14], the RSI power increases with the NLD power because the NLD power is a limiting factor of the channel estimation which leads to a bad estimation of the channel.
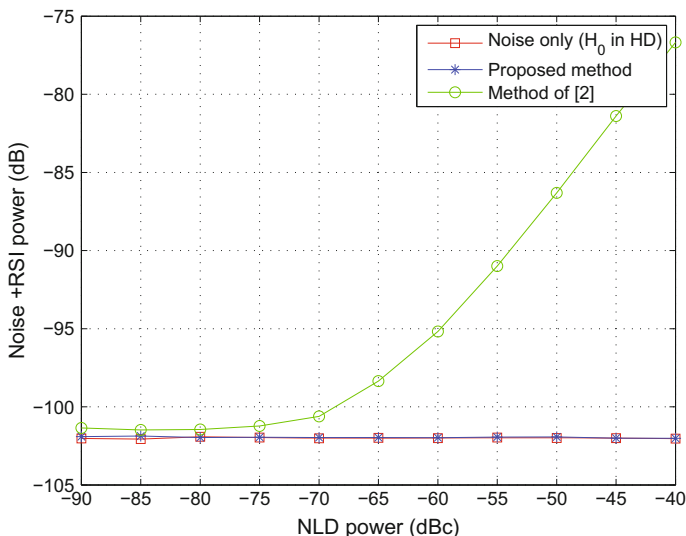
Fig. 18 The power of $\hat{Y}(m)$ under H_0 obtained after applying: (1) our proposed method, (2) the method of [14] is applied and (3) under HD mode [29]
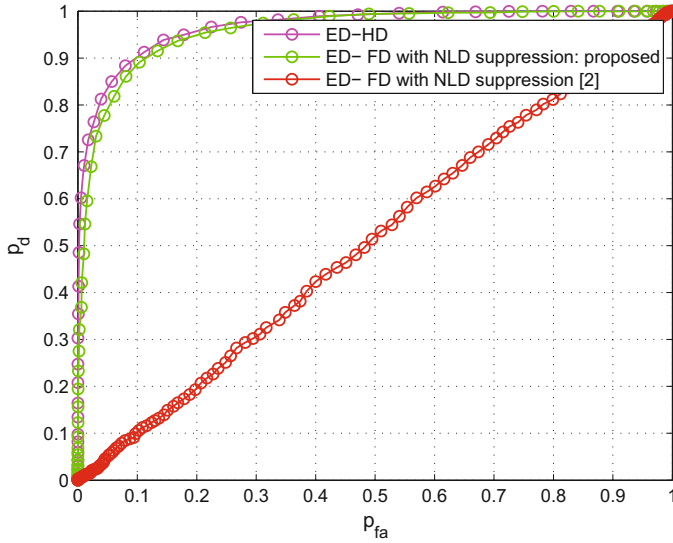
Fig. 19 The ROC curve after applying the proposed technique of NLD suppression [29]
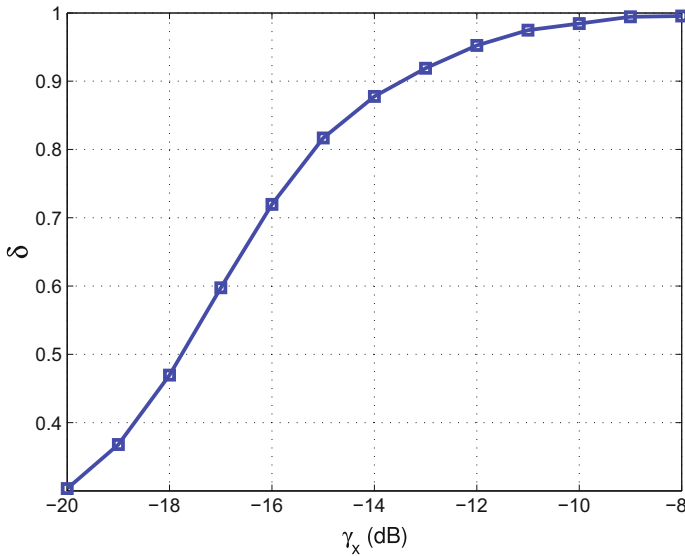
Fig. 20 The gain ratio: $\delta = \frac{p_d^F}{p_d^H}$ for different values of SNR [29]

To show the impact of the NLD on the Spectrum Sensing, Fig. 19 shows the ROC in various situations under $\gamma_x = -10$ dB. The simulations parameters in this figure are similar to those of Fig. 18, only the NLD power is 45 dB under the linear compo-

nent according to NI 5791 indications [28]. The method of [14] leads to a linear ROC, which means that no meaningful information about the PU status can be obtained. By referring to Fig. 18, the RSI power is of -82 dB for a NLD power of -45 dBc, which means that γ_d in this case is about 20 dB. This high RSI power leads to a harmful loss of performance (see Fig. 16). From the other hand, our method makes the ROC in FD mode almost colinear with that of the ROC of HD mode, which means that all SI and receiver impairments is mitigated.

Figure 20 shows the PDR for a target $\alpha = 0.1$ and $p_d^H = 0.9$. The ratio δ increases with the SNR. At a low SNR of -10 dB, δ becomes closed to 1, so that a negligible performance loss is happen. As the SNR decreases the detection process in FD mode becomes more sensitive to the RSI power.

Conclusion

In this chapter, we address the Spectrum Sensing for Half and Full-Duplex Cognitive Radio.

For Half-Duplex mode, new Spectrum Sensing detectors based on the Cumulative Power Spectral Density (CPSD) were proposed. Our proposed detectors verify the linearity of the CPSD shape of the received signal. This shape is almost linear if and only if the PU is absent.

Hard and soft decision schemes are used to combine the CPSD measures, which are derived based on the two symmetric parts of the Power Spectral Density. False alarm and detection probabilities were derived analytically under both Gaussian and Rayleigh flat fading channels. The simulation results show the performance superiority of our detectors comparing to other detectors especially the energy detector.

In addition, simulation results show that the proposed detectors are less affected by the noise uncertainty than the energy detector. However, to avoid the impact of the noise uncertainty, the measured CPSD is normalized by the estimated energy of the received signal. By doing this, we make our detectors independent from noise variance.

In a future work, we are looking to enhance the Power Spectral Density estimator, and to extend our algorithms to deal with Multiple Antennas Spectrum Sensing problem.

On the other hand, we deal with the Full-Duplex CR by addressing the impact of the Residual Self Interference on the Spectrum Sensing for Full-Duplex Cognitive Radio. An analytic relation is derived relating the residual self interference with the probabilities of detection and false alarm under Full-Duplex and Half-Duplex modes. Furthermore, a new method is proposed to mitigate an important receiver impairment, which is the Non-Linear Distortion of the Low Noise Amplifier. This method shows its efficiency, leading the Spectrum Sensing performance in Full-Duplex mode to be closed to that under Half-Duplex mode.

References

1. Mitolal, J.: Cognitive radio: making software radios more personal. IEEE Personal Commun. **6**, 13–18 (1999)
2. Yucek, T., Arslan, H.: A survey of spectrum sensing algorithms for cognitive radio applications. IEEE Commun. Surv. Tutorials. **11**, 116–130 (2009)
3. Spectrum Policy Task Force Report, E.D.N. 02-135, Federal Communications Commission, Washington, DC (2002)
4. Digham, F., Alouini, M.-S., Simon, K.: On the energy detection of unknown signals over fading channels. IEEE Trans. Commun. **55**, 21–24 (2007)
5. Naraghi-Poor, M., Ikuma, T.: Autocorrelation-based spectrum sensing for cognitive radio. IEEE Trans. Veh. Technol. **59**, 718–733 (2010)
6. Dandawate, A., Giannakis, G.: Statistical tests for presence of cyclostationarity. IEEE Trans. Signal Process. **10**, 2355–2369 (1994)
7. Cabric, D., Tkachenko, A., Brodersen, R.W.: Spectrum sensing measurements of pilot, energy, and collaborative detection. In: The IEEE Military Communications Conference (MILCOM), pp. 1–7 (2006)
8. Nasser, A., Mansour, A., Yao, K.C., Charara, H., Chaitou, M.: Efficient spectrum sensing approaches based on waveform detection. In: Third International Conference on e-Technologies and Networks for Development (ICeND), pp. 13–17 (2014)
9. Tandra, R., Sahai, A., Andrade, A.G.: SNR walls for signal detection. IEEE J. Sel. Top. Signal Process. **2**, 4–17 (2008)
10. Atapattu, S., Tellambura, C., Jiang, H.: Energy Detection for Spectrum Sensing in Cognitive Radio. Springer (2014)
11. Cheng, W., Zhang, X., Zhang, H.: Full duplex spectrum sensing in non-time-slotted cognitive radio networks. In: The IEEE Military Communications Conference (MILCOM), pp. 1029–1034 (2011)
12. Derakhshani, M., Le-Ngoc, T., Nasiri-Kenari, M.: Efficient cooperative cyclostationary spectrum sensing in cognitive radios at low SNR regimes. IEEE Trans. Wirel. Commun. **10**, 3754–3765 (2011)
13. Zhu, Y., Liu, J., Feng, Z., Zhang, P.: Sensing performance of efficient cyclostationary detector with multiple antennas in multipath fading and lognormal shadowing environments. J. Commun. Netw. **16**, 162–171 (2014)
14. Elsayed, A., Eltawil, A.M.: All-digital self-interference cancellation technique for full-duplex systems. IEEE Trans. Wirel. Commun. **7**, 291–294 (2015)
15. Everett, E., Sahai, A., Sabharwal, A.: Passive self-interference suppression for full-duplex infrastructure nodes. IEEE Trans. Wirel. Commun. **13**, 680–694 (2014)
16. Nasser, A., Mansour, A., Yao, K.C., Chaitou, M., Charara, H.: Spectrum sensing based on cumulative power spectral density. EURASIP J. Adv. Signal Process. (Under Review)
17. Iouna, C., Mansour, A., Quinquiz, A., Radoi, E.: Digital Signal Processing using MATLAB. Wiley (2014)
18. Akyildiz, T.F., Lo, B.F., Balakrishnan, R.: Cooperative spectrum sensing in cognitive radio networks: a survey. Phys. Commun. **4**, 40–62 (2011)
19. Draft Standard for Wireless Regional Area Networks: Fed. Commun. Commission, Washington, DC (2008)
20. Sahai, A., Patel, G., Dick, C., Sabharwal, A.: IEEE Trans. Veh. Technol. **62**, 3494–4510 (2013)
21. Bliss, D.W., Hancock, T.M., Schniter, P.: Asilomar conference on signals, systems and computers. In: Proceeding Asilomar Conference on Signals, Systems and Computer, pp. 34–39 (2012)
22. Ahmed, E., Eltawil, A.M., Sabharwal, A.: Rate gain region and design tradeoffs for full-duplex wireless communications. IEEE Trans. Wirel. Commun. **7**, 3556–3565 (2013)
23. Ahmed, E., Eltawil, A.M., Sabharwal, A.: Self-interference cancellation with nonlinear distortion suppression for full-duplex systems. In: Proceeding Asilomar Conference on Signals, Systems and Computers, vol. II, pp. 1199–1203 (2013)

24. Heo, J., Ju, H., Park, S., Kim, E., Hong, D.: Simultaneous sensing and transmission in cognitive radio. IEEE Trans. Wirel. Commun. **13**, 149–160 (2014)
25. Rihonen, T., Wichman, R.: Energy detection in full-duplex cognitive radios under residual self-interference. In: 9th International Conference on Cognitive Radio Oriented Wireless Networks (CROWNCOM), pp. 57–60 (2014)
26. Afifi, W., Krunz, M.: Adaptive transmission-reception-sensing strategy for cognitive radios with full-duplex capabilities. In: International Symposium on Dynamic Spectrum Access Networks (DYSPAN). pp. 149–160 (2014)
27. Afifi, W., Krunz, M.: Incorporating self-interference suppression for full-duplex operation in opportunistic spectrum access systems. IEEE Trans. Wirel. Commun. **14**, 2180–2191 (2015)
28. NI 5791R: User Manual and Specifications. Nat. Instrum., Austin, TX, USA. http://www.ni.com/pdf/manuals/373845d.pdf
29. Nasser, A., Mansour, A., Yao, K.C., Charara, H., Chaitou, M.: Spectrum sensing for full-duplex cognitive radio systems. In: 11th International Conference on Cognitive Radio Oriented Wireless Networks (CROWNCOM), France, vol. 55, pp. 363–374 (2016)
30. Schenk, T.: RF Imperfections in High-rate Wireless Systems, Impact and Digital Compensation. Springer, New York, NY, USA, McGraw-Hill, New York, NY, USA (2008)
31. Kang, Y., Kim, K., Park, H.: Efficient DFT-based channel estimation for OFDM systems on multipath channels. IET Commun. **1**, 197–202 (2007)

Dynamic Spectrum Access Algorithms Without Common Control Channels

Jen-Feng Huang and Chi-Tao Chiang

In this chapter, we systematically address the recently important dynamic spectrum access algorithms without common control channels. This kind of dynamic spectrum access algorithms is one of the key MAC layer functions in the *cognitive radio networks* (CRNs). In CRNs, *secondary users* (SUs) are only allowed to exchange their (control) messages on idle licensed channels (called *available channels*) to avoid interfering with *primary user*s (PUs). In normal multiple channels networks, fixed common control channels are used to exchange necessary information, e.g., hopping sequences and synchronization signals. However, implementing the common control channels is a challenge in CRNs because the availability of channels is time-varying and position-varying. Thus, more and more algorithms such as JS [1], CRSEQ [2], etc. are developed on special hopping rules to make SUs exchange their necessary messages on common available channels as soon as possible. The chapter covers the dynamic spectrum access algorithms without common control channels, the metrics of the algorithms, and necessary mathematical background in the following sections.

Introduction for Dynamic Spectrum Access Algorithm

A part of the spectrum has low utilization such as 2G or DTV when we are facing spectrum shortage problem. Thus, spectrum utilization enhancement techniques are focused by researchers in the recent years.[1] One of them, *cognitive radio network* (CRN), becomes a key technique to solve low utilization problem of the spectrum.

---

[1]2008–2016.

J.-F. Huang (✉) · C.-T. Chiang
Industrial Technology Research Institute (ITRI), Zhudong, Hsinchu, Taiwan
e-mail: sampage100@gmail.com

In the CRNs, there are two types of roles: *primary user*s (PUs) and *secondary user*s (SUs). When PUs don't transceive data on its licensed channels, SUs are allowed transceiving data on the channels which aren't used by PUs (called *available channel*s). In other words, SUs can transceive data on the PUs' licensed channels if they don't interfere with PUs. In the past, two communicating nodes can negotiate their communication channels via common control channels. However, implementing fixedly common control channels in CRNs is impractical and ineffective because channels may be occupied by PUs in any time. Thus, no common control channels based dynamic spectrum access algorithms become a research trend in CRNs.

Instead of common control channel solutions, SUs follow special channel hopping algorithms to hop on a common available channel as soon as possible. Intuitively, the worst case and average case of rendezvous time of two SUs are concerned, where *rendezvous* means that two communication SUs can hop on a same channel.

- *Maximum Time-to-Rendezvous* (MTTR): the maximum time for a pair of SUs to rendezvous on a common available channel. An algorithm with MTTR = α means that SUs could rendezvous on a common available channel within α timeslots if there is at least one common available channel for them.
- *Average Time-to-Rendezvous* (ATTR): the average time for a pair of SUs to rendezvous on a common available channel.

Besides, another concerned metric is the robustness of two SUs' communication, named *degree of rendezvous*, which is defined as follow.

- *Degree of Rendezvous* (i.e., *Rendezvous Diversity*): the minimum number of channels on which a pair of SUs can rendezvous, i.e., the minimum number of channels on which any two hopping sequences can rendezvous. Two SUs cannot exchange control messages if their hopping sequences do not rendezvous on a common available channel. Maximizing the degree of rendezvous can increase the probability of rendezvous on common available channels and reduce the impact of PU long time blocking problem.

Based on above design goals, there are several channel hopping algorithms are proposed. Each SU hops on channels based on the hopping sequence sequentially for rendezvous. Once two SUs hop to the same available channel in the same timeslot, SU can communicate with each other. In the literatures, the models of channel hopping algorithms are classified according to the following three criteria.

- *Heterogeneous/homogeneous role*. Heterogeneous role algorithms require role pre-assignment (i.e., every SU has a pre-assigned role as either a sender or a receiver, like the master and slave in Bluetooth) prior to the beginning of channel hopping sequences, while homogeneous role algorithms do not.
- *Symmetric/asymmetric available channel model*. Under symmetric available channel model, all SUs have an identical available channel set, while under

asymmetric available channel model, SUs might have diverse available channel sets.

- *Synchronous/asynchronous*. Synchronous algorithms are used under the assumption that all SUs can start their own channel hopping sequences at the same global time, while asynchronous algorithms can be used without the assumption.

In this chapter, we discuss heterogeneous role and homogeneous role algorithms under symmetric available channel model and asymmetric available channel model in later sections.

Background

Most dynamic spectrum access algorithms without common control channels are based on channel hopping technique. For ease of discussion, we formulate *channel hopping* (CH) system before introducing these algorithms.

A Channel Hopping System

A SU consists of several basic CR components, referred to Fig. 1. The basic components include spectrum sensing, spectrum sharing, spectrum access algorithm, spectrum access radio and scheduling. Each SU performs spectrum sensing

Fig. 1 Basic components of a SU

Basic components of a SU
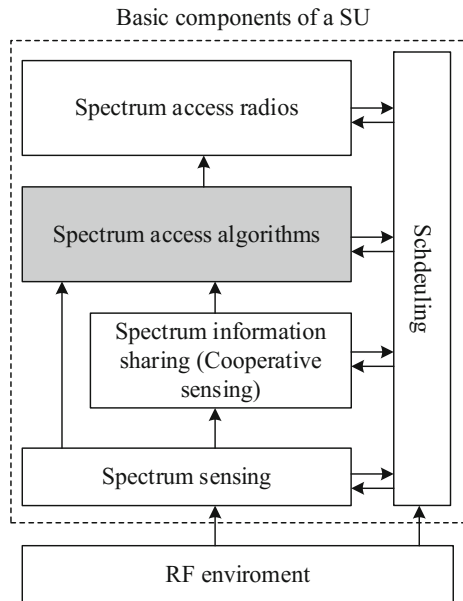
to sense channels and gets available channels set (*available set* for short). The SU shares sensing results with other SUs or PUs for interference avoidance and accesses available channels based on the spectrum access algorithms and spectrum access radios. (Note that some SUs may not have the spectrum information sharing component when they don't perform cooperative spectrum sensing.) Because the availability of channels is time-varying and position-varying, the channel hopping algorithm becomes a popular mechanism in the development of spectrum access algorithm component. In this chapter, we focus on the delay bounded channel hopping algorithms (marked gray part of Fig. 1).

Suppose that there are N licensed channels in a CRN, labelled as $0, 1, \ldots, N-1$. An available set is a subset of N licensed channels. Besides, each SU is equipped with one half-duplex CR radio transceiver. A *CH system* is divided into multiple timeslots, and several timeslots make a CH period. Each SU shall visit the channels according to the CH sequence $\mathbf{U}$ (which is determined by the sequence hopping algorithm $\mathcal{A}$) in the CH period. We represent the CH sequence $\mathbf{U}$ with CH period $= m$ timeslots as follow: $\mathbf{U} = (u_0, u_1, \ldots, u_{m-1})$, where $u_i \in [0, N-1]$ represents the channel visited by SUs in the ith timeslot of a CH period.

Below, we define the rendezvous property of two CH sequences in a CH system. Given two CH sequences $\mathbf{U} = (u_0, u_1, \ldots, u_{m-1})$ and $\mathbf{V} = (v_0, v_1, \ldots, v_{m-1})$. Let $\mathcal{G}(\mathbf{U}, i)$ and $\mathcal{G}(\mathbf{V}, j)$ denote the global time slots of the ith timeslot of $\mathbf{U}$'s CH period and the jth timeslot of $\mathbf{V}$'s CH period, respectively. CH sequences $\mathbf{U}$ and $\mathbf{V}$ have *rendezvous property* if $\mathbf{U}$ and $\mathbf{V}$ have the same element at the same global time slot t. That is there exists integers i and j such that $u_i = v_j$ and $t = \mathcal{G}(\mathbf{U}, i) = \mathcal{G}(\mathbf{V}, j)$. Clearly, if any two SUs hop channels based on sequences $\mathbf{U}$ and $\mathbf{V}$, respectively, they can hop to the same channel (i.e., *rendezvous*).

A CH system is classified two types: synchronous CH system and asynchronous CH system. A *synchronous CH system* is that each SU's local clock synchronizes with the global clock of the CH system, otherwise as an *asynchronous CH system*. In asynchronous CH system, the slot boundaries of SUs are aligned. In distributed networks, maintaining a system global clock is a difficult and impractical task. Therefore we focus the asynchronous CH algorithms. In the section "Introduction for Dynamic Spectrum Access Algorithm", we have mentioned the important metric MTTR. An algorithm has bounded MTTR implies that, for all $0 \leq i, j < m$, the rotation function $\mathcal{R}(\mathbf{U}, i)$ of a CH sequence $\mathbf{U} \in \mathcal{A}$ and the $\mathcal{R}(\mathbf{V}, j)$ of a CH sequence $\mathbf{U} \in \mathcal{A}$ have rendezvous property. The rotation function $\mathcal{R}(\mathbf{U}, i)$ is the sequence generated by shifting i elements of sequence $\mathbf{U}$ left in a circular fashion. For example, as Fig. 2a, if $\mathbf{U} = (1, 2, 0, 2, 1)$, then $\mathcal{R}(\mathbf{U}, 1) = (2, 0, 2, 1, 1)$. Now we give an example of the algorithm which has bounded MTTR. Assume that there is an algorithm which generates CH sequences $\mathbf{U} = (1, 2, 0, 2, 1)$ and $\mathbf{V} = (1, 2, 0, 2, 1)$. All of rotation cases of $\mathbf{U}$ and $\mathbf{V}$ are listed in Fig. 2a, b. In Fig. 2c, we check $\mathcal{R}(\mathbf{U}, 0)$ and $\mathcal{R}(\mathbf{V}, j)$, for all j, and their rendezvous property. Gray grids are rendezvous channels of CH sequences $\mathcal{R}(\mathbf{V}, j)$ and $\mathcal{R}(\mathbf{U}, 0)$. When $j = 0$, i.e., $\mathcal{R}(\mathbf{U}, 0) = \mathcal{R}(\mathbf{V}, 0) = (1, 2, 0, 2, 1)$, two sequences have rendezvous property. When $j > 0$, two sequences also have

(a) | **(b)** | **(c)**

R(**U**, 0) | 1 2 0 2 1
R(**U**, 1) | 2 0 2 1 1
R(**U**, 2) | 0 2 1 1 2
R(**U**, 3) | 2 1 1 2 0
R(**U**, 4) | 1 1 2 0 2

R(**V**, 0) | 1 2 0 2 1
R(**V**, 1) | 2 0 2 1 1
R(**V**, 2) | 0 2 1 1 2
R(**V**, 3) | 2 1 1 2 0
R(**V**, 4) | 1 1 2 0 2

R(**U**, 0) | 1 2 0 2 1
R(**V**, 0) | 1 2 0 2 1
R(**V**, 1) | 2 0 2 1 1
R(**V**, 2) | 0 2 1 1 2
R(**V**, 3) | 2 1 1 2 0
R(**V**, 4) | 1 1 2 0 2

Fig. 2 An example of the bounded MTTR. **a** All rotation cases of **U**. **b** All rotation cases of **V**. **c** Rendezvous property of $\mathcal{R}$(**U**, 0) and $\mathcal{R}$(**V**, j), for all $0 \le j < m$

rendezvous property. For example, the case of $\mathcal{R}$(**U**, 0) rendezvous on channel 2. Besides, if this rendezvous channel is available channel, this algorithm has bounded MTTR (in actual, this channel may not be an available channel).

To solve long time blocking problem, designing an algorithm which has bounded MTTR is necessary in CRNs. Throughout this chapter, nations are defined in Table 1. We introduce quorum system in section "Quorum System" and some basic theorems of CH system in section "Basic Theorems" before we introduce the existing algorithms.

Table 1 Notations

Notations	Means	(Referred to) Sections
N	Number of licensed channels	Section "A Channel Hopping System"
N_c	Number of common available channels	Section "Symmetric Available Channel Model"
$\mathcal{A}_i$	Channel hopping algorithm i	Section "A Channel Hopping System"
$\mathcal{G}$(**U**, i)	Global timeslots of the ith timeslot of sequence **U**	Section "A Channel Hopping System"
$\mathcal{R}$(**U**, i)	A sequence generated by shifting i elements of sequence **U** left in a circular fashion	Section "A Channel Hopping System"
$\mathcal{Q}$	A quorum system	Section "Quorum System"
$\mathcal{D}$	A different set	Section "Quorum System"
Z_k	An integer set, which includes element $\{0, 1, 2, \ldots, k - 1\}$	Section "Homogeneous Role Algorithms"
Z	An integer set	Section "Homogeneous Role Algorithms"
A_I	The available channels set of SU_I	Section "Homogeneous Role Algorithms"
p	The smallest prime which is greater or equal to N	Section "Homogeneous Role Algorithms"
P	The smallest prime number larger than N	Section "Homogeneous Role Algorithms"
d	Number of original ID bits of a SU	Section "Homogeneous Role Algorithms"

Quorum System

The quorum system has intersection property and channel hopping algorithms can guarantee the bounded MTTR by the aid of the property. The definition of a quorum system is stated below:

Definition 1 *Let* $\varepsilon = \{e_0, e_1, \ldots e_{N-1}\}$ *be a set of elements. A quorum system* $\mathcal{Q} \subset 2^\varepsilon$ *is a set of subsets of ε such that every two subsets intersect. Each $Q_i \in \mathcal{Q}$ is called a* quorum.

For example, let quorum system $\mathcal{Q} = \{Q_1, Q_2, Q_3\}$, where $Q_1 = \{1, 3, 5\}$, $Q_2 = \{2, 3, 4\}$, and $Q_3 = \{1, 2, 5\}$. Clearly, arbitrary two quorums intersect, e.g., the Q_1 and Q_2 have intersection element 3.

Definition 2 *A cyclic quorum system constructed by* $\mathcal{Q} = \{Q_0, Q_1, \ldots, Q_{N-1}\}$, *where* $Q_i = \{q_0 + i \pmod N, q_1 + i \pmod N, \ldots, q_{k-1} + i \pmod N\}$, *For all $i = 0$,* 1, …, N – 1.

The cyclic quorum system is constructed by special quorums, called (N, k)-different set. The (N, k)-different set has following property:

Theorem 1 *Given an N-element universal set $U = \{0, 1, \ldots, N - 1\}$. A subset* $\mathcal{D} = \{d_1, d_2, \ldots, d_k\} \subset U$, *where $d_i \in U$ and $k \leq N$, is called a (N, k)-different set if for every $e \neq 0 \bmod N$ there exists at least one ordered pair of elements (d_i, d_j) such that $d_i - d_j = e \bmod N$.*

For example, $\mathcal{Q} = \{Q_0, Q_1, Q_2, \ldots, Q_6\} = \{\{0, 1, 3\}, \{1, 2, 4\}, \{2, 3, 5\}, \{3, 4, 6\}, \{4, 5, 0\}, \{5, 6, 1\}, \{6, 0, 2\}\}$ is a cyclic quorum system constructed by a $(7, 3)$-cyclic different set $\mathcal{D} = \{d_1, d_2, d_3\} = \{0, 1, 3\}$. Arbitrary intersection of two cyclic quorums in $\mathcal{Q}$, namely $\{0, 1, 3\}$, $\{1, 2, 4\}$, $\{2, 3, 5\}$, $\{3, 4, 6\}$, $\{4, 5, 0\}$, $\{5, 6, 1\}$ and $\{6, 0, 2\}$, is not an empty set.

Now we give a simple example to demonstrate application of the quorum system, quorum channel hopping algorithm (QCH) [3]. This algorithm is developed for synchronous scenarios. At beginning, a SU creates a quorum system $\mathcal{Q}$. The SU randomly chooses a quorum $Q_i \in \mathcal{Q}$ over Z_k $(= \{0, 1, \ldots, k - 1\})$. In global timeslot t, the SU hops on channel $(t/k \bmod N)$ if slot $(t \bmod k) \in \mathcal{Q}_i$. Otherwise the SU hops randomly on a channel. For example, referred to Fig. 3, suppose that there are 3 channels and a quorum system $\mathcal{Q} = \{\{0, 1\}, \{0, 2\}, \{1, 2\}\}$ over Z_3. If SU_1 chooses the quorum $\{0, 1\}$, SU_1 hops on channel 0 in timeslot 0 and 1, and channel 1 in timeslot 3 ($3 \bmod 3 = 0$) and 4 ($4 \bmod 3 = 1$), and channel 2 in timeslot 6 ($6 \bmod 3 = 0$) and 7 ($7 \bmod 3 = 1$). In other slots, SU_1 hops on a random channel. As SU_1, SU_2 hops on channel 0 in timeslot 0 and 2, channel 1 in timeslot 3 and 5, and channel 2 in timeslot 6 and 8 if SU_2 chooses the quorum $\{0, 2\}$. Now we focus on channel 0. SU_1 and SU_2 can rendezvous on channel 0 in timeslot 0, because any two quorums intersect. In the same way, SU_1 and SU_2 also rendezvous on other channels. Thus, if two SUs have a common available channel, the SUs can rendezvous and exchange data on the channel. Clearly, the QCH has bounded MTTR $= kN$.

Fig. 3 An example of QCH

SU$_1$	0	0	r	1	1	r	2	2	r
SU$_2$	0	r	0	1	r	1	2	r	2
SU$_3$	r	0	0	r	1	1	r	2	2

Global Timeslot: 0 1 2 3 4 5 6 7 8

Basic Theorems

Except the example of QCH, which is applied the quorum system, any asynchronous CH system, which has bonded MTTR, is a cyclic quorum system in actual, referred to Theorem 2.

Theorem 2 *Suppose a channel hopping algorithm has bounded MTTR α and SUs can rendezvous on channel i. Let set Q_i collect the slot index, where SUs visit channel i. Besides, all elements in the Q_i should mod α. We can find that the set Q_i is a cyclic quorum in Z_α.*

Proof Without loss of generality, we assume $Q = \{q_1, q_2, …, q_k\}$ and $Q_j = \{(q_1 + j)$ mod α, $(q_2 + j)$ mod α, …, $(q_k + j)$ mod $\alpha\}$. Suppose that the set Q and Q_j are not cyclic quorums in Z_α. Because the channel hopping algorithm has bounded MTTR, for all j, SUs can rendezvous on channel i when one of SU's slot delays j slots with another one. That is channel $i = q_o$ in $Q = (q_p + j)$ mod α in Q_j, where $1 \leq o, p \leq k$, a contradiction.   □

Although we assume that the slot boundaries of SUs are aligned in asynchronous CH system, the system works in the scenarios, where the slot boundaries of SUs are misaligned. Details can be found in Theorem 3.

Theorem 3 *If a channel hopping algorithm $\mathcal{A}$ with MTTR = α and degree of rendezvous = m in asynchronous model (where slot boundaries are aligned), sequences generated by the algorithm $\mathcal{A}$ must overlap by at least m/2 slots during α slots when the time slot boundaries are aligned or misaligned* [3].

Proof Suppose that the unit of a slot is 1 and SU$_X$'s local clock is ahead of SU$_Y$'s one slot by an arbitrary amount of time, $(i + \delta)$, where $i \in Z$ and $0 \leq \delta < 1$. Besides, SU$_X$ and SU$_Y$ apply algorithm $\mathcal{A}$ to generate channel hopping sequences **U** and **V**, respectively. We consider two cases.

Case 1, when slot boundaries are aligned (i.e., $\delta = 0$): According the definition of MTTR in asynchronous model. Sequences **U** and **V** overlap at least m slots, which $> m/2$, referred to Fig. 4a.

Case 2, when slot boundaries are misaligned (i.e., $0 < \delta < 1$): Considering two sub cases, $\delta \leq 1/2$ and $\delta > 1/2$. For the first case, $\delta \leq 1/2$, $\mathcal{A}$ can rendezvous in slot i in asynchronous model (where slot boundaries are aligned). Because $\delta \leq 1/2$, the offset of slot i of SU$_X$ and slot $i' = i + \delta$ of SU$_Y$ $\leq 1/2$. Therefore, the overlapping part of **U** and **V** $> 1/2$ slot, referred to Fig. 4b. Because there are at least
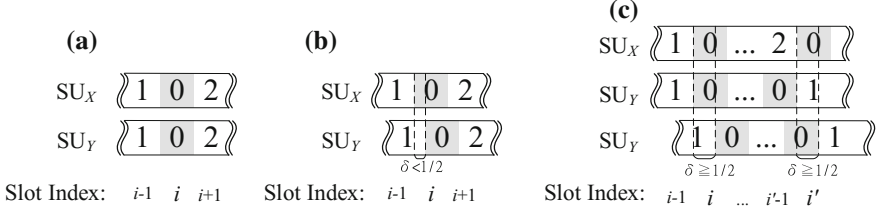
Fig. 4 Theorem 3. **a** Case where $\delta = 0$. **b** Case where $\delta < 1/2$. **c** Case where $\delta \geq 1/2$

m overlapping slots, the overlapping part of **U** and **V** $> m/2$ slots. In the second case, because $\mathcal{A}$ has bounded MTTR, **U** and **V** can rendezvous in slot i when local clock drift among the two sequences is k (i.e., $\mathbf{V} = \mathcal{R}(\mathbf{U}, k)$), where $k \in Z$. Besides, there exists a slot $i' = i + j$ such that **U** and **V** can rendezvous when time offset of them is $k + 1$ (i.e., $\mathbf{V} = \mathcal{R}(\mathbf{U}, k + 1)$). Clearly, when $\delta \geq 1/2$, their overlapping part $> 1/2$ slot in slot i', referred to Fig. 4c. Because there are at least m overlapping slots, the overlapping part of two sequences $> m/2$ slots. The theorem holds. □

According to Theorem 3, we know that if existing algorithms communication's slot is equal to half of CH system's slot, the algorithms which only operated in the scenarios, where SUs' slot boundaries are aligned, can also support the scenarios, where SUs' slot boundaries are misaligned. This is an important theorem. The algorithm is too complicated to design when the slot boundaries of SUs are misaligned. Below, we show another important concept which is used in later sections.

Theorem 4 *Let an $(i \times j)$-length circular queue be filled with an $(i \times j)$-length sequence. We randomly pick a started element of the $(i \times j)$-length circular queue and fill the elements of the circular queue into an $(i \times j)$ matrix in a row major manner. For any two $(i \times j)$ matrices M_1 and M_2 which may start from different elements of the circular queue, given an arbitrary column of M_1, $[e_1, e_2, ..., e_j]^T$, we can find a column which is $[e_{(1+c)} \bmod j, e_{(2+c)} \bmod j, ..., e_{(j+c)} \bmod j]^T$ in M_2, where $c \in Z$.*

Since the proof procedure is trivial, we omit here. We use the following example to explain Theorem 4. There is a 2×4-length circular queue and the queue is filled with 0 to 7, as Fig. 5a. The 2×4 matrix M_1 is constructed by the queue which starts from element 0, and the 2×4 matrix M_2 is constructed by the queue which starts from element 3, as Fig. 5b, c. We arbitrarily pick up a column of M_1, column 0. The column 0 is $[e_0 = 0, e_1 = 2, e_2 = 4, e_3 = 6]^T$ and we can find $c = 2$ (i.e., $[e_{(0+2)} \bmod 4, e_{(1+2)} \bmod 4, e_{(2+2)} \bmod 4, e_{(3+2)} \bmod 4]^T = [4, 6, 0, 2]^T$) in column 1 of M_2. In the other word, in the case, a column of matrices only contains all odd elements or all even elements in asynchronous environments. It is impossible to find the case where odd elements and even elements are filled in a single column. Theorem 4 is very powerful to design a hopping sequence which has bounded MTTR in asynchronous environment.

(a)

stared element
of (b)

(b)

e_1	0	1
e_2	2	3
e_3	4	5
e_4	6	7

(c)

3	4	e_3
5	6	e_4
7	0	e_1
1	2	e_2

stared element
of (c)
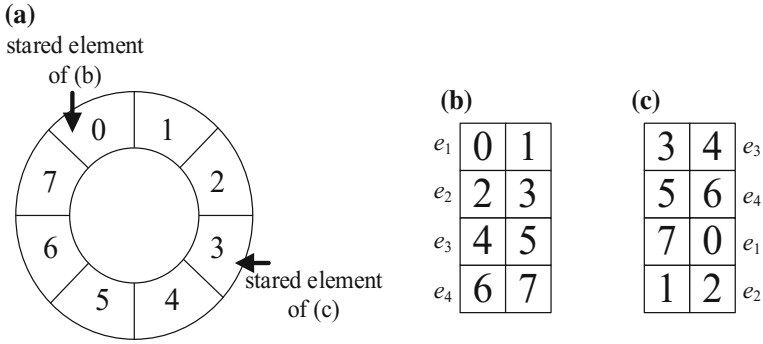
Fig. 5 Property of $(i \times j)$ matrix in row major manner

Symmetric Available Channel Model

Sometimes, the communication range of SUs may be much smaller than one of PUs. In such this scenarios, available channel set detected by SUs are almost same. Thus, some algorithms are developed to suit for the symmetric available channel model. In this section, we assume that SUs' common available channels are same and that the number of the channels is N_c.

Homogeneous Role Algorithms

In this section, the SUs perform the same channel hopping algorithm no matter the role of SUs (sender or receiver). Four algorithms, $\mathcal{A}_{Async-ETCH}$ [4], $\mathcal{A}_{GOS}$ [5], $\mathcal{A}_{DRSEQ}$ [6], and $\mathcal{A}_{A-QCH}$ [3] are introduced as follows.

Async-ETCH [4]

$\mathcal{A}_{Async-ETCH}$ includes $N_c - 1$ rounds. Each round i is divided into N_c parts. A part j has 1 pilot and 2 subsequences. Formally, part j sequence = (jth channel of subsequence i | subsequence i | subsequence i). A subsequence i is $(0 \times (i + 1)$, $1 \times (i + 1)$, $2 \times i + 1)$, …, $(N_c - 1) \times (i + 1))$ in Z_{Nc}, where $0 \leq i \leq N_c - 2$. Z_{Nc} denotes the integer set including element $\{0, 1, 2, …, (N_c - 1)\}$. If there exists an element of the subsequence $i \geq N_c$, the element i should be replaced to i module N_c. For example, when $N_c = 3$, $\mathcal{A}_{Async-ETCH}$ generates two subsequences $(0, 1, 2)$ and $(0, 2, 1)$, referred to Fig. 6a. Round 0 of $\mathcal{A}_{Async-ETCH}$ is (**0**, 0, 1, 2, 0, 1, 2, **1**, 0, 1, 2, 0, 1, 2, **2**, 0, 1, 2, 0, 1, 2,) containing 3 parts. The boldface letters are pilots. Other rounds can be derived in the same way, referred to Fig. 6b.

$\mathcal{A}_{Async-ETCH}$ has bounded MTTR $= (N_c - 2)(2N_c + 1)$. We consider two cases. In the first case, where the pilots (gray grids) of two SUs are misaligned, and the local slot offset of SU_X and SU_Y is 1 in Fig. 6c. Non-pilots shall stay in the same channel in different parts of a round, but pilots shall hop on different channels in

(a)

subsequence 0 | 0 | 1 | 2 |
subsequence 1 | 0 | 2 | 1 |

(b)

|◄---- part 0 ----►|◄---- part 1 ----►|◄---- part 2 ----►|

round 0 | 0 | 0 | 1 | 2 | 0 | 1 | 2 | 1 | 0 | 1 | 2 | 0 | 1 | 2 | 2 | 0 | 1 | 2 | 0 | 1 | 2 |

round 1 | 0 | 0 | 2 | 1 | 0 | 2 | 1 | 2 | 0 | 2 | 1 | 0 | 2 | 1 | 1 | 0 | 2 | 1 | 0 | 2 | 1 |

(c)

|◄------------------------------------- CH period -------------------------------------►|

|◄------------ round **0** ------------►|◄------------ round **1** ------------►|

Async-ETCH$_X$ 0 0 1 2 0 1 2 1 0 1 2 0 1 2 2 0 1 2 0 1 2 | 0 0 2 1 0 2 1 2 0 2 1 0 2 1 1 0 2 1 0 2 1 0

stay   stay   stay   stay   stay   stay

Async-ETCH$_Y$ 0 0 1 2 0 1 2 1 0 1 2 0 1 2 2 0 1 2 0 1 2 | 0 0 2 1 0 2 1 2 0 2 1 0 2 1 1 0 2 1 0 2 1

|◄------------ round **0** ------------►|◄------------ round **1** ------------►|

|◄------------------------------------- CH period -------------------------------------►|

Slot Index: 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42

(d)

|◄------------------------------------- CH period -------------------------------------►|

|◄------------ round **0** ------------►|◄------------ round **1** ------------►|

Async-ETCH$_X$ 0 0 1 2 0 1 2 1 0 1 2 0 1 2 2 0 1 2 0 1 2 | 0 0 2 1 0 2 1 2 0 2 1 0 2 1 1 0 2 1 0 2 1 0

↕↕↕↕↕↕ ↕↕↕↕↕↕ ↕ ↕   ↕ ↕   ↕ ↕

Async-ETCH$_Y$ 0 0 1 2 0 1 2 1 0 1 2 0 1 2 2 0 1 2 0 1 2 | 0 0 2 1 0 2 1 2 0 2 1 0 2 1 1 0 2 1 0 2 1 1

|◄------------ round **0** ------------►|◄------------ round **1** ------------►|

|◄------------------------------------- CH period -------------------------------------►|

Slot Index: 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42

Fig. 6 An example of $\mathcal{A}_{Async-ETCH}$. **a** Async-ETCH's subsequences. **b** Async-ETCH's rounds. **c** The pilots (*gray grids*) of two SUs are misaligned. **d** The pilots (*gray grids*) of two SUs are aligned

different parts of a round. Thus, SU$_X$'s non-pilots will rendezvous with SU$_Y$'s pilot. In the other case, where the pilots of two SUs are aligned, e.g., the local slot offset of SU$_X$ and SU$_Y$ is 7 in Fig. 6d. Because the pilots are aligned, subsequences of two SUs are aligned. Thus, they can rendezvous on the first channel of the subsequences (i.e., channel 0).

Below, we introduce $\mathcal{A}_{GOS}$. $\mathcal{A}_{GOS}$ has shorter MTTR (= $N_c (N_c + 1)$) than $\mathcal{A}_{Async-ETCH}$.

GOS [5]

A GOS sequence has a CH period $N_c (N_c + 1)$ slots. SU shall generate a permutation of common available channels. A GOS sequence includes N_c rounds. Each round i includes 1 leading bit and a permutation, where the leading bit is the ith bits of the permutation. For example, SU$_X$ selects a permutation of these 5 channels, (2, 1, 4, 0, 3), over 5 licensed channels. Thus, a GOS sequence includes 5 same permutations, (2, 1, 4, 0, 3). The first round has leading bit 2. In the same way with other rounds, the complete sequence of GOS is (**2**, 2, 1, 4, 0, 3) | (**1**, 2, 1, 4, 0, 3) | (**4**, 2, 1, 4, 0, 3) | (**0**, 2, 1, 4, 0, 3) | (**3**, 2, 1, 4, 0, 3), referred to Fig. 7a.

The MTTR of $\mathcal{A}_{GOS}$ is $N_c (N_c + 1)$. For example, when $N_c = 5$, the local time offset of SU$_X$ and SU$_Y$ is 2, referred to Fig. 7a. They can rendezvous on channel 1 in slot 8. In case, where two SUs' leading bits aren't aligned, the leading bit shall rendezvous a non-leading bit. The leading bit changes in each round but the non-leading bit keeps the same in each round. Thus, they can rendezvous, referred
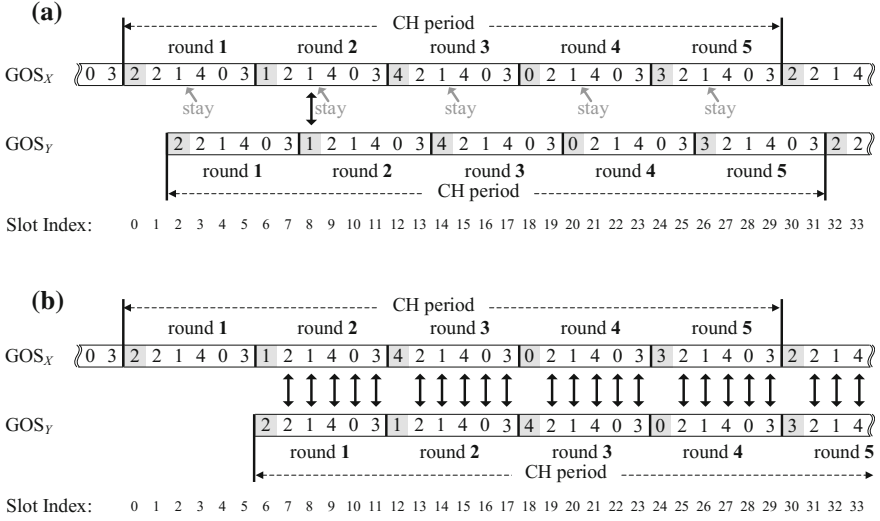
Fig. 7 An example of $\mathcal{A}_{GOS}$. **a** Slot offset of local slots of SU_X and SU_Y is 2. **b** Slot offset of local slots of SU_X and SU_Y is 6

to Fig. 7a. In the case, where two SUs' leading bits are aligned, they can rendezvous on non-leading bits, referred to Fig. 7b.

$\mathcal{A}_{DRSEQ}$ provides shorter MTTR ($= 2N_c + 1$) than $\mathcal{A}_{GOS}$. The $\mathcal{A}_{DRSEQ}$ is described below:

DRSEQ [6]

A DRSEQ sequence is generated by concatenating three subsequences, X_1, X_2, and X_3. $X_1 = (0, 1, \ldots, N_c - 1)$, $X_c = (N_c - 1, N_c - 2, \ldots, 0)$, and $X_3 = (0)$.[2] It's MTTR $= 2N_c + 1$. For example, referred to Fig. 8, when the number of SUs' common available channels $N_c = 5$. The SU_X generates DRSEQ sequence $DRSEQ_X = \{0, 1, 2, 3, 4, 4, 2, 1, 0, 0\}$. SU_Y generates its sequence $DRSEQ_Y$ in the same way. We can find that SUs can rendezvous in gray slots during $2 \times 5 + 1 = 11$ slots. Note that channels 0 to 5 are common available channels for two SUs. Thereby they SU_X and SU_Y exchange their data on the channels.

Actually, the optimal MTTR of homogeneous role algorithms in symmetric available channel model is 1. For example, let SU hop on fixed channel, the MTTR is 1. However, it is not a good algorithm because its degree of rendezvous is 1. Considering scenarios, when SUs send data at the same time on fixed common available channel, the channel will be blocked. Thus, the degree of rendezvous is also an important metric for algorithms. Now, we introduce $\mathcal{A}_{A-QCH}$ and its degree of rendezvous > 1.

---

[2]For ease of discussion, we modify DRSEQ sequence. The original sequence of DRSEQ is (X_1, X_3, X_2), and X_3 is a channel.

	Slot Index
DRSEQ$_X$	0 0 0 1 2 3 4 4 3 2 1 0 0 0 1 2 3 4 4 3 2 1 0 0 0
DRSEQ$_Y$	0 1 2 3 4 4 3 2 1 0 0 0 1 2 3 4 4 3 2 1 0 0 0
DRSEQ$_Y$	0 1 2 3 4 4 3 2 1 0 0 0 1 2 3 4 4 3 2 1 0 0
DRSEQ$_Y$	0 1 2 3 4 4 3 2 1 0 0 0 1 2 3 4 4 3 2 1 0
DRSEQ$_Y$	0 1 2 3 4 4 3 2 1 0 0 0 1 2 3 4 4 3 2 1
DRSEQ$_Y$	0 1 2 3 4 4 3 2 1 0 0 0 1 2 3 4 4 3 2
DRSEQ$_Y$	0 1 2 3 4 4 3 2 1 0 0 0 1 2 3 4 4 3
DRSEQ$_Y$	0 1 2 3 4 4 3 2 1 0 0 0 1 2 3 4 4
DRSEQ$_Y$	0 1 2 3 4 4 3 2 1 0 0 0 1 2 3 4
DRSEQ$_Y$	0 1 2 3 4 4 3 2 1 0 0 0 1 2 3
DRSEQ$_Y$	0 1 2 3 4 4 3 2 1 0 0 0 1 2
DRSEQ$_Y$	0 1 2 3 4 4 3 2 1 0 0 0 1
Slot Index:	0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22

Fig. 8 An example of $\mathcal{A}_{DRSEQ}$

A-QCH [3]

$\mathcal{A}_{A-QCH}$ is a cyclic quorum-based algorithm. Assume that there two common available channels 0 and 1 are in the system. SUs use A-QCH to create a two cyclic quorum systems $\mathcal{Q}_1$ and $\mathcal{Q}_2$ in the first step, where $Q_1 \in \mathcal{Q}_1$, $Q_2 \in \mathcal{Q}_2$, and $Q_1 \bigcap Q_2 = \varnothing$. For example, $\{0, 2, 3\}$ is a cyclic quorum in Z_6, and $\{1, 4, 5\}$ is also a cyclic quorum in Z_6. The two cyclic quorums are disjoint. In the second step, if SU is in its local slots which mod MTTR $\in Q_1$, the SU hops on the common available channel 0; otherwise the SU hops on the common available channel 1. The MTTR of $\mathcal{A}_{A-QCH}$ is maximal element $+ 1$ of $Q_1 \cup Q_2$.

Assume that SU_X and SU_Y generate two cyclic quorums Q_1 and Q_2, where $Q_1 = \{0, 2, 3\}$ and $Q_2 = \{1, 4, 5\}$. The MTTR of $\mathcal{A}_{A-QCH}$ is 6. For each SU_X's slot which mods $6 = 0$ or 2 or 3, SU_X hops on the channel 0. SU_X hops on the channel 1 in its local slots which mod $6 = 1$ or 4 or 5. If the SU_Y creates A-QCH sequence in the same way, they can rendezvous on channels 1 and 2 because $Q_1 \in \mathcal{Q}_1$ ($Q_2 \in \mathcal{Q}_2$) is a cyclic quorum, referred to Fig. 9.

Fig. 9 An example of $\mathcal{A}_{A-QCH}$

A-QCH$_X$	0 1 0 0 1 1 0 1 0 0 1 1
A-QCH$_Y$	0 1 0 0 1 1 0 1 0 0 1
A-QCH$_Y$	0 1 0 0 1 1 0 1 0 0
A-QCH$_Y$	0 1 0 0 1 1 0 1 0
A-QCH$_Y$	0 1 0 0 1 1 0 1
A-QCH$_Y$	0 1 0 0 1 1 0
Slot Index:	0 1 2 3 4 5 6 7 8 9 10 11

Asymmetric Available Channel Model

For developing the channel hopping algorithms to apply in general cases, the asymmetric available channel model becomes a popular research model. In asymmetric available channel issue, a SU performs spectrum sensing to determine its available (channels) set among N licensed channels, and hops on its available set to exchange data with other SUs.

Heterogeneous Role Algorithms

In heterogeneous role algorithms, the SUs are divided into two types, sender and receiver. Each type of SUs has its channel hopping algorithms. If two SUs act the same types, they don't have bounded MTTR in following algorithms.

In this section, we introduce two algorithms, $\mathcal{A}_{A-MOCH}$ [3] and $\mathcal{A}_{ARCH}$ [7]. Their MCTTRs are N^2, where N is the number of channels.

A-MOCH [3]

$\mathcal{A}_{A-MOCH}$ utilizes the *Latin square* and *identical-row square* to rendezvous on the same channel.

Definition 3 *A* Latin square *is a* $n \times n$ *matrix and each element* $e_{i,j} \in Z_n$. *Arbitrary column or row only exactly contains* n *different elements, respectively.*

Take a 3×3 Latin square as an example, referred to Fig. 10a. Column 0 has 3 different elements $\{1, 0, 2\}$ and row 2 also has 3 different elements $\{2, 1, 0\}$.
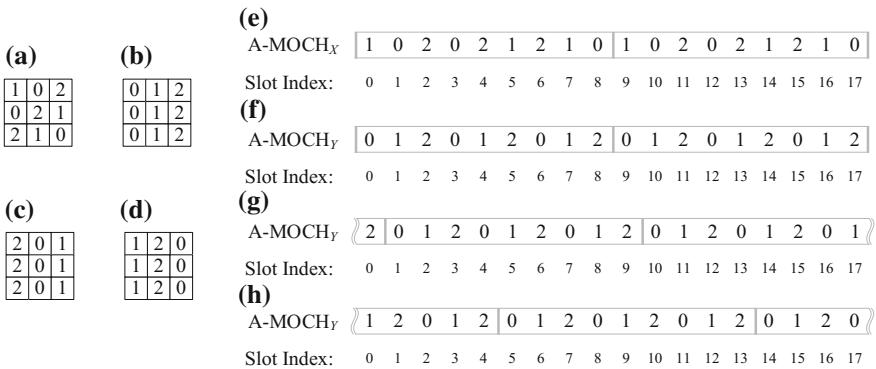

Fig. 10 The $\mathcal{A}_{A-MOCH}$. **a** Latin square of the sender. **b** Identical-row square of the receiver. **c** $N \times N$ identical-row square which starts at global slot 1. **d** $N \times N$ identical-row square starting from global slot 5. **e** A-MOCH$_X$ sequence starting from global slot 0. **f** A-MOCH$_Y$ sequence starting from global slot 0. **g** A-MOCH$_Y$ sequence generated by (**c**). **h** A-MOCH$_Y$ sequence generated by (**d**)

A *identical-row* square is a $n \times n$ matrix and each element $e_{i,j} \in Z_n$. Arbitrary two rows of the identical-row square are identical, referred to Fig. 10b.

The first step of $\mathcal{A}_{A-MOCH}$ is to create two squares, $N \times N$ Latin square and $N \times N$ identical-row square, where N is the number of licensed channels. A SU which acts as a sender (receiver) constructs its hopping sequence according to Latin square (identical-row square). In the second step, a sequence (i.e., a CH period) of a SU which acts as a sender (receiver) is generated by concatenating rows of Latin square (identical-row square). The MTTR of the $\mathcal{A}_{A-MOCH}$ is N^2.

Giving an example of $\mathcal{A}_{A-MOCH}$, let SU_X be a sender and SU_Y be a receiver. SU_X and SU_Y perform A-MOCH$_X$ sequence and A-MOCH$_Y$ sequence over 3 licensed channels, {0, 1, 2}, respectively. SU_X and SU_Y use the Latin square and identical-row square, shown in Fig. 10a, b, respectively. A-MOCH$_X$ sequence, shown in Fig. 10e, is generated through concatenating rows of square in Fig. 10a. In the same way, SU_Y generates its CH sequence A-MOCH$_Y$, referring to Fig. 10f. Clearly, they rendezvous on channels 2, 0, 1 in slots 2, 3, 7, respectively, during the MTTR (= 9 slots) of $\mathcal{A}_{A-MOCH}$ when they are synchronous. Because they can rendezvous on all channels, they can exchange their data if they have common available channels. Note that each column in Latin square intersects each column in an identical-row square. For example, column 0 of Fig. 10a, b are intersecting on element 0 (i.e., channel 0). An identical-row square has different elements in each column. Thus, two squares shall intersect on all elements. Besides, we consider the synchronous case. Assume that A-MOCH$_Y$ starts from global slot 1. They rendezvous on channels 0, 1, 2 in slots 1, 5, 6, respectively. This is not a lucky case. As mentioned into Theorem 4, any starting clock of an A-MOCH$_Y$, referred to Fig. 10f, g, can be filled with an $N \times N$ identical-row square, referred to Fig. 10b, c. Thus, two SUs can rendezvous. Considering another example, when SU_Y starts from global slot 5, the square is also an identical-row square, referred to Fig. 10d, h. Thus, two SUs can rendezvous all channels (because a Latin square always maps to an identical-row square).

ARCH [7]

In $\mathcal{A}_{ARCH}$, each SU constructs a sequence $\mathbf{U} = (0, 1, 2, \ldots, N-1)$. SU_X (as a sender) and SU_Y (as a receiver) choose randomly channels, respectively. Assume that SU_X and SU_Y pick up the channels s and r, respectively. SU_X generates its hopping sequence termed ARCH$_X$ = ($\mathbf{U}$, s) | ($\mathbf{U}$, (s + 1) mod N) | ... | ($\mathbf{U}$, (s + N − 1) mod N), and SU_Y creates its hopping sequence ARCH$_Y$, which is N copies $\mathcal{R}(\mathbf{U}, r)$s. Clearly, the length of ARCH$_X$ and ARCH$_Y$ are N^2. For example, SU_X and SU_Y pick up initial channels 2 and 1 over 3 licensed channels, respectively. Thus, the sender's sequence is (2, 0, 1, 0, 1, 2, 1, 2, 0). The receiver's sequence is (1, 2, 0, 1, 2, 0, 1, 2, 0), referred to Fig. 11a.

Obviously, the $\mathcal{A}_{ARCH}$'s MTTR is N^2. It is not difficult to find that SU_Y hops on the same channel every N slots, and SU_X changes its channel every N slots until the SU_X changes all channels once. Thus, SU_X and SU_Y can rendezvous on their common available channels in NN-slots. For example, when the common available channel is 1, referred to Fig. 11b, the sender hops to all channels in slots 1, 4, 7, and
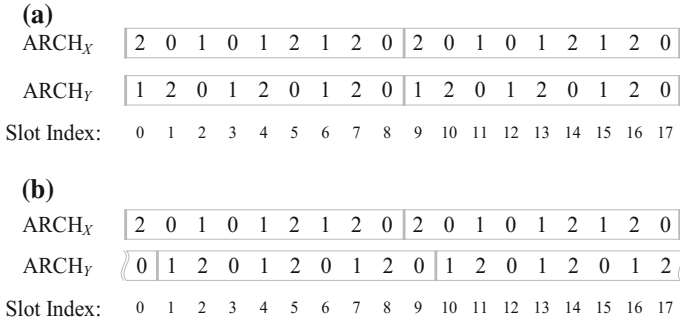
(a)

| ARCH$_X$ | 2 | 0 | 1 | 0 | 1 | 2 | 1 | 2 | 0 | 2 | 0 | 1 | 0 | 1 | 2 | 1 | 2 | 0 |

| ARCH$_Y$ | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 |

Slot Index: 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17

(b)

| ARCH$_X$ | 2 | 0 | 1 | 0 | 1 | 2 | 1 | 2 | 0 | 2 | 0 | 1 | 0 | 1 | 2 | 1 | 2 | 0 |

| ARCH$_Y$ | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 |

Slot Index: 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17

Fig. 11 The $\mathcal{A}_{ARCH}$. **a** ARCH$_Y$ sequence which starts from the global slot 0. **b** ARCH$_Y$ sequence which starts from the global slot 1

receiver stays on the same channel 1 in slots 1, 4, 7. Thus, they can exchange their data in slot 4.

We concern the limitation of MTTR of algorithms in asymmetric available channel model and asynchronous model. Actually, the limitation of MTTR is N^2

Theorem 5 *The MTTR of any algorithms in asymmetric available channel model and asynchronous model is at least N^2* [3].

Proof Assume that algorithm $\mathcal{A}_u$ and algorithm $\mathcal{A}_v$ generate sequences **U** and **V**, respectively, where $\mathcal{A}_u$ may be equal to $\mathcal{A}_v$. Additionally, we assume that the rendezvous channel of the $\mathcal{A}_u$ and $\mathcal{A}_v$ is h and MTTR $= \alpha$. Let h^U collect slot index of channel h, and h^V collect slot index of channel h. The elements in h^U and h^V should mod. Clearly, h^U and h^V must be a quorum in Z_α. According to [8], $|h^V| \geq \sqrt{T}$. Arbitrary channel may be a common available channel, thus $T = \sum_0^{N-1} |h^v| \geq N\sqrt{T}$. Therefore, MTTR $T \geq N^2$. □

As the above mentions in Theorem 5, $\mathcal{A}_{A-MOCH}$ and $\mathcal{A}_{ARCH}$ have optimal MTTR.

Homogeneous Role Algorithms

Although the heterogeneous role algorithms have lower MTTR than homogeneous role algorithms, they are impractical in ad hoc networks. In ad hoc networks, roles of SUs usually cannot be determined in a centralized way. In homogeneous role algorithms, SUs don't choose roles (because they perform the same algorithms) and they have bounded MTTR. In this section, homogeneous role algorithms are introduced. In $\mathcal{A}_{AHW}$ [9], authors used SUs' IDs to generate channel hopping sequences.
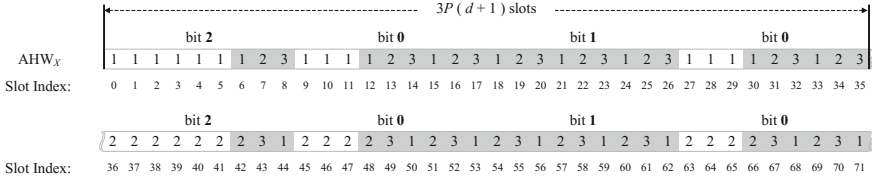
Fig. 12 An example of $_{AHW}$ when SU$_X$'s ID = 010

AHW [9]

$\mathcal{A}_{AHW}$ has three elementary sequences, bit 0 (wait-hop-hop pattern), bit 1 (hop-hop-hop pattern), and bit 2 (wait-wait-hop pattern). An elementary sequence includes $3P$ elements. The target channel of the elementary sequence is depended on an initial value α and a jump value r, where $1 \leq r \leq P$ and $1 \leq a \leq N$. For example, assume that a SU determines $a = 1$, $r = 1$, and $N = P = 3$. The bit 0 of the SU includes $P = 3$ identical channels ($r = 1$, $r = 1$, $r = 1$), and two hop patterns ($a = 1$, $a + r = 1$, $a + 2r = 3$). In the same way, SU generates its sequence for bit 0, bit 1, and bit 2. Each SU has unique id, maybe a 48-bits MAC address, and adds one bit, called bit 2, before its bits. Considering Fig. 12, SU$_X$'s ID is 010, and then adds bit 2 before 010. The newly ID of SU$_X$ is 2010. SU$_X$ can use the newly ID to generate its hopping sequence according to three elementary sequences. Clearly, a complete ID includes $3P(d + 1)$ slots, where d is the number of original ID bits of a SU. After $3P(d + 1)$ slots, the value a should be added 1. SU generates another $3P(d + 1)$ slots in the same way. After SU repeats at most N rounds the process, SUs can rendezvous on a common available channel.

The algorithm based on ID has drawback. Since SUs' unique IDs are usually MAC addresses (48 bits). The algorithm has long MTTR. Below, we introduce non-ID based algorithms, $\mathcal{A}_{CRSEQ}$ [2] and $\mathcal{A}_{A-QCH}$, [10] as following.

CRSEQ [2]

A CH period of $\mathcal{A}_{CRSEQ}$ is $(3p - 1)\,p$ timeslots, where p is the smallest prime greater than or equal to the number of licensed channels N. The CRSEQ sequence consists of p $(3p - 1)$-length subsequences. In the first $(2p - 1)$ slots of $(3p - 1)$-length subsequence, a SU jumps on different channel sequentially. A SU stays on a same channel in the remaining p slots. The starting channel of the first $(2p - 1)$ slots in the ith $(3p - 1)$-length subsequence is determined by triangular numbers, $T_i + 1 = \frac{i(i + 1)}{2} + 1$, where $i = 0, 1, 2, \ldots$. For the remaining p slots in the ith $(3p - 1)$-length subsequence, a SU stays on the same channel i. Take Fig. 13 as an
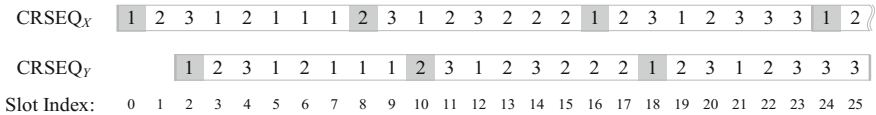

Fig. 13 CRSEQ sequence when $N = 3$

example. Assume that there are 3 licensed channels 1, 2, and 3. SU_X generates the hopping sequence $CRSEQ_X$. The first element of the $(3p-1)$-length subsequences is a triangular number (i.e., gray grids). In the first $(2p-1)$ slots, SUs jump on different channels (1, 2, 3, 1, 2). In the remaining p slots, SUs stay on the same channels (1, 1, 1). $\mathcal{A}_{CRSEQ}$ has bounded MTTR = $(3p-1)\,p$. For example, referred to Fig. 13, the local slot offset of SU_X and SU_Y is 2. They can rendezvous on channels 1, 2, 3, in slots 7, 15, 23, respectively.

DRDS [10]

Recall $\mathcal{A}_{A-QCH}$, the concept of $\mathcal{A}_{DRDS}$ is similar to $\mathcal{A}_{A-QCH}$ (mentioned in section "Homogeneous Role Algorithms"). In the asymmetric available channel model, an arbitrary licensed channel may be an unavailable channel. Therefore, $\mathcal{A}_{DRDS}$ extends the $\mathcal{A}_{A-QCH}$ to make all licensed channels can rendezvous once in the bounded timeslots. In $\mathcal{A}_{DRDS}$, authors construct different sets to generate rendezvous property. The definition of different set had been introduced in Theorem 1. Clearly, a different set $\mathcal{D}_i$ is a cyclic quorum. According to Theorem 2, we know that if a channel hopping algorithm has bounded MTTR, then the slots of a rendezvous channel i is a cyclic quorum. Consequently, we can construct an algorithm, which has bounded MTTR, in the asymmetric available channel model if N disjoint cyclic quorums can be found.

$\mathcal{A}_{DRDS}$ provides a way to generate p different sets in Z_{3p^2}. The MTTR of the algorithm can be bounded in $3p^2$, where p is the smallest prime lager or equal to N, referred to Algorithm 1. Now, according to Algorithm 1, we show an example when $N = 3$, which is a prime. The Algorithm 1 generates 3 different sets $\mathcal{D}_0$, $\mathcal{D}_1$, and $\mathcal{D}_2$ in Z_{27}. $\mathcal{D}_0 = \{0, 1, 2, 3, 6, 13, 16, 22, 25\}$, $\mathcal{D}_1 = \{5, 8, 9, 10, 11, 12, 15, 21, 24\}$, and $\mathcal{D}_2 = \{4, 7, 14, 17, 18, 19, 20, 23, 26\}$. After generating p different sets, a SU should hop on channel i when the SU in its local slot $t \bmod (3p^2) \in \mathcal{D}_i$. Take Fig. 14 as an example. There are 3 licensed channels 0, 1, 2. After applying $\mathcal{A}_{DRDS}$, SU_X hops on the channel 0 in its local slots 0, 1, 2, 3, 6, 13, 16, 22, 25, 27 (mod 27 = 0), 28 (mod 27 = 1), … (i.e., gray grids), which are based on $\mathcal{D}_0$, and hops on the channel 1 in its local slots 5, 8, 9, 10, 11, 12, 15, 21, 24, … (i.e., blue grids), which are based on $\mathcal{D}_1$. Since $\mathcal{D}_0$, $\mathcal{D}_1$, and $\mathcal{D}_2$ are cyclic quorums, any two SUs can rendezvous on channels 0, 1, 2 no matter that SUs are synchronous or asynchronous. For example as illustrated in Fig. 14, when the local clock drift of SU_X
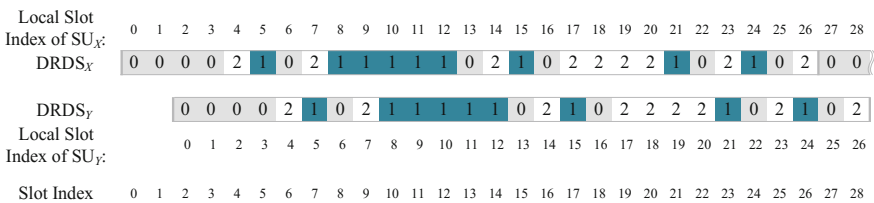

Fig. 14 An example of $\mathcal{A}_{DRDS}$ when $N = 3$

and SU_Y is 2, they can rendezvous on channels 0, 1, 2 in global slots 3, 11, 20, respectively.

Algorithm 1 Generate p different sets in Z_{3p^2}

 1: $S = \emptyset$
 2: **for** ($i = 1$ to $p - 1$) **do**
 3:      Let different set $\mathcal{D}_i = Z_{3p_i+p} \setminus Z_{3p_i}$;
 4:      **for** ($i = 1$ to $p - 1$) **do**
 5:          Let $q_j = j^2$ and $r_{i,j} = \dfrac{(i-q_j)(p+1)}{2}$ mod p;
 6:          Let $t_{j0} = 3pj + p + r_i j$
 7:          Let $t_{j1} = 3pj + 2p + r_i j$
 8:          $\mathcal{D}_i$ collects t_{j0} and t_{j1}
 9:      **end for**
10: **end for**
11: **return** p different sets $\mathcal{D}_0$ to $\mathcal{D}_p$;

The MTTR of above two algorithms are not associated the amount of common available channels. Now we introduce $\mathcal{A}_{JS}$ [1], which can reduce its MTTR when the number of common available channels increases.

JS [1]

A JS hopping sequence consists of several $3P$-length subsequences. A $3P$-length subsequence includes two P-length jump patterns and one P-length stay pattern, where P is the smallest prime larger than N. A jump pattern is a sequence $(a, ((a + r - 1)$ mod $P)$ mod $N + 1$, …, $((a + 2r - 1)$ mod $P)$ mod $N + 1$, …, $((a + (N - 1)r - 1)$ mod $P)$ mod $N + 1)$ and a stay pattern is a sequence $(r, r, …, r)$. In the first $3P$-length sub-sequence, a and r are randomly picked by SUs from range $[0, P - 1]$ and $[1, N]$, respectively. r increases 1 every $3P$ slots. If $r > N$, the value of r mods N. The value of a increases 1 every N $3P$-length sub-sequences, referred to Fig. 15. Finally, if a channel is an unavailable channel in $\mathcal{A}_{JS}$, the SU randomly chooses an available channel to replace the unavailable channel.

We give an example to explain the operation of $\mathcal{A}_{JS}$ over 6 licensed channels. Assume that the initial values of a, r, and P are 3, 2, and 7, respectively. According above description of $\mathcal{A}_{JS}$, the first $3P$-length subsequence is (3, 5, 1, 2, 4, 6, 1) | (3, 5, 1, 2, 4, 6, 1) | (2, 2, 2, 2, 2, 2, 2). The second $3P$-length subsequence is (3, 6, 2, 5, 1, 4, 1) | (3, 6, 2, 5, 1, 4, 1) | (3, 3, 3, 3, 3, 3, 3). Now we further assume that the SU_X's available set $A_X = \{1, 2, 6\}$. The unavailable channels in the JS sequence will be replaced by channels 1, 2 and 6. The first $3P$-length subsequence may be further modified to be (6, 6, 1, 2, 6, 6, 1) | (1, 2, 1, 2, 2, 6, 1) | (2, 2, 2, 2, 2, 2, 2), referred to Fig. 16. The MTTR of $\mathcal{A}_{JS}$ is $3NP(P - N_c) + 3P$.

In the following section, we will introduce another algorithm, $\mathcal{A}_{FRCH}$, whose MTTR ($= N(2N + 1)$) is very close to the optimal MTTR ($= N^2$).
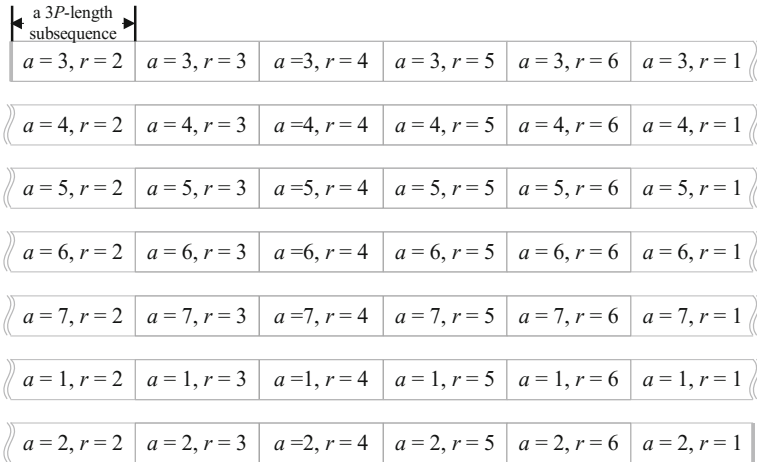
a=3, r=2	a=3, r=3	a=3, r=4	a=3, r=5	a=3, r=6	a=3, r=1
a=4, r=2	a=4, r=3	a=4, r=4	a=4, r=5	a=4, r=6	a=4, r=1
a=5, r=2	a=5, r=3	a=5, r=4	a=5, r=5	a=5, r=6	a=5, r=1
a=6, r=2	a=6, r=3	a=6, r=4	a=6, r=5	a=6, r=6	a=6, r=1
a=7, r=2	a=7, r=3	a=7, r=4	a=7, r=5	a=7, r=6	a=7, r=1
a=1, r=2	a=1, r=3	a=1, r=4	a=1, r=5	a=1, r=6	a=1, r=1
a=2, r=2	a=2, r=3	a=2, r=4	a=2, r=5	a=2, r=6	a=2, r=1

a 3P-length subsequence

Fig. 15 An example of value a and r of $\mathcal{A}_{JS}$

| JS$_X$ | 6 | 6 | 1 | 2 | 6 | 6 | 1 | 1 | 2 | 1 | 2 | 2 | 6 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| Slot Index: | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |

| | 2 | 6 | 2 | 1 | 1 | 2 | 1 | 2 | 6 | 2 | 6 | 1 | 6 | 1 | 2 | 6 | 2 | 1 | 2 | 1 | 1 |
| Slot Index: | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 |
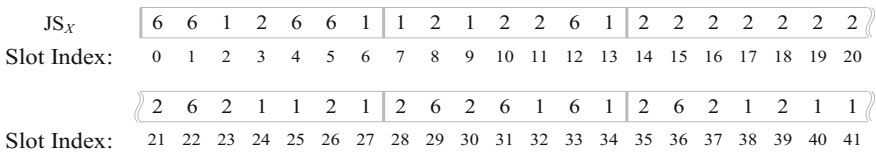
Fig. 16 An example of $\mathcal{A}_{JS}$

FRCH [11]

In $\mathcal{A}_{FRCH}$, authors proposed a frequency hopping algorithm based on $\mathcal{A}_{DRSEQ}$. In the first step of $\mathcal{A}_{FRCH}$, SU$_X$ generates a DRSEQ sequence, termed as DRSEQ$_X$. The DRSEQ$_X$ is concatenated by three subsequences, X_1, X_2, and X_3, where $X_1 = (0, 1, \ldots, N-1)$, $X_2 = (N-1, N-2, \ldots, 0)$ and $X_3 = (0)$. For example, the DRSEQ$_X$ = (0, 1, 2, 3, 4, 3, 2, 1, 0, 0) when $N = 5$. A CH period of FRCH includes N rounds DRSEQ sequences, i.e., $N(2N+1)$ slots, referred to Fig. 17a. In the second step, SU$_X$ determines its available set A$_X$ by means of scanning all licensed channels. If all channels are available, SU$_X$ does not replace any channels, referred to Fig. 17a. Otherwise, SU$_X$ replaces each unavailable channel of SU$_X$ in DRSEQ$_X$ with an available channel of SU$_X$. For example, the available set of SU$_X$ is {0, 3, 4} and unavailable channels of the DRSEQ$_X$ of SU$_X$ in the first round are replaced by channel 0. Unavailable channels in DRSEQ$_X$ of SU$_X$ are replaced with available channels 3 and 4 in the second round and the third round, respectively, referred to Fig. 17b.

The MTTR of $\mathcal{A}_{FRCH}$ is $N(2N+1)$. Suppose that SU$_X$'s available set $A_X = \{0, 3, 4\}$ and SU$_Y$'s available set $A_Y = \{2, 3, 4\}$ over $N = 5$. The local slot drift of two SUs is 5. We can find that two SUs can rendezvous on their common available
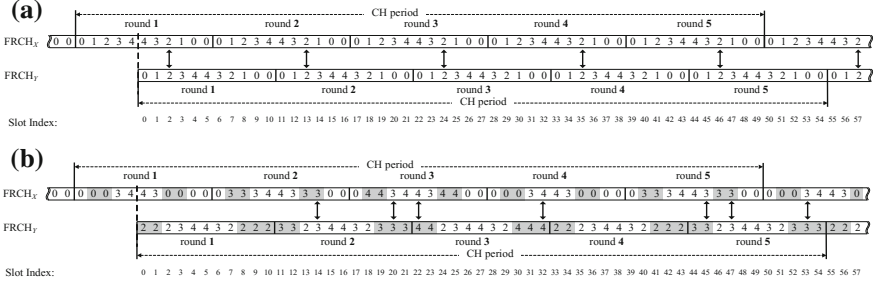
Fig. 17 Sequences of $\mathcal{A}_{FRCH}$ when $N = 5$. **a** FRCH$_X$ and FRCH$_Y$ of SU$_X$ and SU$_Y$ when channels are available. **b** FRCH$_X$ and FRCH$_Y$ of SU$_X$ and SU$_Y$ when the available sets of SU$_X$ and SU$_Y$ are {0, 3, 4} and {2, 3, 4}, respectively

channels 3 and 4, referred to Fig. 17b. Note that gray grids shown in Fig. 17b mean the timeslots which are visiting unavailable channels before replacement. Thus, the same ith slot (i.e., slot i, $i + (2N + 1)$, $i + 2(2N + 1)$, …, $i + N(2N + 1)$) marked as gray grids in Fig. 17b shall be replaced with different available channels for every round. If its target SU stays in the same available channels (white girds) in each round, they can rendezvous. Figure 17b shows the example. Because channel 1 is unavailable for SU$_X$ in time slot 3, the channel shall be replaced with an available channel. SU$_X$ shall replace it with a different available channel every round ($(2 \times 5 + 1) = 11$ slots). Thus, in slots 3, 14, 25, 36, 47, SU$_X$ shall hop all available channels once. If SU$_Y$ stays on the same available channel 3 in slots 3, 14, 25, 36, 47, they rendezvous on common available channel 3 in slots 14 and 47. If there exists an unavailable for SUs, two sequences always have these "gray-white" grids pairs when $N \neq (5r - 1)/2$, where $r \geq 3$, such that they can rendezvous on common available channels.[3] If no existing an unavailable for SUs, they rendezvous on at least one channel because $\mathcal{A}_{DRSEQ}$ makes SUs rendezvous on at least one channel, referred to section "Homogeneous Role Algorithms".

T-CH [12]

A T-CH sequence is generated by concatenating rows in the special $\left(2p + \left[\frac{P}{2}\right]\right) \times N$ N triangular number matrix M sequentially, where p is the smallest prime which is greater or equal to N. The detail of the triangular number matrix M is described below. A triangular number matrix M contains two kinds of columns: *jumping column* and *staying column*. Each *jumping column* is filled with N distinct elements, while each *staying column* is filled with the same element.

- Jumping columns: In a triangular number matrix M, column 0 to column $\left(N + \left[\frac{N}{2}\right] - 1\right)$ are jumping columns. Jumping columns are generated by the aid of triangular numbers $T_i = \frac{i(i + 1)}{2}$, where $i \in Z$. For each jumping column j,

---

[3]For ease of discussion, we assume that $\alpha = 0$, referred to Table 2.

where $0 \leq j \leq N + \left[\frac{N}{2}\right] - 1$, it is filled with N elements, $(T_j \bmod N)$, $((T_j + 1) \bmod N)$, …, and $((T_j + N - 1) \bmod N)$.

- Staying columns: In a triangular number matrix M, column $N + \left[\frac{N}{2}\right]$ to column $\left(2N + \left[\frac{N}{2}\right] - 1\right)$ are staying columns. For each staying column j with $j = N + \left[\frac{N}{2}\right] + i$ and $0 \leq i \leq N - 1$, it has N elements with the identical value i.

We give an example to show a jumping column and a staying column of a triangular number matrix M. When $N = 5$, column 0 to 6 are jumping columns and column 7 to 11 are staying columns. Column 0 of M is filled with 0 ($= T_0 \bmod 5 = \frac{0(0+1)}{2} \bmod 5$), 1 ($= T_0 + 1 \bmod 5$), 2 ($= T_0 + 2 \bmod 5$), 3 ($= T_0 + 3 \bmod 5$), 4 ($= T_0 + 4 \bmod 5$). Column 1 of M is filled with 1 ($= T_1 \bmod 5 = \frac{1(1+1)}{2} \bmod 5$), 2 ($= T_1 + 1 \bmod 5$), 3 ($= T_1 + 2 \bmod 5$), 4 ($= T_1 + 3 \bmod 5$), 0 ($= T_1 + 4 \bmod 5$). Other 6 jumping columns can be constructed in the same way, referred to Fig. 18a. In the staying column, column 7 $\left(= N + \left[\frac{N}{2}\right] + 0\right)$ is filled with 0, referred to Fig. 18a.

After constructing the triangular matrix M, each SU generates the T-CH sequence by concatenating rows in matrix M in Fig. 18a sequentially, referred to Fig. 18d. The CH period is equal to the triangular number matrix's size, $\left(2N + \left[\frac{N}{2}\right]\right) \times N$. Actually, the MTTR of T-CH is also equal to $\left(2N + \left[\frac{N}{2}\right]\right) \times N$ because T-CH can rendezvous on all channels during the $\left(2N + \left[\frac{N}{2}\right]\right) \times N$ slots.

Assume that there are two SUs, SU_X and SU_Y, exchanging their data, and SU_X (SU_Y) performs the sequence T-CH$_X$ (T-CH$_Y$) according to the triangular number matrix M_X (M_Y). According to Theorem 4, no matter what amount of the time shifts between the starting time of T-CH$_X$ and T-CH$_Y$, there exists an execution time of column $i + c$ of M_Y such that it overlaps the execution time of column i of M_X for all i. In case $c > \left[\frac{N}{2}\right]$, there exists at least N columns such that column i of M_X and
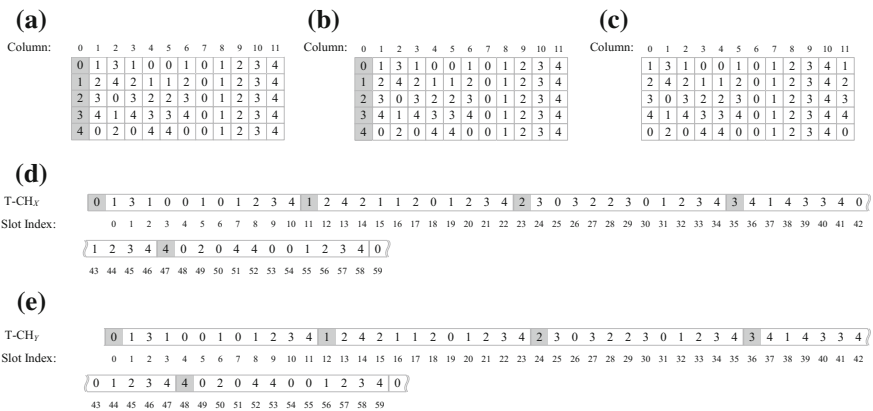

Fig. 18 $\mathcal{A}_{T-CH}$'s triangular matrix M and sequence when $N = 5$. **a** Triangular matrix M_X of SU$_X$. **b** Triangular matrix M_Y of SU$_Y$. **c** Delay one slot for triangular matrix M_X. **d** T-CH$_X$ sequence. **e** T-CH$_Y$ sequence

column $i + c$ of M_Y are not jumping columns or staying columns. Therefore they can rendezvous on all channels. In case $c \leq \left[\frac{N}{2}\right]$, there exists an column $i + c$ of M_Y such that it and column i of M_X are identical because the first $N + \left[\frac{N}{2}\right]$ columns are triangular numbers.

Take Fig. 18d, e as an example. Assume that local slot offset of between SU_X and SU_Y is 1. They can rendezvous on channels 0, 1, 2, 3, and 4 in global slots 4, 16, 28, 40, and 52, respectively. If the available sets of SU_X and SU_Y are $\{0, 1, 2, 3\}$ and $\{3, 4\}$, respectively, they can rendezvous on common available channel 3 in global slot 40.

Summary

Now we compare dynamic spectrum access algorithms without common control channels and with fixedly common control channels. There are three potential solutions for constructing fixedly common control channels, *dedicated common control channel* and *ultra wideband*. In the first solution (the dedicated common control channels), common control channels are usually allocated to unlicensed channels, unused bands, or the licensed channels which are dedicated to SUs. If common control channels are allocated to unlicensed channels, transmissions on unlicensed channels may be inferred from other devices, e.g., WIFI or blue-tooth. Additionally, unused bands in licensed channels (e.g., guard bands) are proposed [13], called O-CCC. However, the O-CCC can only provide few bandwidth. When the number of SUs increases, the CRNs will congest. In the second solution, messages are separated on several bands in very low power. Thus, PUs don't be interfered by SUs' control messages. Clearly, SUs' communication range are limited due to low power. It's a fetal problem. Two SUs can exchange their data but cannot exchange their control messages when the distance of the SUs is long.

Fixedly common control channels can provide timely messages exchange but they have the above problems. Some emergency networks with low bandwidth requests suit for common control channels which is based on above dedicated common control channel or ultra wideband. However, in the networks, they usually have dedicated channels to transmit their data. That is to say that they don't use PUs' licensed channels. SUs in CRNs usually cannot have dedicated channels or have large number of entries; therefore they should exchange their data by hopping algorithms.

We have summarized recent channel hopping algorithms in Table 2. The algorithms are classified into Homogeneous Role Algorithms (HoR Algo. for short), Heterogeneous Role Algorithms (HeR Algo. for short), Symmetric Available Channel Model (SAC Model for short), and Asymmetric Available Channel Model (AAC Model for short).

Table 2 shows that the shortest MTTR of HeR Algo. in AAC model is N^2 slots and that of HoR Algo. in AAC model is $N(2N + 1)$ slots. Most algorithms have

Table 2 Asynchronous symmetric-role channel hopping algorithms under asymmetric-available-channel model

Algorithm		Deg. of rendezvous	MTTR
HoR Algo. in SAC Model	Asyn-ETCH [4]	1	$(N-2)(2N+1)$
	GOS [5]	1	$N(N+1)$
	DRSEQ [6]	1	$2N+1$
	A-QCH [3]	2	Maximal element +1 of $(Q_1 \cup Q_2)$
	SeqR [22]	1	$N(N+1)$
HeR Algo. in AAC Model	A-MOCH [3]	N	N^2
	ARCH [7]	N	N^2
HoR Algo. in AAC Model	FRCH [11]	1	$N(2N+1)^{\dagger}$
	T-CH [12]	N	$2p^2 + \left[\frac{p}{2}\right] \times p$
	JS [1]	N	$3NP(P-N_c)+3P \geq 3N^3-3N^2+3N$
	EJS [23]	N	$4P(P+1-N_c)$
	CRSEQ [2]	N	$3p^2-p^{*}$
	DRDS [10]	N	$3p^2$
	AHW [9]	G_{min}	$3p(d+1)(K+1-N_c) \geq (d+1)(1.5N^2)$
	HH [19]	N	$3p^2$
	SARCH [7]	N	$8N^2 + 8N$
	Sym-ACH [24]	N	$6dN^2$
	D-CH [12]	N	$(d+1)(N^2+N)$

$\dagger$ $N \neq ((5 + 2\alpha) \times r - 1)/2$ for all integer $\alpha \geq 0$ and odd integer $r \geq 3$. p the smallest prime which is larger or equal to N. P the smallest prime which is larger than N. N_c the number of common available channels; G_{min} the minimum number of common available channels between two SUs. d the length of the ID string of a SU; K the size of minimum available channel set. Q_1, Q_2 cyclic quorums

optimal degree of rendezvous to avoid serious collision. But there are still many research issues in spectrum access algorithm.

- *Security problem*. Most sequences are construed by deterministic channel hopping algorithms. Therefore the jammers can jam the target SUs easily [14–18].
- Muti-channel parallel transmission: Most recent algorithms are based on the transmission on a single channel in a slot. However, more and more devices transmit on muti-channels simultaneously.
- *K-synchronous scenarios*. Most existing algorithms are designed for synchronous or asynchronous model. However, when devices are equipped with GPS or use IEEE 1588 protocol, they can support synchronous in a bounded error K, called K-synchronous scenarios. Clearly, the MTTR of algorithms can be future reduced in this model.

- *Different sensing capabilities for SUs*. In real CRNs, SUs may have different capabilities. Thus, their sensing channels can be different. Designing an algorithm which has bounded MTTR for this scenario is an important issue [19–21].

In this chapter, we have introduced symmetric/asymmetric available channel models, heterogeneous/homogeneous role algorithms. The algorithms have bounded MTTR to avoid long time delay in CRNs. State of the art have nearly optimal MTTR. Thus, there may only be a marginal benefit to reducing MTTR of existing algorithms in current models. However, the above mentioned three challenges need to be investigated in the future.

References

1. Liu, H., Lin, Z., Chu, X., Leung, Y.-W.: Jump-stay rendezvous algorithm for cognitive radio networks. IEEE Trans. Parallel Distrib. Syst. **23**(10), 1867–1881 (2012)
2. Shin, J., Yang, D., Kim, C.: A channel rendezvous scheme for cognitive radio networks. IEEE Commun. Lett. **14**(10), 954–956 (2010)
3. Bian, K., Park, J.-M., Chen, R.: Control channel establishment in cognitive radio networks using channel hopping. IEEE J. Sel. Areas Commun. **29**(4), 689–703 (2011)
4. Zhang, Y., Li, Q., Yu, G., Wang, B.: ETCH: Efficient channel hopping for communication rendezvous in dynamic spectrum access networks. In: IEEE INFOCOM, 2011 (2011)
5. Theis, N.C., Thomas, R.W., DaSilva, L.A.: Rendezvous for cognitive radios. IEEE Trans. Mobile Comput. **10**(2), 216–227 (2010)
6. Yang, D., Shin, J., Kim, C.: Deterministic rendezvous scheme in multichannel access networks. IET Elec. Lett. **46**(20), 1402–1404 (2010)
7. Chang, G.-Y., Teng, W.-H., Chen, H.-Y., Sheu, J.-P.: Novel channel-hopping schemes for cognitive radio networks. IEEE Trans. Mobile Comput. **13**(2), 407–421 (2014)
8. Jiang, J.-R., Tseng, Y.-C., Hsu, C.-S., Lai, T.-H.: Quorum-based asynchronous power-saving protocols for IEEE 802.11 ad hoc networks. ACM Mobile Netw. Appl. **10**(1–2), 169–181 (2005)
9. Chuang, I.-H., Wu, H.-Y., Lee, K.-R., Kuo, Y.H.: Alternate hop-and-wait channel rendezvous method for cognitive radio networks. In: IEEE INFOCOM, 2013 (2013)
10. Gu, Z., Hua, Q.-S., Wang, A.Y., Lau, F.C.: Nearly optimal asynchronous blind rendezvous algorithm for cognitive radio networks. In: IEEE SECON, 2013 (2013)
11. Chang, G.-Y., Huang, J.-F.: A fast rendezvous channel-hopping algorithm for cognitive radio networks. IEEE Commun. Lett. **17**(7), 1475–1487 (2013)
12. Chang, G.-Y., Huang, J.-F., Wang, Y.S.: Matrix-based channel hopping algorithms for cognitive radio networks. IEEE Trans. Wirel. Commun. **14**(5), 2755–2768 (2015)
13. Chowdhury, K.R., Akyldiz, I.F.: OFDM-based common control channel design for cognitive radio ad hoc networks. IEEE Trans. Mobile Comput. **10**(2), 228–238 (2011)
14. Huang, J.-F., Chang, G.-Y., Huang, J.-X.: Anti-jamming Rendezvous scheme for cognitive radio networks. IEEE Trans. Mobile Comput. (in press) (2016)
15. Chang, G.-Y., Huang, J.-F., Wu, Z.-H.: A frequency hopping algorithm against jamming attacks under asynchronous environments. In: IEEE GLOBECOM, 2014 (2014)
16. Lee, E.-K., Oh, S., Gerla, M.: Timely and robust key establishment under jamming attack in critical wireless networks. In: IEEE MILCOM, 2011 (2011)
17. Strasser, M., Capkun, S., Capkun, S., Cagalj, M.: Jamming-resistant key establishment using uncoordinated frequency hopping. In: IEEE S&P, 2008 (2008)

18. Wang, Q., Xu, P., Ren, K., Li, M.: Delay-bounded adaptive UFH-based anti-jamming wireless communication. In: IEEE INFOCOM, 2011 (2011)
19. Chen, L., Bian, K., Chen, L., Liu, C., Park, J.-M.J., Li, X.: A group-theoretic framework for rendezvous in heterogeneous cognitive radio networks. In: ACM MobiHoc, 2014 (2014)
20. Wu, C.-C., Wu, S.-H.: On bridging the gap between homogeneous and heterogeneous rendezvous schemes for cognitive radios. In: ACM MobiHoc, 2013 (2013)
21. Wu, S.-H., Wu, C.-C., Hon, W.-K., Shin, K.G.: Rendezvous for heterogeneous spectrum-agile devices. In: IEEE INFOCOM, 2014 (2014)
22. Salameh, H.B., Krunz, M., Younis, O.: Mac protocol for opportunistic cognitive radio networks with soft guarantees. IEEE Trans. Mobile Comput. **8**(10), 1339–1352 (2009)
23. Lin, Z., Liu, H., Chu, X., Leung, Y.W.: Enhanced jump-stay rendezvous algorithm for cognitive radio networks. IEEE Commun. Lett. **17**(9), 1742–1745 (2013)
24. Bian, K., Park, J.-M.: Maximizing Rendezvous diversity in rendezvous protocols for decentralized cognitive radio networks. IEEE Trans. Mobile Comput. **12**(7), 1294–1307 (2012)

Quantified Dynamic Spectrum Access Paradigm

Nilesh Khambekar

Introduction

Dynamic spectrum sharing paradigm is essential for improving spectrum utilization and meeting the ever-increasing demand for radio frequency (RF) spectrum. The dynamic spectrum sharing paradigm however brings in new challenges on technical, regulatory, and business fronts. For example, it is not trivial to understand *how much* spectrum is available for sharing in the time, space, and frequency dimensions. For effective spectrum sharing, non-harmful interference needs to be ensured to the receivers in the system given the non-deterministic propagation and dynamic spectrum-access conditions. Due to the aggregate interference effects, dynamic propagation conditions, and software defined capabilities, the regulation of dynamic spectrum-access constraints is a complex issue. Furthermore, from a business perspective, it is also important to be able to flexibly and efficiently share or trade the spectrum in addition to solving the core technical and regulatory issues.

In order to address the challenges for the adoption of the new paradigm, we investigate what constitutes the use of spectrum and emphasize the need to characterize the use of spectrum by each of the transmitters and receivers in the space, time, and frequency dimensions. We highlight the lack of the ability to quantify the performance of recovery and exploitation of the underutilized spectrum. We present a Methodology to characterize and quantify the USE of spectrum in the space, time, and frequency dimensions (MUSE). In order to articulate, characterize, and quantify the use of spectrum, spectrum-space is discretized in the space, time, and frequency dimensions. We define the QDSA framework for defining and enforcing quantified spectrum-access rights and present signal-processing techniques in order to implement it in real-time.

N. Khambekar (✉)
Computer Science and Engineering, University at Buffalo, Suny, USA
e-mail: nvk3@buffalo.edu
URL: http://www.cs.buffalo.edu/~nvk3

The chapter is organized as follows. In section "Motivation", we underscore the need for a methodology to characterize and quantify the use of spectrum and the performance of spectrum management functions. In section "MUSE: Characterizing and Quantifying the Use of Spectrum", we present the mechanisms for characterizing and quantifying the use of spectrum in the space, time, and frequency dimensions using MUSE. In section "MUSE: Illustration and Discussion", we explain the methodology with a few examples and discuss the impact of the key factors while applying MUSE. In section "Applying MUSE: Maximizing the Use of Spectrum", we apply MUSE to spectrum management functions and demonstrate maximizing the use of spectrum. In section "Quantified Dynamic Spectrum Access Paradigm", we describe QDSA framework that defines and enforces quantified spectrum-access rights and quantified dynamic spectrum management. In section "Towards Real–Time Quantified Dynamic Spectrum-Access", We present signal processing techniques for RF-environment characterization and illustrate defining and enforcing quantified spectrum-access rights in real-time. In section "Benefits of the Quantified Dynamic Spectrum Sharing Paradigm", we present the benefits of quantified dynamic spectrum sharing paradigm in terms of addressing the technical, operational, regulatory, and business challenges. Section "Chapter Summary" summarizes and concludes the chapter.

Motivation

The dynamic spectrum sharing approaches have been evolving since the past decade [1–3]. Depending on the degree of sharing, the various spectrum sharing approaches fall into exclusive spectrum use, static spectrum sharing, dynamic spectrum sharing, and pure spectrum sharing categories [3]. Dynamic spectrum sharing differs from pure spectrum sharing in the sense that under pure spectrum sharing all services have equal spectrum-access priority. Zhao et al. classified spectrum sharing approaches into open sharing model, dynamic exclusive use model, and hierarchical access model [2]. The hierarchical access model could be further categorized into spectrum underlay model, non-prioritized spectrum overlay model, and prioritized spectrum overlay model. Spectrum underlay model imposes tight constraints on secondary spectrum-access in order to protect the spectrum-access rights of the incumbents. Under non-prioritized spectrum overlay model, a secondary spectrum-access is granted on a first come, first served basis while ensuring non-harmful interference to the receivers of the incumbent services. Under prioritized spectrum overlay model, certain services are assigned priority access privileges and the secondary access by these services is protected. Other non-prioritized secondary spectrum accesses are required to vacate if a priority user wishes to access spectrum. The proposed 3.5 GHz Citizens Broadband Radio Service (CBRS) [4] is an example of prioritized spectrum overlay model.

In terms of articulating the spectrum access rights, the spectrum sharing mechanisms primarily resort to statically or dynamically defining a spatio-temporal bound-

ary along with a *fixed* set of constraints. In this regard, the case study of dynamic spectrum sharing in UHF bands has brought out several technical, regulatory, and business difficulties.

In Nov. 2008, Federal Communications Commission (FCC) released a Notice of Proposed Rule Making (NPRM) to allow the unlicensed radios to operate in the TV bands without causing harmful interference to the incumbent services [5]. The Opportunistic Spectrum Access (OSA) of the unused UHF bands received a wide commercial interest for several potential wireless services; However, the performance estimation studies of OSA have revealed that the amount of the *implied* available spectrum is *very limited* to meet the increasing demand for RF spectrum [6–8]. Moreover, the secondary users cannot ensure desired quality of service necessary for the business cases due to the *secondary* rights for accessing the spectrum. On the other hand, incumbents do not have any incentive for sharing the spectrum. Furthermore, the secondary access to the spectrum is very hard to regulate. Considering interference aggregation effects, dynamic nature of propagation conditions, and dynamic spectrum-access scenarios, the primary owners of the spectrum need a way to confirm that their receivers are not subjected to harmful interference and the service experience is not degraded. This requires the ability to reliably estimate the interference margin at the receivers and accordingly infer the maximum transmit-power at the secondary transmitter positions. Furthermore, the behavior of software defined radio devices could be altered with software changes and thus the service is exposed to attacks from the secondary users of the spectrum. In order to ensure protection of the spectrum rights, the spectrum-access constraints need to be *enforceable*.

We observe that the decisions for exercising spectrum-access in case of OSA are based on detection of primary transmitter signal using a certain specified radio sensitivity. In this case, the decision for spectrum-access is binary in nature. This gives rise to '*not enough spectrum for secondary usage*' if the policy for shared spectrum-access is conservative and '*no guarantee for ensuring service quality*' if the shared spectrum-access policy is aggressive. The binary nature of the spectrum-access decision cannot protect the spectrum rights of incumbents and requires the spectrum-access policy to be increasingly conservative to guard against interference aggregation. Therefore, when multiple secondary transmitters exercise spectrum-access, we need to *quantitatively articulate the spectrum-access rights*. This helps maximizing spectrum-access opportunity without causing harmful interference to cochannel receivers. If technical and regulatory problems are solved, more and more incumbents will have an incentive to share the spatially, temporally, and spectrally unexploited spectrum.

Figure 1 illustrates the need for a methodology to characterize and quantify the use of spectrum under dynamic spectrum sharing paradigm with the aid of a question-map. The question-map enumerates the quantitative decisions involved in the process of investigating the weaknesses of a spectrum sharing mechanism, comparing various algorithms and architectures for recovery and exploitation of the spectrum, and optimizing the spectrum sharing opportunities.

Analyzing the Use of Spectrum

Objective: Maximizing the use of spectrum

- *How much is the opportunity for sharing the spectrum?*
  - How much of the spectrum is underutilized in the space, time, and frequency dimensions?
  - How much of the spectrum could be shared using a certain sharing model?
  - How much of the available-spectrum is rendered inaccessible due to certain conservative assumptions?
- *Is the recovery low?*
  - How much of the available-spectrum is recovered using a certain technique?
  - How much of the available-spectrum is lost due to false positives?
  - How much is the benefit of a certain fusion scheme over another?
  - How much is the impact of lack of knowledge of the propagation conditions on the recovery of the underutilized spectrum?
- *Is the exploitation low?*
  - How much of the spectrum is not exploited? Why?
  - How much is the impact of directional transceivers?
  - How much is the spectrum used by a service? How to maximize the number of satisfied spectrum-accesses?

Defining and Regulating a Dynamic Spectrum-access Policy

- How to dynamically define the spectrum-access rights based on the real-time RF-environment conditions for maximizing the use of spectrum?
- How to quantify violation of the spectrum-access rights occurred for a certain wireless network during a certain time period within a certain region?
- How to quantify violation of the spectrum-access rights by a specific transmitter?

Fig. 1 Example questions in case of optimizing a typical dynamic spectrum sharing scenario. The questions shade light on the various quantitative decisions involved with regard to spectrum sharing and spectrum management. The question-map emphasizes on the need for a methodology to characterize and quantify the use of spectrum in order to effectively manage the use of spectrum

Traditionally the performance of spectrum recovery is measured in terms of the throughput for the secondary users and outage probability [9–11]. The performance of detection of spectrum holes is also captured in terms of probability of missed detection and false positives [12–14]. However, this characterization of the performance is in the context of spectrum sharing constraints defined by a certain spectrum sharing model or in terms of system-level objectives. In order to maximize the use of spectrum, we need a methodology that can characterize the performance of the recovery and exploitation of the underutilized spectrum in the space, time, and frequency dimensions.

The existing methodologies to define the use of spectrum and quantify its efficiency are based on the static spectrum assignment paradigm and are not suitable for the dynamic spectrum sharing paradigm. ITU defined *spectrum utilization factor* as product of the frequency bandwidth, geometric space, and the time denied to other potential users [15]. However, spectrum utilization factor does not represent *actual usage*. For example, if a licensed user does not perform any transmissions, the spectrum is still considered to be *used*. It also cannot quantify the use of spectrum under

spatial overlap of wireless services. The IEEE 1900.5.2 draft standard captures spectrum usage in terms of transceiver-model parameters and applies standard methods for ensuring compatibility between the spectrum sharing networks [16]. Thus, the approach helps to ensure compatibility; however, it cannot *characterize and quantify* the use of spectrum and the performance of spectrum management functions.

Finally, from a business perspective, the ability to qualitatively and quantitatively interpret a spectrum sharing opportunity in a certain frequency band within a geographical region of interest is essential in order to evaluate its business potential. With the change in paradigm, businesses need the ability to control the use of spectrum at a fine granularity in order to maximize fine granular spectrum-reuse opportunities. With spectrum as a quantified resource perspective, the spectrum trade conversation could be on the following lines: "*I have 'x' units of spectrum right now, I have given 'y' units of spectrum to somebody and have 'z' units of spare spectrum which I can share.*" Also, the quantification of the use of spectrum would provide insight into the business implications of a dynamically identified spectrum-access opportunity in terms of the service quality, range, and user experience.

MUSE: Characterizing and Quantifying the Use of Spectrum

In order to define a methodology that enables us to characterize the use of spectrum in the space, time, and frequency dimensions, we first look into what constitutes the use of spectrum.

How Is Spectrum Consumed?

Traditionally, we assume that spectrum is consumed by the transmitters; however, the spectrum is *also* consumed by the receivers by constraining the RF-power from other transmitters. We note that for guaranteeing successful reception, protection is traditionally accomplished in terms of guard-bands, separation distances, and constraints on operational hours. Thus, the presence of receivers enforces limits on the interference-power in the space, time, and frequency dimensions. When the access to spectrum is exclusive in the space, time, and frequency dimensions, the spectrum consumed by the receivers need not be separately considered [15].

System Model

We consider a generic system with multiple heterogeneous *spatially-overlapping*[1] wireless services sharing the RF-spectrum. We define a *RF-link* as *zero*[2] or one transmitters and one or more receivers exercising spectrum-access. A *RF-network* represents an aggregate of RF-links. We refer to the aggregate of RF-networks sharing a spectrum space in the time, space, and frequency dimensions within a geographical region of interest as a *RF-system*. We consider that a multiple RF-systems are sharing the spectrum in the time, space, and frequency dimensions within the geographical region of interest.

We seek to capture spectrum-access at the lowest granularity. In this regard, RF-link represents the lowest granularity of spectrum-access.

Under the system model, we consider that the transceivers optionally employ directional transmission and reception in order to minimize interference. A receiver can withstand a certain interference when the received Signal to Interference and Noise Ratio (SINR) is greater than a receiver-specific threshold,[3] β.

Let P_{MAX} represent the maximum permissible power at any point and P_{MIN} represent the minimum power at any point in the system. P_{MIN} could be chosen to be an arbitrary low value below the ambient noise floor. The difference between the maximum permissible power and the arbitrary minimum power represents the maximum spectrum consumption, P_{CMAX}, at a point and it is given by

$$P_{CMAX} = P_{MAX} - P_{MIN}. \tag{1}$$

MUSE: Definitions

Transmitter-occupancy: We define transmitter-occupancy as the amount of spectrum consumed by a transmitter *at a point* in terms of RF-power occupied at the point.

Receiver-liability: We define receiver-liability as the amount of spectrum consumed by a receiver *at a point* in terms of the constraint imposed on the RF-power that can be exercised at the point by a potential or an existing transmitter. Thus, it represents *liability to the receiver* in order to protect the receiver from harmful interference.

---

[1]Without allowing spatial-overlap of wireless services, spectrum sharing may lead to spatial fragmentation of coverage for a wireless service. Furthermore, as discussed in the previous section, imposing a spatial boundary on spectrum sharing leads to suboptimal spectrum sharing.

[2]This is to include the use of spectrum by the receive-only systems; for example, radio astronomy telescopes.

[3]The threshold, β, represents the quality of a receiver and incorporates receiver-noise and other receiver technology imperfections. Thus, β models the receiver-performance under the MUSE methodology.
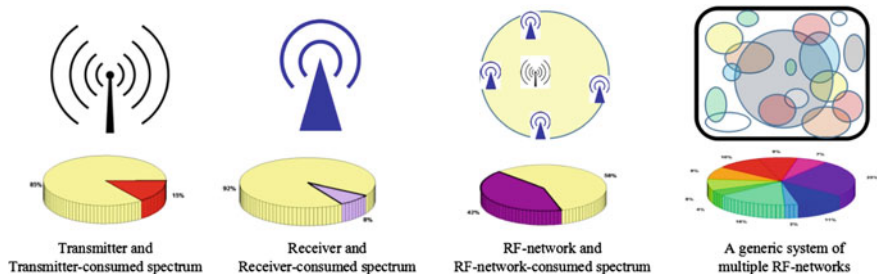
Fig. 2 RF-entities and associated spectrum-consumption spaces. The figure shows various entities within a generic system of wireless services sharing RF-spectrum: a transmitter, a receiver, and a RF-network. The rightmost picture shows a generalized spectrum sharing scenario with multiple *spatially-overlapping* heterogeneous wireless services sharing spectrum in the time, space, and frequency dimensions. The generalized topology emphasizes on the significance of spectrum sharing among heterogeneous wireless services without defining spatial, temporal, and spectral boundaries

Discretized Spectrum-space: The spectrum consumed by a transmitter or a receiver is *continuous* in the space, time, and frequency dimensions. In order to facilitate characterization and quantification of the use of the spectrum within a geographical region, we *divide* the total spectrum-space into discrete units and characterize the spectrum consumed by the transmitters and receivers in the unit spectrum-spaces. We refer to this discretized view of the spectrum in the space, time, and frequency dimensions as *discretized spectrum-space*.

A unit spectrum-space: A unit spectrum-space represents the spectrum within an unit area, in a unit time-quanta, and a unit frequency band.

RF-entity: We use an RF-entity as a generic term for an entity exercising spectrum-access. A RF-entity may represent an individual transmitter, an individual receiver, a RF-Link, a RF-network, or a RF-system.

A spectrum-consumption space: A spectrum-consumption space captures the spectrum consumption by a RF-entity in the discretized spectrum-space. The unit of a spectrum-consumption space is Wm^2. Figure 2 shows different RF-entities and the associated spectrum-consumption spaces.

The total spectrum-space: The total spectrum-space represents the spectrum in the space, time, and frequency dimensions within a geographical region of interest. Let the geographical region be discretized into $\hat{A}$ unit-regions, $\hat{B}$ unit-frequency-bands, and $\hat{T}$ unit-time-quanta. The total spectrum-space is given by

$$\Psi_{Total} = P_{CMAX} \, \hat{T}\hat{A}\hat{B}. \tag{2}$$

We identify following attributes with a unit-spectrum-space to characterize and quantify spectrum-consumption spaces.

Unit-spectrum-space Occupancy: We define unit-spectrum-space occupancy as the amount of spectrum consumed by all transmitters in a unit-spectrum-space.

Unit-spectrum-space Liability: We define unit-spectrum-space liability as the amount of spectrum consumed by all receivers in a unit-spectrum-space.

Unit-spectrum-space Opportunity: We define unit-spectrum-space opportunity as the amount of spectrum available for consumption in a unit-spectrum-space.

Quantifying Spectrum Consumption

Transmitter-occupancy

The power received from a transmitter t_n at a point ρ in the spatial dimension is given by

$$P_{r_\rho}(t_n) = P_{t_n} min\left\{1, L(d(t_n, \rho)^{-\alpha})\right\}, \tag{3}$$

where P_{t_n} is the transmit power of the transmitter and $d(t_n, \rho)$ is the distance between the transmitter t_n and the point ρ in the space. α is the path-loss exponent[4] and $L(d(t_n, \rho)^{-\alpha})$ denotes the path-loss factor.[5] Thus, (3) represents transmitter-occupancy of t_n at the point ρ in the geographical region.

Spectrum-occupancy

The aggregate power received at a point ρ is given by

$$\bar{P}_\rho = \sum_n P_{r_\rho}(t_n) + W_\rho, \tag{4}$$

where W_ρ is the average ambient noise power at ρ. Thus, (4) represents the spectrum-occupancy at the point ρ in the geographical region.

Unit-spectrum-space occupancy

Let us consider a unit-spectrum-space defined by unit-region χ, time-quanta τ, frequency-band ν. We define unit-spectrum-space occupancy, $\omega(\chi, \tau, \nu)$, as the spectrum occupancy at the *sample point* $\rho_0 \in \chi$, in the frequency band ν, at an instant within the time-quanta τ. Therefore,

$$\omega(\chi, \tau, \nu) = \bar{P}_{\rho_0}. \tag{5}$$

The unit for unit-spectrum-space occupancy is W.

Receiver interference-margin

Let $r_{n,m}$ be the mth receiver of the nth RF-link. The amount of interference power receiver $r_{n,m}$ can tolerate, that is the interference-margin for $r_{n,m}$, is given by

---

[4]The path-loss exponent can be a function of n and m in order to capture fine-grained spatial variations in the path-loss.

[5]The choice of propagation model would typically determine the path-loss.

$$\check{P}_{r_{n,m}} = \frac{P_{r_{n,m}}(t_n)}{\beta_{n,m}} - W_{r_{n,m}}. \tag{6}$$

Here, $W_{r_{n,m}}$ is the average ambient noise at the position of $r_{n,m}$. The unit of interference-margin is W.

We can view interference-margin $\check{P}_{r_{n,m}}$ as the upper-bound on the transmit power of an interferer at a *spatial separation of zero*. We characterize the limit on the interference power at a point ρ in space in terms of the receiver-imposed upper bound on the interference power.

$$\acute{I}(r_{n,m}, \rho) = \check{P}_{r_{n,m}} min\{1, L(d(\rho, r_{n,m})^\alpha)\}, \tag{7}$$

where $d(\rho, r_{n,m})$ is the distance between the receiver $r_{n,m}$ and the point ρ. We note that the receiver-imposed constraint on the interference power increases with increasing separation.

Receiver-imposed interference-opportunity
Let $\check{I}(r_{n,m}, \rho)$ denote the *aggregate interference power*[6] seen at a distant point ρ; then the *interference-opportunity* imposed by this receiver at ρ is given by the difference between the upper bound on the tolerable interference power and the aggregate interference power.

$$\ddot{I}(r_{n,m}, \rho) = \acute{I}(r_{n,m}, \rho) - \check{I}(r_{n,m}, \rho). \tag{8}$$

We note that when $\ddot{I}(r_{n,m}, \rho)$ is negative, the receiver $r_{n,m}$ is experiencing harmful interference.

Spectrum-opportunity
By combining the limits on the maximum interference power imposed by all the receivers, from all the RF-links in the system, we obtain the net interference-opportunity at a point ρ as

$$\bar{I}_\rho = \min_n(\min_m(\ddot{I}(r_{n,m}, \rho))). \tag{9}$$

We term the net interference-opportunity at a point as *spectrum-opportunity*.

Unit-spectrum-space opportunity
We define unit-spectrum-space opportunity, $\gamma(\chi, \tau, \nu)$, as the spectrum-opportunity at the sample point $\rho_0 \in \chi$, in frequency band ν, at an instant within the time-quanta τ. Therefore,

$$\gamma(\chi, \tau, \nu) := \bar{I}_{\rho_0}. \tag{10}$$

The unit for unit-spectrum-space opportunity is W. Here, $\bar{I}_{\rho_0}$ is computed within the time and frequency limits of the unit-spectrum-space.

---

[6]That is, the aggregate RF-power received at ρ from all the interference sources for the receiver $r_{n,m}$.

Unit-spectrum-space liability
We obtain unit-spectrum-space liability, that is the spectrum consumed by all the receivers in a unit-spectrum-space by subtracting the unit-spectrum-space occupancy and unit-spectrum-space opportunity from the maximum spectrum-consumption. Therefore,

$$\phi(\chi, \tau, \nu) = P_{CMAX} - (\omega(\chi, \tau, \nu) + \gamma(\chi, \tau, \nu)). \tag{11}$$

The unit for unit-spectrum-space liability is W.

Characterizing the spectrum consumed by a RF-entity at a point enables characterizing the spectrum-consumption space associated with the RF-entity. In this regard, we characterize the spectrum consumed by an individual transceiver in a unit-spectrum-space.

Receiver-liability
We get receiver-liability, that is the amount of spectrum consumed by an *individual* receiver at a point by subtracting the aggregate transmitter-occupancy and the interference-opportunity caused by the receiver from the maximum spectrum-consumption. Therefore,

$$L_\rho(r_{n,m}) = P_{CMAX} - (\bar{P}_\rho + \ddot{I}(r_{n,m}, \rho)). \tag{12}$$

The unit of receiver-liability is W.

Transmitter-occupancy in a unit-spectrum-space is given by transmitter-occupancy at the sample point $\rho_0 \in \chi$, in frequency band ν, at an instant within the time-quanta τ. Therefore,

$$\omega_{t_n}(\chi, \tau, \nu) \coloneqq P_{r_{\rho_0}}(t_n). \tag{13}$$

Receiver-liability in a unit-spectrum-space is given by receiver-liability at the sample point $\rho_0 \in \chi$, in frequency band ν, at an instant within the time-quanta τ. Therefore,

$$\phi_{r_{n,m}}(\chi, \tau, \nu) \coloneqq L_{\rho_0}(r_{n,m}). \tag{14}$$

Quantifying a Spectrum Consumption Space

A spectrum-consumption space associated with a RF-entity is quantified by aggregating the spectrum consumed by the RF-entity across all the unit-spectrum-spaces within a geographical region. We identify a few spectrum-consumption spaces towards maximizing the use of spectrum in Table 1 and quantify these spectrum spaces in this subsection.

Table 1 Example spectrum-consumption spaces

Spectrum consumption space	Description	Significance
Transmitter-consumed spectrum	It represents the spectrum-space consumed by a specified transmitter	It can be used in the context of *defining and enforcing* spectrum-access rights for a single transmitter
Receiver-consumed spectrum	It represents the spectrum-space consumed by a specified receiver	It can be used in the context of *defining and enforcing* spectrum-access rights for a single receiver
Utilized-spectrum	It represents the spectrum-space consumed by all the transmitters in the system	It can be used in the context of analysis and optimization of the spectrum consumed by transmitters
Forbidden-spectrum	It represents the spectrum-space consumed by all the receivers in the system	It can be used in the context of analysis and optimization of the spectrum consumed by receivers
Available-spectrum	It represents the spectrum-space not consumed by all the transmitters and receivers in the system and therefore *available*[a] for consumption	It can be used in the context of analysis of the potential of spectrum sharing and for assigning spectrum-access footprints

[a]In a spectrum sharing scenario, the spectrum-sharing policy defines certain constraints which determine what spectrum can be exercised for shared-access. We distinguish the spectrum implied available by a spectrum-sharing policy, that is, the implied-available spectrum, from the available-spectrum

Transmitter-consumed spectrum: The spectrum consumed by a transmitter within a geographical region is obtained by aggregating transmitter-occupancy across the unit-spectrum-spaces. Therefore,

$$\Omega(t_n) = \sum_{k=1}^{\hat{B}} \sum_{j=1}^{\hat{T}} \sum_{i=1}^{\hat{A}} \omega_{t_n}(\chi_i, \tau_j, \nu_k). \tag{15}$$

Receiver-consumed spectrum: The spectrum consumed by a receiver within a geographical region is obtained by aggregating receiver-liability across the unit-spectrum-spaces. Therefore,

$$\Phi(r_{n,m}) = \sum_{k=1}^{\hat{B}} \sum_{j=1}^{\hat{T}} \sum_{i=1}^{\hat{A}} \phi_{r_{n,m}}(\chi_i, \tau_j, \nu_k). \tag{16}$$

Utilized-spectrum ($\Psi_{utilized}$): We define *utilized-spectrum* as the spectrum consumed by all the transmitters within a geographical region. Utilized-spectrum is obtained by summing the unit-spectrum-space occupancy across all the unit-spectrum-spaces. Therefore,

$$\Psi_{utilized} = \sum_{k=1}^{\hat{B}} \sum_{j=1}^{\hat{T}} \sum_{i=1}^{\hat{A}} \omega(\chi_i, \tau_j, \nu_k). \tag{17}$$

Forbidden-spectrum ($\Psi_{forbidden}$): We define *forbidden-spectrum* as the spectrum consumed by all receivers within a geographical region. The forbidden-spectrum is quantified by aggregating unit-spectrum-space reliability across all the unit-spectrum-spaces. Therefore,

$$\Psi_{forbidden} = \sum_{k=1}^{\hat{B}} \sum_{j=1}^{\hat{T}} \sum_{i=1}^{\hat{A}} \phi(\chi_i, \tau_j, \nu_k). \tag{18}$$

Available-spectrum ($\Psi_{available}$): We define available-spectrum as the spectrum not consumed transmitters and receivers and therefore *available* for consumption. The total available-spectrum within a geographical region is obtained by summing unit-spectrum-space opportunity across all the unit-spectrum-spaces. Therefore,

$$\Psi_{available} = \sum_{k=1}^{\hat{B}} \sum_{j=1}^{\hat{T}} \sum_{i=1}^{\hat{A}} \gamma(\chi_i, \tau_j, \nu_k). \tag{19}$$

For *completeness*,[7] we express the relationship between these spectrum-consumption spaces. The spectrum consumption in a unit-spectrum-space is specified in terms of the unit-spectrum-space occupancy, unit-spectrum-space opportunity, and unit-spectrum-space liability. From (11),

$$\omega(\chi, \tau, \nu) + \phi(\chi, \tau, \nu) + \gamma(\chi, \tau, \nu) = P_{CMAX}. \tag{20}$$

Summing over all the $\hat{A}$ unit-regions in the geographical-region, $\hat{B}$ frequency-bands, $\hat{T}$ unit-time quanta, we get following relation between utilized-spectrum, forbidden-spectrum, and available-spectrum.

$$\Psi_{utilized} + \Psi_{forbidden} + \Psi_{available} = \Psi_{Total}. \tag{21}$$

Quantifying other spectrum-consumption spaces
One can identify a spectrum-consumption space with regard to the desired objective and quantify the spectrum-consumption space to facilitate analysis and optimization.

---

[7]In fact, this relationship follows from the definition of unit-spectrum-space reliability.

For example, one can quantify the harmful interference caused by a single transmitter to the cochannel receivers. This can be useful in terms of regulation of a spectrum-access policy.

Characterizing and Quantifying Performance of the Spectrum Management Functions

Similar to characterization of the use of spectrum in terms of spectrum-consumption spaces, we characterize the performance of a spectrum management function in terms of the *spectrum management function (SMF) spaces*.

Let us consider an attribute θ characterizing the performance of a spectrum-management function in the discretized spectrum-space. For example, the error in the estimated unit-spectrum-space opportunity represents the performance of spectrum recovery. Similarly, the amount of spectrum not exploited in a unit-spectrum-space can capture weaknesses of a spectrum exploitation mechanism. Table 3 describes the SMF spaces associated with spectrum sharing, spectrum recovery, and spectrum exploitation functions.

We characterize the SMF attribute θ in a *unit-spectrum-space* defined by unit-region χ, time-quanta τ, frequency-band ν in terms of the SMF attribute at the sample point $\rho_0 \in \chi$, in the frequency band ν, at an instant within the time-quanta τ. Therefore,

$$\theta(\chi, \tau, \nu) = \theta_{\rho_0}. \tag{22}$$

The SMF space within a geographical region is obtained by summing the SMF attribute across the unit-spectrum-spaces. Therefore,

$$\Theta = \sum_{k=1}^{\hat{B}} \sum_{j=1}^{\hat{T}} \sum_{i=1}^{\hat{A}} \theta(\chi_i, \tau_j, \nu_k). \tag{23}$$

MUSE: Illustration and Discussion

We start our illustration of the methodology with an abstract view of spectrum-use at a point. The total spectrum at a point is determined by P_{MAX} and P_{MIN}.

- If there are no transmitters and receivers in the system, transmitter-occupancy and receiver-liability at this point are zero; the spectrum-opportunity will be maximum, that is, P_{CMAX}. The spectrum-opportunity represents maximum RF-power that can be used by a future transmitters while ensuring non-harmful interference at the receivers. This scenario is captured by the leftmost bar in Fig. 3.
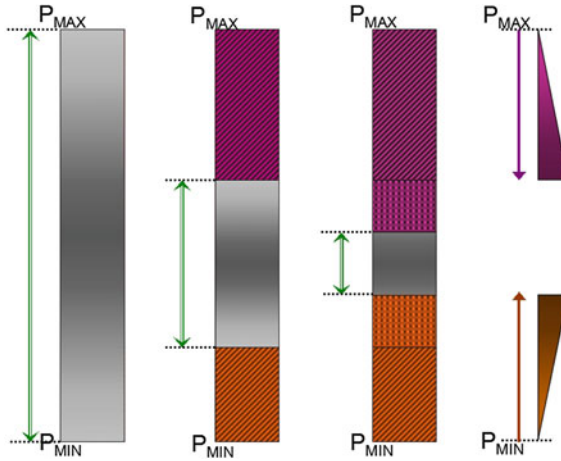
Fig. 3 The use of spectrum at a point. The *leftmost bar* captures the maximum ($P_{MAX} - P_{MIN}$) spectrum-opportunity (shown with *green double arrow*) at a point when no transceivers are present. The *middle bar* shows the spectrum consumed by a transmitter and its receiver. The *rightmost bar* shows the spectrum consumed by two pairs of transceivers. Here, we note that the spectrum-occupancy grows from P_{MIN} towards P_{MAX} while spectrum-liability, representing a constraint on the occupiable RF-power, grows from P_{MAX} towards P_{MIN}. The spectrum-opportunity goes on reducing as the transceivers consume more and more of the spectrum at a point

- If we add a transmitter-receiver pair, we can observe nonzero transmitter-occupancy and receiver-liability. A key thing to note is that receiver-liability is the limit on the maximum RF-power at a point and grows from P_{MAX} towards P_{MIN}. Thus, the higher the minimum SINR for successful reception, the higher is the receiver-liability. The middle bar in Fig. 3 shows this scenario and we can observe that the spectrum-opportunity has been reduced due to the constraint imposed by receiver.

- As more and more transceivers are added to the system, the spectrum-opportunity decreases. This scenario is shown in the rightmost bar in Fig. 3. In this case, the spectrum-occupancy captures the aggregate value of the transmitter occupancy from the individual transmitters. With regard to receivers, different receivers impose different constraints on the RF-power sourced from the point. The effective constraint at this point is determined by the receiver having the highest receiver-liability at the point.

Next, we consider a simplistic setup and illustrate the spectrum consumed by the transceivers at an arbitrary point in the system. Let us consider a 4.3×3.7 km geographical region. We assume, P_{MAX} is 1 W or 30 dBm; P_{MIN} is -200 dBm. The ambient noise floor is assumed to be -106 dBm (for channel bandwidth of 6 MHz). Large-scale path-loss model is used with path-loss exponent of 3.5. The minimum desired SINR at receiver for successful reception, β, is assumed to be 3 dB. A transmitter is positioned at (1000, 1200), a receiver is positioned at (1000, 2100), and

Table 2 Spectrum consumption at a point under three scenarios

S/N	Transmit power (dBm)	Receiver position	Receiver SINR (dB)	Transmitter-occupancy (dBm)	Spectrum-opportunity (dBm)	Receiver-liability (dBm)
I	−24	(1000, 2100)	12.0	−132.6	11.84	29.93 (984 mW)
II	6	(1000, 2100)	42.0	−102.6	30	−200 (0 mW)
III	6	(1000, 2500)	17.5	−102.6	19.0	29.64 (920 mW)

spectrum consumption is quantified at the point (2250, 1800). Scenario I in Table 2 shows the spectrum consumption at this point in terms of transmitter-occupancy, receiver-liability, and spectrum-opportunity.

Scenario II, we change the transmitter power from −24 to 6 dBm; accordingly, the spectrum consumption by the transmitter at point (2250, 1800) changes from −132.6 to −102.6 dBm. With regard to the spectrum consumption by the receiver, we observe that SINR at the receiver is significantly improved and consequently the tolerance for interference is improved. Thus, the receiver can withstand interference of 30 dBm generated from position (2250, 1800) without experiencing harmfully interference. Since, the spectrum-opportunity is maximum (30 dBm or P_{MAX}) in case II, the spectrum consumption by the receiver is minimal (−200 dBm or P_{MIN}).

In Scenario III, we move the receiver farther from its transmitter; this results in reduced SINR and consequently reduced tolerance to interference. Thus, the spectrum-opportunity caused by the receiver at (2250, 1800) is lowered from 30 to 19 dBm and the spectrum consumed by the receiver increases to 920 mW.

Quantification of a Spectrum Consumption Space

After characterizing spectrum consumption at an arbitrary point, we move to characterizing a spectrum-consumption space within a geographical region of interest. Let us consider a 4.3×3.7 km geographical region with 676 hexagonal unit regions with each side 100 m long. Let the maximum RF-power at a point, P_{MAX} in the unit regions be 30 dBm, that is 1 W. Let P_{MIN} be −200 dBm. Let us consider 6 MHz spectral range as unit bandwidth and a 10 s time period as unit time. In this scenario, the maximum spectrum consumption in the geographical region, in a 6 MHz spectral band, in a 10 s time period is 676 Wm^2 as given by (2). We model the propagation conditions by large-scale path-loss model with the path-loss exponent is 3.5.

Fig. 4 Spectrum
consumption space of an
individual transmitter. The
figure shows spatial
distribution of the
transmitter-occupancy in the
unit-spectrum-spaces within
a geographical region. Thus,
it captures the spectrum
consumed by a transmitter
within the geographical
region. The transmitter is
shown by a *solid square* and
the receiver is shown by a
non-solid square
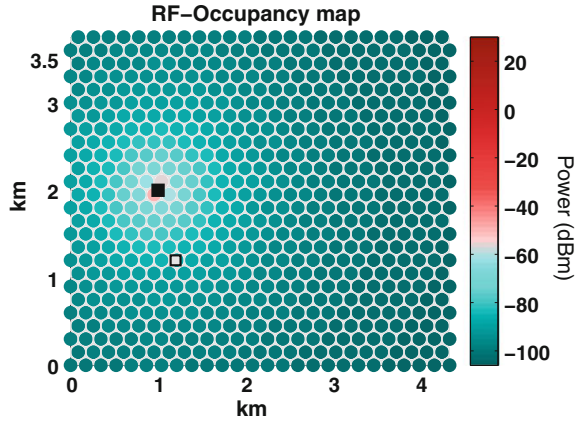
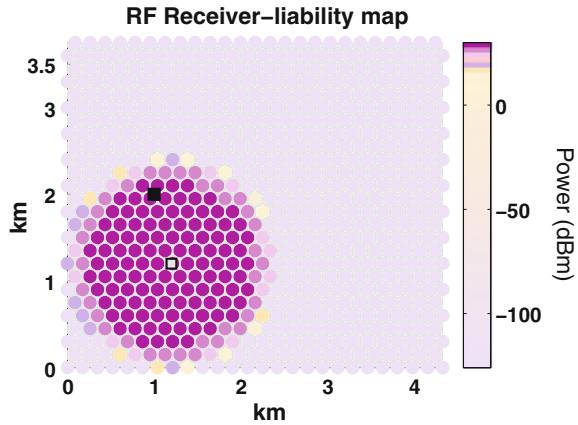## Spectrum Consumed by a Transmitter

First, we will look into the spectrum-consumption space for an individual transmitter. Figure 4 illustrates the spectrum consumed by a transmitter within a geographical region according to (15). The transmitter is located at (1000, 2000) and is exercising omnidirectional transmission with transmit power of 15 dBm. The spectrum consumed by the transmitter is 1.8×10^{-8} Wm^2 (2.7×10^{-9} % of the total spectrum space).

Spectrum Consumed by a Receiver

Figure 5 illustrates the spectrum-consumption space for a receiver according to (16). The receiver is located at (1200, 1200) and is exercising omnidirectional reception requiring minimum SINR of 6 dB and the actual experienced SINR at the receiver is 33 dB. We note that as the distance from a receiver increases, a cochannel transmitter can exercise higher transmission power. Thus, the liability for ensuring non-harmful interference to the receiver goes down with the distance from the receiver. The spectrum consumed by the receiver is 112.4 Wm^2 (16.6 % of the total spectrum space).

The spectrum consumed by a RF-entity is the sum of the spectrum consumed by all the transmitters and receivers within the RF-entity. Thus, the spectrum consumed by the RF-link in this case is 112.4 Wm^2. The spectrum consumed by a RF-link has been considered as a parameter for scheduling RF-links in order to minimize spectrum consumption and improve the performance of scheduling in [17].

Fig. 5 Spectrum consumption space of an individual receiver. The figure shows the spatial distribution of receiver-liability in the unit-spectrum-spaces within a geographical region. Thus, it captures the spectrum consumed by a receiver within the geographical region. The transmitter is shown by a *solid square* and the receiver is shown by a *non-solid square*

Available-Spectrum

The spectrum not consumed by the transmitters and receivers is the available-spectrum within a geographical region. Figure 6 depicts spatial distribution of unit-spectrum-space opportunity given by (19) for the above topology. The available-spectrum space is 563.6 Wm^2 (83.4 % of the total spectrum space).
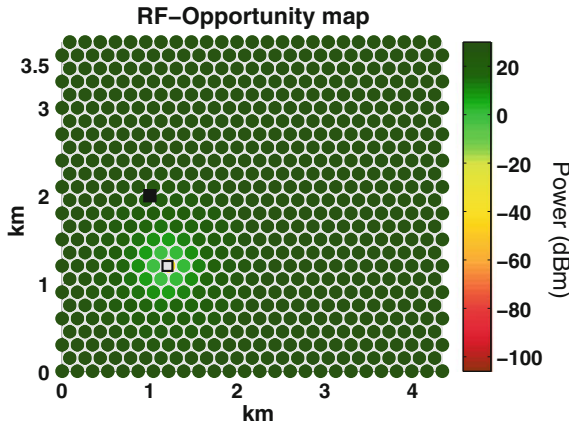
Fig. 6 Available-spectrum within a geographical region. The figure shows the spatial distribution of unit-spectrum-space opportunity within a geographical region. Thus, it captures the available-spectrum within the geographical region. The transmitter is shown by a *solid shape* and the receiver is shown by a *non-solid shape*. We observe that the unit-spectrum-space opportunity near the receiver is lower in order to ensure non-harmful interference at the receiver

Quantification of a Spectrum-Management Space

The performance of spectrum sharing depends primarily on the spectrum-sharing policy which defines what spectrum can be *considered available* for exploitation, the *optional*[8] spectrum recovery function which defines how efficiently the available spectrum is recovered, and the spectrum exploitation function which influences how efficiently the recovered available spectrum is assigned for satisfying spectrum-access requests. In this regard, Table 3 identifies a few example spectrum-management spaces.

Quantifying these spectrum-management spaces enables comparison, analysis, and optimization of spectrum sharing. Let us consider the recovery of the underutilized spectrum by estimating the unit-spectrum-space opportunities within a geographical region using a dedicated RF-sensor network. The RF-sensors sense the RF-environment in order to detect the presence of cochannel transmitters, geolocate the transmitters, and estimate the transmit-power of the transmitters. A missed-detection, false-positive, an error in geolocation implies error in the estimated spectrum-opportunity. A negative error implies spectrum opportunity is lost in the unit-spectrum-space and a positive error may potentially lead to harmful interference. Figure 7 shows the performance of spectrum recovery in term of estimation of unit-spectrum-space opportunity given by (10). We observe that 630.7 W (99.3 % of the total spectrum) of the available-spectrum within the geographical region has been recovered; 12.7 W (2 % of the total spectrum) of the available-spectrum has been lost and 8.4 W (1.3 % of the total spectrum) of the not available-spectrum has been erroneously considered available for exploitation [18].

Considerations While Applying MUSE

Granularity of Unit-spectrum-space The granularity of spectrum sharing identifies the smallest portion of spectrum-space for which spectrum-access rights could be defined and enforced. In this regard, a unit-spectrum-space represents the lowest granularity of spectrum sharing. Thus, the granularity of spectrum sharing plays a key role in determining the sampling rate in the space, time, and frequency dimensions [19, 20]. In favor of standardization, an alternate perspective could be choosing the unit-spectrum-space dimensions; thus, the unit-spectrum-space granularity could determine the spectrum-sharing granularity. In this case, we suppose the *population-density* and the *propagation environment characteristics* can play a key role in determining the spatial granularity of a unit-spectrum-space. The temporal granularity for a unit-spectrum-space can be considered to depend upon the traffic

---

[8]The management of spectrum may vary across different spectrum sharing models. Market based approach to spectrum sharing presumes a spectrum pool while overlay approach requires recovering the underutilized spectrum.

Table 3 Example spectrum-management spaces

Function	Spectrum-management space	Description
Spectrum sharing	Implied-available spectrum space	It represents the portion of available spectrum implied accessible under the constraints imposed by a spectrum-sharing policy
	Implied-guard spectrum space	It represents the portion of available spectrum that has been (usually intentionally) treated as not available spectrum
	Implied-incursed spectrum space	It represents the portion of non available spectrum that has been erroneously treated as available spectrum
Spectrum recovery	Recovered-available spectrum space	It represents the portion of implied-available spectrum that has been recovered for exploitation
	Lost-available spectrum space	It represents the portion of implied-available spectrum that has been erroneously treated as not available for exploitation
	Potentially-incursed spectrum space	It represents the portion of non available spectrum that has been erroneously treated as available for exploitation
Spectrum exploitation	Exploited-available spectrum space	It represents the portion of the recovered-available spectrum consumed by transmitters and receivers
	Unexploited-available spectrum space	It represents the portion of the recovered-available spectrum not consumed by transmitters and receivers
	Incursed spectrum space	It represents the portion of the non available spectrum consumed by transmitters and receivers

characteristics. The transceiver technology and its frequency agility would typically drive the granularity of a unit-spectrum-space in the frequency dimension.

Stochastic Modeling of Spectrum Consumption

With MUSE, we calculate the spectrum-use using a set of sample-points in the spatio-temporal-spectral spectrum-space instead of employing a stochastic model.
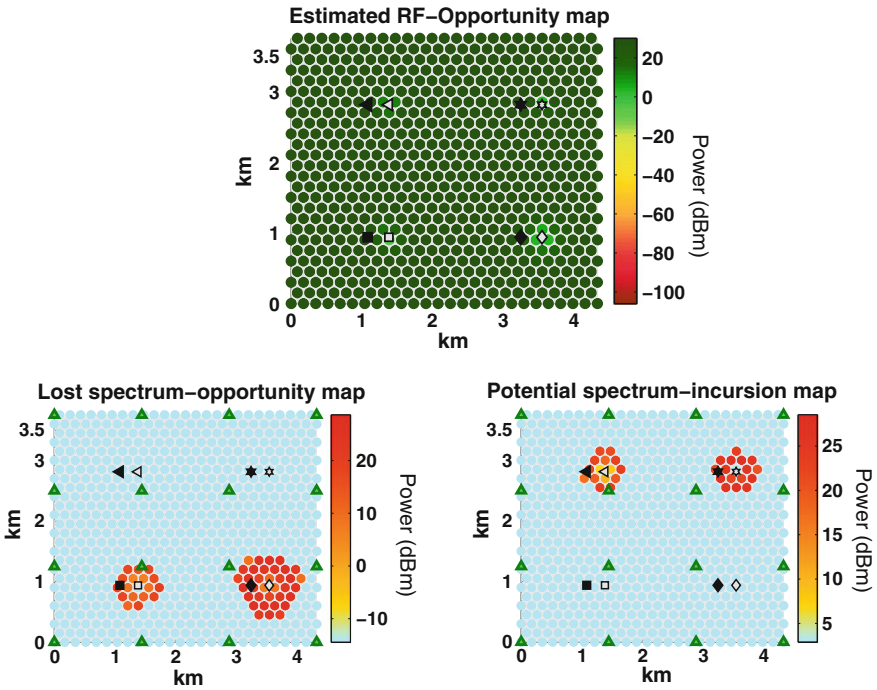
Fig. 7 Performance analysis of recovering the available-spectrum. A dedicated RF-sensor network with 16 RF-sensors estimates the unit-spectrum-space opportunities exploiting signal-cyclostationarity across the unit-spectrum-spaces within a geographical region [18]. The errors in the detection, geolocation, and transmit-power estimation result into lost-available spectrum and potentially-incursed spectrum. The *top plot* characterizes the estimated unit-spectrum-space opportunities. The *bottom two plots* capture the spatial distribution of lost-spectrum-opportunity and potential-spectrum-incursion within the geographical region. The RF-sensors are shown by upward-pointing *triangle* markers. The other 4 markers represent 4 transceiver pairs; a *solid* marker shows a transmitter and an *unfilled* marker shows a receiver

It is possible to enrich the representation of the spectrum consumption in a unit-spectrum-space using stochastic models.

Spectrum Use in the Code Dimension MUSE is agnostic of the waveforms employed by the RF-entities. Thus, it does not capture the spectrum use in the *code* dimension of spectrum-access.

Applying MUSE: Maximizing the Use of Spectrum

Now, we revisit the question-map from Fig. 1 and investigate solutions for the primary goal of maximizing the use of spectrum. The ability to characterize the use of

spectrum provides an insight into the opportunities to maximize the use of spectrum. By articulating the spectrum rights in terms of quantified use of the spectrum, we get the ability to precisely *control* the use of spectrum.

In order to maximize the use of spectrum, we emphasize on

- maximizing the available spectrum-space: This is accomplished by maximizing the spectrum sharing opportunities and maximizing the recovery of the underutilized spectrum.
- maximizing the exploitation of the available spectrum-space.

Maximizing the Available Spectrum-Space

For maximizing the spectrum sharing opportunities, we need to investigate into the spectrum measurements. The spectrum measurements data from [21] illustrates slower to faster degrees of variations in the occupied RF-power. Figure 8 from [21] captures the fast variations in the occupied RF-power over time. For maximizing the use of spectrum, we need to take into account variations in the use of spectrum in the space, time, and frequency dimensions. The case-study of OSA has shown that conservative assumptions based on the worst-case conditions severely limit the spectrum available for sharing [8]. Hence, in order to maximize the spectrum available for sharing, it is necessary to characterize the use of spectrum in real-time.

Next, once we observe a significant quantity of the underutilized spectrum within a geographical region, we seek to reuse of the underutilized spectrum using dynamic spectrum sharing paradigm. A *spectrum-sharing policy*[9] under a spectrum sharing model plays a central role in shaping the spectrum available for sharing. The constraints defined in a spectrum-sharing policy may suggest a guard-space in time, frequency, or space dimensions. This implies a certain amount of available spectrum is rendered un-exploitable. In [22], we identified that using the worst-case propagation conditions and using the worst-case receiver positions due to the lack of knowledge of the receiver positions as the key weaknesses of OSA spectrum-sharing policy. Consequently, the minimum sensitivity and the maximum transmit-power constraints imposed on the secondary users tend to be very conservative and the available-spectrum that can be exploited by the secondary users is severely (less than 1 %) limited [22]. In [17], we addressed optimizing the spectrum available for sharing by optimizing the constraints defined under a spectrum-sharing policy while ensuing non-harmful interference to the primary and secondary users.

---

[9]Here, we distinguish these spectrum-access constraints that imply the spectrum available for sharing from the spectrum-access constraints on an individual RF-entity while exercising a spectrum-access. We call the former one as *spectrum-sharing policy* and the later one as *spectrum-access policy*.
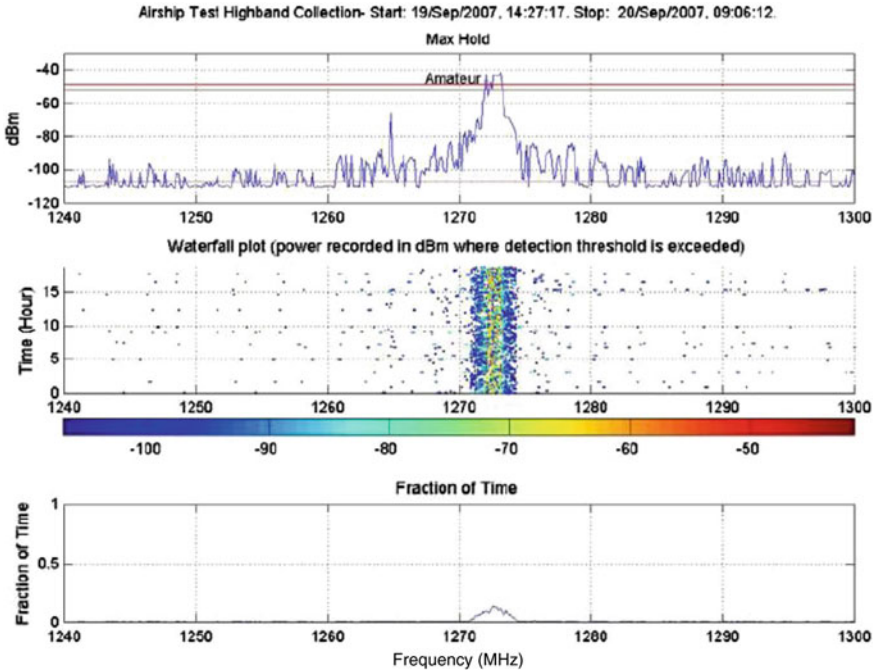
Fig. 8 Spectrum occupancy measurements at Loring Commerce Center, Limestone, Maine during September 18–20, 2007 [21]. The measurements illustrate the fast variations in the spectrum occupancy over time

Maximizing the Exploitation of the Available Spectrum-Space

The actual spectrum exploited by a transceiver depends on the *spectrum-access policy*. In order to dynamically define an optimal spectrum-access policy, we need the ability to recover the spectrum implied available by a spectrum-sharing policy. In [23], we illustrate recovery of the underutilized spectrum using a dedicated external RF-sensor network. In contrast to the traditional approach wherein spectrum holes are inferred by employing detection of the transmitter-signal, this approach estimates unit-spectrum-space opportunity given by (10) across the unit-spectrum-spaces within a geographical region. The RF-sensors characterize the fine-grained propagation environment, estimate the spectrum-access parameters of the transceivers exploiting signal cyclostationarity, and thereby estimate the available spectrum-space.

In order to optimally ***exploit the recovered spectrum*** for satisfying the spectrum-accesses from multiple spatially-overlapping heterogeneous wireless networks, we emphasize the need for defining a quantified spectrum-access policy. A quantified spectrum-access policy identifies spectrum-access rights in terms of how much of the spectrum can be consumed by an RF-entity within each of the unit-spectrum-

spaces within a geographical region. Thus, when multiple spatially-overlapping RF-entities are sharing the spectrum, non-harmful interference could be ensured to the receivers in the system. We note that using the RF-sensor network, the dynamically defined spectrum-access rights can be enforced in real-time by estimating the spectrum-consumption spaces for the individual transmitters.

Defining quantified spectrum-access rights requires us to develop spectrum assignment schemes that can *quantitatively* control the spectrum-footprint assigned to each of the transceivers. A quantified approach to spectrum exploitation essentially transforms the spectrum-scheduling and spectrum-allocation problems into a problem of optimizing the spectrum-consumption spaces for a set of spectrum-access requests. In [17], we present maximizing the number of spectrum-access requests based on *spectrum-consumption by an RF-entity*.

We can further improve the amount of spectrum available for sharing by *reducing* the spectrum-consumption by RF-entities. Directional transmission and reception helps to improve SINR; reduce the spectrum consumed by transceivers and increase the available-spectrum [17]. Also, with active role by incumbents, primary transmission power could be increased to enhance SINR at the primary receivers, minimize the receiver consumed spectrum, and maximize the spectrum available for sharing with the secondary services [17].

Dealing with RF-Environment Dynamicity

The key challenge for optimizing spectrum sharing potential is the *dynamicity of RF-environment*. The propagation conditions may vary quite fast at the order of every few 100 m. A certain frequency band may not be available at a certain location due to a reappearing primary user. The link quality for secondary access may change due to another mobile secondary user. Here, we note that traditionally wireless network services have been *conservatively handcrafted* to ensure minimum performance under the worst-case conditions. Dynamic spectrum sharing model forces us to come of the constrained setup and develop dynamic responses to the unknown possibilities in terms of the RF environmental and access conditions. Such dynamic response is possible only via learning the RF-environment and synthesizing behavior, decisions, and actions in advance.

The spectrum-discretization approach enables bringing in learning and adaptation to the spectrum management functions. For example, in order to dynamically define the spectrum access rights, we investigate fine granular characterization of the shadowing profile within a geographical region [18]. Real-time characterization of the shadowing loss within a fine granular region helps to control the guard space at the finer granularity and maximize the spectrum available for sharing under dynamic RF environments. In section "Towards Real-Time Quantified Dynamic Spectrum-Access", we present the signal processing techniques for RF environment characterization and real-time quantified dynamic spectrum access.

With regard to adaptation of the spectrum-access, we need the spectrum-aggregation abilities that can provision redundancy for RF-links. By characterizing the unit-spectrum-space opportunity in the spatial dimension, we can infer *spectrum-connectivity* across the adjacent unit-regions within a geographical region using a certain frequency band. We can also combine the spectrum-access opportunities across multiple frequency bands and infer aggregate RF-connectivity map as shown in Fig. 9. Such fine granular characterization of the use of spectrum is useful in the optimization of scheduling, spectrum assignment, and routing. Thus, analysis of the use of spectrum provides rich information and improves the ability to adapt in case of spectrum mobility events.
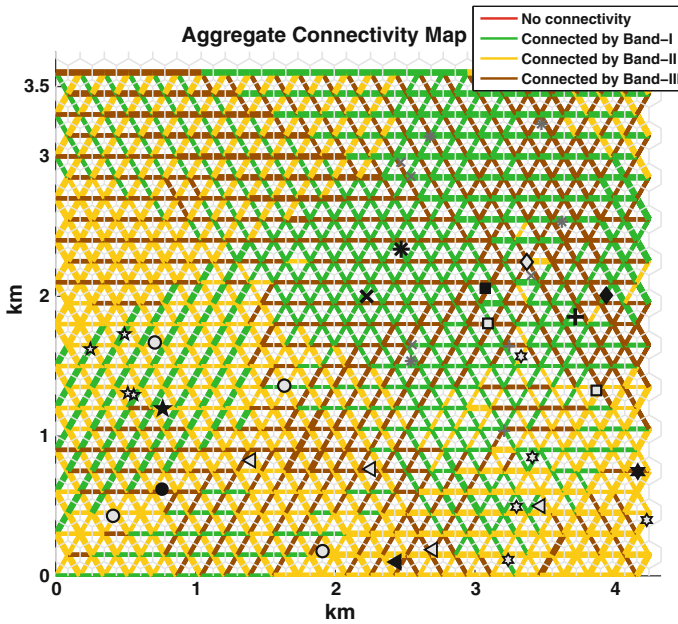
Fig. 9 Multiple-band RF-connectivity map showing the degree to which adjacent unit spectrum-spaces can connect using a new RF-link. Transmitters and receivers in the same network have the same shape; the transmitter is *solid*. The particular frequency band is encoded through the color of the connecting *lines*, and the *line color* is determined by the best available connectivity. The map reveals exploitable spectrum opportunities in the spatial and frequency dimensions. For this particular set of networks, the spectrum opportunities within the band I (*green*), band II (*yellow*), and band III (*brown*) are easily discerned. Moreover, the directional variation in spectrum opportunity is taken into account, so that the best channel to use depends on the spatial orientation of the to-be-added transmitter-receiver pairs

Quantified Dynamic Spectrum Access Paradigm

Quantified Dynamic Spectrum Access (QDSA) is an abstract framework for dynamic spectrum-access. Under QDSA, it is possible to implement various types of spectrum sharing models. QDSA provides the ability to share or trade RF spectrum in a flexible and quantifiable manner similar to a commodity. QDSA defines following primitives to accomplish quantified sharing of the spectrum.

Primitives for Quantified Dynamic Spectrum Access

RF-entity: An RF-entity is a generic term used for an entity exercising spectrum-access. A RF-entity may represent an individual transmitter, an individual receiver, a RF-Link, a RF-network, or a RF-system.

Priority of Spectrum-Access: It captures the relative priority among RF-entities. Thus, the primitive abstracts out the various spectrum sharing models such as prioritized dynamic spectrum sharing, non-prioritized spectrum sharing, pure spectrum sharing.

Unit-spectrum-space Occupancy: The net spectrum consumed by all the transmitters in the system in a unit spectrum-space.

Unit-spectrum-space Liability: The net spectrum consumed by all the receivers in the system in a unit spectrum-space.

Unit-spectrum-space Opportunity: The amount of spectrum available for consumption in a unit spectrum-space.

Quantified spectrum-access policy: Spectrum-access policy enumerates the spectrum-access constraints on a RF-entity exercising spectrum access. The spectrum-access constraints specify the bounds on spectrum consumption in each of the unit spectrum-spaces. Thus, a spectrum-access policy quantifies the spectrum used by a RF-entity. Based on the assigned spectrum-consumption values in each of the unit spectrum-spaces, the individual transceiver parameters are chosen.

Granularity of Spectrum-access policy: It captures the granularity of an RF-entity for which spectrum-access rights are defined and enforced.

Granularity of Spectrum Sharing: It captures the smallest portion of spectrum-space for which spectrum-access rights could be defined and enforced.

A quantified spectrum-access policy is defined dynamically based on the dynamic spectrum-access attributes and propagation conditions. In the next subsection, we describe the core spectrum management functions for defining and regulating a quantified dynamic spectrum-access policy.

Defining and Regulating a Quantified Dynamic Spectrum-Access Policy

Defining a real-time quantified spectrum-access policy needs estimating the *available* spectrum-space. In [23, 24], we present estimating the spectrum consumed by multiple spatially-overlapping heterogeneous RF-systems under shadowing and multipath conditions. The key signal processing functions necessary for estimating transmitter and receiver spectrum consumption are

- estimating transmitter position and directionality.
- estimating mean path loss exponent and shadowing loss in the unit spectrum spaces within the geographical region of the interest.
- estimating SINR at the receiver positions.

Regulating a spectrum-access policy is essential for guaranteeing protection of spectrum rights of spectrum users. In this regard, spectrum consumption estimation also enables estimating the spectrum consumption footprints of the RF-entities and detecting violations against assigned quantified spectrum-access policy.

An Example Spectrum Management Infrastructure for Quantified Dynamic Spectrum Sharing

We define an example framework for real-time quantified dynamic spectrum access. The key elements of such a framework are

- Spectrum Sensing Infrastructure (SSI) that learns the fine granular propagation conditions and estimates the transceiver spectrum-access parameters.
- Spectrum-consumption Analysis Infrastructure (SAI) uses the known and estimated transceiver spectrum-access parameters and learned fine-grained propagation model parameters to estimate spectrum-consumption spaces for individual transceivers and the available spectrum space.
- Spectrum-access Policy infrastructure (SPI) that manages spectrum-access requests from RF-entities. It receives the desired spectrum-access parameters for RF-entities of a given spectrum-access request and assigns a quantified spectrum-access policy that ensures non-harmful interference with the RF-entities while satisfying the minimum desired spectrum-access attributes. SPI also detects violation of the policy based on information from SAI regarding the quantification of spectrum consumption by the RF-entities.
- Spectrum-access Management Infrastructure (SMI) uses the available spectrum-consumption space information in order to schedule and assign spectrum-consumption footprints to the individual transceivers of a spectrum-access request.
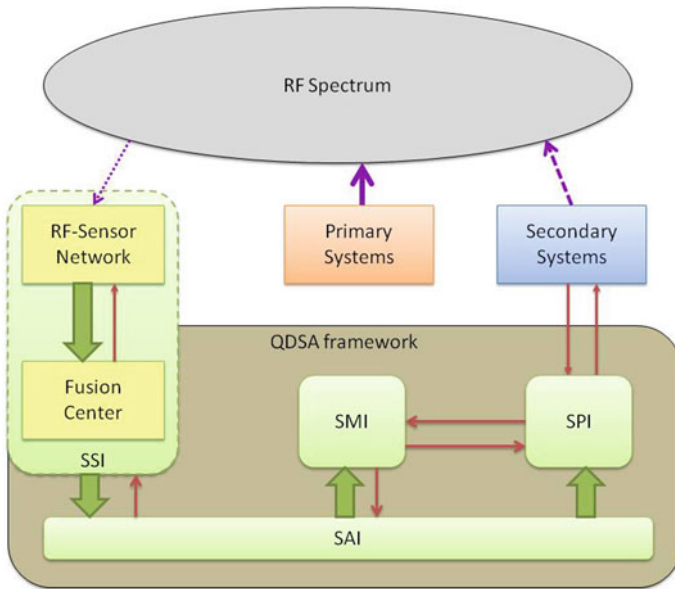
Fig. 10 An example spectrum management infrastructure for quantified dynamic spectrum sharing. The Spectrum Sensing infrastructure (SSI) performs RF-environment characterization using an external RF-sensor network. The unit spectrum-space opportunities are estimated by Spectrum-consumption Analysis infrastructure (SAI). The Spectrum-access Management Infrastructure (SMI) manages scheduling and assignment of the harvested spectrum. The Spectrum-access Policy Infrastructure (SPI) manages spectrum-access requests and assigns quantified spectrum-access policies to the RF-entities

Figure 10 shows an example spectrum management infrastructure for quantified dynamic spectrum sharing and Fig. 11 illustrates a scenario of defining and enforcing a quantified dynamic spectrum-access policy.

Quantified Dynamic Spectrum Management

In this section, we present a quantified approach to spectrum management and illustrate how it helps to analyze and optimize spectrum sharing opportunities.

Dynamic spectrum sharing paradigm essentially calls for fundamental changes in spectrum management. With exclusive spectrum allocation and static spectrum sharing, spectrum management has been centered around spectrum assignment for long durations and at specific geographical locations. The spectrum-access constraints in this case do not need to vary within the fine-granular spaces and time-intervals.

With dynamic spectrum sharing paradigm, we identify two core spectrum management functions: *producing* spectrum-resource and *consuming* spectrum-resource.
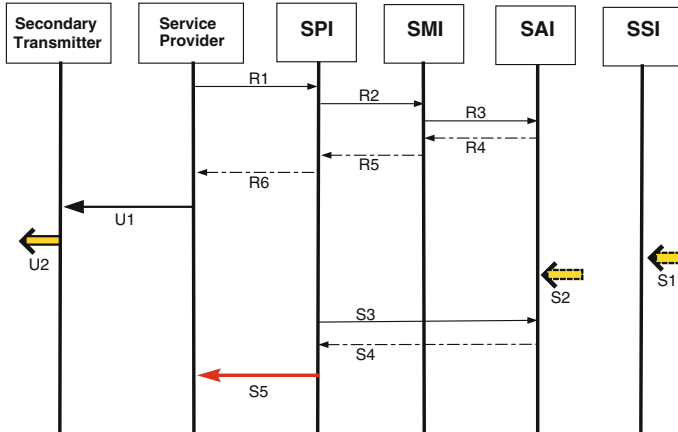
Fig. 11 Illustration of defining and enforcing a spectrum-access policy. A service-provider requests a spectrum-access footprint from SPI along with the information about position and capabilities of the transceivers. This is shown with *arrows R1–R6*. A service-provider assigns a partial time to one of the secondary transmitters (*Arrow U1*). The secondary transmitter fails to conform to the assigned quantified spectrum-access policy (*Arrow U2*). This scenario is detected with transmitter spectrum consumption estimation (*Arrows S1–S4*) and a regulatory action is taken (*Arrow S5*)

In a broader context, considering the various spectrum sharing approaches, we loosely call these two functions as spectrum harvesting and spectrum exploitation respectively.

Quantified Spectrum Harvesting
Traditionally the underutilized spectrum is identified by detection of the primary transmitter signal. In this case, the performance of spectrum sensing is measured in terms of probability of missed detection and false positives [25]. Quantified spectrum harvesting characterizes the RF-environment in order to estimate unit-spectrum-space opportunity and thus enables to choose optimum spectrum-access parameters. One of the key benefits of quantified approach to spectrum harvesting is aggregation of unit-spectrum-space opportunities across multiple frequency bands. *Spectrum aggregation* helps better scheduling of spectrum-access requests and enables advanced routing of the spectrum connections. Furthermore, it helps to build a RF-connectivity map that can be used for provisioning redundancy for RF-links and thus an ability to adapt in case of spectrum mobility events as shown in Fig. 9.

Quantified Spectrum Exploitation
When quantified spectrum-access footprints are assigned to RF-entities, it enables us to understand how efficiently the harvested spectrum is exploited. It also enables us to quantify the amount of harmful interference caused to the cochannel receivers by an individual transmitter or a collection of transmitters. Thus, appropriate spectrum assignment schemes could be developed that can *quantitatively* control the spectrum-footprint assigned to each of the transmitters. With a quantified approach to spectrum

consumption, the traditional spectrum-scheduling and spectrum-allocation problems are transformed into a problem of optimizing the spectrum-consumption spaces for a set of spectrum-access requests [26]. This enables us to efficiently exploit the harvested available spectrum.

Maximizing Spectrum Harvesting and Exploitation

With quantified spectrum management, it is now feasible to characterize the impact of various spectrum-access parameters and spectrum management functions, optimize the spectrum consumption, and maximize the number of coexisting wireless services.

The Impact of Spectrum-Access Policy
The available spectrum can get severely limited by the spectrum-access constraints defined by a SAM. In [22], we identified following the key weaknesses of OSA spectrum-access policy

- the lack of knowledge of the receiver positions
- the lack of knowledge of the propagation conditions

As a result, the minimum sensitivity and the maximum transmit-power constraints imposed on the secondary users tend to be very conservative and less than 1 % of the available spectrum can be harvested by the secondary users [22].

Improving Spectrum Harvesting Performance
With quantified spectrum harvesting, it is possible to quantify the lost available spectrum with false positives and *potentially* harmfully-interfered spectrum with the missed detection. The performance of spectrum harvesting can further be improved learning of the fine-grained propagation model parameters which helps to improve estimation of unit spectrum-space opportunities [18].

Improving Spectrum Exploitation Performance
Spectrum allocation and scheduling performance could be studied and optimized based on the amount of spectrum consumed, the amount of harmfully-interfered spectrum, the number of scheduled spectrum-access requests. In [17], we perform several experiments to maximize the spectrum exploitation performance. Figure 12 shows the effect of network range on the number of satisfied spectrum-access requests and the unexploited or available spectrum.

Further Optimizing Spectrum Consumption by Transceivers
Directional transmission and reception helps to improve SINR and thus reduce the spectrum consumption by transceivers. Also, with active role by incumbents, primary transmission power could be increased to enhance SINR at the primary receivers, minimize the receiver consumed spectrum, and maximize the spectrum available for sharing with the secondary services.
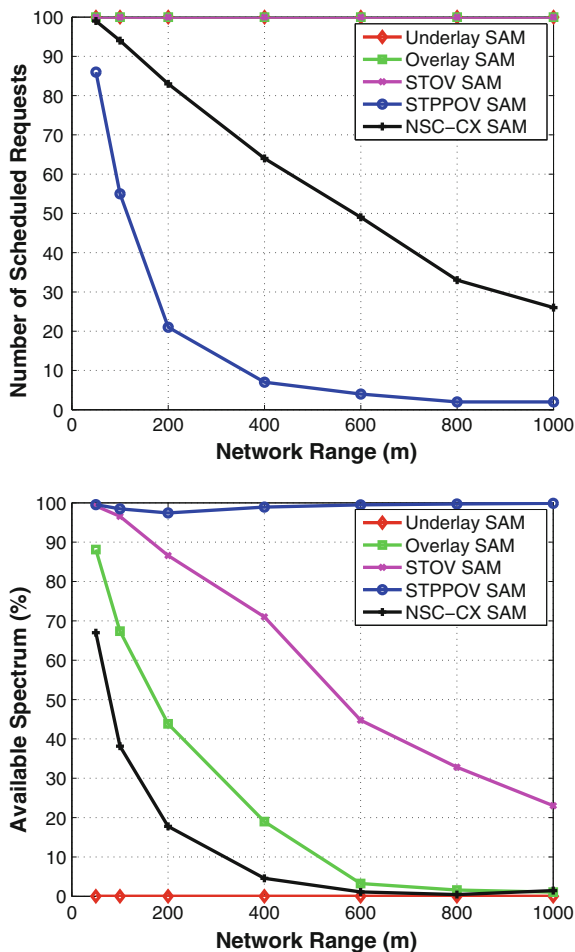
Fig. 12 The performance comparison of multiple spectrum access mechanisms (SAMs) based on the number of satisfied spectrum-access requests and the amount of unexploited spectrum. The experiment involves assigning quantified spectrum-access footprints to 100 spatially overlapping spectrum-access requests within a 4.3×3.7 km geographical region. The experiment assumes a single frequency band. The experiment setup, SAM algorithms, and the findings are described in [17]. We note that quantified spectrum sharing provides key insight into spectrum consumption and enables optimizing the available spectrum for maximizing the number of scheduled spectrum-access requests. More specifically, in this experiment, we observe that fine granular access with smaller network-ranges leads to higher available spectrum and a higher number of satisfied spectrum-access requests [17]

Towards Real-Time Quantified Dynamic Spectrum-Access

QDSA paradigm enables multiple spatially-overlapping RF-systems to efficiently share spectrum in the space, time, and frequency dimensions. One of the key challenges for real-time spectrum sharing is dynamicity of the propagation conditions and spectrum-access parameters. In this section, we address the problem of defining and enforcing spectrum-access rights based on real-time RF-environment conditions and realistic spectrum-access scenarios. Our approach is based on articulating the spectrum-access rights in terms of actual spectrum use in the space, time, and frequency dimensions by an individual transceiver. We divide the spectrum-space into discrete unit-spectrum spaces and estimate the use of spectrum by an individual transceiver in the space, time, and frequency dimensions. The estimation of used spectrum enables estimating the amount of spectrum that is potentially available for sharing with further RF users. A spectrum-access mechanism (SAM) can then define the spectrum-access rights for future spectrum-access requests in terms of the allowed quantified use of spectrum by each of the transceivers. The defined spectrum-access rights can be enforced with estimation of the actual use of spectrum.

In order to estimate the use of spectrum in real time, we employ an external dedicated RF-sensor network. The RF-sensors learn the fine-grained RF-environment and estimate the spectrum-access attributes of the transmitters. To passively estimate the spectrum-access attributes in the presence of cochannel interference, we employ detection, location estimation, and transmit-power estimation algorithms that exploit signal cyclostationarity [27]. The estimates of the spectrum-access parameters by multiple RF-sensors are fused to *estimate the use of spectrum*[10] in the unit-spectrum-spaces within a geographical region.

Estimating the Use of the Spectrum

Using MUSE to characterize the use of the spectrum, the spectrum consumed by individual transceivers and the spectrum available for new users can be estimated.

Sub-Problems

The spectrum consumed by transmitters in the space, time, and frequency dimensions is dependent on the actual (as against the maximum) transmit power and the antenna directionality employed during transmission. Similarly, the spectrum con-

---

[10]As described here, the problem of spectrum-estimation use involves several sub-problems. In the interest of illustrating how these sub-components come together to accomplish real-time characterization of the use of spectrum, we do not discuss the individual sub-problems in detail. In [24], we presented cochannel interference-tolerant algorithms for the purpose of signal detection, received power estimation, and TDOA estimation and provided illustrations of these algorithms.

sumption by receivers in the space, time, and frequency dimensions is dependent on the minimum signal to interference and noise ratio (SINR) required for successful reception and the antenna directionality employed during reception.

In order to passively estimate the spectrum used by all transmitters, it is necessary to

- detect the active transmitters.
- estimate the position of the transmitters.
- estimate the transmit-powers and the radiation patterns.
- estimate the transmitter occupancy *at sample points* across all the unit-spectrum spaces using (4) and (5).

We assume the receiver parameters are specified during the spectrum-access request to the spectrum-access management infrastructure. This is to avoid worst-case assumptions regarding the receiver positions and spectrum-access parameters. In order to estimate the available spectrum, it is necessary to

- estimate the received-power from all the cochannel transmitters at all the cochannel receivers.
- estimate the SINR at each of the cochannel receivers.
- estimate the spectrum-opportunity *at sample points* across all the unit-spectrum-spaces using (9) and (10).

Estimating the Transmitter Spectrum-Access Attributes

The enforcement of the spectrum-access rights requires a passive technique for estimating use of the spectrum. For validating enforcement of spectrum-access footprints, DSA requires a passive technique for estimating use of the spectrum. The techniques exploiting signal cyclostationarity can be applied in passive manner, are tolerant to noise and cochannel interference, and provide good performance under low SINR conditions [24, 27, 28].

A cyclostationary process has statistical properties that cyclically over time [27]. In most of the modern communication systems, periodicity is induced due to signal processing operations like sampling, multiplexing, modulation and coding. The *cycle frequency* of a signal characterizes this periodicity induced in the signal. In our earlier work [24], we presented algorithms exploiting signal cyclostationarity for the purpose of signal detection, received power estimation, and TDOA estimation. Here, we briefly describe the approach in the context of estimation of the location and transmit-power of the transmitters.

Detecting a Transmitter Signal:
For signal detection, we exploit the second-order statistics of each signal through the spectral correlation function (SCF) [24]. The technique only requires the knowledge of a cycle frequency of the transmitted signal.

Estimating the Received Signal Power:
Exploiting the knowledge of SCF for the unit-power version of the transmitter-signal, a least-squares estimation problem is formulated in order to estimate the received signal power in the presence of cochannel signals.

Estimating the Transmit-Power:
In order to estimate the transmit power, each RF sensor estimates the received power from the transmitter. Using the estimated position of the transmitter, the estimated mean path-loss exponent (PLE), and the estimated shadowing loss, the transmit-power is then estimated.

Estimating the Position of a Transmitter:
We employ a method of maximizing the measured SCF by phase alignment for obtaining a TDOA estimate [29]. The method requires knowledge of the transmitter cycle frequency and synchronization between the involved pairs of RF-sensors. For the purposes of TDOA estimation, we choose the candidate RF-sensor pairs based on their *estimated* received power at the RF-sensors. We employ a least-squares position estimation technique in order to estimate position of the transmitter based on a set of TDOA estimates. TDOA estimation is sensitive to multipath and hence it is subject to errors. Due to an overdetermined system of equations, with least squares solution, the error in the location estimation is minimized [18].

Characterizing the Propagation Environment

Estimating the transmit-power from the received-power estimate requires the knowledge of the propagation environment. Since the transmit-power estimation is very sensitive to the path-loss exponent and shadowing, real-time enforcement of a dynamic spectrum-access policy cannot rely on the assumed propagation parameters. In this regard, we estimate the mean path-loss and shadowing variance at a fine granularity using a dense RF-sensor network.

We divide the geographical region into multiple fine-grained unit-sections and consider a log-normal shadow fading environment within each unit-section. To facilitate real-time characterization in the presence of cochannel interference, the path-loss and shadowing variance within *each* of the unit-sections are estimated using monitoring signals with known transmit-power and exploiting signal-cyclostationarity [18].

Estimating Spectrum Occupancy and Spectrum Opportunity

The fusion center uses the estimated spectrum-access parameters of all the cochannel transmitters and estimates the spectrum-occupancy and spectrum-opportunity at the sample-points in each of unit-regions in the geographical area under interest.

Censoring: The spectrum-occupancy estimate quality depends on the performance of the detection, received-power estimation, and geolocation subalgorithms. In this regard, the RF-sensors that are far away from a certain transmitter or the RF-sensors that are very close to the cochannel interference source cannot accurately estimate the transmitter spectrum-access parameters and therefore introduce errors in the estimation of spectrum-occupancy and spectrum-opportunity [18]. To improve the estimation performance, we employ *estimated-SINR*-based censoring of the position estimates and the received-power estimates from each of the RF-sensors.

Incorporating Directionality: Using the shadowing profile information, the estimated transmitter spectrum-access parameters from an RF-sensor, and the known receiver spectrum-access parameters, the fusion center estimates spectrum-occupancy and spectrum-opportunity perceived by *each* of the non-censored RF-sensors. This is especially helpful considering directional transmission. In this case, the received power from a directional transmitter at the individual RF-sensors is different and the spectrum-consumption footprint for the directional transmitter can be accordingly estimated. The directionality of the receiver antennas is considered while estimating the spectrum-opportunity within the unit-regions.

Fusion: For each unit-region, we have multiple spectrum-occupancy and spectrum-opportunity estimates from each of the RF-sensors. In order to facilitate choosing a conservative or an aggressive estimate, we define a *guard-margin factor* and choose a single value from the set of estimates. The value of the guard-margin factor ranges from -1 to 1. The lower boundary represents the most conservative behavior (selecting the minimum spectrum-opportunity estimate from the set) and the upper boundary represents the most aggressive behavior (selecting the maximum spectrum-opportunity estimate from the set).

Illustration

Setup

Let us consider a 4.3×3.7 km geographical region. We estimate the utilized and available spectrum in a single 6 MHz wide frequency band within a unit-time-quantum. The maximum power at any point P_{MAX} is considered 30 dBm or 1 W. The minimum power at any point P_{MIN} is considered to be -125 dBm. The minimum desired SINR for successful reception is considered to be 3 dBm. The ambient noise floor is -106 dBm considering channel bandwidth of 6 MHz. The geographical region is divided into 676 hexagonal cells with side of 100 m. Thus, the total spectrum in the geographical region is 676 Wm^2.

Within each Monte Carlo trial, a network topology with multiple transmitters and receivers is generated. We simulate the large-scale fading effects and the transmitter signals at the physical layer. We implement the algorithms for detection, location estimation, and transmit-power estimation exploiting cyclostationarity in software.

The errors in estimation of the use of spectrum can be captured at various levels. For example, detection errors (missed detections and false positives), TDOA-estimation errors, position-estimation errors, received-power estimation errors, transmit-power estimation errors, and finally the spectrum-occupancy and spectrum-opportunity estimation errors. In this chapter, we primarily illustrate the resulting spectrum-occupancy and spectrum-opportunity estimation errors.

We note that a positive error in the estimated spectrum-opportunity implies loss of actually available spectrum while a negative error may lead to potential harmful interference at some of the receivers in the system. We capture these two effects in terms of the *lost-available spectrum* and *potentially-degraded spectrum*.

Estimating the Spectrum-Access Footprint of a Single Transmitter

Figure 13 illustrates the spectrum consumed by a transmitter within a geographical region. The transmitter is located at (1000, 2000) and is exercising omnidirectional transmission with transmit power of 15 dBm. We can estimate the RF-power from the transmitter at any of the unit-regions in accordance with the estimated fine-grained shadowing profile. For example, at (1000, 2400), the power received from this transmitter is estimated to be −66 dBm. We note that exploiting signal cyclostationarity enables us to estimate the RF power at a point when multiple transmitters are simultaneously exercising spectrum-access in the same frequency band within a geographical region. The estimated use of spectrum by a specific transmitter can be applied for validating a spectrum-access policy.

Fig. 13 Estimating the spectrum-consumption space of an individual transmitter using a RF-sensor network. The figure shows spatial distribution of the transmitter-occupancy in the unit-spectrum-spaces within a geographical region. The transmitter is shown by a *solid square* and the receiver is shown by a *non-solid square*
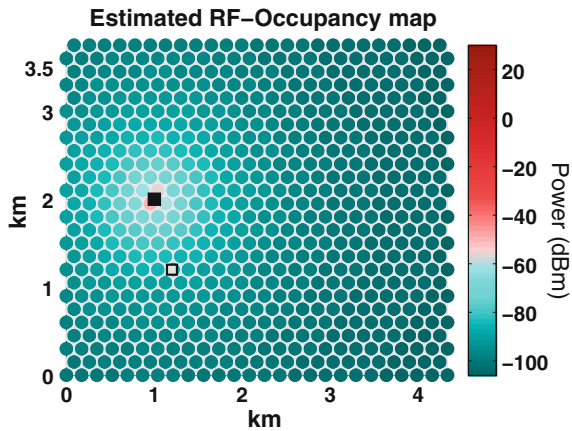
Estimating the Available Spectrum

Next, we estimate the available spectrum in case of multiple cochannel transmitters. We note that the SINR at the RF-sensors can be poor with respect to many of the transmitters due to proximity with other cochannel transmitters; therefore, when there are a large number cochannel transmitters, accurate spectrum consumption estimation requires a large number of RF sensors.

Figure 14 shows the estimation performance with 16 transceiver-pairs and 169 RF-sensors. As the accuracy of spectrum-opportunity estimation depends on the receiver-SINR, we vary SINR at a receiver by varying the receiver's distance from its
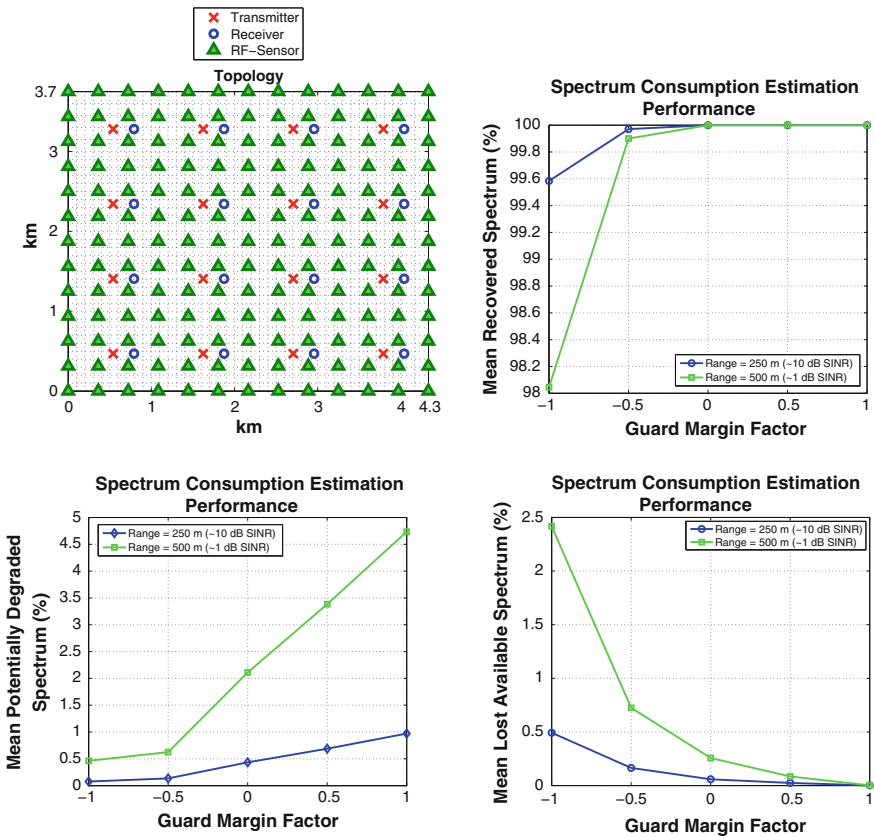
Fig. 14 Estimation of the available spectrum. The RF-sensors estimate the spectrum-access parameters of all cochannel transmitters and estimate spectrum opportunity in the unit regions within the geographical region. The lost-available spectrum and potentially-degraded spectrum capture the positive and negative errors in the estimation of spectrum opportunity, respectively. The 16 cochannel transmitters have distinct cyclostationary signatures. When the SINR at the receivers is lower, the spatial footprint of the receiver-consumed spectrum is larger and the spectrum-opportunity estimation errors are pronounced

transmitter. In Fig. 14, range refers to the distance between a receiver and its transmitter. The 16 transmitters employ distinct cyclostationary signatures, which is achieved through signal design [30]. We observe that 169 RF-sensors used with 16 cochannel networks accomplish reasonable spectrum consumption estimation performance with less than 3 % error assuming no shadowing. Shadow fading introduces significant errors into estimation of the transmit-power and consequently in the estimation of spectrum occupancy and spectrum opportunity. Therefore, we characterize the shadowing profile at a fine granularity.

Characterizing the Shadowing Profile

With a dense RF-sensor network, we characterize the shadowing losses in the fine-grained unit-sections of the geographical region as shown in Fig. 15. We incorporate the mean PLE and the fine-grained shadowing losses while estimating transmit-power for all the transmitters. The shadowing profile is also used in the estimation of spectrum-occupancy and spectrum-opportunity in each of the unit-regions of the geographical region.

From Fig. 16, we observe significant errors in the transmit-power estimation while not employing characterization of the shadowing profile (refer to legend, '*with PLE estimation only*'). In this case, we deployed a RF sensor network with 36 sensors to estimate mean PLE across the geographical region. With the learning approach (refer to legend, '*Learning with 676 RF-sensors*'), the spatial variations in the shadowing loss are estimated with reasonable accuracy and it lowers the errors in transmit power estimation at the RF-sensors.

In comparison with the MUSE-based approach to estimating the use of spectrum, the simplistic primary user transmitter-signal-detection approach identifies much less underutilized spectrum due to constraints on the minimum sensitivity and

Fig. 15 Fine-grained characterization of shadowing. A dense RF-sensor network is applied for estimating mean path-loss index and shadowing variance within the fine-grained sections of a geographical region. The characterization of fine-grained shadowing loss helps to accurately estimate use of the spectrum in the unit-spectrum spaces
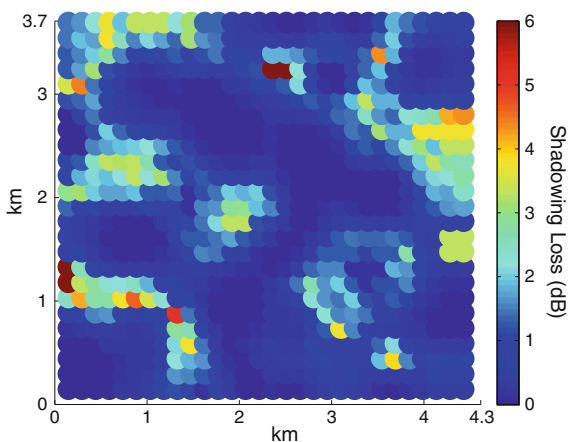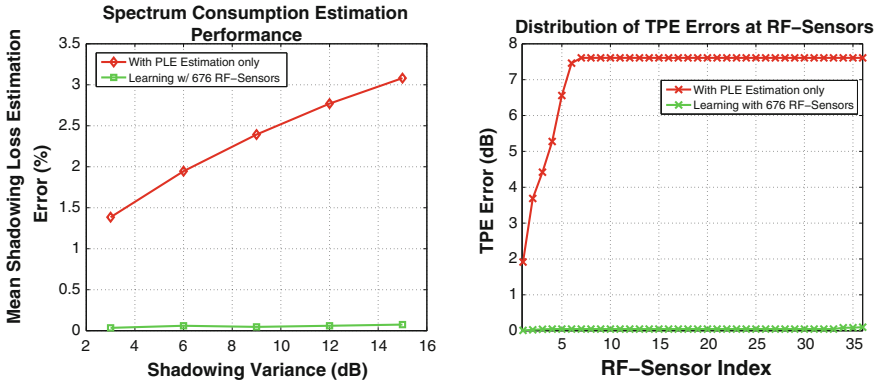
Fig. 16 Characterizing the shadowing losses within a geographical region helps to improve the transmit-power estimation performance. Transmit-power is estimated from the received-power using the estimated mean path-loss exponent (PLE) and the shadowing loss. When shadowing losses are not characterized at a fine granularity (case: 'with PLE estimation only'), it impacts the performance of transmit-power estimation and consequently that of estimating spectrum occupancy and spectrum opportunity. When a dense RF-sensor network is employed (case:'Learning with 676 RF-sensors'), the transmit-power estimation (TPE) performance is improved

maximum secondary user transmit-power defined to statically handle the worst-case RF-environment conditions and spectrum-access scenarios [22, 26].

We acknowledge that the number of RF-sensors required to characterize the propagation environment increases with terrain complexity. In order to reduce the number of RF-sensors, the terrain information from contour and/or satellite maps may be helpful. We pursue this in future work.

Defining and Enforcing a Quantified Spectrum Access Policy

Fine-grained estimation of the use of spectrum in the space, time, and frequency dimension enables spectrum sharing with dynamic spectrum-access rights. Figure 17 describes the overall approach for defining a policy with quantified spectrum-access rights in real time. A spectrum-sharing model may choose to add a guard-margin to the estimated spectrum-opportunity. A spectrum-access mechanism (SAM) may further control the spectrum consumed by the to-be-added transceivers and thereby increase the overall number of spectrum accesses. Thus, the spectrum-access rights for the transceivers are defined based on the real-time spectrum-access opportunity, spectrum-sharing constraints, and spectrum-access etiquette. The rights are articulated in terms of allowed use of the spectrum in the space, time, and frequency dimensions and accordingly spectrum-access parameters for the transceivers can be inferred.
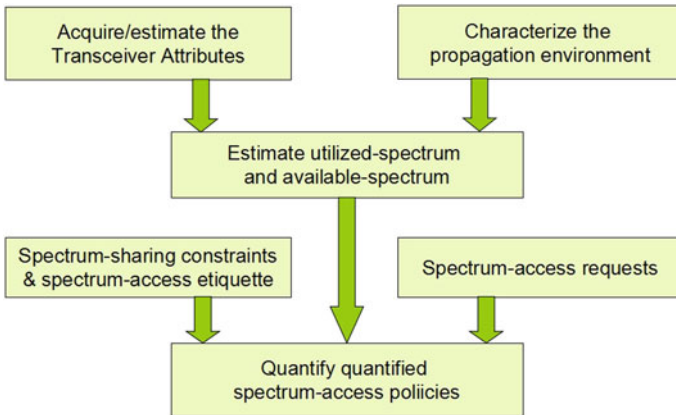
Fig. 17 Illustration of defining spectrum-access rights based on real-time use of the spectrum. By passively estimating the spectrum-access attributes of the transmitters and by characterizing the propagation environment, the use of spectrum by transmitters and the available spectrum can be estimated. Based on spectrum-sharing constraints and etiquette, quantified spectrum-access rights can be defined and enforced in real time

Using the estimation of the spectrum-opportunity across the unit-regions in a geographical region:

- we can choose the best frequency band based on the spectrum opportunity at the potential transmitter location across multiple bands.
- we can define the transmit power based on the spectrum opportunity at the transmitter location and make more efficient use of spectrum while ensuring non-harmful interference to all the receivers in the system.
- we can precisely control sharing of the spectrum among multiple networks. For example, let us consider spectrum opportunity at an arbitrary point to be -20 dBm. We can allow spectrum access to a single transmitter such that the RF power from this transmitter at this location is -20 dBm or we can allow two or more transmitters (with lower power levels) access to the spectrum while ensuring the same constraint.

Thus, estimating the spectrum-opportunity enables efficient allocation and scheduling options for spectrum management.

Benefits of the Quantified Dynamic Spectrum Sharing Paradigm

Quantified dynamic spectrum sharing paradigm offers several benefits from technical, operational, regulatory, and business perspectives.

Benefits Towards Spectrum Management

- MUSE helps to characterize and quantify the use of spectrum at the desired granularity in the space, time, and frequency dimensions. MUSE helps to query how much spectrum is consumed by a single transceiver or any logical collection of the transceivers.
- MUSE helps to compare, analyze, and optimize the performance of spectrum management functions. For example, it is possible to quantitatively analyze performance of ability to recover the underutilized spectrum of various spectrum sensing algorithms (like energy-detection, cyclostationary feature detection) or various cooperative spectrum sensing infrastructures based on the recovered spectrum space, lost-available spectrum space, and potentially-incursed spectrum space.
- MUSE can help to estimate the available spectrum and the exploited spectrum. Thus, it offers the ability to define the spectrum-access rights based on the real-time RF-environment conditions. Using the real-time RF-environment conditions helps to get rid of conservative assumptions and make an efficient use of the spectrum.
- MUSE and the presented spectrum-discretization approach facilitates adaptation of the spectrum management functions under dynamic RF environment conditions and dynamic spectrum-access scenarios.

Benefits Towards Dynamic Spectrum Access

- MUSE enables us to articulate, define, and enforce spectrum-access rights in terms of the use of spectrum by the individual transceivers.
- From an operator's perspective, the guard space could be effectively controlled. The discretized spectrum management approach enables us to easily map a guard margin value to the amount of the inexercisable spectrum. Thus, depending on the user-scenario, spectrum sharing behavior could be changed with visibility into the implied availability of the spectrum.
- Another advantage from an operational perspective is controlling the granularity of spectrum sharing. With discretized approach to spectrum management, the dimensions of a unit-spectrum-space imply the granularity of sharing of the spectrum resource.
- With characterization of spectrum-access opportunity in the space, time, and frequency, MUSE provides the ability to share spectrum without defining a boundary across spectrum uses.
- The discretized spectrum management can be applied independent of the spectrum sharing model. Thus, it can be applied in case of the completely dynamic spectrum sharing model like pure spectrum sharing model or even in case of a conservative spectrum sharing model like static spectrum sharing model.

- From a regulatory perspective, MUSE offers the ability to enforce a spectrum-access policy and ensure protection of the spectrum rights of the users. As the spectrum-access rights are identified at the granularity of a single transceiver, the violations by a particular transmitter, or the harmful interference for the individual receivers could be characterized and quantified.

Benefits Towards Spectrum Trade

- The quantified approach brings in simplicity in spectrum trade. It enables easier understanding and interpretation of the outcomes; thus, it requires less skills of its users.
- The quantified approach enables to investigate the amount of the spectrum that can shared and evaluate the potential for a business opportunity.
- From a business development perspective, spectrum sharing models devised using a quantified approach enable spatial overlap of multiple RF-systems and avoid spatial fragmentation of coverage. This is important for defining new services exercising shared spectrum-access rights.
- Aggregation of fine granular spectrum sharing opportunities gives incentives for spectrum-owners to extract more value out of their underutilized spectrum; a bigger spectrum-pool is attractive for secondary users as well. Thus, characterization of the fine granular spectrum-access opportunities enables building a bigger spectrum-resource pool.

Finally, a note on the real-time quantified dynamic spectrum access, we encourage defining and enforcing spectrum access rights in real-time. Although this requires a dedicated spectrum management infrastructure, it potentially brings in new business models along with flexible and efficient use of the spectrum and an ability for automated regulation of the dynamic spectrum-accesses.

Chapter Summary

In this chapter, we introduced a quantified approach to spectrum sharing wherein spectrum-access rights are defined at the granularity of an individual transceiver. In order to articulate, characterize, and quantify the use of spectrum, spectrum-space is discretized in the space, time, and frequency dimensions. From a technical perspective, discretization of the spectrum-space enables quantifying, estimating, analyzing, and optimizing the spectrum consumed by the transceivers and the available spectrum. A quantified dynamic spectrum-access policy facilitates regulation of spectrum-access and protection of spectrum rights. QDSA paradigm makes it possible for an operator to control and optimize the granularity of spectrum sharing and spectrum-access.

In addition to solving technical and regulatory obstacles to dynamic spectrum sharing, the quantified dynamic spectrum sharing approach offers several benefits from a business perspective. The quantified approach transforms spectrum into a commodity that can be exchanged with service providers, and a resource that can be precisely controlled for making an efficient use. It facilitates easier interpretation and requires less skills of its users. Furthermore, quantified dynamic spectrum sharing enables spatial overlap of multiple RF-systems while protecting their spectrum rights. This is important for avoiding spatial fragmentation and efficiently sharing spectrum enabling a large number of fine-grained spectrum-accesses within a geographical region. The quantified approach to spectrum management facilitates building a bigger spectrum pool with spectrum aggregation and thus enables building attractive business models for dynamic spectrum access.

References

1. Akyildiz, I.F., Lee, W.Y., Vuran, M.C., Mohanty, S.: NeXt generation/dynamic spectrum access/cognitive radio wireless networks: a survey. Comput. Netw. J. (Elsevier) **50**, 2127–2159 (2006)
2. Zhao, Q., Swami, A.: A survey of dynamic spectrum access: signal processing and networking perspectives. In: IEEE International Conference on Acoustics, Speech and Signal Processing, Apr 2007
3. Wireless Innovation Forum: Dynamic Spectrum Sharing Annual Report 2014. http://groups.winnforum.org/d/do/7881, Jan 2015
4. Federal Communications Commission (FCC): Proposes Creation of New Citizens Broadband Radio Service in 3.5 GHz. Further Notice of Proposed Rulemaking 14-49, Apr 2014
5. Federal Communications Commission: In the Matter of Unlicensed Operation in the TV Broadcast Bands: Second Report and Order and Memorandum Opinion and Order. FCC. Std. 08-260, Nov 2008
6. Harrison, K.I., Mishra, S.M., Sahai, A.: How much white space capacity is there? In: New Frontiers in Dynamic Spectrum Access Networks, DySPAN (2010)
7. Mishra, S.M.: Maximizing available spectrum for cognitive radios. Ph.D. dissertation, EECS Dept., Univ. of California, Berkeley, Jan 2010
8. Kone, V., Yang, L., Yang, X., Zhao, B.Y., Zheng, H.: On the feasibility of effective opportunistic spectrum access. In: Internet Measurement Conference (IMC) (2010)
9. Visotky, E., Kuffner, S., Peterson, R.: On collaborative detection of TV transmissions in support of dynamic spectrum sharing. In: New Frontiers in Dynamic Spectrum Access Networks, Nov 2005
10. Ghurumuruhan, G., Li, Y.: Agility improvement through cooperative diversity in cognitive radio. In: IEEE Global Telecommunications Conference, Nov 2005
11. Mishra, S.M., Sahai, A., Broderson, R.W.: Cooperative sensing among cognitive radios. In: IEEE International Conference on Communications, June 2006
12. Raman, C., Kalyanam, J., Seskar, I., Mandayam, N.: Distributed spatio-temporal spectrum sensing: an experimental study. In: Proceedings of Asilomar Conference Signals, Systems, and Computers, Nov 2007
13. Ganesan, G., Li, Y., Bing, B., Li, S.: Spatiotemporal sensing in cognitive radio networks. IEEE J. Sel. Areas Commun. **26**, 52 (2008)
14. Do, T., Mark, B.L.: Joint spatial-temporal spectrum sensing for cognitive radio networks. In: Proceedings of 43rd Conference on Information Systems and Sciences, Mar 2009

15. International Telecommunication Union Radiocommunication Sector (ITU-R): Definition of Spectrum Use and Efficiency of a Radio System. ITU-R Recommendation SM-1046-2 (2011)
16. IEEE 1900.5.2 Workgroup: Standard Method for Modeling Spectrum Consumption. http://grouper.ieee.org/groups/dyspan/5/index.htm (2015)
17. Khambekar, N., Spooner, C.M., Chaudhary, V.: On improving serviceability with quantified dynamic spectrum access. In: IEEE Symposium on Dynamic Spectrum Access Networks (DySPAN), Apr 2014
18. Khambekar, N., Spooner, C.M., Chaudhary, V.: Extended Report: Spectrum Consumption Estimation. http://www.cse.buffalo.edu/~nvk3/sce.pdf (2014)
19. Khambekar, N., Spooner, C.M., Chaudhary, V.: MUSE: A Methodology for Characterizing and Quantifying the Use of Spectrum. http://arxiv.org/abs/1508.02307 (2015)
20. Khambekar, N., Spooner, C.M., Chaudhary, V.: MUSE: A methodology for quantifying spectrum usage. In: IEEE Global Communications Conference (GLOBECOM), (2016)
21. Erpek, T., Lofquist, M., Patton, K.: Spectrum Occupancy Measurements: Loring commerce centre, Limestone, Maine, 18–20 Sept 2007. Shared Spectrum Company Report (2007)
22. Khambekar, N., Spooner, C.M., Chaudhary, V.: Characterization of the missed spectrum-access opportunities under dynamic spectrum sharing. In: IEEE COMSNETS (2015)
23. Khambekar, N., Chaudhary, V., Spooner, C.M.: Estimating the use of spectrum for defining and enforcing the spectrum-access rights. In: IEEE MILCOM (2015)
24. Spooner, C.M., Khambekar, N.: A signal-processing perspective on signal-statistics exploitation in cognitive radio. In: Invited Position Paper for the IEEE International Conference on Computing, Networking, and Communications (ICNC) (2012)
25. Liang, Y.C., Zeng, Y.H., Peh, E., Hoang, A.T.: Sensing-throughput tradeoff for cognitive radio networks. IEEE Trans. Wirel. Commun. **7**(4) (2008)
26. Khambekar, N., Spooner, C.M., Chaudhary, V.: Maximizing Spectrum Availability and Exploitation: How to Maximize Spectrum Sharing Benefits to the Incumbents? http://arxiv.org/abs/1407.7134 (2014)
27. Gardner, W.A.: Statistical Spectrum Analysis: A Nonprobabilistic Theory. Prentice Hall, Englewood Cliffs, NJ (1987)
28. Spooner, C.M., Nicholls, R.B.: Spectrum sensing based on spectral correlation, Ch. 18. In: Fette, B. (ed.) Cognitive Radio Technology, 2nd edn. Academic Press (2009)
29. Gardner, W.A., Chen, C.K.: Signal-selective time-difference-of arrival estimation for passive location of man-made signal sources in highly corruptive environments, Part 1: Theory and method. IEEE Trans. Signal Process. **40**(5), 1168–1184 (1992)
30. Sutton, P.D., Nolan, K.E., Doyle, L.E.: Cyclostationary signatures for rendezvous in ofdm-based dynamic spectrum access networks. New Frontiers in Dynamic Spectrum Access Networks (DySPAN), Apr 2007

A Case Study of Cognitive Radio Networks: Secure Spectrum Management for Positive Train Control Operations

K.R. Damindra S. Bandara, Anthony Melaragno, Duminda Wijesekera and Paulo Costa

Introduction

Positive Train Controller (PTC) is a radio-based control system designed to improve the safety of train operations. Its safety objectives include avoiding train-to-train collisions and train derailments and ensuring railroad worker safety. Once implemented, PTC will guarantee safe train operations by ensuring proper separation between trains traveling on the same track, and by providing automatic braking when the train is moving at speeds higher than the mandated train speed for the track.

Failure to follow speed restrictions might incur in major accidents as illustrated by the train derailment in Philadelphia in May 2015 [1]. In that particular event, the train was traveling at more than twice the mandated speed, which was considered the primary cause of the derailment. In another example, the Amtrak crash that killed two railroad workers in April 2016 was a result of the train operator not following the safety rules imposed upon him to protect workers [2]. In systems without automated control, the operator is solely responsible for following operational standards and procedures. Automated train control systems such as PTC add an extra, ultimate layer of safety of the crew, passengers, and railroad workers. If PTC was implemented, neither accident would have happened.

K.R.D.S. Bandara (✉) · A. Melaragno · D. Wijesekera · P. Costa
George Mason University, Fairfax, VA, USA
e-mail: kbandara@masonlive.gmu.edu

A. Melaragno
e-mail: amelarag@masonlive.gmu.edu

D. Wijesekera
e-mail: dwijesek@gmu.edu

P. Costa
e-mail: pcosta@gmu.edu

For the initial case study of this Chapter, we adopted a particular PTC system called I-ETMS. After much consideration, the railroad community has agreed to use two separate radio networks to provide the stated design objectives of PTC; the signaling network and the wayside interface network. The signaling network provides two-way communication between a train and the command center that coordinates train movements, referred to as the back office. The fundamental premise of PTC is that any train governed by PTC should receive an authorization to enter the next track segment, referred to as a block in railroad terminology. In the envisioned radio network, this authority is communicated using a radio, referred to as a signaling point. The signaling point communicates with the back office that in turn consults the so-called dispatching center (residing at a central location) that provide permissions (and therefore schedules) for requesting trains. Also, the signaling network relay messages between the train and the back office to provide functionality such as sending the train location reports to the back office and sending train consists (geometry and load characteristics) from the back office to the train. The wayside interface network monitors and transmits the status of wayside devices (e.g., switch alignments, broken rail detectors or flood detector, etc.) through its wayside interface units (WIU). Each node of the WIU network uses radios periodically to broadcast the status of its associated wayside devices to trains and, optionally, to send copies of these status messages to the back office.

For PTC-controlled trains to obtain the safety benefits enabled by the signaling network and the WIU network, they must consistently communicate with both networks. These trains must have two-way data radios designed to process commands submitted by the signaling network and to ensure full adherence to the constraints imposed by WIU radio broadcasts. Standards are emerging to regulate radio communications between entities of the described architecture [3, 4]. For the PTC communication system American railroads use 217–219 MHz (uplink) and 221–222 MHz (downlink) [5]. This limited bandwidth should be appropriately shared between the WIU network and the signaling network and all trains that communicate inside every radio network cell. Due to the bandwidth limitations, proper cell design and dynamic channel management is required to avoid interference between allocated channels. Interference decreases the signal to noise ratio (SNR) for received signals and increases the associated bit error rate (BER).

PTC is a safety critical application, and thus operates under fail safe regulations. As an example, if something happens out of normal behavior the train controller will automatically apply brakes and slow down or stop the train. Since PTC uses wireless communications to obtain safety information and status from the infrastructure, an attacker can shut down the train operation simply by jamming the link. In other words, the extra safety precautions inherent to PTC operations brings vulnerabilities that might allow malicious players or even environmental factors (e.g. some natural phenomena interfering with the transmission) to trigger unwanted disruptions. Addressing these vulnerabilities is the key motivation for the architecture proposed in this chapter.

An important aspect of our architecture is the use of cognitive radio (CR) technology to bring built-in intelligence into communication radios, making them capable

of identifying malicious activities and of dynamically changing radio parameters in response to such events. The CR technology we propose is slightly different from the conventional dynamic spectrum allocation CR because PTC leverages the fact that a limited part of the spectrum was already assigned for train operations. For example, when there is a need for channel hopping, infrastructure CR units (either at signaling points or WIUs) first check whether there are channels available in the 220 MHz frequency band. It only resorts to channels in other bands (160 and 900 MHz can be used for train communication) when no channels are available at 220 MHz bandwidth. This behavior significantly enhances performance, reliability, and responsiveness of the CR network.

In this chapter, we present a multi-tiered cognitive radio network architecture that is being developed to address the spectrum allocation for PTC operations. In addition to efficient spectrum sharing, the proposed CR architecture is also capable of detecting spectrum abuse, detecting any potential cyber threats to PTC operations, and improving the security of radio communication between the PTC nodes. We are developing a prototype of the CR network in which radio nodes are implemented using Ettus [6] N210 Universal Software Radio Peripheral (USRP)s and experiments are performed in an electromagnetically controlled environment provided by noise cancellation chambers (i.e. Faraday cages). The evaluation tests of the prototype include measuring the performance of the CR network on the PTC timing constraints, as a means to ensure that the added functionality does not prevent the whole system from being compatible with these limitations.

This chapter is organized as follows. Section "Related Work" discusses the most pertinent work related to this area. Section "Cognitive Radio Network" describes the PTC cognitive radio network. Section "Cognitive Radio Architecture" explains the internal architecture of cognitive radios, including their components. Section "Implementation and Testing" provides the design, implementation and testing of the cognitive radios and section "Conclusions" conveys our conclusions.

Related Work

The Positive Train Control System

In 1990, Union Pacific and General Electric partnered to develop a train safety system called Precision Train Control. The system was based on the concept of a Moving Block, which is a safety zone encompassing a train as well as a distance from both its front and its back that depends on the location and speed of the train. Precision Train Control is the precursor to the current PTC concept.

Before PTC, most train lines solely relied on the human crew to obey safety rules in operating trains, which in turn had to rely on voice radio to receive authorizations to enter blocks and to inform when they left a block. The problem with this approach is its ineffectiveness in preventing factors such as negligence of a crew member from

causing incidents and accidents. PTC provides a safety system that will automatically stop the train when an operator fails to act on time. After the Chatsworth train collision [7] of 2008, The Federal Government mandated to implement PTC in all the train lines within the USA by the end of 2015.

Hartong's thesis [8] suggests developing a block-based movement authority system for PTC. In his proposed approach, a train line is divided into fixed size blocks, and PTC will allow only one train to occupy a block at a time. To enter a block, the train must send a request in advance to the back office, and can only enter after receiving the associated response, with both the request and the response transmitted via voice radio. To provide the response, the back office coordinates with the dispatching center (currently uses an automated dispatching system to help human operators) which has the details of where all the trains are residing at a time. Based on this information, the back office decides if it is safe to allow the train to enter the next block.

Abadie's thesis [9] proposed a risk engine for PTC to avoid high-risk environments through adjustments in train operations. Abadie's thesis describes the risks of communication link availability due to environmental factors and adversarial attempts due to SDR vulnerabilities. Abadie's work only considers risk as a function of availability. Consequently, his work should be enhanced to incorporate factors such as congestion in the radio bandwidth and integrity and confidentiality when calculating the operational risk. The cognitive radio approach that we describe in this chapter evaluates risk as a function of both availability and integrity and provides dynamic reconfiguration to mitigate it.

Cognitive Radio

Mitola [10] introduced the concept of a Cognitive Radio (CR) that integrates agent-base control, natural language processing, and machine learning. He also proposed using CRs for spectrum pooling, the principle that allows multiple users to share a radio frequency range. Currently, spectrum pooling is used for secondary users who do not own a licensed spectrum to gain access to parts of the licensed spectrum not used by primary users. When the authorized (primary) user claims a channel, secondary users must vacate them [11].

Dynamic Spectrum Access using CRs reduce the congestion in unlicensed bandwidths because unlicensed users get opportunistic access to licensed spectrum. It also allows vendors to manage their wireless applications without purchasing licensed bandwidth for the user communications. Therefore, dynamic spectrum allocation has been used widely by different vendors in various applications including vehicular networks. In [12], Chen presents the use of dynamic spectrum allocation in vehicular networks and shows the use of creating a cooperative CR network between vehicles in fleets like Google driver-less cars. These cars can use a CR network to exchange information such as traffic conditions with other vehicles to increase driving efficiency and improve safety.

Singh [13] presents the use of CRs in Vehicular *ad hoc* networks (VANET) for Vehicle to Vehicle (V2V) and Vehicle to Infrastructure (V2I) communications. He shows how the *Intelligent Transportation System's frequency spectrum* can be overcrowded when more vehicles use VANET applications and how using CRs to use other application's spectrum as a secondary user can avoid congestion and increase the safety of traveling.

Many research publications exist in the area of CR for vehicular applications. All these applications focus on providing opportunistic licensed spectrum access to unauthorized users using CR. Our usage of CR network differs from these as we propose to use CRs to manage spectrum dynamically within the bands allocated for PTC usage based on traffic patterns and to use spectrum characteristics to ascertain if the legitimate spectrum users are under attack.

Ammana et al. [14] show how to use CR to allocate spectrum dynamically for PTC to improve its usage. They list necessary PTC spectrum requirements and show how CR can be used to meet them. Their work concentrates on detecting and switching to better channels to maintain link availability. They also discuss how to support interoperability between train tracks from different companies by synchronizing spectrum parameters and protocols. Although they have explained how CR technology can benefit in PTC environment, they have not presented any implementation. Further, their design does not comply with PTC specifications. Our work differs from Ammana's work by developing a CR network that operates according to an enhanced PTC protocol. Our CR network holistically manages the spectrum efficiently in addition to providing a blueprint for a CR-based IDS system that can prevent spectrum abuse by unlicensed radios.

Standard CR approaches attempt to find white spaces in the unused licensed spectrum, focusing on factors such as how spectrum sensing is performed, which algorithms can be used, their associated level of security and others relevant to dynamic spectrum allocation [11, 15]. A key goal of these approaches is to allow the lawful use of the licensed spectrum as secondary users. Thus, the proposed implementations include spectrum sensing, generally relying on artificial intelligence techniques to detect the most appropriate frequency channel and frequency hopping. Our CR differs from these, and from other CRs known to us, in the following aspects:

1. Operates according to PTC protocol
2. Allows for dynamic spectrum management within a licensed spectrum band
3. Provides cryptography-based security and detects any malicious activities

In [16], we described the primary design of this cognitive radio network. In this book chapter, we present the detailed design of the proposed cognitive radio network, its communication protocols, parts of its implementation, and associated testing and evaluation. We believe our research can contribute to the existing cognitive radio research as we are exploring the applicability of moving the threat detection and security enhancement capability of an application to the radio layer using the concept of a cognitive radio.
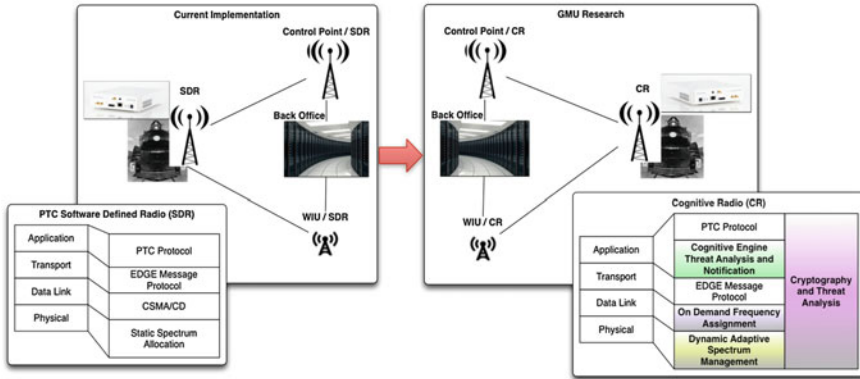
Fig. 1 Cognitive radio network

Cognitive Radio Network

Figure 1 shows the architecture of our proposed cognitive radio network. In our model all PTC Software Defined Radio (SDR) (i.e. for trains, signaling points and WIUs) will be converted to cognitive radios. These CRs will efficiently share spectrum, detect spectrum abuse, and any potential cyber threats. Our architecture follows a Open System Interconnection (OSI) structure. Security and threat detection span the physical layer and the application layer. The cognitive radio will change the current PTC radio in the following ways:

- **Physical**: Detect anomalies in the signal and change radio parameters dynamically to improve spectrum sharing efficiency.
- **Data Link**: Assign frequencies based on the train traffic density and train speed, as higher speed trains will communicate with more radios at a higher rate.
- **Application**: Detect potential threats to the radio network.

Cognitive Radio Architecture

As shown in Fig. 2, each cognitive radio in our system is designed to have two tiers. The lower tier consists of cognitive engines, which are designed to manage spectrum and ensure secure communication. These lower tier cognitive engines identify any risks to their local operations and inform it to the master Cognitive Engine (CE). The master CE collects information from lower tier CE and assesses the risk to PTC operations.
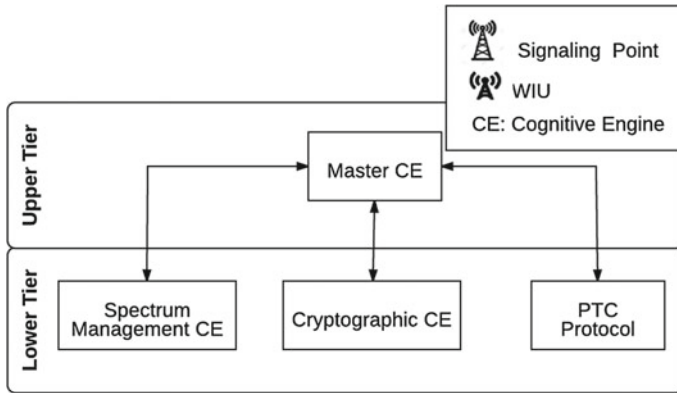
Fig. 2 Internal architecture of a cognitive radio node

The spectrum management CE is responsible for ensuring the reliability of the communication link. It does that by performing the following functions:

1. Measures the variations in the *Signal to Noise Ratio (SNR)* and in the *Bit Error Rate (BER)* of the signal, and informs any abnormality to the master CE.
2. Manages spectrum congestion based on the train priority, train density, and bandwidth availability.
3. Changes the frequency, power, and modulation schema dynamically.
4. Ensures proper message transmission and reception.

The cryptographic CE ensures the security of the communications between PTC entities. Its primary functions are:

1. Change the cryptographic keys periodically to protect broadcast messages from malicious attacks.
2. Analyze messages to detect any potential impostures, replays, bandwidth exhaustion attacks, and unauthorized radios broadcasting in this limited access bandwidth (one way of achieving this is by identifying the location of the user).
3. Report suspicious activities to the master CE.

The master CE holistically evaluates the operational risk, determines a potential response and communicates the necessary actions to the appropriate lower tier cognitive engines. The master CE bases these decisions on factors such as Doppler effect, environmental conditions (e.g., precipitation), foliage, and multi-path effects, as well as historical information related to them. Also, the master CE shares appropriate information among its peers. Taken together, CRs create a secure cognitive radio network for PTC.

The Master Cognitive Engine

The two major functions of the master CE are to assess any possible risks affecting the secure message communication of PTC operations and to take the appropriate actions to minimize it by using information collected from the lower tier cognitive engines. For example, when the cryptographic CE detects a sequence of CRC failures it does not have sufficient information to decide if this failure is an intentional message modification or an unintentional corruption due to environmental effects. Conversely, when the master CE learns about failures, can compare the signal quality information it gets from the spectrum management CE against historical data, and use that as a basis to distinguish between intentional or unintentional corruption while reducing the number of false positives.

Once a risk is detected, the master CE determines appropriate mitigation actions. The risk reduction strategy includes the following couple of actions.

1. **Inform the risk to the PTC controller**: The master CE can attribute the loss of information to the disturbance in the communication link and radio-based attacks to the PTC controller or any other logging mechanisms. In all cases, it will let the PTC controller knows its assessment.
2. **Change radio parameters**: If the loss was due to disturbances in the radio channel and if the communication link appears unavailable, important messages from the signaling stations and critical infrastructure may not arrive at the PTC controller on time. The master CE can assess the information that it gets from the spectrum management CE to categorize the loss as a selective jamming or interference, if that is the case, the master CE can change the transmission frequency to a better channel. Similarly, the master CE can modify the power and modulation parameters dynamically to provide a better communication medium for the PTC transmission. After changing the radio settings, the master cognitive will inform it to the other nodes in the PTC network.

Spectrum Management Cognitive Engine

The primary objective of the spectrum management CE is to maintain the desirable quality of service in the radio communication between various PTC nodes (i.e., Trains, WIUs, and Signaling Points). Its main functions are to detect any variations in the spectrum quality, manage spectrum congestion, and dynamically reconfigure the radio parameters such as frequency, transmission power and modulation. Also, it performs receiving and transmitting messages from other nodes.

As shown in Fig. 3 the functionality of the spectrum management CE is implemented as a collection of separate functions. These functions act independently but
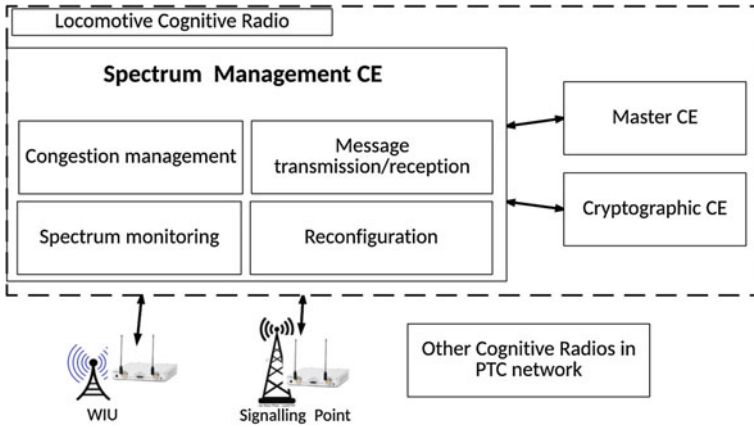
Fig. 3 Functionality of the spectrum management CE

complement each other. For example, the spectrum congestion management functionality makes use of the spectrum sensing, dynamic parameter reconfiguration, and message transmission/reception functions.

- **Rail Traffic Congestion Management**: The objective here is to assign frequency channels dynamically to reduce the spectrum congestion when the train density is high. More details about this function are provided in section "Enhancing the PTC Protocol with Dynamic Spectrum Allocation". This function leverages all the other functions to provide proper congestion management.
- **Spectrum Sensing**: Spectrum sensing detects any abnormalities in the signal such as drops in SNR or BER, and failure to receive messages within the specified time interval. It then communicates these abnormalities to the master CE.
- **Reconfiguration**: Based on the decisions from the master CE, the spectrum management CE will modify radio parameters such as center frequency, channel bandwidth, transmission power, or modulation scheme. This capability plays a primary role in implementing the frequency congestion management function. When the spectrum management CE receives a request to change any radio parameter, it checks if the modification fits within the accepted range and, if positive, proceeds with the changes. Otherwise, it will return an alert to the master CE.
- **Message Transmission/Reception**: The spectrum management CE is also responsible for the transmission of messages from one CR to another. It periodically reads, encodes, modulates, and transmits messages from all the CE components that need to be transmitted to other nodes. At the receiving end, when a message arrives, it demodulates, decodes and passes to the application level CE.

Frequency Management

PTC operations use 217–219 MHz spectrum bands for the up-link and 221–222 MHz spectrum bands for the down-link. This limited bandwidth should be appropriately shared between the WIU and the signaling networks to provide safe navigation.

In [17] and [18], we have analyzed the PTC communication network to ascertain the sufficiency of bandwidth. Our results show that the static allocation of frequency channels for PTC radios is inefficient because it limits the maximum number of trains that can be operated at the same time. Section "Constraints on Radio Bandwidth" lists the constraints on frequency planning for the PTC-based radio network.

Constraints on Radio Bandwidth

To perform spectral analysis, we made the following assumptions:

1. Train tracks follow a linear path
2. Communication between trains and the back office is continuous. That is, the so-called *dark territories* do not exist.

We consider the following to be the primary constraints:

1. The total bandwidth required for the signaling and WIU channels must be less than the total available bandwidth. In other words, if b is the bandwidth of a channel, N is the total number of channels (including both the signaling and WIU channels), g is the guard band and B is the total bandwidth, then:

$$N * b + (N + 1) * g \leq B \qquad (1)$$

and if the number of control channels is N_c and the number of WIU channels is N_w, then:

$$N \geq N_w + N_c \qquad (2)$$

Bit rate of the channel is a function of its bandwidth and modulation scheme. The maximum possible modulation is a function of the SNR. Hence, if br is the channel bit rate, mod is the modulation scheme, SNR is the signal to noise ratio, the transmission band-pass filter has a roll off of α and Channel bandwidth b, then:

$$b = rate_{symbol} * (1 - \alpha) \qquad (3)$$

and

$$rate_{symbol} = br/bits\_symbol \qquad (4)$$

where *bits_symbol* is the number of bits per symbol.
Using Eqs. 3 and 4,

$$br = (b \cdot bits\_symbol)/(1 - \alpha) \qquad (5)$$

The number of bits per symbol (BPS) depends on the modulation scheme used. For Binary Phase Shift Keying (BPSK), the number of bits per symbol is 1 and for Quadrature Phase Shift Keying (QPSK), the number of bits per symbol is 2. The maximum possible modulation scheme depends on the Signal to Noise ratio (SNR) at the receiver. Using Shannon's theorem the channel capacity C [19] can be calculated by:

$$C = b * log_2(1 + S/N)$$

where b is the channel bandwidth.
Using Hartley's Law for a M-ary error-less channel, it can be shown that:

$$C = 2 * b * log_2(M)$$

Comparing Shannon's theorem with Hartley's law would yield:

$$b * log_2(1 + S/N) = 2 * b * log_2(M)$$

$$M = \sqrt{1 + S/N} \qquad (6)$$

2. In general, wayside devices are located close to each other, which causes interference avoidance between neighboring WIUs to be an issue. On the other hand, WIUs that are close can be grouped and connected to a single tower, which will then transmit the beacons from each WIU interleaved in time. The number of WIUs that can be attached to a tower should be less than the maximum number of beacons interleaved to the transmit signal. Therefore, if r is the beacon rate of a WIU, p is the WIU packet size, and M is number of WIUs that can be connected to one tower, then:

$$r * p * M \leq br \qquad (7)$$

3. Bit rate of a signaling channel should be sufficient to propagate control signaling between the back office and the train. Typically, trains exchange data (such as position information and track consists) with the back office. Also, trains exchange handover information with the back office during the handover process. Therefore, the control packet rate is estimated as:

   - Signaling message rate when handover happens: dispatching messages ($signaling_{ho}$)
   - Control rate at normal operations ($signaling_{normal}$)

If the speeds of trains that the signaling point has to support is $V[]$, control packet rate is $rate_{signaling}$ and control packet size is pc, then:

$$rate_{signaling} = max(rate_{signaling\_ho}, rate_{signaling\_normal})$$

The bit rate of a control point br should be greater than the maximum signaling traffic rate:

$pc * rate_{signaling} \leq br$

However, it is also true that:

$rate_{signaling} = f(v)$

Therefore,

$$for \quad all \quad v : V[] ==> pc * f(v) \leq br \qquad (8)$$

4. Because train routes follow a linear path, all signaling points and WIU towers also follow a linear geometry as shown in Fig. 4. In such a formation, there can be more than one WIU tower located in the cell area of a signaling point. Therefore, if the maximum frequency to avoid adjacent channel interference is $freq_{max}$, the number of signaling points in the track is n, and the number of WIU towers in one signaling cell area is m, then the following statements are true.

- The interference between adjacent signaling point transmissions should be minimized:

$$for \quad all \quad i : [0 : n] ==> |f(C_i) - f(C_{i+1})| \geq freq_{min} \qquad (9)$$

- The interference between WIU towers in the same signaling point area should be minimized:
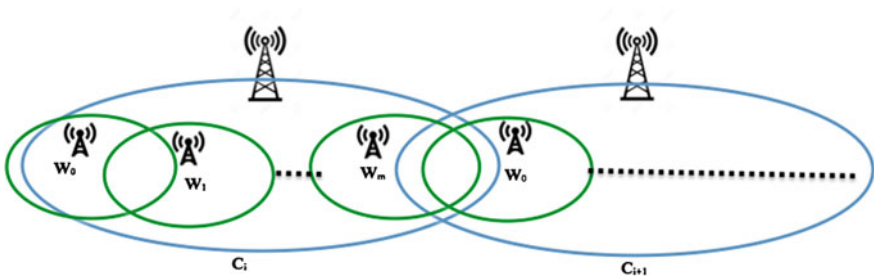
Fig. 4 WIU and signaling point locations

$$for \quad all \quad j, k : [0 : m] \quad \& \quad j \neq k ==> |f(W_i, C_i) - f(W_k, C_i)| \geq freq_{min}$$
(10)

- The interference between signaling point transmission frequency and the WIU tower frequencies in that cell area should be minimized:

$$for \quad all \quad j : [0 : m] ==> |f(C_i) - f(W_j, C_i)| \geq freq_{min}$$
(11)

- Interference between the WIU towers at the boundary of two signaling cells should be minimized:

$$|f(W_m, C_i) - f(W_0, C_{i+1})| \geq freq_{min}$$
(12)

Enhancing the PTC Protocol with Dynamic Spectrum Allocation

To avoid spectrum congestion and to provide more efficient spectrum management, channels should be dynamically allocated. In [18], we provided a preliminary schema for cell design and dynamic channel allocation that accommodates PTC's varying bandwidth needs. Because WIU locations are static, the cell architecture for a group of WIUs can be predetermined. The proportion of channels allocated between the WIU network and the signaling network depends on the WIU density and the average train arrival density. Once the optimal number of signaling channels are determined by the radio network, they are dynamically allocated for a train to communicate based on the train priority, train speed, train density, and the channel availability.

Our extended protocol uses dynamic channel allocation, dynamic modulation change, and transmission power control to allocate signaling channels efficiently. This functionality is being implemented as a function in the spectrum management CE. Once the master CE received a message that follows the enhanced PTC protocol it will be handed over to the spectrum management CE. The functionality of this module varies based on the part each CR node plays on executing the protocol.

Figure 5 shows the behavior of the PTC protocol in CR module of a signaling point. The main functionality can be described as follows. The signaling point tracks the location of the train. Once the signaling point receives a movement authority request, in addition to checking the block availability, it also checks if the next block the train wants to enter is within the same signaling cell area. If the block is available and within the same signaling area, then movement authority is granted. If the block is available, but the block is in a different cell area, the signaling point communicates with the signaling point to which the next block belongs, to determine if there are sufficient channels to accommodate the incoming train. If available, movement authority is granted. Otherwise, it checks the train priority. If the train has high priority, the signaling point reduces the channel bandwidth for all non-priority trains and accepts the incoming high priority train. If the train is not a priority train, the movement authorization will be rejected.
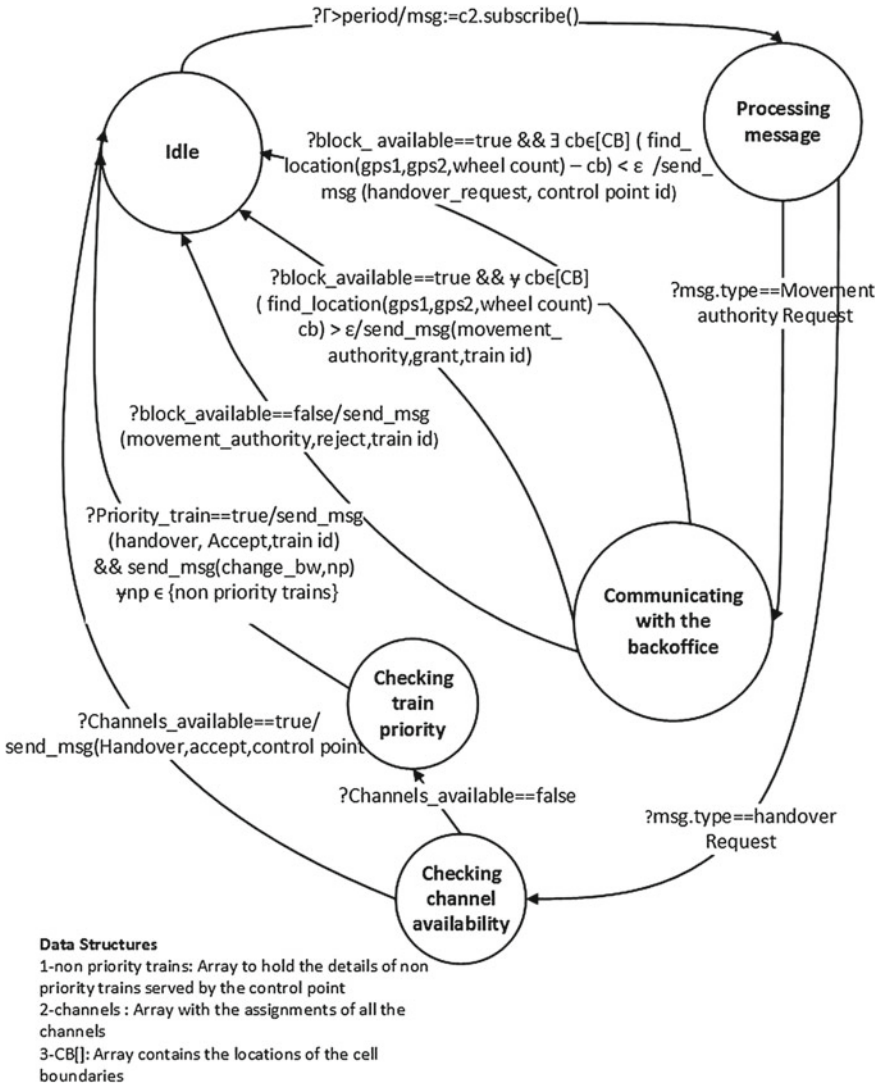
Fig. 5 Channel allocation functionality in signaling point

The proposed functionality shown in Fig. 5 goes beyond those mandated and already part of the system design. However, it has been analyzed, and a successful safety case has been made in [18]. Consequently, accepting the additions we propose will require re-making the safety case with these added suggestions. One of the immediate needs we see is to reexamine the processing requirements that these enhancements may bring to the vital real-time controller design of the system that resides on the trains.

Cryptographic Cognitive Engine

The functionality of the cryptographic cognitive engine is to generate cryptographic material, verify message integrity, and communicate detected threats to the master cognitive engine.

Cryptographic Key Generation

The integrity of all PTC messages should be preserved so that it is not possible for an attacker to tamper with a message. PTC specifies a 32-bit HMAC field to detect any tampering attempts. It further uses a 4-bit time stamp to avoid replay attacks. However, since the HMAC and the 4-bit time stamp repeats every 16 s (i.e., because the range of the time stamp is 0000(0) to 1111(15)), an attacker can replay a status message every second until it matches with the correct time stamp. For example, when the train is expecting a *STOP* status, an attacker can replay an earlier message with *PROCEED* status and can cause the train to collide with another train.

The replay attack can be minimized by changing the seed of the hash function frequently. To modify the key based on a key chain, we decided to use an algorithm based on the Time Efficient Stream Loss-tolerant Authentication (TESLA) [20]. That decision was based on its low computational and transmission overhead, as well as on its high tolerance to packet losses.

TESLA Protocol

In TESLA, the sender first decides on the number of keys N and computes a chain of keys based on the one-way key generation algorithm shown in Eq. 13. This algorithm is inspired by Lamport's One-Time Password Scheme [21].

$$K_0 = IV \quad \&\& \quad \forall i > 0 \quad K_i = F(K_{i-1}) \tag{13}$$

The total transmission period is divided into N periods, and each period is assigned with a key starting from the last key of the key chain. That is, the first time slot is assigned with the key k_N, the second slot with the key k_{N-1} and so on. The transmitter uses the key that is assigned to the current timeslot and uses it to generate the hash for transmission. Once the receiver gets a message, it puts in a buffer, since it does not have the key to authenticate the associated packet. A short time later, the sender discloses the key and the receiver validates the packet based on that key. The approach of deriving the key chain using a forward function and using it backward is meant to avoid any attacker from obtaining the next key using the disclosed key.

Our use case is different than the general broadcast use cases where TESLA is more appropriate. The main differences are listed below.

1. Our primary objective is to preserve the integrity and avoid replay attacks. The TESLA algorithm's primary objective is to ensure authentication of the sender.
2. In our use case, it is safe to assume all trains to be trustworthy.
3. Trains can store a chain of keys in their on-board database
4. Messages should follow PTC packets specification. Therefore, transmitting the key has to be done as an additional transmission and this will add a huge overhead.

To account for these differences and their effects, in our work the TESLA algorithm is enhanced to suit for the PTC use case.

Enhanced TESLA Protocol

The main differences between TESLA and our protocol are:

1. In our protocol, the transmitter and receiver generate the key chain. Therefore, no key transmission is required.
2. Our protocol changes both the salt and the hash generation algorithm randomly with time.

Seed chain is derived using a Lamport scheme similar to Tesla. For our initial experiment, we used SHA1 as the function to generate the seed chain. Starting with an Integrity Value (IV), we derived the seed chain.

The derivation of algorithms in enhanced TESLA follows a similar approach to the seed derivation. Algorithm 1 describes the process used in deriving the algorithm. The derivation process starts with an algorithm derivation seed being provided as part of the initialization vector. This algorithm derivation seed is then hashed and rehashed successively following the same approach as traditional TESLA in deriving salts. Each generated result is then operated upon via the modulus operation with the modulus exponent equated to the total number of available hashing algorithms. The process continues for the entire time range and during each time step within the communication period.

$$seed_0 = IV_A \&\&$$
$$\forall \quad i > 0 \quad (seed_i = hash(seed_{i-1}) \quad \&\& \quad AlgoSelect(i) = seed(i)\%n) \tag{14}$$

The derived seed chain and the equation sequence are loaded to the train's on-board database. The train's CR uses the seed and the algorithm assigned for the message generation time to proceed with integrity signing and verification. The complete Enhanced Tesla algorithm is shown in Algorithm 1.

Symbols

Tx = Transmitter
Rx = Receiver
AlgoSelect = Algorithm Selection
IV = Initialization Vector
n = length of the salt chain and the algorithm chain

Preconditions

1. An Algorithm seed, IV_A and salt derivation seed IV_S, is provided as part of the bootstrap process
2. The communication devices are utilizing indirect time synchronization such as the Global Positioning System (GPS)

Algorithm

1. $\rightarrow Tx, Rx : IV_S, IV_A$ Securely
2. The generate salt chain using Eq. 13 where $F() = hash()$ and algorithm chain using Eq. 14.
3. Assign salt values and algorithms to time slots such that
   $\forall i = 1 : n, Time\_slot_i \leftarrow salt_{n-i}, Algorithm_{n-i}$

Algorithm 1: Enhanced TESLA protocol

Threat Analysis

Another functionality provided by the cryptographic CE is to analyze potential threats to message communication. Figure 6 shows the threat detection procedure of the current threat module. Our threat analysis module uses the Cyclical Redundancy Check (CRC) and hash validation to identify attacks. For each message, the threat analysis module checks the CRC and the hash value. If both the CRC and the hash values are correct, the message is identified as a proper message. If either the CRC or the hash value is incorrect, it compares the message with the previous messages to assess whether there is a potential replay attack. The current threat module can detect the following three types of attacks.

1. **Type 1—Replay Attack**
   In a replay attack, the attacker captures a previously sent message and transmits it later to the intended receiver. The original PTC specification defines the usage of a static salt value for the hash function, which makes it harder to detect a replay attack. In contrast, our proposed cognitive radio architecture includes changing the cryptographic seed value with time, which makes it much more efficient in detecting the replay attempts since when the replayed message arrives, the system will likely be using a different hash key. The threat analysis component is described in more detail in [22].
   As shown in Eq. 15, a message is identified as a potential replay if the message is syntactically correct, has a valid CRC value, but the hash value does not match
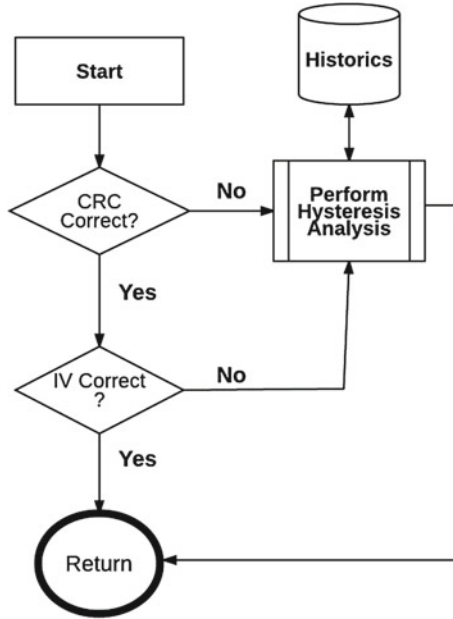
Fig. 6 Detection process

the hash value corresponding to the key used by the system at the time of the message reception. We then check the hash value with the stored historic values to see whether the hash is generated using a previous salt value, which indicates a potential replay attack.

$$ReplayDetected : (CRC = VALID)\&(hash = INVALID)\& \\ \&(M = SyntacticallyCorrect)\&\exists t < t_{current} \quad hash_t = VALID \tag{15}$$

2. **Type 2—Message Corruption Attack**
   In the message corruption attack, the attacker captures a message, modifies its content and transmits it back to the intended recipient. This attack is a special case of a message modification attack performed by an unsophisticated attacker that does not possess the ability to modify the message content while avoiding a CRC corruption.
   As shown in Eq. 16, the message corruption is detected by CRC invalidation. Our current message corruption detection algorithm does not have sufficient intelligent to distinguish between intentional message corruption or unintentional message corruption.

$$MessageCorruption : (CRC = INVALID)\&(hash = VALID|INVALID) \\ \&(M = SyntacticallyCorrect) \tag{16}$$

3. **Type 3—Message Guessing Attack**

   Similar to the message corruption attack, the attacker tries to modify the content of the message and send it back to the receiver. However, in this case, the attacker is more sophisticated and capable of producing modified messages that do not fail the CRC check. We assume that such an attacker has information about the CRC generation and key generation algorithms, but does not have the initial seed value to generate the key chain. Therefore, the attacker guesses the initial key, generates key chains, and uses it for message transmission.

   Because the attacker has knowledge of all the algorithms, he can generate a syntactically correct message. However, as mentioned earlier our cryptographic cognitive engine changes the keys frequently, forcing the attacker to guess the key used at the period in which the modified message will be received. As a result, the attacker's modified message will fail the integrity validation process. The logic used to detect this kind of attacks as shown in Eq. 17, in which hash invalidation happens while the message is syntactically correct, it has a valid time stamp and a valid CRC value.

$$
\begin{aligned}
GuessingAttack : (CRC = VALID)\&(hash = INVALID) \\
\&CurrentTimestamp\&(M = SyntacticallyCorrect)
\end{aligned}
\tag{17}
$$

Key Management

Before a train starts a planned journey, the cryptographic seeds for hash generation are loaded to the train's on-board database. When a train approaches a WIU or a signaling point, it identifies the WIU or the signaling point and uses the corresponding seed in its on-board database for that specific WIU or signaling point. This behavior is illustrated in Algorithm 2.

If the pre-loaded seed values are compromised for a particular WIU or signaling point, then a new seed value is generated as shown in Algorithm 3. In this scenario, the back office will produce an emergency seed value, encrypt is using a Back Office Private Key associated with the specific Wayside Interface Unit (WIU) or geographical region and transmit it to the compromised entity. Once an emergency re-key event is securely broadcast using the signaling network, the associated Public Key is sent to the locomotive so its CR module can generate and validate the integrity values for message transmission.

Locomotive (L) enters the block and performs a GetWIUStatus Request (G) or Timed Beacon Request (T)for a WIU (W) and sends a message (M) and Status (S).

a. $L \rightarrow W : \{CRC|IV_{Algo,Salt}|M\{G,T|WIU_{ID}\}\}$
The WIU_{ID} is used internal by the Locomotive to lookup the salt value and algorithm to use
b. $W \rightarrow L : \{CRC|IV_{Algo,Salt}|M\{S\}\}$
The WIU is provided with an integrity vector (IV) using the signaling network, which contains the seed values to generate the salt sequence and algorithm sequence

Algorithm 2: Normal Protocol Exchange: GetWIUStatus

Locomotive (L) is either entering a block or within a block and is sent a rekey request (R) for WIU(W) by the secure signaling network (S) for a specific activation time.

a. $S \rightarrow L : \{CRC|IV|R\{WIU_{ID}|WIU_{PublicKey}|Time_{Activate}\}\}$
The Signaling network sends the specific WIU to rekey as well as the public key to decrypt the encrypted new IV (which contains the seed values to generate the salt and algorithm sequence) and the activation time to the Locomotive and start using the salts.
b. $S \rightarrow W : \{CRC|IV|\{M\{Salts|Times\}\}$
The Signaling Network sends the Beacon the salts and times associated with the salts.
c. $L \rightarrow S : \{CRC|IV|\{M\{Time_{Activated}\}\}$
The Locomotive informs the signaling network that the new key and algorithm chains have been activated at the exact time according to the Locomotive.
d. $S \rightarrow W : \{CRC|IV|\{M\{Time_{L_{Activated}}\}\}$
The Signaling network sends the Locomotive activation time to the WIU.

Algorithm 3: Emergency Reseed Event

Implementation and Testing

The detailed design architecture of our cognitive radio is shown in Fig. 7. As the diagram show, we have implemented the cognitive radio as separate modules within the application layer, in which each module runs as separate sub-processes implemented in Python.

For the communication between the physical layer and the sub-processes of the cognitive radio and the communication between the sub-processes, we used an open source middleware called REDIS [23]. REDIS is a publication subscription service that can be used as an *in memory* message passing.

To develop the radio physical layer, we used GNURadio [24]. GNURadio is an open source software development toolkit that provides signal processing blocks required to implement software defined radios. Signal processing blocks in GNURadio are implemented in C++. Connections between signal processing blocks are programmed using Python. We developed two GNURadio blocks that provide the link
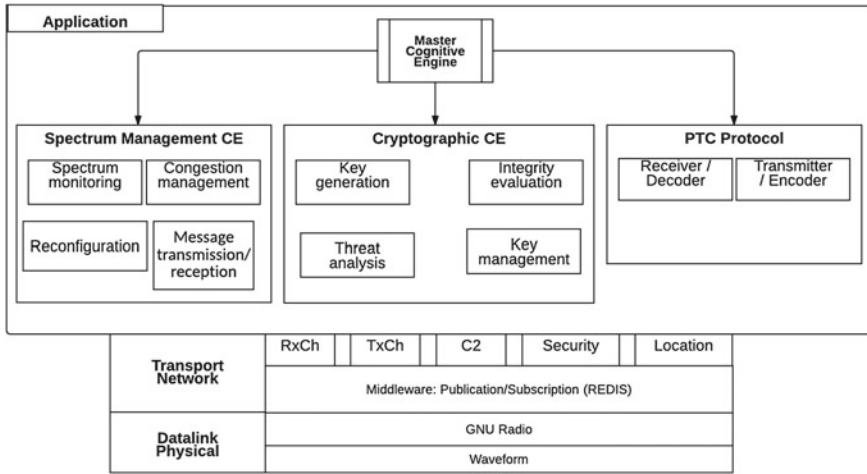
Fig. 7 Design architecture of the cognitive radio

between the physical layer to the REDIS middleware which are *redis source v* and *redis sink v* explained in section "Physical Layer Implementation".

Physical Layer Implementation

Each cognitive radio has a transmitter and a receiver that runs in parallel using two different center frequencies to avoid interference between the transmitted wave and the received wave. In our first experiment, we developed a point-to-point communication between two radios, replicating the communication between a Locomotive radio and the WIU radio. For this experiment, we used the 900 MHz bandwidth for the communication link from locomotive radio to the WIU radio and 950 MHz bandwidth for the communication from WIU radio to the locomotive radio.

Figure 8 shows the transmitter path. The application layer cognitive engines publish the messages that they want to transmit to a REDIS channel. The *Redis source v* block subscribes to that REDIS channel and passes the messages to the packet encoder. The packet encoder block converts it to a bit stream, add the header details and pass it to the modulator (we used a QPSK modulator for this experiment). The modulated signal is then passed to the USRP source to transmit using the radio medium.

Figure 9 shows the receiver path. Due to the characteristics of the QPSK modulation, the receiver requires precise time and phase synchronization. In the current receiver path, we implemented a *polyphase clock synchronization* for clock synchronization and a *costas loop* to synchronize and lock to the correct phase and frequency. Further, a *15-tap constant modulus algorithm (CMA) equalizer* is used to remove any
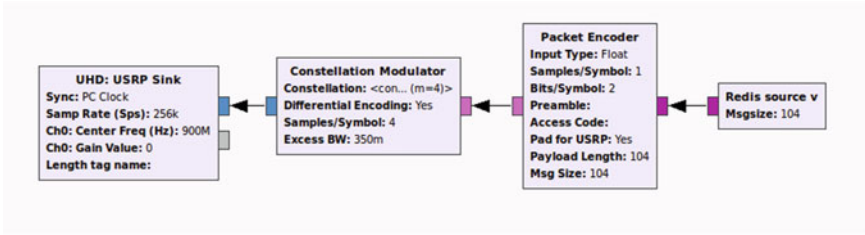
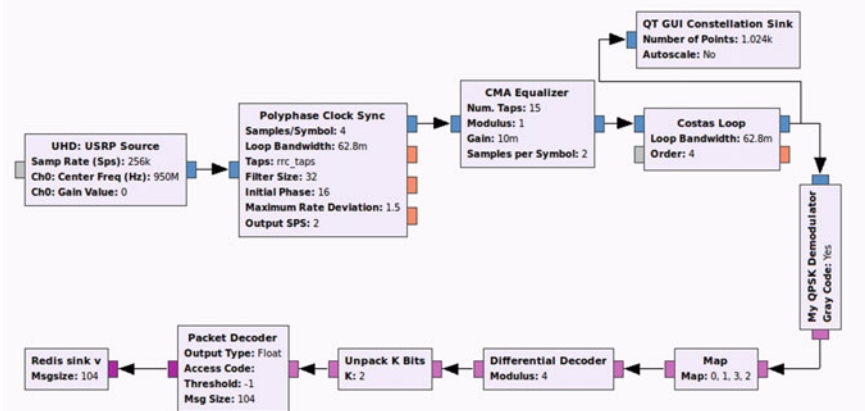**Fig. 8** Transmitter implementation in GNURadio

Fig. 9 Receiver implementation in GNURadio

multipath effect. Apart from the time synchronization, the *polyphase clock synchronization* block removes inter-symbol interference and down samples the signal to 1 sps [25]. Once properly synchronized, the signal is demodulated, decoded and written to a REDIS channel. This is done by the *Redis sink v* block. The application level cognitive engines can read this message from the REDIS channel.

REDIS Middleware as the Transport Layer

In REDIS, we assigned different channels for communications between modules as shown below.

1. **RxCh**: Receives all messages from GNU Radio. Once a message is demodulated and decoded at the physical layer receive path, a command message is published to RxCh. To get the message the Cryptographic CE subscribes to the RxCh. RxCh provides the interface between the receive path physical layer and the application layer.

2. **TxCh**: If a sub-process wants to transmit a message to a different cognitive node (e.g., transmission from the locomotive to the WIU), the message is written to the TxCh. GNURadio transmits path subscribes to the TxCh, gets the message, encodes, modulates, and transmits it.

3. **C2**: Handles message exchanges between different sub-cognitive engine processes.

4. **Security**: Relays message status, including any detected threats to the master cognitive engine.

5. **Location**: Periodically sends the location of the locomotive based on exact measurements of track and Global Positioning System (GPS) location. The location attribute provides valuable information in determining the correct position if the system comes under attack.

Application Layer: Cognitive Radio

We developed a prototype of the cryptographic CE and tested it in a scenario emulating the communication between the locomotive and the WIU. The CE is capable of dynamically changing the cryptographic salts and algorithms, integrity validation and detecting threats such as message modification, replay attacks and randomly guessing the cryptographic keys.

Figure 10 shows the application front end, which provides the user with the ability to customize its behavior in diverse ways. For example, the user can define general parameters, such as setting whether a WIU or a locomotive, selecting cryptographic parameters (e.g. specifying the initial seed value, defining the time interval between cryptographic roll over periods, etc.) and issuing commands (e.g., sending "GETWIUStatus" messages, etc.). In this prototype, the threat analysis module is implemented at the WIU.

Figure 11 shows components, interconnections, and the message flow of the designed cryptographic CE. The message flow can be explained as follows.

1. A message is received from GNU Radio, demodulated and decoded by spectrum management CE and placed into the RxCh.

2. The cryptographic CE subscribes to RxCh, gets the value, and evaluates $\{CRC|IV|M\}$ whether the CRC and IV are valid. Then, depending on the type of message, it is either forwarded to the C2 or the Security channel.

3. If CRC and IV in a packet are found to be invalid, then the threat module performs a deeper inspection of the packet. If a threat is determined, as described in section "Threat Analysis", then it informs the detected threat to the master cognitive engine. The master cognitive engine can fuse Location and time information and relay the threat results to the Back Office.

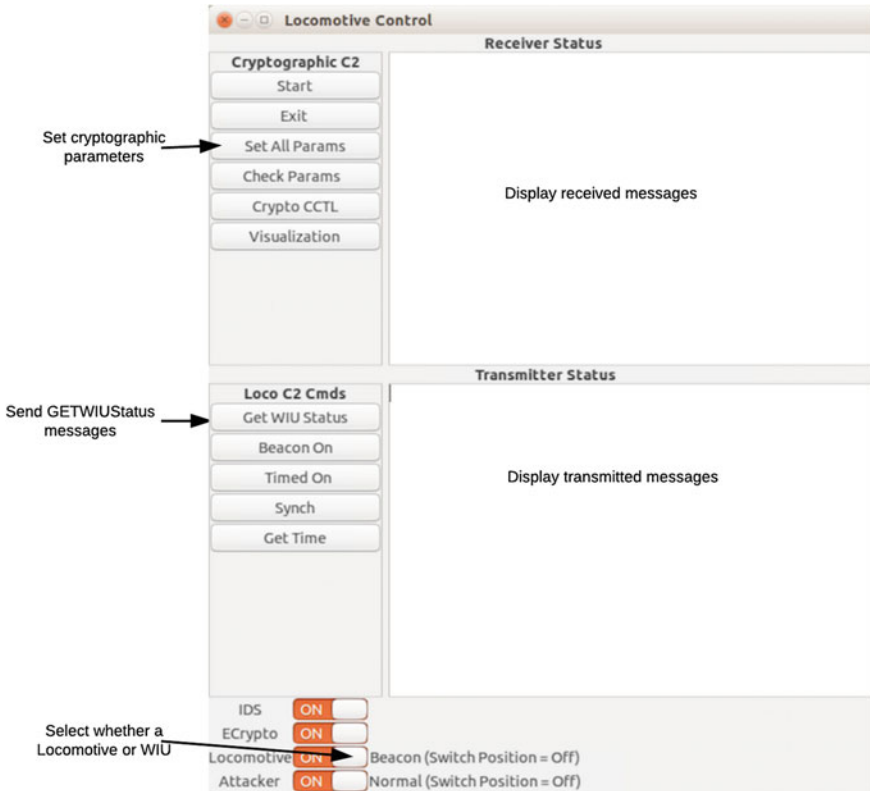4. If the cognitive radio needs to transmit messages, it will add the message to the TxCh.

Fig. 10 Application front end

Testing

We tested the prototype CR using two Ettus N210 USRPs [6], one simulating the locomotive radio and another simulating the WIU radio. We conducted the experiments in a radio chamber (Faraday cage) to isolate the radio transmission from ambient noise. The test setup is shown in Fig. 12.

When setting up the experiment, we first defined the cryptographic key chains (i.e. the salt sequence and the algorithm sequence) for the locomotive and the WIU using the application front end described in section "Application Layer: Cognitive Radio". After generating the cryptographic material, we started the 'GETWIUStatus' function at the locomotive. This initial experiment consists of sending 500 'GETWIUStatus' messages separated by 10 s intervals. These application level messages are encoded and modulated using QPSK modulation and transmitted via the USRP. Once a message is received at the WIU USRP, the unit first demodulates, decodes, and submits it to the threat analysis module. The threat module evaluates the message with respect to the American Railway Association (AAR) specification [3],
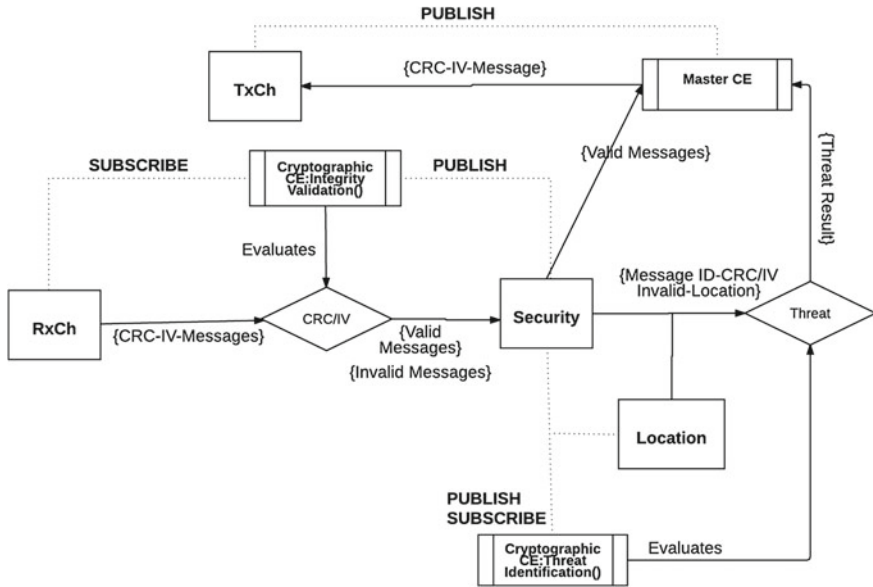
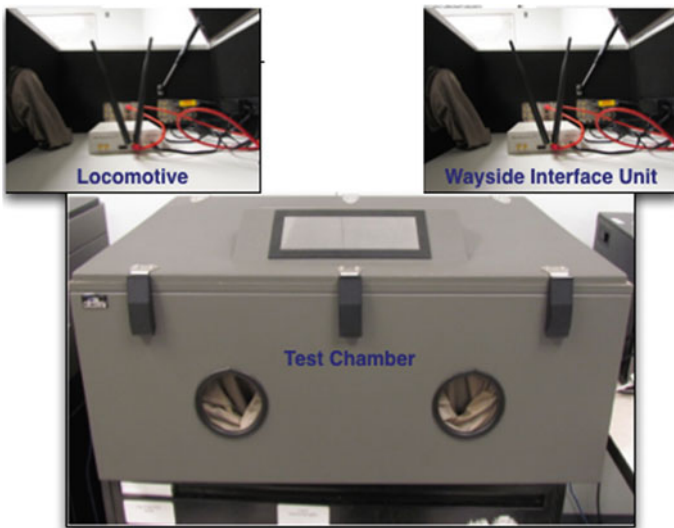**Fig. 11** Cryptographic CE components, interconnections, and message flow

Fig. 12 Experimental setup

as explained in section "Threat Analysis". If the message is correct, the CR will generate 5 WIU status beacons, send it to the message transmission function, which will transmit it to the locomotive radio.

If the message does not follow the AAR specifications, the threat analysis module evaluates the message to categorize between the attack types mentioned in section "Threat Analysis". The threat analysis module records the status of all the messages to an SQL database. These stored data is used by the threat analysis module later to detect replay attacks.

Experimental Validation

The following experiments were conducted to determine the effectiveness of the cryptographic key generation and threat analysis components.

- **Normal Operation**: Experimental and performance results describing the regular communications between a WIU and a locomotive.
- **Normal varying cryptographic rollover period**: Experimental and performance results illustrating the effect of the *roll over* cryptographic time boundaries as salts are expiring.
- **Message Guessing Detection**: Experimental and performance results describing the capability to detect an attacker guessing at a cryptographic salt to create WIU messages.
- **Message Replay Detection**: Experimental and performance results representing replay attack against the WIU.
- **Message Corruption**: Experimental and performance results describing the capability to detect corrupted messages.

Test Case 1—Normal Operation

This test was conducted to measure the detection rate of the threat module during the normal operations using enhanced TESLA. We created cryptographic material as described in section "Cryptographic Cognitive Engine". The keys are generated for a total duration of 1 day and the time is divided into 40 time segments, where each segment lasts for 2160 s or 36 min. 40 salts were generated and assigned to each time slot according to enhanced TESLA algorithm. Similarly, the algorithms are assigned to each time slot. The transmitter and receiver salt chain and algorithm generation are started approximately at the same time.

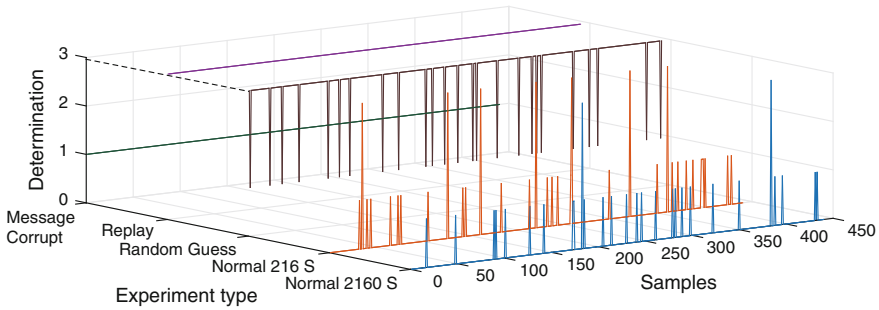The detection pattern is shown in Fig. 13 as *Normal 2160 s*.

Fig. 13 Threat module determination for different test cases

Test Case 2—Normal Varying Cryptographic Rollover Period

The objective of this test is to determine the effect of multiple cryptographic rollover periods on salts and algorithm changing procedure because the number of false positives increases near the rollover boundaries.

The experiment consists of transmitting and receiving approximately 500 messages at a rate of one message every 10 s. The cryptographic rollover periods were then shortened until the bit error rate increased. Figure 13 *Normal 216 s* plots the effects of decreasing the rollover period to 216 s, in which the false positive rate increased.

The experimental data lead to the conclusion that the false positive rate depends on the key generation starting time in the transmitter and the receiver and on the time between the cryptographic rollover.

Test Case 3—Message Guessing Detection

The objective of this test was to validate the capability to detect an attacker who can derive algorithms and seeds. The test assumes that the attacker has the complete implementation details of the locomotive, and can derive algorithms and seeds similar to a valid locomotive. The attacker guesses the IV values, initializes their radio, and begins transmission. The result shown in Fig. 13 *Random Guess* illustrates that the threat module detects the attacker either as a guessing attacker or miss-categorized as a replay attacker.

Test Case 4—Message Replay Detection

The objective of this test was to validate the detection capability during a replay attack. Due to the limitations of the testing setup, the locomotive acts as the attacker. The first message is copied and then replayed 500 times. The replay transmission

rate is one replayed message every 10 s. The result is shown in Fig. 13 labeled as
Replay.

Test Case 5—Message Corruption Detection

The objective of this test was to determine the threat module's capability of detect-
ing an intentional data modification or environmental data corruption by evaluating
the CRC. In this scenario, the attacker's radio corrupts a field in the message. The
receiver then invalidates the message when performing the CRC check. Results are
shown in Fig. 13 labeled as *Message Corrupt*.

Analyzing Experimental Outcomes

Table 1 summarizes the accuracy and the error rates for the test cases mentioned
above. In this table, accuracy refers to the system's ability to classify the type of op-
eration correctly, while error rate includes any miss-categorization. Types of miss-
categorization include false positives (normal operation events categorized as at-
tacks), false negatives (i.e., attacks categorized as normal operation events), and
miss-categorized attacks (i.e. events correctly classified as attacks but within a wrong
category). In normal operations when the time between the cryptographic rollover
period is 2160 s, the threat detection module has an accuracy of 93.8 %. This ac-
curacy is reduced to 88.6 % when the cryptographic rollover period is reduced to
216 s. That is, a tenfold reduction of the rollover period causes the system to in-
crease its false alarm rate roughly twofold (i.e., from 6.2 % to 11.4 %). When the
attacker plays a guessing attack, the threat detection module's accuracy is 93.1 %.
However, note that the detection rate (ability to detect an attack) is 100 %, which

Table 1 Threat module determination counts

Test scenario	Determination count				Accuracy (%)	Error rate (%)
	Valid	CRC corrupt	Replay	Random Guess		
Normal operation (Normal 2160 s)—cryptographic rollover period 2160 s	394	24	0	2	93.8	6.2
Normal operation (Normal 216 s)—cryptographic rollover period 216 s	391	28	0	22	88.6	11.4
Random Guess	0	30	0	408	93.1	6.9
Replay	0	0	463	0	100	0
CRC corrupt	0	442	0	0	100	0

Error rate false positive/false negative/miss-categorization

means all attacks events are categorized as such, although in 6.9 % of the time the random guess attacks are miss-categorized as replay attacks. In the CRC corruption attack and in replay attack the system has an accuracy of 100 %.

The results obtained from experiments lead to the conclusion that cryptographic CE's accuracy is highly correlated to time synchronization. In the experiment, an accurate time source was not available and time synchronization was off on average by a few seconds. Decreasing the time segment intervals between cryptographic material increases the number of false positives. The experiment reducing the cryptographic utilization time from (2160 s to 216 s), shown in Table 1 demonstrated a direct impact on accuracy. A reference time and an accurate timing source are needed for cryptographic key generation.
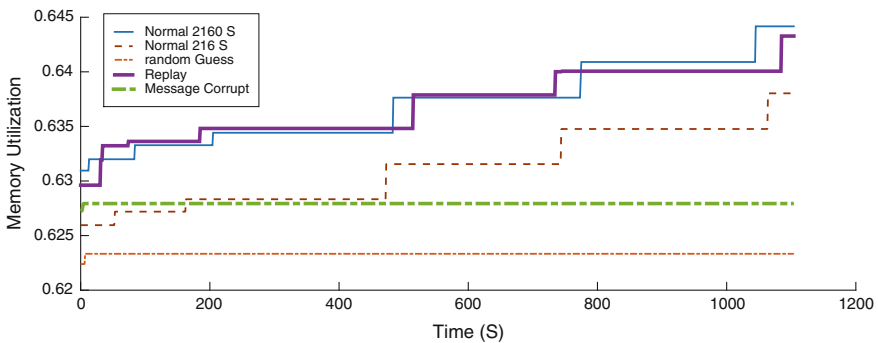
Fig. 14 Memory utilization variation

Table 2 CPU and memory utilization

	Memory utilization (%)		CPU utilization (%)	
	Mean	Standard deviation	Mean	Standard deviation
Normal operation (Normal 40)—cryptographic rollover period 2160 s	0.6605	0.0031	0.0171	0.1229
Normal operation (Normal 400)—cryptographic rollover period 216 s	0.6325	0.0034	0.0171	0.1229
Random Guess	0.6300	5.9377e-05	0.0045	0.0635
Replay attack	0.6369	0.0029	0.0225	0.1620
CRC corruption	0.6279	3.7909e-05	0.0018	0.0402

For all the test case the Central Processing Unit (CPU) utilization by the WIU application, which includes key generation and threat analysis was mostly zero, although there were spikes of 0.9 %. Additionally, the replay test case shows the occasional CPU use of 1.8 %. Figure 14 shows the memory utilization of the WIU application with time. The memory utilization for all the test cases is around 0.6 %. Normal operation and random guess attacks caused a slight increase in this number, where for all the other cases the memory utilization remained constant. More details about the memory and CPU utilization can be found in Table 2.

Conclusions

We designed a multi-tiered cognitive radio network to enhance the efficiency of spectrum sharing and the security of the radio communication for PTC operations. In our architecture, all PTC nodes are expected to use cognitive radios to share spectrum, analyze different risks and, if needed, provide mitigating actions. The cognitive radios that we implemented provide better spectrum management for PTC because their design is customized to behave according to PTC specifications, traffic requirements, and spectrum demand, while considering that the PTC network as a whole can determine the bandwidth allocation. Further, our cognitive radio system recognizes the cryptographic safety of message broadcasts by changing the integrity seed with time and providing threat analysis for detecting any malicious activities.

This Chapter provided the detailed design architecture of our multi-tiered cognitive radio network and the implementation of the cryptographic CE, which were implemented and tested using ETTUS USRPs. Our experimental results suggested that both dynamic key generation procedure and physical level signal recovery procedure require precise time synchronization. Further, the results indicate that the threat analysis component can be extended to an Intrusion Detection System (IDS) customized to PTC radio communication, which can be more efficient in detecting both cryptographic and physical layer vulnerabilities.

We are in the process of implementing the spectrum management CE and the master CE. The capabilities of dynamically changing frequency, power, and modulation are currently being implemented. We plan to conduct more comprehensive experiments on the current setup to better understand the reasons behind the message interruptions that happen when radio parameters are changed, as well as to find and evaluate mitigation strategies for them. The design and the implementation of cognitive radio requires a more careful planning and evaluation than a regular software defined radio, as the former provides much more essential functionality than the latter. The research work presented in this Chapter can be seen as an initial, although important, step in achieving such functionality in a sound and comprehensive fashion.

References

1. CNN: Investigating the philadelphia Amtrak train derailment. http://www.nytimes.com/interactive/2015/05/13/us/investigating-the-philadelphia-amtrak-train-crash.html?_r=0 (2015)
2. CNN: Amtrak restores service after fatal crash; investigation continues. http://www.cnn.com/2016/04/03/us/philadelphia-amtrak-derailment/ (2016)
3. AAR: Interoperable Train Control Wayside Interface Unit Requirements Railway Electronics S9202 (2010)
4. AAR: Office to Locomotive ICD version draft 2.6 S 9352A (2010)
5. Joint Council on Transit Wireless Communications: Positive Train Control-White Paper (2012)
6. Wikipedia: Universal software radio peripheral. https://en.wikipedia.org/wiki/Universal_Software_Radio_Peripheral (2014)
7. Wikipedia: 2008 Chatsworth train collision. https://wikipedia.org/wiki/2008_Chatsworth_train_collision (2015)
8. Hartong, M.: Secure communications based train control (CBTC) operations. Ph.D. thesis, George Mason University, Fairfax, Virginia (2009)
9. Abadie, A.: Combining operational and spectrum characteristics to form a risk model for positive train control communications. Ph.D. thesis, George Mason University (2014)
10. Mithola, J.: Cognitive radio: an integrated agent architecture for software defined radio. Ph.D. thesis, Royal Institute of Technology (KTH), Sweden (2000)
11. Crohas, A.: Practical Implementation of Cognitive Radio System for Dynamic Spectrum Access (2008)
12. Chen, S.: Vehicular dynamic spectrum access: using cognitive radio for automobile networks. Ph.D. thesis, Worcester Polytechnic Institute (2012)
13. Singh, K., Rawat, P., Bonnin, J.M.: Cognitive radio for vehicular ad hoc networks (CR-VANETs): approaches and challenges. EURASIP J. Wirel. Commun. Netw. **1**, 49 (2014). doi:10.1186/1687-1499-2014-49
14. Amanna, A., Gadhiok, M., Price, M.J., Reed, J.H., Siriwongpairat, W.P., Himsoon, T.K.: Rail-CR: cognitive radio for enhanced railway communication. In: ASME, pp. 467–473 (2010). doi:10.1115/JRC2010-36201
15. MacKenzie, A., Reed, J., Athanas, P., Bostian, C., Buehrer, R., DaSilva, L., Ellingson, S., Hou, Y., Hsiao, M., Park, J.-M., Patterson, C., Raman, S., da Silva, C.: Cognitive radio and networking research at Virginia Tech. Proc. IEEE **97**(4), 660–688 (2009). doi:10.1109/JPROC.2009.2013022
16. Bandara, K.R.D.S., Melaragno, T., Fewell, A., Wijesekara, D.: Multi-tiered cognitive radio network for positive train control operations (2016)
17. Bandara, D., Abadie, A., Melaragno, T., Wijesekara, D.: Providing wireless bandwidth for high-speed rail operations. Procedia Technol. **16**, 186–191 (2014). doi:10.1016/j.protcy.2014.10.082
18. Bandara, K.R.D.S., Abadie, A., Wijesekara, D.: Cell planning for high-speed train operations in USA. In: ASME, p V001T03A007 (2015). doi:10.1115/JRC2015-5805
19. Peterson, L.L., Davie, B.S.: Computer Networks: A Systems Approach. Morgan Kaufmann (1996)
20. Perrig, A., Tygar, J.D.: Secure Broadcast Communication in Wired and Wireless Networks, vol. 1. Kluwer Academic Publishers, Assinippi Park, Norwell Massachusetts 020601 USA (2002)
21. Lamport, L.: Password authentication with insecure communication. Commun. ACM **24**(11), 770–772 (1981). doi:10.1145/358790.358797. http://doi.acm.org/10.1145/358790.358797
22. Melaragno, T., Bandara, K.R.D.S., Wijesekara, D., Costa, P.: Rail radio intrusion detection system (RRIDS) for communication based train control (CBTC). In: 2016 IEEE International Conference on Intelligent Rail Transportation (2016)
23. Wikipedia: Redis. https://en.wikipedia.org/wiki/Redis (2015)

24. Corgan, J.: Welcome to GNU Radio. http://gnuradio.org/redmine/projects/gnuradio/wiki (2015)
25. Rondeau, T.: GNU radio-new tutorial 7. http://gnuradio.org/redmine/projects/gnuradio/wiki/Guided_Tutorial_PSK_Demodulation/5 (2014)

Centralized and Distributed Algorithms for Stable Communication Topologies in Cognitive Radio Ad hoc Networks

Natarajan Meghanathan

Introduction

Cognitive radios are software-defined radios that can dynamically adapt their transmission parameters to suit to the available channels (frequencies) in the operating environment [2]. A cognitive radio network (CRN) is a wireless network whose devices are embedded with cognitive radios that can sense the available channels in the neighborhood and switch the communication channel, if needed. Cognitive radios are considered a promising solution to alleviate the problem of spectrum scarcity [23]. A CRN comprises of two types of users: primary users (PUs) and secondary users (SUs); the PUs own licensed channels and the SUs do not own any licensed channel [2]. The SUs employ the available licensed channels of the PUs that are not in use. When the PUs of the currently used channels become active, an SU either relinquishes the channel and switches to any other available PU channel, called spectrum overlay, or reduces its transmission range and continues to use the channel without interfering the licensed PU, called spectrum underlay [23]. For the rest of this chapter, we restrict ourselves to the spectrum overlay approach.

Our focus in this chapter is on a category of CRNs called the cognitive radio ad hoc networks (CRAHNs): a self-organized network comprising of the PU nodes and SU nodes [3]. The PU nodes and SU nodes operate with a limited transmission range and are peers of each other. For simplicity, we assume each PU node to own a distinct licensed channel, identified using the PU node's ID. A PU channel is considered to be available for use by an SU node if the two nodes are within the transmission range of each other and that the PU node is turned OFF (i.e., not active). Two SU nodes are connected with a link if they share at least one available PU channel in their common neighborhood. Depending on the activity status of the

N. Meghanathan (✉)
Department of Computer Science, Jackson State University,
Mailbox 18839, Jackson, MS 39217, USA
e-mail: natarajan.meghanathan@jsums.edu

PU nodes, the set of PU channels available for the SU nodes to use changes dynamically with time. Thus, even if the SU nodes and PU nodes are static, communication topologies (like paths and trees) that connect the SU nodes may have to be frequently reconfigured based on the PU channels' availability in the SU-SU neighborhoods. The routing solutions for CRAHNs proposed so far have focused mainly on discovering optimal routes with respect to traditional metrics like hop count [22], end-to-end delay, with the channel switch delay considered an integral component of end-to-end delays [5, 6, 13], throughput [19], energy consumption [18] and etc. Not much work has been done on determining stable paths or trees whose lifetime is longer.

Our objectives in this chapter are two-fold: (i) To develop a generic centralized benchmarking algorithm that can be used to arrive at upper bounds for the lifetime of any communication topology that spans the SU nodes. We call the algorithm as the Maximum Lifetime Communication Topology (MLCT) algorithm. Given that there is a polynomial-time algorithm or heuristic to determine the communication topology of interest (say, shortest path tree, minimum spanning tree, connected dominating set, etc.) that spans the entire network of SU nodes, the objective of the MLCT algorithm is to find a stable sequence of instances of the communication topology for the lifetime of the network. In this chapter, we choose shortest path trees as the representative communication topology of interest, as these trees are widely used to determine minimum hop paths from a source node to every other node in a network (a standard routing problem in any network), including CRAHNs. As our focus here is to develop a benchmarking algorithm, we take a centralized approach and assume the availability of any relevant information that could contribute towards arriving at the best solution. Accordingly, we assume that the availability status (ON or OFF) of every PU channel at any time instant within the duration of the network session is known before-hand. (ii) To develop a local spectrum knowledge-based distributed routing protocol that uses the number of common available PU channels for an SU-SU edge as the basis to determine stable end-to-end SU-SU paths that will go through fewer path transitions. Referred to as MCPUR (Maximum Common Primary User-channel Routing) protocol, we model a time-variant CRAHN of SUs with links between any two SUs if they have at least one common PU channel available for use in their neighborhood and the weight of an edge is the number of such common PU channels available for use. MCPUR prefers to choose an SU-SU source-destination (s-d) path with the largest value for the sum of the number of common PU channels available for use across each of its constituent edges. Our hypothesis is that such an s-d path is likely to exist for a longer time (and incur fewer broadcast route discoveries) as the end nodes of the constituent SU-SU edges are more likely to have at least one common available PU channel that can be used to complete the transmission and reception of data packets.

The rest of the chapter is organized as follows: section "Related Work and Our Contributions" discusses related work on routing solutions for CRAHNs and highlights our contributions through this chapter. Section "Generic Algorithm to Determine Maximum Lifetime Communication Topology (MLCT)" presents the MLCT algorithm to determine stable sequence of any network-wide

communication topology and its application to determine maximum lifetime shortest path trees (MLSPTs). We also prove the correctness of the algorithm, analyze its run-time complexity as well as illustrate its working to determine MLSPTs through an example. Section "Simulation Study of the MLCT Algorithm" presents simulation results comparing the performance of the MLSPTs and minimum hop shortest path trees (MHSPTs) with respect to tree lifetime and height. Section "Design of the MCPUR Protocol" presents the design of the proposed MCPUR protocol for CRAHNs. Section "Simulation Study of the Maximum Common Primary User Channel-Based Routing (MCPUR) Protocol" presents a detailed simulation study of the MCPUR protocol vis-a-vis a shortest path routing (SPR) protocol. Section "Conclusions" concludes the chapter and outlines directions for future research. Throughout the chapter, we use the terms 'route' and 'path', 'node' and 'vertex' as well as 'edge' and 'link' interchangeably. They mean the same.

Related Work and Our Contributions

Routing solutions for CRAHNs are either full spectrum knowledge based or local spectrum knowledge based. Tables 1 and 2 respectively summarize the previous works available for full-spectrum and local-spectrum based routing solutions for CRAHNs and also exhibits where our proposed protocols centralized and distributed solutions fit in. The solutions based on the full spectrum knowledge assume each SU node to be completely aware of all the available PU channels in the network and choose optimal routes with respect to either minimum number of hops per SU-SU path [5], maximum conflict-free assignment [20] of PU channels or minimum number of channel switches per SU-SU path [21]; there is bound to be switching of channels when none of the PU channels available for the end nodes of an SU-SU link are the same as the preferred PU channels for one or both the end nodes to which they stay tuned by default for transmission and reception. Source-based routing is a commonly used routing strategy for full spectrum

Table 1 Comparison of our centralized MLCT algorithm with existing full-spectrum knowledge based routing solutions

Solution strategy	Time instant of topology used	Route maintenance	References
Min. # of hops per SU-SU path	Current only	LORA	Cheng et al. [5]
Max. conflict-free assignment of PU channels	Current only	LORA	Xie et al. [20]
Min. # of channel switches per SU-SU path	Current only	LORA	Xin et al. [21]
Max. lifetime shortest path trees	Current plus predicted future	LORA	This chapter

Table 2 Comparison of our distributed MCPUR protocol with existing local-spectrum knowledge based routing solutions

Solution strategy	Metric optimized	Stability-aware	References
Minimum power routing protocol	Energy consumption for broadcast discovery, transmission, channel switching and reception across all the SU-SU edges	No	Ma et al. [13]
Minimum bandwidth-footprint protocol	Sum of the interference areas of the SU nodes on the path	No	Lo [12]
Minimum access delay protocol	Sum of the delays incurred by the SU nodes to access a PU channel	No	Ma et al. [13]
Minimum queuing delay protocol	Sum of the queuing delays at the intermediate SU nodes	No	Cheng et al. [6]
Maximum common primary user channel-based routing (MCPUR) protocol	Sum of the number of common PU channel for the SU nodes part of the SU-SU edges of a path	Yes	This chapter

knowledge-based solutions [4]; here, the source uses link state advertisements that capture the weight of each link as a function of the number of common PU channels for every SU-SU node pair, the strength of the PU channels, access/switching delays and etc. The source then locally constructs a network graph of all the SU nodes and sets the link weight to the metric being optimized and runs the Dijkstra algorithm [7] to determine the targeted optimal path. Hou et al. [9] use a Mixed Integer Non-Linear Programming formulation to maximize the spectrum reuse factor and reduce the overall bandwidth usage.

Routing solutions based on the notions of layered graph [21] and colored graph [24] were also proposed for CRAHNs. Under the layered approach, the number of layers is the number of channels: each SU node is represented by a vertex (representing the node) and one sub vertex per PU channel; there are three types of edges—horizontal (connects sub vertices of the same logical layers), access (connects a vertex with each of its sub vertices) and vertical (connects the sub vertices in various layers for the same SU node). Under the colored graph approach, the number of colors is the number of channels: each vertex represents an SU node and there is an edge between two vertices; two vertices are connected by edges that represent the common channels between the two SU nodes. Once the graph is set, an optimal route from the desired source to the targeted destination. However, both these approaches incur significant computation and control overhead and are more suitable only for CRAHNs with low PU dynamics.

All of the above full spectrum knowledge-based solutions take a centralized approach like we took in this chapter; however, the full spectrum knowledge for the current time instant alone cannot be used to arrive at benchmarks for the routing

metric, if one intends to use the LORA strategy [1] of staying on with a route as long as it exists. In order to arrive at benchmark values for the routing metric and still use the LORA strategy, it is imperative that the routing algorithm knows the future topology changes or at least predict them and make use of. The approach taken in our chapter to make use of the knowledge about future topology changes (i.e., availability of the PU channels) to arrive at optimal benchmark values for the maximum tree lifetime could be construed as a strategy for developing distributed algorithms to determine stable communication topologies.

The routing solutions based on the local spectrum knowledge are distributed solutions that make use of the spectrum information collected in the neighborhood through the common control channel [18]. The routing solutions based on local spectrum knowledge could be classified according to the class of metric they are targeted to optimize in a distributed fashion. The minimum power routing protocol [13] discovers SU-SU paths that minimize the energy consumption incurred due to broadcast discovery, transmission, channel switching and reception. The bandwidth footprint (BFP) minimization-based routing protocol [12] discovers *s-d* (source-destination) routes that minimally impact the interference area of the SU nodes (called the bandwidth footprint) that are part of the ongoing *s-d* sessions. While analyzing the tradeoffs between farthest neighbor routing (FNR) and nearest neighbor routing (NNR) for CRAHNs, it was observed that FNR provides better end-to-end channel utilization and reliability, whereas NNR is relatively more energy-efficient [22]. With regards to minimizing delay-sensitive routing protocols for CRAHNs, Ma et al. [13] developed protocols to minimize the channel switching and access delays and Cheng et al. [6] developed protocol to minimize the queuing delays at the intermediate nodes. Ding et al. [8] proposed a decentralized algorithm to solve the power and spectrum allocation problem for two common cooperative transmission schemes: decode-and-forward (DF) and amplify-and-forward (AF), based on convex optimization and arithmetic-geometric mean approximation techniques. Though these schemes tend to maximize throughput, their stability with respect to PU channel switching has not been well-studied.

Joshi et al. [11] advocate searching for an alternate PU channel (rather than discovering a new SU-SU path) if a currently used PU channel simply becomes unavailable for an SU-SU link. Such an approach is intended to minimize the routing overhead and prolong network lifetime. However, the underlying route selection strategy is not based on selecting SU-SU links with a larger number of common PU channels. We argue that an alternate PU channel is more likely to become available only if the underlying SU-SU link originally has a larger number of common PU channels; merely attempting to discover an alternate PU channel need not truly minimize the number of path discoveries.

The proposed MLCT algorithm could be adapted to determine stable communication topologies that could also incur lower values for power consumption, interference or channel switch delay. For a mobile graph [14], the weight of an SU-SU edge is simply the sum of the weights of the edge in the constituent static graphs. The weights of the SU-SU edges in the static graphs have to be just suitably

modeled to suit to the communication topology of interest. For example, to determine stable minimum power consumption routes, one could model the weight of an SU-SU edge at a particular time instant as the sum of the energy lost due to reception, transmission (function of the distance between the two end SU nodes at that time instant) and channel switching (depending on the preferred PU channels of the two end SU nodes at that time instant), and simply run the Dijkstra algorithm [7] on the long-living connected mobile graphs to determine a sequence of stable minimum power consumption *s-d* paths such that the energy lost due to broadcast route discovery (route transitions) is the global minimum. Likewise, to determine stable minimum interference routes, one could model the weight of an SU-SU edge for a static graph to be the number of peer SU nodes that are also in the neighborhood of the PU channel used by the edge for transmission and reception, and simply run the Dijkstra algorithm on the long-living connected mobile graphs to determine a sequence of stable minimum interference *s-d* paths such that the number of route transitions is the global minimum.

To the best of our knowledge, we have not come across any work that has proposed a benchmarking algorithm to determine stable communication topologies for CRAHNs. In earlier works, separate benchmarking algorithms have been proposed based on the idea of taking graph intersections to determine stable sequence of unicast paths, multicast Steiner trees and broadcast connected dominating sets [16] for mobile ad hoc networks (MANETs) and to determine stable sequence of data gathering trees [17] for wireless mobile sensor networks (WMSNs). The characteristic of both MANETs and WMSNs is that the nodes are mobile and it is the mobility of the nodes that triggers the topology changes. On the other hand, nodes in the CRAHNs considered in this research are static and it is the availability of the PU channels that changes dynamically with time, triggering changes in the communication topology of interest. Moreover, the benchmarking algorithms proposed for MANETs and WMSNs focus on a specific communication topology of interest (like an *s-d* path, multicast tree, data gathering tree, etc.) and are not proposed as generic algorithms. In this chapter, we propose the MLCT algorithm as a generic algorithm that can be used to determine a stable sequence of instances of long-living mobile graphs such that the average lifetime of these mobile graphs will be the upper bound (benchmark) for the average lifetime incurred for any communication topology of interest that spans the entire network of SU nodes (be it a shortest path tree, minimum spanning tree, minimum connected dominating set, etc.); the shortest path tree has been just chosen as a representative communication topology of interest to demonstrate the working of the MLCT algorithm and its analysis.

Also, as can be seen from the above discussion of related work, there is no single work that has proposed the use of the number of common available PU channels (in the mutually intersecting neighborhood of the SU nodes constituting the links of a source-destination path) as the criteria to select stable routes that are likely to go through fewer path transitions. Though there have been attempts at individually minimizing either the end-to-end delay per packet or energy consumption, MCPUR is the first such attempt to discover stable routes that can contribute towards

accomplishing optimal values for both end-to-end delay per packet and energy consumption as well as provide better quality of service due to less frequent changes in the paths traversed by the data packets.

Generic Algorithm to Determine Maximum Lifetime Communication Topology (MLCT)

Network Model and Assumptions

We assume a static and centralized setup of the CRAHN comprising of the primary user nodes (PU nodes) and secondary user nodes (SU nodes), each of which are uniquely identified by an ID. Each PU node is assumed to own a licensed channel that has a unique frequency and is identified with the ID of the PU node itself. An SU node identifies a PU channel on the basis of the latter's ID. The SU nodes and PU nodes are assumed to be randomly distributed in a 2-dimensional network. The transmission range (R) for both the categories of nodes is assumed to be identical. Using the node locations and transmission range R, we could identify the PU node neighbors for each SU node. A PU node is said to be a neighbor of an SU node if the Euclidean distance between them is within the transmission range. Accordingly, we say that a PU channel is available for use by an SU node at a particular time instant only if the corresponding PU node is in the neighborhood of the SU node (which is always the case, as the nodes are static) and that the PU node is turned OFF and not using its licensed channel at that time instant.

Each PU channel is independently turned OFF and ON (alternately) for the entire network session. To begin with, a PU channel is randomly turned ON or OFF and continues to stay so for a time period randomly chosen from the range $[0…MAX_{Random\_ON}]$ or $[0…MAX_{Random\_OFF}]$ respectively. After this, the PU channel changes its status from ON to OFF or OFF to ON (depending on the case) and chooses to be in the new status for a random time period selected from the above range. This procedure is repeated throughout the network session.

We model two types of SU graphs: (i) A time-invariant SU graph $G_i(SU)$ comprises of only the SU nodes and there exists an SU-SU edge only if the Euclidean distance between the end-to-end SU nodes of the edge is within the transmission range, R. (ii) A time-variant SU graph $G_t(SU)$ that comprises of both the SU nodes and PU nodes; there exists an SU-SU edge only if the edge exists in $G_i(SU)$ and there exists at least one idle PU node (that is turned OFF and its licensed channel is available for use) in the neighborhood of both the end-to-end SU nodes of the edge. We sample the network periodically for every t_{sample} seconds. We model a time-variant mobile SU graph $G_{t…t+k}(SU) = G_t(SU) \cap G_{t+1}(SU) \cap … \cap G_{t+k}(SU)$ such that $G_{t…t+k}(SU)$ is connected (i.e., spans all the SU nodes), wherein $G_t(SU)$, $G_{t+1}(SU)$, $G_{t+2}(SU)$, …, $G_{t+k}(SU)$ are respectively the static SU graphs at time instants sampled at time instants $t, t + 1, t + 2, …,$ and $t + k$.

MLCT Algorithm

Given the knowledge of the future availability of the PU channels, the MLCT algorithm works as follows: At a time instant t for which we want to find a stable communication topology that spans the entire network of SU nodes (say a shortest path tree rooted at source SU node s), we check if the static SU graph $G_t(SU)$ is connected (i.e., the SU nodes are reachable from each other through one or more paths). We can check the connectivity of an SU graph (static graph or mobile graph) using the BFS algorithm. If all the vertices of the graph can be visited by running BFS from any arbitrarily chosen node, then the graph is connected; otherwise, not. If $G_t(SU)$ is connected, the network-wide communication topology of interest exists at time instant t. We then proceed to the next sampling time instant $t + 1$ and take the intersection of the graphs $G_t(SU)$ and $G_{t+1}(SU)$ and refer to the intersection graph as the mobile graph $G_{t...t+1}(SU) = G_t(SU) \cap G_{t+1}(SU)$. If $G_{t...t+1}(SU)$ is connected, we continue to proceed to the next sampling time instant $t + 2$ and check whether $G_{t...t+2}(SU) = G_t(SU) \cap G_{t+1}(SU) \cap G_{t+2}(SU)$ is connected. We continue like this until we come across sampling time instants $t + k$ and $t + k+1$ such that $G_{t...t+k}(SU)$ is connected and $G_{t...t+k+1}(SU)$ is not connected. If $G_{t...t+k}(SU)$ is connected, it implies the communication topology of interest (say a shortest path tree rooted at source node s) exists in each of the static graph snapshots $G_t(SU)$, $G_{t+1}(SU)$, ..., $G_{t+k}(SU)$. We could use that communication topology across all of these time instants t, $t + 1$, ..., $t + k$. We repeat the above procedure starting from time instant $t + k + 1$ to find the next stable mobile graph and communication topology of the sequence and continue likewise until the end of the network session.

The MLCT algorithm finds for us a stable sequence of instances of the mobile graph and the communication topology of interest (say the shortest path tree) such that the number of transitions from one instance of the topology to another in the sequence is the global minimum. The average lifetime of the mobile graphs in the stable sequence found by the MLCT algorithm would serve as an upper bound (benchmark) for communication topologies (be it a shortest path tree, spanning tree, connected dominating set, etc.) that span all the SU nodes found by any centralized or distributed algorithm for CRAHNs.

Figure 1 presents the pseudo code of the MLCT algorithm with respect to determining a stable sequence of mobile graphs and the corresponding instances of shortest path trees. As one can see, the MLCT algorithm is generic and could be used to find a stable sequence of any communication topology that spans the entire CRAHN network of SU nodes, provided there exists an underlying algorithm or heuristic to determine that communication topology (e.g., shortest path tree, minimum spanning tree, connected dominating set, etc.) on a network graph. The tradeoff we anticipate is that the stable sequence determined for the communication topology of interest may not be optimal with respect to the metric for which it is known for. For example, the shortest path tree determined on a sequence of mobile graphs $G_{1...j}(SU)$, $G_{j+1...k+1}(SU)$, ... $G_{r+1...T}(SU)$ may not have the same average

----
Input: Static graphs G_t(SU) for all the sampling time instants $t \in 1....T$ of the network session; Source node s
Output: Stable sequence of mobile graphs and maximum lifetime shortest path trees (MLSPTs)
Auxiliary variables: i, j
Initialization: $i = 1, j = 1$, MLSPT$_{1...T}$(SU) = ϕ; Stable_G$_{1...T}$(SU) = ϕ

Begin *MLCT Algorithm*
1   while ($i \leq T$) do
2   Find a mobile graph $G_{i...j}$(SU) = G_i(SU) $\cap$ G_{i+1}(SU) $\cap$ ... $\cap$ G_j(SU) such that $G_{i...j}$(SU) is connected and $G_{i...j+1}$(SU) is not connected.
3   Stable_G$_{1...T}$(SU) = Stable_G$_{1...T}$(SU) U {$G_{i...j}$(SU)}
4   Find a shortest path tree SPT$_{i...j}$(SU) = BFS($G_{i...j}$(SU), s)
5   MLSPT$_{1...T}$(SU) = MLSPT$_{1...T}$(SU) U {SPT$_{i...j}$(SU)}
6          $i = j+1$
7   end while
8   return MLSPT$_{1...T}$(SU) and Stable_G$_{1...T}$(SU)
End *MLCT Algorithm*

Fig. 1 Pseudo code of the maximum lifetime communication topology algorithm and its application to determine stable sequence of maximum lifetime shortest path trees (MLSPTs)

tree height (a measure of the minimum number of hops for a path between any two SU nodes) compared to the average height of a sequence of shortest path trees determined on the individual static graphs G_1(SU), G_2(SU), …, G_T(SU) where time instants 1 and T are respectively the beginning and ending time instants of the network session. This is because, when we aim for and determine a mobile graph $G_{t...t+k}$(SU) that exists for the longest time, the probability of finding an SU-SU edge in the mobile graph decreases and we mostly end up with the minimal number of SU-SU edges that would be needed to keep the SU nodes connected. A common shortest path tree determined on such a mobile graph $G_{t...t+k}$(SU) might have a height larger than the average height of the individual shortest path trees determined on each of the static graphs G_t(SU), G_{t+1}(SU), …, G_{t+k}(SU).

Time Complexity of the MLCT Algorithm

The time complexity of the MLCT algorithm depends on the time complexity to determine the mobile graphs and the number of times we run the BFS algorithm to determine the connectivity of the static graphs and mobile graphs as well as on the time complexity of the algorithm or heuristic employed to determine the underlying communication topology of interest. The number of edges in any static SU graph cannot be more than the number of edges in the time-invariant SU graph G_i(SU). Let E be the number of edges in G_i(SU) and this is also the maximum number of

edges in a static graph $G_t(SU)$ at any time instant t as well as in a mobile graph $G_{t...t+k}(SU)$ at any time instant $t + k$. The complexity of running BFS on an SU graph of V nodes is $\Theta(V + E)$, where V and E are respectively the number of SU nodes and SU-SU edges in the graph.

Given a mobile graph $G_{t...t+k}(SU)$, we arrive at mobile graph $G_{t...t+k+1}(SU)$ by simply checking whether each of the edges in $G_{t...t+k}(SU)$ exists in the static graph $G_{t+k+1}(SU)$. $G_{t...t+k+1}(SU)$ contains only those edges of $G_{t...t+k}(SU)$ that also exist in $G_{t+k+1}(SU)$. Edges in $G_{t+k+1}(SU)$ that do not exist in $G_{t...t+k}(SU)$ are not considered for inclusion in $G_{t...t+k+1}(SU)$. Thus, the time complexity to determine a mobile graph $G_{t...t+k+1}(SU)$ is the time complexity to determine the mobile graph $G_{t...t+k}(SU)$ plus the number of edges in $G_{t...t+k}(SU)$, which could be at most E. In each iteration of the algorithm, we either move from mobile graph $G_{t...t+k}(SU)$ to $G_{t...t+k+1}(SU)$ if the latter is connected or start the algorithm fresh from $G_{t+k+1}(SU)$ if $G_{t...t+k+1}(SU)$ is not connected. Either way, the algorithm advances by one time instant ($t + k$ to $t + k + 1$) in each iteration. Hence, the number of iterations of the algorithm is T, the duration of the network session (also the number of static graph samples generated). In each such iteration, we spend $\Theta(E)$ time to form a mobile graph. At the worst case, we may have to run the BFS algorithm twice in each iteration: once to check whether a mobile graph $G_{t...t+k}(SU)$ is connected and (if it is not connected) to check whether the static graph $G_{t+k}(SU)$ is connected. Hence, the time complexity to determine the mobile graphs and check for the connectivity of the mobile graphs/static graphs across all the iterations is $\Theta((E + V + E) * T)$.

In the case of maximum lifetime shortest path trees (MLSPTs), the underlying algorithm used to determine the communication topology of interest is also BFS. We run the BFS algorithm for the communication topology once in each iteration: either on the mobile graph $G_{t...t+k}(SU)$ if it is connected or on the static graph $G_{t+k}(SU)$ if the former is not connected. Thus, the time complexity incurred to run the BFS algorithm (as the algorithm to determine the communication topology of interest) across all the T iterations is $\Theta(T * (V + E))$. Putting together all the time complexities, we arrive at the overall-time complexity of the MLCT algorithm to determine the MLSPTs to be: $\Theta((E + V + E) * T) + \Theta(T * (V + E)) = \Theta(T * (V + E))$. Note that this will also be the same time-complexity to determine a sequence of minimum-hop shortest path trees by running the $\Theta(V + E)$-BFS algorithm on each of the static graph snapshots. Thus, the MLCT algorithm could determine the stable MLSPTs without any substantial increase in the run-time complexity.

In general, if $\Theta(X)$ is the run-time complexity to determine the communication topology of interest on a particular SU graph, the overall-time complexity to run the MLCT algorithm to determine a stable sequence of the communication topology of interest would be: $\Theta(T * (V + E + X))$. For example, if one wants to determine a sequence of stable minimum spanning trees using the MLCT algorithm, the overall-time complexity would be: $\Theta(T * (V + E + ElogE))$, where $\Theta(ElogE)$ is the time-complexity to determine minimum spanning tree on a graph of E edges (in turn dependent on the number of vertices) according to the well-known Kruskal's algorithm [7]. Note that the weight of an edge in a mobile graph $G_{t...t+k}(SU)$ is the

sum of the weights of the edge in the constituent static graphs $G_t(SU)$, $G_{t+1}(SU)$, …, $G_{t+k}(SU)$.

To determine network-wide communication topologies with unit values for the edge weights (for example: minimum hop shortest path trees and minimum connected dominating sets), the sequence of stable mobile graphs, Stable_$G_{1\ldots T}(SU)$, obtained by running the MLCT algorithm once for a particular input sequence of the static SU graphs covering the entire duration of the network session would be sufficient to determine each of such network-wide communication topologies; one could simply run the appropriate algorithms to determine those communication topologies on the *same* Stable_$G_{1\ldots T}(SU)$.

Example to Illustrate the Execution of the MLCT Algorithm

Figures 2, 3 and 4 illustrate an informative example for the execution of the MLCT algorithm to determine a stable sequence of shortest path trees across five time instants 1, 2, …, 5. The SU nodes are represented as larger circles (white background) with the node ID inside the circles; the PU nodes are represented as smaller circles (color background: either blue or red) and node IDs are not assigned. In the static graphs at time instants 1, 2, …, 5, an active PU node (i.e., is turned ON) is shown as a red-colored node and an idle PU node (i.e., is turned OFF) is shown as a
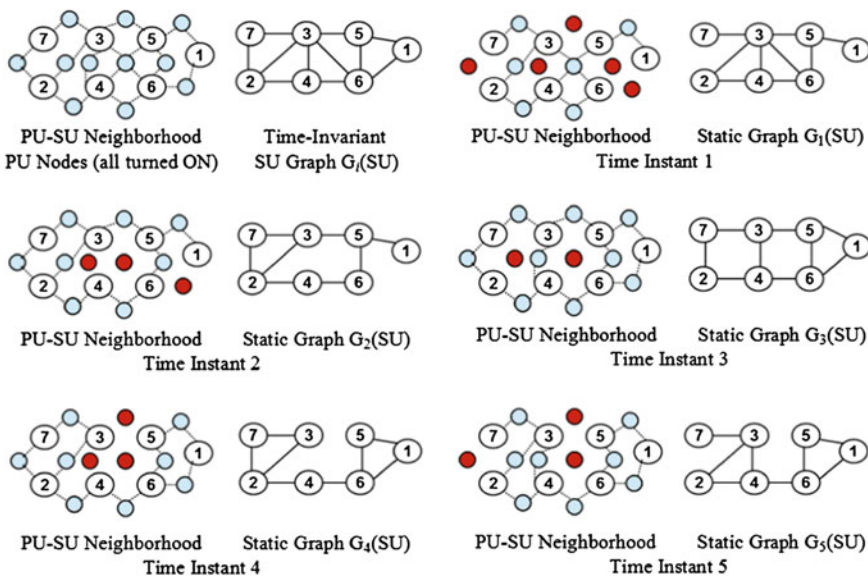

Fig. 2 Example to illustrate the sequence of static SU graphs of a cognitive radio Ad hoc network for five consecutive time instants
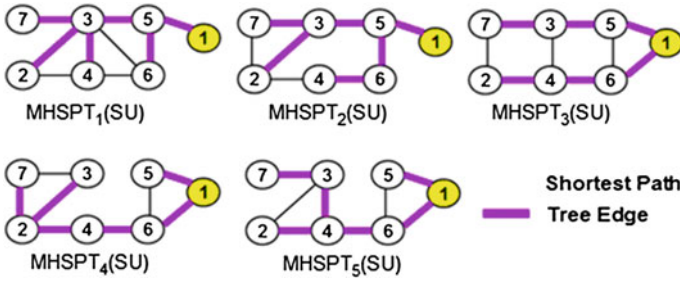
Fig. 3 Sequence of minimum hop shortest path trees (MHSPTs) on the static SU graphs of a cognitive radio Ad hoc network shown in Fig. 2

blue-colored node. The dotted links represent the presence of a PU node within the transmission range of the SU node(s). The time-invariant initial graph of SU nodes represents the network of SU nodes with all possible links: there exists a link between any two SU nodes in this graph if the two SU nodes are neighbors. In the static SU graphs (see Fig. 2) for time instants 1, 2, …, 5, there exists an SU-SU edge only if the edge exists in the time-invariant SU graph as well as there exists an available/idle PU node within the transmission range of the two SU nodes at that time instant. Depending on the availability status of the PU nodes, one or more SU-SU edges in the time-invariant SU graph may not be present in the static SU graphs captured in the time instants 1, 2, …, 5. We observe the static SU graphs captured at time instants 1, 2, …, 5 to differ depending on the availability of the PU nodes.

Let node 1 be the source node rooted at which we are interested in finding the shortest path tree. The shortest path tree on an individual graph or mobile graph is determined by running the BFS algorithm, starting from node 1. Just for simplicity and uniformity in the example illustrated, we break the tie to explore the neighborhood of the vertices in the increasing order of their IDs as well as we visit the neighbors of a particular vertex in the increasing order of their IDs. However, in the simulations (section "Simulation Study of the MLCT Algorithm"), we randomly break the ties to visit the neighbors. In the example graphs of Fig. 3, we observe that a shortest path tree for a particular SU graph (captured at a specific time instant) does not exist in the subsequent SU graphs and needs to be reconfigured for every time instant, leading to a total of four transitions (tree changes) when we aim for minimum-hop/minimum-height shortest path trees (MHSPTs) when we want to determine one at a particular time instant. The time-averaged height of the shortest path trees across these five time instants is $(3 + 3 + 3 + 4 + 4)/5 = 3.4$.

In Fig. 4, we observe the MLCT algorithm to be able to find a mobile graph that spans the time instants 1, 2 and 3 (but does not exist further) and a mobile graph that spans the time instants 4 and 5; thus there is a need for only one tree transition when we use the MLCT algorithm to determine a stable sequence of shortest path trees across time instants 1, 2, …, 5. The time-averaged height of the maximum lifetime shortest path trees (MLSPTs) is $(4 + 4 + 4 + 5 + 5)/5 = 4.4$. Though we
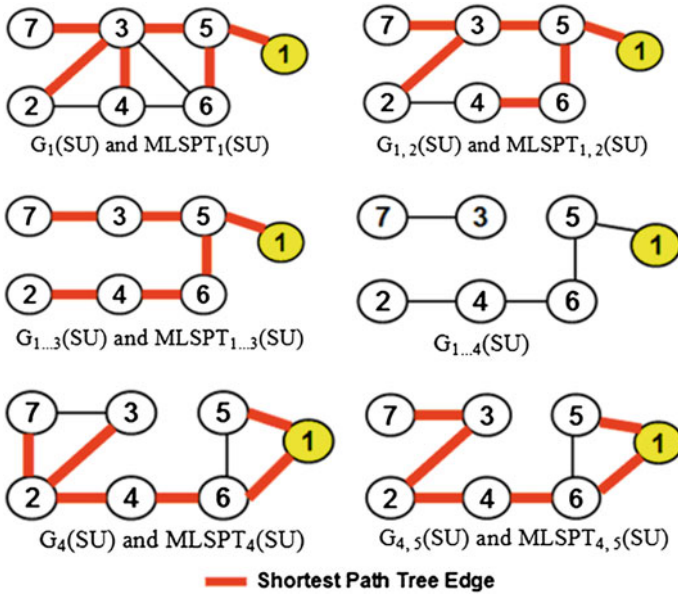
Fig. 4 Sequence of stable maximum lifetime shortest path trees (MLSPTs) on the static SU graphs of a cognitive radio Ad hoc network shown in Fig. 2

observe the average tree height of the MLSPTs (4.4) to be appreciably larger than the average tree height of the MHSPTs (3.4)—indicating the possibility of a -hop count tradeoff, we do not observe any such appreciable tradeoffs in the simulation results (see sections "Average Tree Lifetime" and "Average Tree Height").

Proof of Correctness of the MLCT Algorithm

We now prove that the MLCT algorithm does discover a stable sequence of the mobile graphs and the corresponding instances of the communication topology of interest such that the number of transitions from one instance of the topology to another in the sequence is the global minimum (i.e., optimum). We use the approach of proof by contradiction. Let m be the number of transitions (change from one instance of the topology to another) in the sequence of instances of the communication topology determined by the MLCT algorithm. Let there be a hypothetical algorithm that determines a sequence of instances of the communication topology such that the number of transitions is m' wherein $m' < m$. If such a hypothetical algorithm exists, for m to be greater than m', (without loss of generality) there should be a mobile graph $G_{t+p...t+s}(SU)$ for which the MLCT algorithm should have incurred at least one transition in the time span from $t + p$ to $t + s$, where t is the beginning time instant of the network session and $p < s$ are the

increments to t indicating sampling time instants $t + p$ and $t + s$ respectively, but that the hypothetical algorithm did not incur any transition in the communication topology from $t + p$ to $t + s$. If the hypothetical algorithm were to find a network-wide communication topology that exists from time instants $t + p$ to $t + s$, the mobile graph $G_{t+p...t+s}$(SU) must be connected. But, the MLCT algorithm underwent at least one transition during the time period $t + p$ to $t + s$ because the mobile graph $G_{t+p...t+s}$(SU) was not connected. This implies the hypothetical algorithm cannot find an instance of the mobile graph and the communication topology of interest whose duration of existence is larger than that determined by the MLCT algorithm. Hence, the sequence of instances of the communication topology determined by the hypothetical algorithm should have underwent at least the same number of transitions as the sequence of instances of the communication topology determined by the MLCT algorithm, which means $m' \geq m$. This is a contradiction to the initial hypothesis that $m' < m$ and hence the hypothesis cannot be true. Thus, the MLCT algorithm determines a stable sequence of the mobile graphs and the corresponding instances of the communication topology of interest such that the number of topology transitions is the global minimum (optimum).

Simulation Study of the MLCT Algorithm

We developed a discrete-event simulator in Java for Cognitive Radio Ad hoc Networks (CRAHNs) and implemented the Maximum Lifetime Communication Topology (MLCT) algorithm and adapted it to determine a sequence of maximum lifetime shortest path trees (MLSPTs) among the SU nodes. We also implemented the algorithm to determine a sequence of minimum hop shortest path trees (MHSPTs) that determines a MHSPT on a particular static SU graph and uses it in the static SU graphs for the subsequent time instants as long as the MHSPT exists. Such an approach is referred to as the Least Overhead Routing Approach, LORA, commonly used by the distributed routing protocols for ad hoc networks [15].

The network topology is sampled for every 0.25 s to generate the static SU graphs. Simulations are run for a total of 1000 s; thus, the number of static SU graphs constructed is 4001, including the static graph of SU nodes at time 0. The source node for a particular MHSPT or MLSPT is chosen randomly from the set of SU nodes. As MHSPTs follow the LORA strategy, a chosen source node for a particular MHSPT is continued to be used as the source node for the entire lifetime of the MHSPT. For the MLCT-MLSPT algorithm, the source node for a particular mobile graph $G_{t...t+k}$(SU) is randomly chosen among the SU nodes. According to the MLCT algorithm, when we find out that a mobile graph that existed from time instant t to $t + k$ no longer exists at time instant $t + k+1$, we run the BFS algorithm on the mobile graph $G_{t...t+k}$(SU), starting from a randomly chosen source SU node and the shortest path tree determined in $G_{t...t+k}$(SU) is considered to be the maximum lifetime shortest path tree (rooted at the particular source node) used for each of the time instants t to $t + k$.

The network is of dimensions 1000 × 1000 m. The number of SU nodes is fixed at 50; the number of PU nodes is varied from 25, 40, 50, 75 to 100 nodes. Accordingly, we define the PU-SU ratio to be the ratio of the number of PU nodes to the number of SU nodes. For the above said values of PU nodes and SU nodes, the PU-SU ratio values range from 0.5 to 2.0. The PU nodes and SU nodes are identified with a unique ID; the PU channels are identified based on the ID of the PU node that holds their license. Each PU node is assumed to own a channel of distinct frequency. The transmission range of the SU nodes as well as the PU nodes is fixed to be 250 m.

A PU node is considered to be in the neighborhood of an SU node if the Euclidean distance between them is within the transmission range (250 m). We construct a time-invariant graph of the SU nodes—G_i(SU) that indicates the set of all possible edges between any two SU nodes in the network: an SU-SU edge exists in G_i(SU) if the Euclidean distance between the two SU nodes is within the transmission range (250 m). At any time instant, an SU-SU edge exists only if the edge exists in G_i(SU) and at least one PU channel is available in their mutually intersecting neighborhood (i.e., a PU node in their mutually intersecting neighborhood is idle—turned OFF) at that time instant. The values of the parameters MAX_{Random_ON} and MAX_{Random_OFF} are each: 5, 10 and 20 s. A particular simulation is thus run for a fixed value of PU-SU ratio, MAX_{Random_ON} and MAX_{Random_OFF}, leading to a total of 5 * 3 * 3 = 45 scenarios for the operating conditions. The results (shown in Figs. 5 and 6) are average values obtained from 20 runs of the simulations for each combination of values for PU-SU ratio, MAX_{Random_ON} and MAX_{Random_OFF}.

The performance metrics analyzed in the simulations are: (i) *Average Tree Lifetime*: This is the average of the lifetime of the shortest path trees determined according to the MLCT-MLSPT and MHSPT algorithms for particular values of PU-SU ratio, MAX_{Random_ON} and MAX_{Random_OFF}. The larger the value for the average tree lifetime, more stable is the tree determined by the corresponding algorithm. (ii) *Average Tree Height*: This is the time-averaged height of the shortest path trees determined for the sequence of MLSPTs and MHSPTs. The tree height is a measure of the diameter of the network (the maximum of the minimum number of hops for a path between any two nodes).
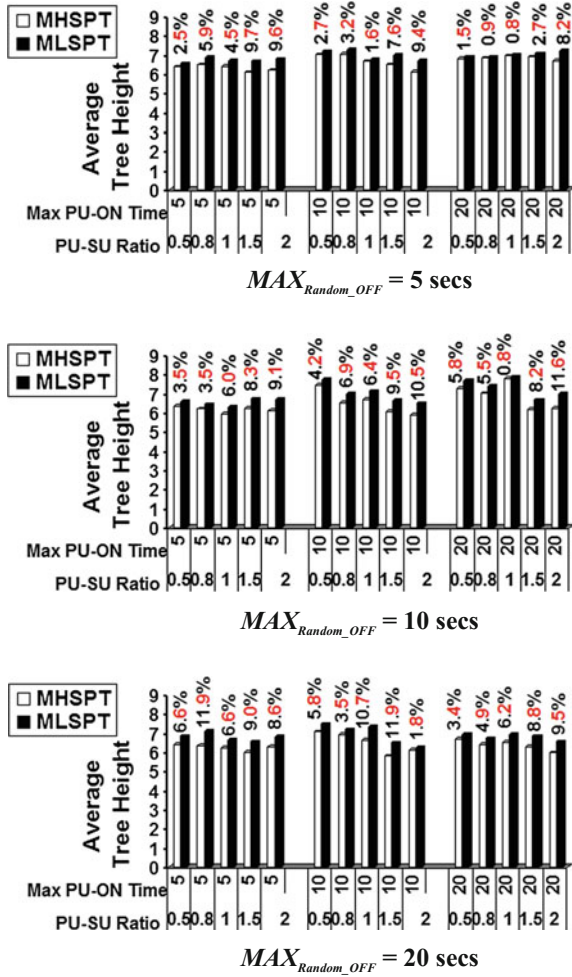
Average Tree Lifetime

We observe that the MLSPTs incur significantly larger values of tree lifetime compared to the MHSPTs for all the scenarios. The average tree lifetime incurred for the MLSPTs form benchmarks for the average lifetime that could be incurred for any centralized or distributed algorithm developed to determine shortest path trees spanning the SU nodes (or for that matter any network-wide communication topology spanning all the SU nodes) of a CRAHN. The percentage values shown in Fig. 5 are based on the ratio of the lifetime of the MLSPTs to that of the MHSPTs,

Fig. 5 Lifetime of maximum lifetime shortest path trees and minimum hop shortest path trees
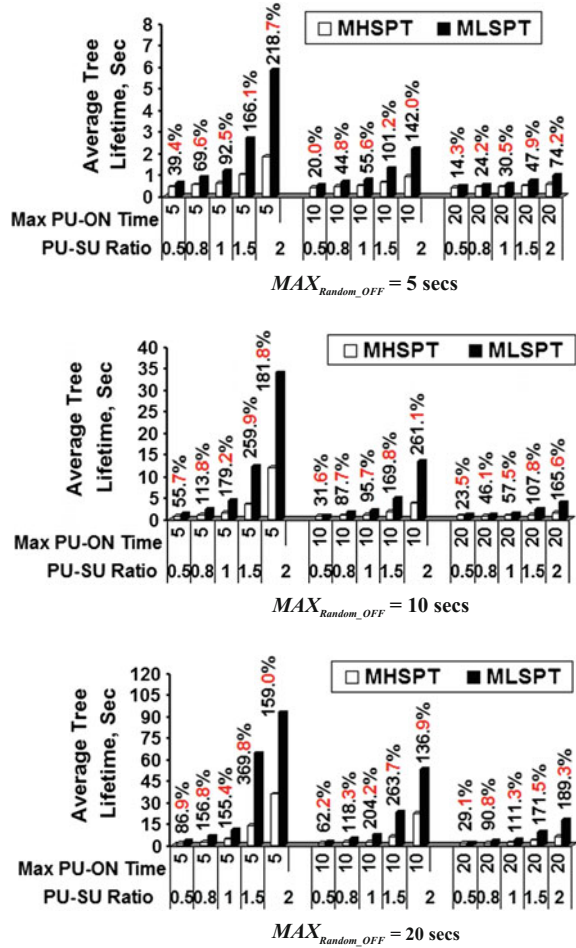
indicating the percentage difference in the lifetimes of the two trees. For a fixed value of the MAX_{Random_ON} and MAX_{Random_OFF} parameters, the average lifetime for both the MLSPTs and MHSPTs increases as the PU-SU ratio increases; nevertheless, the rate at which the average lifetime of the MLSPTs increases is significantly faster than the rate at which the average lifetime of the MHSPTs increases. In other words, the difference in the average lifetime incurred by the MLSPTs and MHSPTs increases as the PU-SU ratio increases. This is attributed to the larger number of PU channels (with increase in the number of PU nodes) accessible to the SU nodes. As a result, the SU-SU links are likely to exist for a longer lifetime. The MLCT algorithm leverages the a priori future knowledge about the PU channels and their availability and is able to find a long-living shortest path tree that exists in several of the static SU graphs spanning a mobile graph; on the

Fig. 6 Height of maximum lifetime shortest path trees and minimum hop shortest path trees

other hand, the MHSPT algorithm could make use of only information about the availability of the PU channels in the particular SU graph at which it is run.

As the PU-SU ratio increases, the average lifetime of the MLSPTs increases and the increase is further enhanced with increase in the available time of the PU channels (i.e., with increase in the value of MAX_{Random_OFF}). We observe that for a MAX_{Random_OFF} value of 5 s, the average lifetime of the MLSPTs increased by a factor of 2–9 (the rate of increase gets smaller for larger values of MAX_{Random_ON}) as the PU-SU ratio is increased from 0.5 to 2.0; on the other hand, for a MAX_{Random_OFF} value of 20 s and a fixed MAX_{Random_ON}, the average lifetime of the MLSPTs increased by a factor of 9–30 (the rate of increase is smaller for larger values of MAX_{Random_ON}) as the PU-SU ratio is increased from 0.5 to 2.0. As stated above, when the PU channels are used by their PU nodes for a longer time and

remain unavailable for the SU nodes, the rate of increase for the average lifetime of the MLSPTs reduces. A similar trend is observed for the MHSPTs too, albeit with a relatively lower magnitude of increase of average tree lifetime when the increase in the PU-SU ratio is complemented with an increase in the available time of the PU channels. As the number of PU channels increases (i.e., increase in the PU-SU ratio) and the availability time of the PU channels increases, we observe an increase in the percentage difference between the lifetimes of the MLSPTs and MHSPTs.

For a fixed MAX_{Random_OFF} value, as the MAX_{Random_ON} value increases, the average lifetime of the MLSPTs and MHSPTs decreases and the difference between the average lifetimes of the two trees narrows down for smaller values of PU-SU ratio. This is because when the number of available PU channels are lower and these channels are occupied for a longer time, the MLCT algorithm could not find a long-living shortest path tree that exists for several time instants. Nevertheless, the difference between the average lifetimes of the two trees is appreciable even under such scenarios (i.e., at smaller values of PU-SU ratio along with larger values of MAX_{Random_ON} and smaller values of MAX_{Random_OFF}): MLSPTs incur at least 14-30 % larger lifetime than that of the MHSPTs.

Average Tree Height

We observe the MLSPTs to incur at most 12 % larger height than MHSPTs (see Fig. 6). There is no definite trend in the difference in the height between the two trees across the simulation conditions. The overall average percentage difference in height between the two trees (when averaged over all the scenarios) is 6.2 %. With the significant gains in the lifetime of the MLSPTs and an inconsequential increase in the height of the trees, we could infer that there is no stability-hop count tradeoff in CRAHNs, unlike the MANETs and mobile sensor networks wherein we have observed a stability-hop count tradeoff [16] and stability-data gathering delay tradeoff (Perkins and Royer [17] respectively. Overall, we observe a slight increase in the percentage difference in the height between the two trees with increase in the PU-SU ratio and/or increase in the availability time of the PU channels. This could be attributed to the existence of connected mobile graphs that span over several time instants, but with the bare minimal edges to keep the graph connected. The MLSPTs determined on such sparse mobile graphs are likely to sustain a relatively larger height (as also evidenced in the example of Figs. 2, 3 and 4). In other words, in scenarios conducive for finding long-living stable trees, the percentage difference in the height of the two trees increases. Nevertheless, the increase in the height of the MLSPTs is not significantly larger than that observed in scenarios that limit the existence of the connected mobile graphs to fewer instants (true for scenarios with larger values of MAX_{Random_ON} and low-moderate values of PU-SU ratio and smaller values of MAX_{Random_OFF}).

Design of the MCPUR Protocol

Network Model

Let N_{PU} and N_{SU} be the number of licensed primary users (PU nodes) and unlicensed secondary users (SU nodes) respectively. Let R be the transmission range for the nodes and is the same for both the PU and SU nodes. Both the PU and SU nodes are assumed to be static. All the nodes are assumed to be uniform-randomly distributed in a 2-dimensional network of dimensions [0…*XMAX*][0…*YMAX*]. We assume an SU node to know the location of the SU nodes and PU nodes in its neighborhood. An SU node and PU node are considered to be neighbors if the Euclidean distance between the two nodes is less than or equal to the transmission range, R. As part of the network start-up procedure, we let each SU node to broadcast its ID and the list of neighbor PU nodes to all the SU nodes within its transmission range, R. Such a broadcast facilitates an SU node to learn the identity of its neighbor SU nodes as well as determine the common PU channels with each of the SU nodes in its neighborhood (using the list of neighbor PU nodes broadcast by its neighbor SU node).

We model a time-invariant SU graph $G_i(SU)$ of the CRAHN as the graph comprising of only the SU nodes; a link exists between two SU nodes u and v in $G_i(SU)$ if and only if the Euclidean distance between the two SU nodes is less than or equal to the transmission range, R. We model a time-variant SU graph $G_t(SU)$ of the CRAHN as a snapshot graph of the CRAHN at time instant t: the graph comprises of both the SU nodes and PU nodes; there is no link between any two PU nodes; there exists a link between two SU nodes u and v in $G_t(SU)$ if and only if there exists a link between u and v in $G_i(SU)$ as well as there exists at least one common PU node whose licensed channel is available in the neighborhood of both u and v for use (i.e., the PU node is turned OFF). We take snapshots of the CRAHN for every t_{sample} seconds and generate the snapshot graph $G_t(SU)$ of the network for every such sampling time instant.

For each of these time instants, we also identify the preferred PU channel that can be assigned for each SU node. The preferred PU channel for an SU node is the available PU channel whose corresponding PU node (is turned OFF) is in the neighborhood of a majority of the neighbors of the SU node. In case of a tie, we arbitrarily assign one among the contending PU channels as the preferred PU channel. We assign a preferred PU channel per SU node (that is common to a majority of the neighbor SU nodes of the SU node) so that the number of channel switches per hop on an end-to-end SU-SU path could be minimized.

We assume there are can be a maximum of N_{PU} distinct PU channels available for the SU nodes to be used as a data channel; a PU channel is simply identified by the ID of the corresponding PU node. An SU node is assumed to be able to identify a PU channel based on its ID. A PU channel is turned ON and OFF alternatively for a random time period uniform-randomly chosen each time from the range [0…MAX_{Random_ON}] and [0…MAX_{Random_OFF}] respectively. In other words, a PU node

decides to stay ON for a time period selected uniform-randomly from the range
$[0…MAX_{Random\_ON}]$; after staying active for the duration of time selected, the PU
node stays OFF for a different time period that is chosen uniform-randomly from
the range $[0…MAX_{Random\_OFF}]$. The node again turns ON and stays active for a
different time period chosen randomly from $[0…MAX_{Random\_ON}]$ and then goes
back to OFF mode. This procedure is followed independently at each PU node
throughout the duration of the network session.

We assume the presence of a common control channel [18] that is available for
use by all the SU nodes. An SU node uses the common control channel to sense the
local spectrum and identify the PU channels that are available for use. The SU
nodes also use the common control channel for the propagation of the RREQ and
RREP packets as well as the PU-UC (PU channel update) and RERR (Route error)
packets as part of route discovery and maintenance respectively.

Route Discovery

Whenever a source SU node (*s*) requires to start a data transfer session to a des-
tination SU node (*d*), the SU node initiates the route discovery process by broad-
casting a Route Request (RREQ) packet to its neighbors through the common
control channel that can be accessed by all the SU nodes. The broadcasts for the *s-*
d pair are identified using a unique ID. The RREQ packet includes three fields that
are updated by the SU nodes on the path along which the packet propagates from
the source to the destination: *Route Vector* field in which each intermediate for-
warding SU node as well as the source/destination SU nodes record their ID;
Preferred PU Channel Vector field that will be updated with the preferred PU
channel at that time instant (one entry for each node) by the source and destination
SU nodes as well as by each of the forwarding nodes of the RREQ packet; *Number
of Common Available PU Channels* field that is updated by a downstream node of
an SU-SU link indicating the number of common available PU channels with the
upstream node of the link (updated by all the intermediate forwarding SU nodes and
the destination SU node).

When an SU node *u* receives the RREQ packet (for the particular broadcast
session of the *s-d* pair) from an SU node *v*, it checks if there is at least one common
PU channel available in their respective neighborhood (i.e., there exists at least one
PU node in the neighborhood of both *u* and *v* and that is currently turned OFF,
available for use by the SU nodes) and also checks if its ID is already not listed in
the Route Vector field of the packet. If only both the above validations return true,
the SU node receiving the RREQ packet updates the Route Vector, the Pre-
ferred PU Channel Vector and the Number of Common Available PU Channels
fields respectively with its SU Node ID, its preferred PU channel and the number of
common PU channels available in the mutually intersecting neighborhood with the

SU node from which the RREQ packet was received, and then rebroadcasts the RREQ packet in its neighborhood of SU nodes. An SU node simply discards an RREQ packet in which its node ID is already recorded (to avoid looping) and/or if there is no common available PU channel to the SU node from which the RREQ packet was received.

The destination SU node (*d*) receives the RREQ packets along one or more paths from the source SU node (*s*). For every RREQ packet received, the destination SU node updates the three fields in the packet with entries corresponding to itself. After waiting to receive the RREQ packets for a certain time (large enough to have received sufficient number of RREQ packets for the particular broadcast cycle), the destination SU node selects the best *s-d* path as follows: The Route Vector field indicates the sequence of SU nodes on the path from the source to destination. Any two adjacent SU nodes in the Route Vector field constitute an edge of the path. The destination SU node simply adds the Number of Common Available PU Channels corresponding to the edges of the route traced by the RREQ packet. The destination chooses the route that yielded the largest value for the sum of the Number of Common Available PU Channels. In case of a tie, the minimum hop path among the contending paths is chosen; if the tie cannot be still broken, one among the contending paths is arbitrarily chosen.

The destination SU node *d* initiates a Route Reply (RREP) packet back to the source SU node *s*. Since all the links are bi-directional, the RREP packet could simply propagate back on the reverse of the chosen *s-d* path. The Route Vector and the preferred PU channel fields of the RREQ packet are copied in the RREP packet. An intermediate SU node receiving the RREP packet records in its routing table the source and destination end SU nodes of the path, the upstream and downstream nodes of the path (identified from the Route Vector field) as well as their preferred PU channels and forwards it further upstream towards the source SU node. The source SU node starts the data session after receiving the RREP packet.

Transfer of Data Packets and Channel Switch

The data packets for an *s-d* session between two SU nodes (*s* and *d*) are routed along the *s-d* path determined according to the route discovery procedure described in section "Route Discovery". An SU node remains tuned to its preferred PU channel for that time instant. To transmit a data packet, the sending node (i.e., the upstream node of a hop) takes the responsibility to keep itself in sync with the preferred PU channel of the receiving node (i.e., the downstream node of the hop). If the preferred PU channels for both the upstream and downstream nodes of a hop are not the same, the upstream node switches its data channel to the preferred PU channel of the downstream node of the hop, and then switches back to its preferred PU channel (after the transmission across the link is completed). Thus, an SU node

on the *s-d* path receives the data packet on its preferred PU channel; but may have to transmit the data packet on a data channel different from its preferred PU channel. This procedure is repeated at every link on the *s-d* path until the data packet reaches its destination *d*.

Route Maintenance

If the preferred PU channel for an SU node changes, it notifies all the upstream SU nodes of the *s-d* paths that it is part of through a PU Channel Update (PU-CU) packet sent on the common control channel. The upstream nodes accordingly update their routing table. Thus, any change in the preferred PU channel for a downstream node is locally handled without much overhead. When an upstream SU node finds out that it has no common available PU channel to its downstream SU node on the *s-d* path, it decides that the route has broken and notifies the source node *s* through a Route Error (RERR) packet. The source node *s* then initiates a fresh broadcast route-reply cycle to determine a new *s-d* path to the destination SU node *d*.

Simulation Study of the Maximum Common Primary User Channel-Based Routing (MCPUR) Protocol

We conducted the simulations in the ns-2 (v. 2.35) discrete-event simulator [10]. As the focus of the chapter is on the network layer (routing protocol design and performance analysis), we assume a perfect PU detection in the physical layer and an ideal MAC layer for transmission and reception of packets across each SU-SU link. The network is of dimensions [0…1000 m] [0…1000 m]. The transmission range of any node (PU node and SU node) is 250 m. We uniform-randomly distribute the PU nodes and SU nodes in the network area. We conduct the simulations with 50 SU nodes and vary the number of PU nodes with values of 25, 40, 50, 75 and 100 (a fixed number of PU nodes for a particular simulation run). Accordingly, we introduce a parameter called the PU-SU ratio that is the ratio of the number of PU nodes to that of the number of SU nodes in the network; for the above said number of SU nodes and PU nodes, the PU-SU ratio ranges from 0.5 to 2.0. Each PU node is alternatively turned ON and OFF for a time period uniform-randomly chosen from the range [0…MAX_{Random_ON}] and [0…MAX_{Random_OFF}] respectively. The values used for each of MAX_{Random_ON} and MAX_{Random_OFF} are: 5, 10 and 20 s. The network snapshot is taken for every 0.25 s; the total simulation time is 1000 s. A total of 30 *s-d* pairs were randomly chosen among the SU nodes. At any time, an SU node maintains a list of PU channels available in its neighborhood and is able to determine the presence/absence of one or more

common PU channels with each of its neighboring SU nodes. As explained in section "Network Model", a comprehensive master list of all the PU neighbors per SU node is broadcast by the SU node at the time of network start-up. We assume the availability of a common control channel for the SU nodes to use for identifying the status of the PU channels as well as perform route discovery and maintenance.

We compare the performance of MCPUR with that of the minimum hop-based shortest path routing (SPR) protocol. SPR is designed and implemented similar to that of MCPUR (as described in section "Design of the MCPUR Protocol"): the only difference is that the destination SU node chooses the minimum hop path traversed by the RREQ packets and sends the RREP on the chosen minimum hop path (ties are broken arbitrarily). Both the SPR and MCPUR routes are considered to be valid as long as there is at least one common PU channel available for each pair of the end vertices of the constituent links of the path.

Performance Metrics

The following performance metrics were analyzed in the simulations:

(i) *SU Network Connectivity*: We measure the probability of connectivity of the network of SU nodes by running the Breadth First Search algorithm [7] on each of the time-variant graph snapshots $G_t(SU)$ and evaluate the fraction (ranges from 0.0 to 1.0) of the total number of time instants the SU network is connected (i.e., there exists a path between any two SU nodes).

(ii) *SU-SU Edge Ratio*: This is the average of the fraction of the total number of SU-SU edges in the time-variant graph snapshots $G_t(SU)$ to that of the time-invariant SU network graph $G_i(SU)$, averaged across all of the connected SU network snapshots for the simulation runs corresponding to a particular combination of values for PU-SU ratio, MAX_{Random_ON} and MAX_{Random_OFF}.

(iii) *Hop Count per Path*: The time-averaged value of the number of hops per path, averaged over all the *s-d* paths and the duration of these paths.

(iv) *# Path Transitions*: We determine the average number of transitions per path (a path transition is defined as the change from one path to another path as a result of a broadcast route-request reply cycle), averaged over all the *s-d* sessions.

The simulation results presented in Figs. 7 through 10 are averaged over 20 runs for each combination of values for PU-SU ratio, MAX_{Random_ON} and MAX_{Random_OFF}. Figure 7 illustrates the SU-SU edge ratio observed for scenarios with probability of SU network connectivity greater than 0. Figures 9 and 10 respectively report the values of hop count per path and # path transitions for probability of SU network connectivity values of 0.20 or above.

Fig. 7 Probability of SU
network connectivity
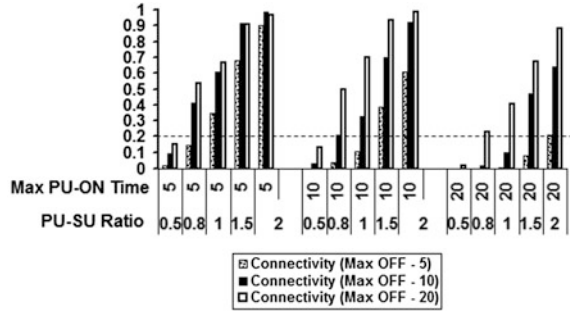
SU Network Connectivity

Figure 7 illustrates the values observed for the probability of SU network con-
nectivity (ranges from 0 to 1). We observe that for a fixed value of MAX_{Random_ON}
and MAX_{Random_OFF}, the probability of SU network connectivity increases with
increase in the PU-SU ratio. Thus, for a given number of SU nodes and a fixed
frequency of availability of the PU channels (fixed average duration of ON and OFF
times for the PU nodes), the probability of connectivity of the SU network increases
with increase in the number of PU nodes. For a fixed value of PU-SU ratio and
MAX_{Random_ON}, the probability of connectivity of the SU network increases with
increase in MAX_{Random_OFF}. This is as expected, as longer the duration of time the
PU nodes are turned OFF, the longer the time the corresponding PU channels are
available for use by the SU nodes. On the other hand, for a fixed value of PU-SU
ratio and MAX_{Random_OFF}, the probability of SU network connectivity decreases
with increase in MAX_{Random_ON} as the PU channels are likely to be unavailable for a
longer time.

SU-SU Edge Ratio

The SU-SU edge ratio (for a particular operating condition of PU-SU ratio,
MAX_{Random_ON} and MAX_{Random_OFF}) is the ratio of the number of edges observed in
the time-variant SU network graphs to that of the time-invariant SU network graph,
averaged over all the time instants for which the time-variant SU network graph is
connected across the simulation runs. For the SU node density considered in this
chapter (50 SU nodes with a transmission range of 250 m distributed in a network
of dimensions 1000 m x 1000 m: on average 10 SU neighbors per node in the
time-invariant SU network graph), though an SU-SU edge ratio of 0.6 is sufficient
to observe a non-zero probability of connectivity between any two SU nodes (i.e.,
probability of SU network connectivity > 0), in order to have at least a 20 %
chance of being connected, we require an SU-SU edge ratio of 0.88 or above (see
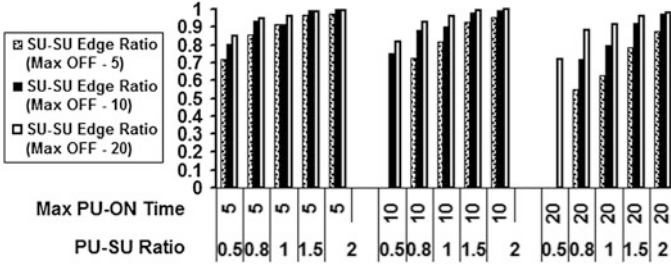Fig. 8).

Fig. 8 SU-SU Edge Ratio

Hop Count Per Path

We observe the hop count per *s-d* path to marginally increase (by about 8 % on average; at most 17 %) with maximum number of common PU channels-based routing compared to the minimum hop count per path encountered with shortest path routing (see Fig. 9). There is no particular operating condition in which we observe a larger difference in the hop counts; nevertheless, we observe the difference in hop count per path to be relatively more pronounced for MAX_{Random_ON} values of 10 and 20. We attribute the marginal increase in the average hop count per path with MCPUR to the protocol aiming at finding paths that have a larger number of common available PU channels in the neighborhood of the constituent SU-SU links of the path and only using the hop count to break any ties. Since MCPUR incurs only at most 17 % larger hop count per path and 8 % on average (compared to minimum hop-based shortest path routing) and a significantly lower number of path transitions (as low as 62 %, as observed in section "Path Transitions"), we conjecture that the performance of MCPUR with respect to hop count will be comparable (if not better) to any other CRAHN routing protocol.

Path Transitions

We define a path transition as change from one path to another. We also consider the discovery of the first *s-d* path as a path transition. We show the path transitions in a logarithmic scale (log to the base 2) in Fig. 10. Only path transitions incurred for SU networks with a probability of connectivity 0.2 or above are reported. We observe MCPUR to consistently incur fewer path transitions than that of SPR for all the conditions. The percentage reduction in the number of path transitions is very much appreciable for almost all the scenarios, especially as the SU network connectivity increases. For a given value of MAX_{Random_ON} and MAX_{Random_OFF}, the difference in the number of path transitions incurred by MCPUR and SPR shows a trend to increase with increase in the PU-SU ratio. Given that we increase the

Fig. 9 Maximum common PU channels-based routing versus shortest path routing: hop count per *s-d* path
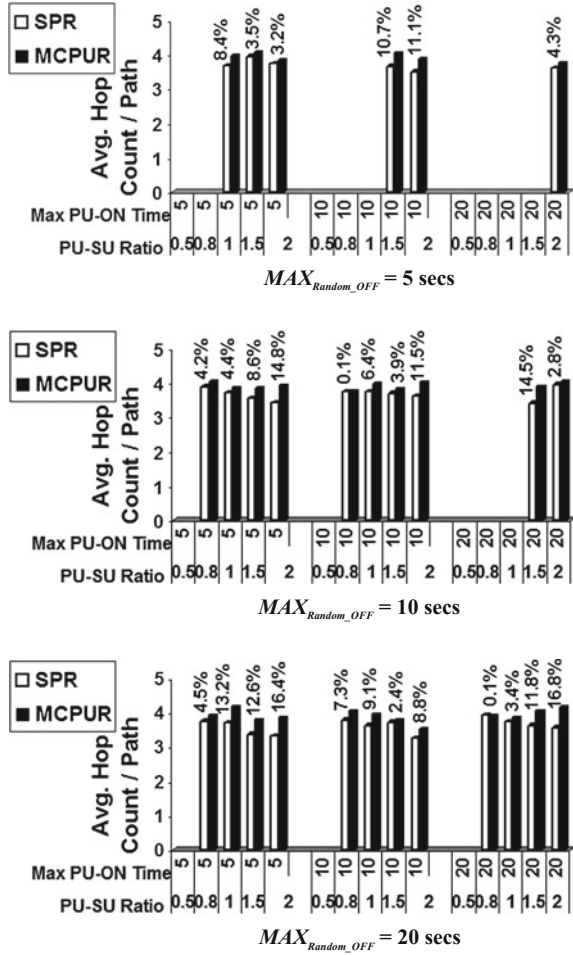
PU-SU ratio by keeping the number of SU nodes a constant and by only increasing the number of PU nodes, we could infer that as the number of available PU channels increases, the two protocols start showing diverse performance. MCPUR makes use of the increase in the number of PU nodes and their unique channels and discovers end-to-end *s-d* paths between two SU nodes that are likely to exist for a longer time. We can also observe that the absolute number of path transitions incurred for both MCPUR and SPR decreases with increase in the PU-SU ratio (for a fixed MAX_{Random_ON} and MAX_{Random_OFF})—indicating that the availability of a larger number of PU nodes and their channels in the neighborhood of the SU nodes lets the end-to-end SU-SU paths to stay for a longer time, even in the case of shortest path routing.

For a given PU-SU ratio and MAX_{Random_ON}, the absolute number of path transitions incurred with both MCPUR and SPR decreases with increase in

Fig. 10 Maximum common PU channels-based routing versus shortest path routing: number of path transitions
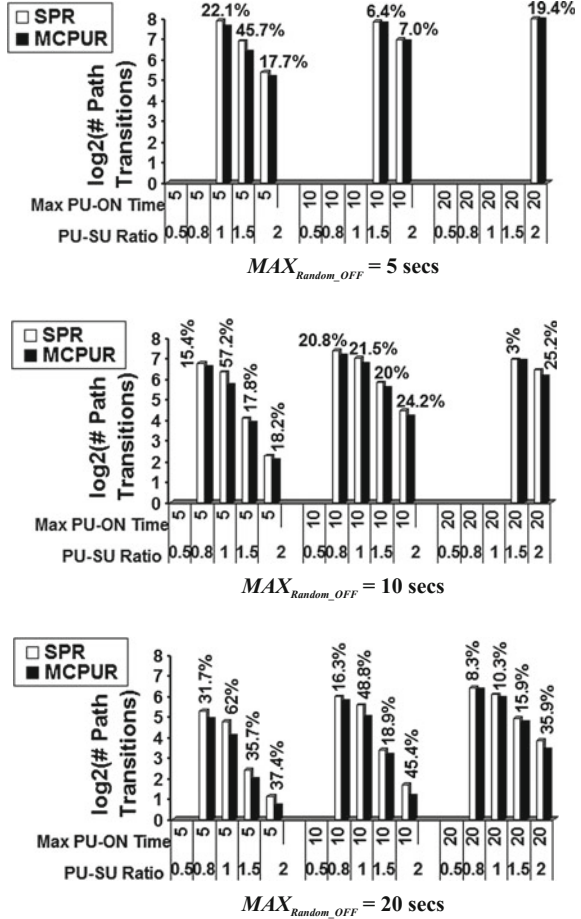
MAX_{Random_OFF}. This is due to the virtue of an increase in the availability time of the PU channels for SU use. When complemented with a larger number of PU channels (an increase in the PU-SU ratio), a larger value of MAX_{Random_OFF} facilitates better SU network connectivity and reduced number of path transitions for both MCPUR and SPR. MCPUR makes use of the increase in the availability time of the PU channels more effectively by discovering *s-d* paths that are relatively more stable than those discovered using SPR.

For a given PU-SU ratio and MAX_{Random_OFF}, the absolute number of path transitions for both SPR and MCPUR increases with increase in the value of MAX_{Random_ON}. This is because, as the PU nodes tend to use their channels for a longer time, the SU nodes could not use those PU channels and have to frequently switch from one PU channel to another in search of an available PU channel. We also observe that for a given PU-SU ratio and MAX_{Random_OFF}, the difference in the absolute number of path transitions for both MCPUR and SPR shows a trend to

decrease with increase in MAX_{Random_ON}. This is because as the PU channels stay active for a longer time, the number of common available PU channels in the neighborhood of the SU nodes decreases; the route selection criteria of MCPUR cannot be that effectively enforced when the number of common available PU channels for use is very limited—as a result, the difference in the number of path transitions for MCPUR and SPR reduces. Nevertheless, we observe MCPUR to be able to incur a significantly lower number of path transitions than that of SPR for larger MAX_{Random_ON} values.

Conclusions

Simulation results for the MLCT algorithm indicate that the average lifetime of the MLSPTs truly serve as benchmarks for average tree lifetime of shortest path trees among the SU nodes in a CRAHN as the MLSPTs incur substantially longer lifetime compared to the MHSPTs without incurring any significant increase in the tree height. The lifetime of the MLSPTs could be anywhere from 14 to 370 % larger than that of the lifetime of the MHSPTs; whereas, the increase in the height for the MLSPTs could be at most 12 %. Thus, there is no -hop count tradeoff when we aim for stable shortest path trees spanning the SU nodes in a CRAHN. Note that the MLCT algorithm uses a greedy strategy of determining a mobile graph that spans for the largest number of time instants starting from the current time instant and running an appropriate algorithm for the communication topology of interest on the mobile graph. We have shown that such a greedy strategy does indeed gives an optimal solution for the minimum number of transitions (and hence a maximum lifetime) among the instances of the communication topology determined for the duration of the network session. Greedy strategies rarely give optimal solutions with exceptions like the Dijkstra algorithm for minimum-weight path trees, Prim's and Kruskal's algorithms for minimum spanning trees, Huffman algorithm for prefix-free variable-bit encoding, etc. [7]. Thus, the MLCT algorithm joins the league of these greedy strategy based algorithms that give an optimal solution for the problem they are meant for.

Simulation results for the MCPUR protocol indicate that the protocol indeed tends to incur significantly fewer path transitions than that of the minimum-hop based shortest path routing (SPR) protocol for all the different combinations of values for the maximum PU channel ON time, OFF time and the number of PU channels. We also observe the hop count of MCPUR routes to be very close to that of the minimum hop count incurred with SPR and the average difference (increase) in hop count is only about 8 % (and the difference could be at most 17 %). On the other hand, MCPUR consistently incurs fewer path transitions than SPR for all the simulation conditions and the difference could be as large as 62 %. We are thus confident that the performance of MCPUR will be comparable to that of any other existing CRAHN routing protocol and at the same time incur a relatively lower number of path transitions. Future work has been planned with regards to extending

MCPUR for spectrum underlay-based CRAHNs (the network graph of which is likely to have unidirectional links) and mobile CRAHNs (where the PU nodes and SU nodes are mobile).

References

1. Abolhasan, M., Wysocki, T., Dutkiewicz, E.: A review of routing protocols for mobile Ad hoc networks. Ad Hoc Netw. **2**(1), 1–22 (2004)
2. Akyildiz, I.F., Lee, W.-Y., Vuran, M.C., Mohanty, S.: Next generation/dynamic spectrum access/cognitive radio wireless networks: a survey. Comput. Netw. **50**(13), 2127–2159 (2006)
3. Akyildiz, I.F., Lee, W.Y., Chowdhury, K.R.: CRAHNs: cognitive radio Ad hoc networks. Ad Hoc Netw. **7**(5), 810–836 (2009)
4. Cesana, M., Cuomo, F., Ekici, E.: Routing in cognitive radio networks: challenges and solutions. Ad Hoc Netw. **9**(3), 228–248 (2011)
5. Cheng, G., Liu, W., Li, Y., Cheng, W.: Spectrum aware on-demand routing in cognitive radio networks. In: Proceedings of the 2nd International Symposium on New Frontiers in Dynamic Spectrum Access Networks, pp. 571–574. IEEE, Dublin, Ireland (2007a)
6. Cheng, G., Liu, W., Li, Y., Cheng, W.: Joint On-demand routing and spectrum assignment in cognitive radio networks. In: Proceedings of the International Conference on Communications, pp. 6499–6503. IEEE, Glasgow, Scotland
7. Cormen, T.H., Leiserson, C.E., Rivest, R.L., Stein, C.: Introduction to Algorithms, 3rd edn. MIT Press, Cambridge, MA, USA (2009)
8. Ding, L., Melodia, T., Batalama, S.N., Matyjas, J.D.: Distributed resource allocation in cognitive and cooperative Ad hoc networks through joint routing, relay selection and spectrum allocation. Comput. Netw. **83**, 315–331 (2015)
9. Hou, Y.T., Shi, Y., Sherali, H.D.: Optimal spectrum sharing for multi-hop software defined radio networks. In: Proceedings of the 26th International Conference on Computer Communications, pp. 1–9. IEEE, Anchorage, AK, USA (2007)
10. Issariyakul, T., Hossain, E.: Introduction to Network Simulator NS2, 2nd edn. Springer, New York City, NY, USA (2012)
11. Joshi, G.P., Nam, S.Y., Kim, C.S., Kim, S.W.: A Cross-layer-based routing protocol for Ad Hoc cognitive radio networks. Int. J. Distrib. Sens. Netw. **2015**(938191), 1–7 (2015)
12. Lo, B.F.: A survey of common control channel design in cognitive radio networks. Phys. Commun. **4**(1), 26–39 (2011)
13. Ma, H., Zheng, L., Ma, X., Iuo, Y.: Spectrum aware routing for multi-hop cognitive radio networks with a single transceiver. In: Proceedings of the 3rd International Conference on Cognitive Radio Oriented Wireless Networks and Communications, pp. 1–6. IEEE, Singapore (2008)
14. Meghanathan, N.: On the of paths, steiner trees and connected dominating sets in mobile Ad hoc networks. Ad Hoc Netw. **6**(5), 744–769 (2008)
15. Meghanathan, N., Mumford, P.: A benchmarking algorithm to determine the sequence of stable data gathering trees for wireless mobile sensor networks. Informatica: Int. J. Comput. Inf. **37**(3), 315–338 (2013)
16. Meghanathan, N.: A location prediction based reactive routing protocol to minimize the number of route discoveries and Hop count per path in mobile Ad hoc networks. Comput. J. **52**(4), 461–482 (2009)
17. Perkins, C.E., Royer, E.M.: Ad hoc On-demand distance vector routing. In: Proceedings of the 2nd Annual International Workshop on Mobile Computing Systems and Applications, pp. 90–100. IEEE, New Orleans, LA, USA (1999)

18. Pyo, C.W., Hasegawa, M.: Minimum weight routing based on a common link control radio for cognitive wireless Ad hoc networks. In: Proceedings of the International Conference on Wireless Communications and Mobile Computing, pp. 399–404. IEEE, Shanghai, China (2007)
19. Shi, Y., Hou, Y.: A distributed optimization algorithm for multi-hop cognitive radio networks. In: Proceedings of the 27th International Conference on Computer Communications, pp. 1292–1300. IEEE, Phoenix, AZ, USA
20. Xie, M., Zhang, W., Wong, K.-K.: A geometric approach to improve spectrum efficiency for cognitive relay networks. IEEE Trans. Wireless Commun. **9**(1), 268–281 (2010)
21. Xin, C., Ma, L., Shen, C.-C.: A path-centric channel assignment framework for cognitive radio wireless networks. Mobile Netw. Appl. **13**(5), 463–476 (2008)
22. Xin, C., Xie, B., Shen, C.-C.: A novel layered graph model for topology formation and routing in dynamic spectrum access networks. In: Proceedings of the 1st International Symposium on New Frontiers in Dynamic Spectrum Access Networks, pp. 308–317. IEEE, Baltimore, MD, USA (2005)
23. Zhao, Q., Sadler, B.: A survey of dynamic spectrum access. IEEE Signal Process. Mag. **24**(3), 79–89 (2007)
24. Zhou, X., Lin, L., Wang, J., Zhang, X.: Cross-layer routing design in cognitive radio networks by colored multi graph model. Wireless Pers. Commun. **49**(1), 123–131 (2009)

Cognitive Radio Testbeds: State of the Art and an Implementation

Selahattin Gokceli, Gunes Karabulut Kurt and Emin Anarim

Introduction

Available spectrum is becoming overly congested with the increasing number of mobile devices and increased data rate expectations. The most critical issue for authorities who govern communication services is the management of current frequency spectrum. If this trend of the mobile devices will continue, spectrum scarcity will be more problematic for the deployment of next generation mobile technologies. As highlighted in Sun et al. [25], Long Term Evolution (LTE) technology has almost reached Shannon's capacity limit in terms of spectral efficiency within the %20 implying that it cannot be further increased significantly. Recently, cognitive radio (CR) technology has been studied widely as an important candidate for the management of dense spectrum usage. CR is a radio technology that can sense spectral activity even in wide spectrum bands and can use the obtained information strategically in wireless networks in order to provide necessary service quality to the user Cabric et al. [4].

CR networks contain two types of users which are primary users (PUs) and secondary users (SUs). PUs have the corresponding license of the spectrum band and they have a higher priority. SUs can only use the spectrum band when SUs do not harm usage of the PUs. CR systems are usually realized in the form of one of three

S. Gokceli (✉) · G. Karabulut Kurt
Wireless Communications Research Laboratory, Department of Electronics
and Communication Engineering, Istanbul Technical University, Istanbul, Turkey
e-mail: gokcelis@itu.edu.tr

G. Karabulut Kurt
e-mail: gkurt@itu.edu.tr

E. Anarim
Signal and Image Processing Group, Department of Electrical and Electronic
Engineering, Bogazici University, Istanbul, Turkey
e-mail: anarim@boun.edu.tr

main approaches which are underlay, overlay and interweave Goldsmith et al. [8]. In the underlay approach, CR user operates as SU and has knowledge of all PUs in the band. In this setting, CR user cannot have powerful signaling which can cause interference to PUs. In the overlay approach, CR user has the knowledge of codewords or messages of PUs and all these are used for cancellation of interfering signal. Lastly, the interweave approach covers opportunistic usage of all vacant spectrum bands without interfering transmissions of PUs.

A considerable increase of spectrum utilization efficiency can be achieved by allocating spectrum to both PUs and SUs without harming to the usage of PUs Liang et al. [15]. In order to provide this improvement, the spectrum usage of PUs must be detected rapidly and properly. Spectrum sensing provides this functionality and it is essential for the quality of CR systems. With spectrum sensing, CR users can detect vacant bands and utilize for communication these bands. When corresponding PUs start transmission, CR users look for another band that does not contain any PUs at that time. A new communication process is initiated when a free band is found.

Spectrum management in CR systems is the main functional block, which includes some steps to detect vacant bands without harming PUs communications and share band within all CR users. As the first step, the vacant band is detected with spectrum sensing. However, there are other steps beyond spectrum sensing in the management process; spectrum decision, spectrum mobility and spectrum sharing. Spectrum decision is a process to choose a free band in accordance with quality of service (QoS) expectations for CR users when several free bands are detected. At spectrum mobility step, CR users leave corresponding bands when PUs appear and switch to new bands. With spectrum sharing, vacant spectrum band is allocated fairly to all CR users, which are active in that band. With combination of these steps, a proper CR system is obtained. As mentioned earlier, spectrum sensing is the main element of spectrum management process. Without proper detection of PU activity, proper realization of CR system cannot be provided.

In this chapter, we present details of CR systems, specifically, we target real-time implementations of these systems. In accordance with this, we will highlight various software defined radio (SDR) implementations of CR systems in the literature. Most popular SDR tools are GNU Radio, LabVIEW and USRP hardware. GNU Radio is an open source software tool which contains firmware of USRP device Tucker and Tagliarini [26]. It is programmed with both C++ and Python and includes various libraries like digital signal processing. LabVIEW is also a software tool which has visual programming property that is different from most of other tools Instruments [10]. LabVIEW is very suitable to SDR implementations. In terms of hardware components, USRP is a hardware that can operate as transmitter or receiver and can realize real-time data transmission with different characteristics that depend on the technical features of the hardware. USRP can be programmed both with GNU Radio and LabVIEW tools.

In this chapter, we will give the details of our spectrum monitoring of 2.4 GHz ISM Band implementation by using SDR technology. Here, we handle spectrum sensing process of CR system implementation and use energy detection based sensing method. With the utilized spectrum monitoring, channel based usage is analyzed

and vacant channels are detected. In such implementations, measuring the performance of spectrum sensing is vital to understand system performance. Frequently, the performance of sensing for energy detection method is measured by checking detection and false alarm probabilities Yucek and Arslan [30]. Here, false alarm implies deciding falsely to band is in usage when band does not include active PUs actually. Detection probability is the probability of detection truly active user in band. By measuring these probabilities, the performance of sensing is measured and the decision about performance is made as good if false alarm probability results in a low value. In our implementation, we tackle these issues and provide beneficial insights about real-time deployments of CR systems, and also focus on channelization aspects.

Importance of Spectrum Sensing for Cognitive Radio

CR systems operate over wireless networks and characteristics of wireless channels where transmissions are realized, must be taken into account in order to understand principals of the CR systems. Transmission channels are very important especially for spectrum sensing. Physical transmission channel is usually modeled as

$$y(t) = h(t) * s(t) + w(t), \tag{1}$$

where $s(t)$, $h(t)$ and $w(t)$ represent transmitted signal, channel impulse response and noise signal respectively. * denotes the convolution operator. According to this model, signal is received with additive noise. In general, for mathematical simplicity, the sampled noise is modeled as Gaussian distributed. Due to fading process captured by $h(t)$, power of the received signal takes different values at distinct time and frequency slots. In the literature, some channel models that capture fading effects have been widely used Goldsmith [7]. An important one among these models is the Rayleigh channel model. Rayleigh model assumes that any direct path between transmitter and receiver pairs does not exist, and reflections from various reflectors are captured by the receiver. As mentioned in Kishore et al. [13], distorting effects of wireless channel caused by fading and shadowing, directly affect the spectrum sensing performance and the detection probability significantly decreases. Furthermore, as explained in Ahmed et al. [1], the presence of the PU cannot be detected and interference can be generated on licensed bands due to multipath fading. Realistic channel models are vital for CR system implementations, algorithms that are suitable to real-life should be selected.

Software Defined Radio Technology

CR technology can provide various improvements on spectrum management with its SDR and intelligent signal processing components Shukla et al. [23]. SDR is a communication system component that realizes some processes (which were processed

by using dedicated hardware in the past) by using a software with the help of an embedded system or a computer. SDR technology provides efficient usage and management of communication resources. In classical hardware based communication systems, modifications on systems can only be applied physically and this causes serious cost disadvantages. However, with SDR, devices can be upgraded by modification of software easily and without additional cost. In this way, the transmission quality of the system can be improved as well. Due to the requirements of protocols that change transmission characteristics dynamically in real-time, SDR technology has a significant potential in tactical and mobile communication technologies. It is expected that SDR technology will be a crucial part of future communication technologies.

In SDR, radio parameters can be defined with software in programmable processing units. Main examples of these units are field-programmable gate arrays (FPGA) and digital signal processors (DSP). SDR technology decreases hardware requirements and provides performance improvement. Programming SDRs offers flexibility between unsuitable radio frequencies and modulation techniques and allows the combination of these techniques Hickling [9]. For example, a radio that operates at 900 MHz can communicate with another radio which operates at 2.4 GHz. Therefore, a common communication structure can be created with SDR technology.

SDR devices offer various commercial benefits. Moreover, the usage of SDR technology has also benefits for research and development process of the next generation technologies. By using SDR devices, a system can be implemented and performance measurements can be obtained before the specialized deployment of the corresponding system. Simulation based performance analysis is the most common method in CR techniques. However, simulations are usually very insufficient in modeling real-time algorithms. For example, channel models like Rayleigh fading may not match realistic conditions. Some channel models can model the environment successfully, yet some techniques such as multiple-input and multiple-output (MIMO) cannot be modeled easily Instruments [10]. Due to advantages like flexible system modeling, configurable system settings, easy deployment of new generation communication algorithms, SDR usage in research efforts is more preferable over simulation based approaches.

In order to exploit the benefits of the SDR technology, especially for real-time SDR implementations, these three factors should be considered Instruments [10]:

1. **Hardware Abstraction**: If abstraction between software source and utilized hardware can be provided, users can understand the necessary technical details of the hardware without dealing system complexity and the system design can be simplified. The scope of the abstraction is very important. Abstraction should be provided at all layers in wide range.

2. **Comprehensive Software Development Environments**: SDR system development tools should only contain heterogeneous processing components. Moreover, they should be capable of joint operation with different computation algorithms and software tools. In this way, development and optimization of a system can be achieved by using proper tools and methods. Another requirement about the software tool is quality of the simulation services. Simulation in used software tool should reflect implementation environment and data processing must be rapid.

3. **Heterogeneous Multiprocessing**: Software tool of the SDR system should support different hardware and libraries. Multiple tools should be considered and system should be configurable. Joint usage of different hardware and libraries should not cause any problems.

Various processes can be implemented by using SDR before transmission; an SDR can define appropriate transmission channels, detect suitable modulation types to the channel, determine appropriate transmission gains and then realize transmission. Similarly, before reception; an SDR can detect energy distribution in the main channel and adjacent channels, define transmission quality, adaptively cancel interference, decode channel modulation and try to fix bit errors in accordance with the desired bit error rate (BER) level Mitola [16]. SDR devices support a wide variety of protocols and appropriate improvements.

As explained earlier, SDR implementation has a unique benefit from channel models perspective when compared to simulation modeling. By using SDR platform, algorithms are evaluated in a realistic environment and effects of wireless channel are investigated. In such experiments, CR's essential functionality, spectrum sensing must perform well to create a robust behavior. In accordance with these details, we implemented a robust SDR based CR system in real-time by measuring wireless channel results by a spectrum analyzer, as will be detailed later.

State of the Art on Cognitive Radio Testbeds

In this section, SDR implementations for CR studies are reviewed with an emphasis on spectrum sensing. Moreover, some leading studies are detailed and importance of SDR technology in CR studies is highlighted in order to emphasize their suitability for real-life.

Spectrum monitoring in CR systems is mostly implemented based on three sensing methods, which are energy detector based sensing, waveform based sensing and cyclostationary based sensing Yucek and Arslan [30]. Energy detector method is based on the comparison of output of the energy detector with a threshold, which is determined according to noise characteristics. Waveform method is based on the comparison of the received signal with a reference copy of the signal when receiver knows the corresponding pattern. Cyclostationary method contains the sensing of the primary signal by using signal features.

Energy detector is more frequently used than other methods because of its advantages like low computational complexity and easy implementation. Moreover, the receiver does not need any priory information about PU. In the literature, energy detector method is widely used in spectrum monitoring applications. In Xu and Alam [27], implementation advantages of the energy detection are shown with LabVIEW and NI PXI hardware. However, this work does not include wideband sensing operations. Similarly, in Yang et al. [28], in order to show flexibility of the SDR technology, a simple spectrum monitoring implementation is given by using LabVIEW and USRP hardware. This study does not cover ISM band sensing. In Denkovski et al.

[6], spectrum measurements of 2.4 GHz ISM band, which are obtained separately by Anritsu MS2690A spectrum analyzer and USRP2 hardware, are compared without channel based sensing. Similarly, in Najafzadeh et al. [17], real-time spectrum monitoring is implemented in 2.4 GHz ISM band by using LabVIEW and USRP. Channel based sensing is not included in this work.

In accordance with the 5G requirements, problems of current spectrum sensing frameworks are tackled and a novel method is proposed in Zhang et al. [31]. Traditional frameworks assume that all SUs should join the sensing process and each one should have good properties which is usually not valid in practice. Moreover, a wide sensing range and frequent detection cause higher costs and increased energy consumption, serious handicaps for 5G systems. To this end, a novel framework is proposed and spectrum sensing is implemented by spectrum agents. This approach overcomes some drawbacks and provides a performance that is suitable for 5G networks.

In Sharma et al. [22], due to vulnerability of using a common channel between SU transmitter and receiver, sequential channel scanning and quorum-based rendezvous methods are implemented without using a common channel with a small testbed by using NI USRP-2921 hardware and software tools LabVIEW and MATLAB. According to their model, SU pairs scan for common channel by knowing idle channel list a priori. Success of the model is demonstrated and supported by numerical results. In Jiang et al. [12], synchronicity between PU and SU nodes is targeted and issues as well as solutions related to asynchronous spectrum sensing are explained. As highlighted, asynchronous sensing is more suitable to realistic scenarios. Due to lack of related studies, issues about asynchronous sensing are handled. Moreover, two sensing schemes for non-cooperative and cooperative sensing are proposed. As shown with simulation results, proposed asynchronous approaches perform very well.

As studied in Soltani et al. [24], distributed implementation is more desirable than using a control channel for multi-hop setting. Therefore, a distributed cognitive radio network (CRN) architecture is modeled and full stack implementation is done in real-time by using GNU Radio and USRP N210. Moreover, a programmable emulation testbed, which includes emulated channels obtained from Radio Frequency Network channel Emulation Simulation Tool (RFnest), is developed and detailed performance measurements are implemented. As shown with results, this approach proposes an effective way with proper performance results.

In Sarijari et al. [20], a complete CR system is designed and implemented in real-time by using GNU Radio and USRP SDR hardware. Implementation consists of spectrum sensing, spectrum management and spectrum decision processes. Study targets spectrum scarcity problem and dynamic spectrum access (DSA) approach is utilized to provide a solution. Design model is proposed as a CR user, which behaves like an SU, aims to use licensed spectrum when PU does not use corresponding spectrum band. Hence, CR user first performs spectrum sensing and decides whether channel is free or not. Then, spectrum management is applied to manage the frequency band with other users. Spectrum decision part is responsible for usage decisions such as when to connect and duration of usage. At the end, data transmission is realized, which is the step that licensed band is used. For spectrum sensing, energy

detection based approach is utilized. Implementation setup consists of four USRPs which are used as SU transmitter, SU receiver, spectrum analyzer and PU transmitter. During implementation, two channels which are determined as 2.488 GHz and 2.492 GHz are used. With experiments, two different cases are observed, which are PU and CR users use the same channel or different channels. Firstly, by utilizing Gaussian minimum shift keying (GMSK) and differential quadrature phase shift keying (DQPSK) modulations, spectrum sensing is performed and CR user can identify PU user, afterwards changes channel. Then, two mentioned cases are measured. Performance is measured in terms of packet reception rate (PPR). According to the first case, PPR decreases around 91.6 % from the first case for SU when PU transmits with 50 % power transmission on the same channel. In second case, regardless of power level, performance decrease is not observed. Therefore, it is shown that spectrum sensing is very crucial for DSA based solutions and licensed user can have robust performance only when spectrum is utilized correctly. Because of these facts, it provides effective results. However, authors do not focus on spectrum sensing because, the main target is a fully functional CR system. Limitation of sensing with two channels is one example of this fact. Further details of spectrum sensing can be analyzed with channel based sensing approach.

Bogale and Vandendorpe [3] focus on noise variance uncertainty problem in the energy detector implementations and by utilizing some sensing algorithms such as max-min signal-to-noise ratio (SNR) signal energy based spectrum sensing algorithms are investigated. Real-time implementation is realized with USRP hardware. As explained in the work, experiments for energy detectors show the existence of noise variance uncertainty, which degrades detection performance. Novel max-min SNR signal energy based spectrum sensing algorithms are investigated provide good properties against this problem and minimize the effects of noise variance uncertainty, adjacent channel interference, carrier frequency offset (CFO) and symbol timing offset. Due to such benefits, these algorithms are utilized in the experiments. A linear combination of oversampled received signal is the core component of the models of the utilized sensing algorithms. Parameters that play significant roles in detection are determined and by using these results, experiments are carried out. Testbed includes USRPs as hardware components and LabVIEW as the software component. Carrier frequency of the transmission is considered as 433.5 MHz. For transmission, orthogonal frequency division multiplexing (OFDM) waveform is used and both transmitter and receiver contains filtering steps. Performance of the system is observed with different tests, which measure different metrics. As expected, increase in SNR causes higher detection probabilities. Moreover, results show that experimented channel performs like an additive white Gauss noise (AWGN) channel. Thus, study performs well and its suitability to real-life is valid due to asynchronicity between hardware, which makes study robust against some distortions such as CFO and timing offset.

A scenario is evaluated in Nir and Scheers [18] where a jammer perturbs the communication between two CR nodes and waveform switch is aimed during the communication. This model is implemented in real-time by using USRP hardware and CogWave software. CogWave is an open-source software which includes various

waveforms. With CogWave, CR node can change the waveform according to transmission quality. Model of study includes two waveforms which are OFDM waveform with detection and avoid (DAA-OFDM) and delay and add direct sequence (DADS). If obtained throughput becomes lower than a threshold, DADS which has low complexity, is changed with multichannel DAA-OFDM which implements energy-based sensing method. Testbed includes 4 USRP B100 and 1 USRP N210 devices and these devices are managed remotely by using an Arduino module. Real-time results are obtained by using this testbed. According to results, DAA-OFDM performs better than DADS in simulated channels. Furthermore, if threshold is higher than the throughput, switch of waveform is successfully implemented. It is clear that such a realistic scenario can be implemented easily with SDR tools and realistic properties can be observed.

Cooperative sensing scheme is implemented in real-time by using USRP N210 devices and GNU Radio software in Yoshimura et al. [29]. Similar to other approaches in the literature, this study also sees combination of CR and SDR technologies as a functional tool to solve spectrum scarcity problem. The study includes the implementation of a sensing algorithm with a model where two SUs perform cooperative sensing. Accordingly, Roy's Largest Root Test (RLRT) algorithm, which is based on calculation of eigenvalues of the received signal covariance matrix, is used as the base algorithm for sensing decisions. Two scenarios are experimented, RLRT with cooperation and without cooperation. Beyond real-time experiments, the scheme's performance is also observed with simulations for all scenarios. Results show that cooperative sensing has significant performance advantages. Cooperative sensing is based on combination of the individual sensing measurements of each CR. With this approach, individual sensing limitations and difficulties can be overcome and sensing performance can be clearly increased.

In Yoshimura et al. [29], two methods are used, centralized and distributed methods. A centralized database manages cooperative sensing process in centralized case, and this management brings additional processing disadvantages to network due to necessity of a certain control mechanism. In distributed case, mentioned limitations are not valid and a centralized data base is not used. This study covers the real-time implementation of an RLRT algorithm based cooperative sensing method as well as simulation validated observations. The study provides comprehensive implementation properties of this scheme and an important performance difference is demonstrated. However, channel based sensing is not handled and observed frequency range is limited. Moreover, implementation details are not mentioned in detail probably due to scope of the study. By handling a more comprehensive spectrum sensing process and the modification of some real-time characteristics, cooperative sensing can be shown as a more suitable method.

In Shahid and Yao [21], methods such as frequency hopping, band hopping, transmission band selection for interference avoidance are analyzed from CR perspective and a USRP and GNU Radio based testbed is created for the implementation of these methods. In the given testbed, both transmission and reception processes are included. According to system model, optimal channel is scanned by a device and if a channel matches expectations, a user can operate in this channel with high transmis-

sion performance. Afterwards, it is modeled that device continues to scan operation and observes other signal interactions in the channel. If the presence interference is validated, then the node can change the channel. Four algorithms are evaluated with testbed, which are frequency hopping, band hopping, band selection, and new band assignment. Frequency hopping algorithm is utilized to prevent interference from harmful users, which could be implemented in security aimed studies. With band hopping, interference signals from other users can be prevented. Similar to energy detection methods, if signal power has a higher value than a threshold, band is changed. Note that during this process, transmission frequency of the CR user is skipped in order to prevent confusion. Band selection algorithm includes first time operations such as scanning the spectrum and deciding proper frequency spectrum. When a band is selected, frequency and band hopping processes are followed with the help of new band assignment algorithm. Real-time evaluation of system model is done by using a USRP, a signal generator, a spectrum analyzer and GNU Radio software. USRP device is used as a CR node, spectrum generator and analyzer are used as PU and monitoring device, respectively. In experiments, three separate scenarios are tested. Band selection scenario handles USRP's band preference after initial scan when PU is not active. In signal detection scenario, signal detection performance of the USRP is tested with the operation of PU. In the last scenario, it is questioned whether USRP can yield to a PU or not. According to obtained results, channel preferences are similar for the first experiment. In the second experiment, signal strength of the PU is a crucial parameter and at 20 dBm gain, success rate of 91 % can be obtained. Authors mention that results are generally good, however, some performance limitations arise due to USRP's hardware characteristics. In order to improve this study further, this fact should be considered, additionally, wideband spectrum sensing could provide more efficient data as mentioned in the paper.

Sarijari et al. [19] include a GNU Radio and USRP based implementation of spectrum sensing with an energy detection method that is applied by calculation of magnitude square of the fast Fourier transform (FFT) output. By the implementation of this energy detection method, a hypothesis test is evaluated and probability of detection errors are calculated. According to the test scenario, channels are sampled 100 times separately and values are compared with thresholds to decide whether channel is free or not. Real-time experiments are evaluated with two USRPs, which are used as a receiver that implements spectrum sensing and a PU transmitter, which transmits data with GMSK modulation, respectively. In these experiments, two different scenarios are measured. In the first case, sensing ability of the receiver is measured without the existence of PU. Accordingly, vacant channels can be found successfully throughout time slots. In the second scenario, a similar model is implemented with the PU transmitter. Accordingly, receiver is again able to detect PU presence. Note that for all experiments, sensing is utilized to a fixed bandwidth, which is the range between 2.5 and 2.508 GHz. Study provides essential observations, however, wideband channel based sensing is not included. Scope of the testbed can be extended and more comprehensive results can be provided.

OFDM based spectrum sensing is analyzed and implemented in real-time by using GNU Radio and USRP in Blad et al. [2]. OFDM based sensing brings some diffi-

culties, which are mostly emerged from channel distortions. Due to these, detection of OFDM signals is a hot topic in the literature and various detectors have been proposed for this process. In this study, these detectors are analyzed and experimented in real-time in order to measure performances in a realistic environment which also includes hardware related issues. Moreover, handled algorithms are also modified to make them suitable to CFO compensation problem. CFO's existence harms autocorrelation properties of the detectors, thus these algorithms are modified accordingly. Methods and standard such as wireless regional area network (WRAN), generalized likelihood ratio test (GLRT) and averaging are evaluated in the paper. Testbed includes two USRPs, which are used as PU and SU, respectively. SU is considered as inactive during the sensing process. 2.4 GHz ISM band is considered and a shielded test environment, which does not contain interference from other sources, is chosen. According to implementation details of SU, firstly SNR estimation is handled. Then time-variant auto-correlation function is computed to be used in three algorithms. At the end, obtained values are compared with corresponding thresholds and decisions are made. For proper SNR values and fair results, receiver gain is kept constant. In experiments, two different scenarios are measured with different metrics. First one is based on an FFT length of 2048 and the other is based on a length of 256. According to the first scenario, the performance of GLRT and WRAN detectors are almost same. With increased sensing time, these performances increase too. Averaging detector performs badly in these experiments. However, it is advantageous when the SNR level is quite low, as in this case it can outperform other detectors. In the second case, results show similar facts, except the difference that a smaller FFT length with fixed CP size increases performances. Here, the WRAN and GLRT detectors show much better performances than the averaging detector when compared to previous case. To sum up, this study provides significant results about the detection of the OFDM signal and shows proper characteristics of the mentioned algorithms to realistic implementations. It is clear that real-time brings additional issues which are hard to model in simulations and analyzing corresponding algorithms in this environment provides more realistic observations.

Kumar and Sai [14] target link adaptation and performance improvement of the wireless networks when CR technology is used. As mentioned earlier, CR has various parameters that affect the system performance. Due to this fact, multi-objective optimization is a suitable approach. Various machine learning techniques are used with CR systems in order to determine efficient parameters. Similar to these approaches, an adaptive parameter adjustment method, which is based on genetic algorithm, is used in this study to provide performance improvement in terms of bit error rate, power, bandwidth and data rate. As different from previous studies on the same issue, real-time implementation of utilized system is realized by using SDR hardware, which supports MIMO configuration, and realistic performance results are obtained. Implementation of the transmitter contains three blocks; adaptive modulation, pilot inserter and genetic algorithm blocks. In adaptive modulation block, selection of parameters are determined based on some performance criterion. Selection of utilized modulation type is the main process. M-ary Quadrature Amplitude Modulation (MQAM) is used and decision of the M is made in accordance with the

quality of the link. If link quality quality is high, higher M such as 16 or 64 is used in order to increase the data rate. Otherwise, M is selected as 4. In the pilot inserter block, pilots are inserted to the transmitted data to correct channel effects and frequency offset. The genetic algorithm block is the essential part of transmission and the genetic algorithm is realized by following necessary steps. According to procedure, some parameters are taken as inputs and a score is given as output by the fitness function. Determined input parameters are bit error ratio, bandwidth, band efficiency, transmission power and data rate. These are modeled with some other parameters due to their dependencies, which are transmit power, modulation type, pulse shaping roll-off factor and symbol rate. By using these parameters, the algorithm is run by following necessary optimization steps of the genetic algorithm. As stated earlier, real-time experiments are implemented in order test performance of the algorithm in a realistic environment. NI RF Signal Generator 5673, NI RF signal Analyzer 5663 modules with NI PXIe-1075 chassis are used as hardware components. Channel is modeled as AWGN and BER is used as the performance metric. According the results, higher M values result in higher BER values. Moreover, system is able to optimize the initial selection and improve the performance. Thus, the study successfully shows the better performance of the adaptation algorithm in real-time and provides feasible results for dynamical parameter adaptation necessity of CR systems by using SDR technology.

In Chen et al. [5], design of the CRN is discussed and core details are explained. Moreover, created model is experimented in real-time by wireless open-access research platform (WARP), which is a functional SDR platform. According to the authors, the usage of machine learning algorithms in CRN has an important future in this research field. However, the algorithms such as matrix completion and decomposition, kernel adaptive filtering bring significant processing overhead, thus this usage requires powerful hardware for these computations. However, utilized hardware in CRN cannot be supported corresponding functions easily and possible cost of such an action is quite high. Thus, approaches for solution of this issue are valuable to deploy proper models. In this study, CRN testbed is modeled with connection of all nodes to the cloud and passing all computations to the cloud. With this configuration, limitations of the implementation of learning algorithm are overcome. In testbed implementation, nodes are connected to a local network. Network is modeled with some computers and computations are processed by these computers, which are also in the local network. As the software tool, MATLAB is used due to its ability of computation various algorithms and its suitability to such cloud based applications. However, MATLAB may cause some computation delays. In order to manage this and Ethernet connection related delays, a latency detection algorithm is designed. With this usage, problems related to these delays are mostly eliminated. Frequency band shifting, which is a core feature of CR, is evaluated in this study with one transmitter SU and one receiver SU. This feature covers the process that SU looks for vacant bands by sensing the spectrum. An important problem that band shifting process brings is to provide a robust communication link between two SUs. In order to overcome this, the authors propose an algorithm where SU applies spectrum sensing. If sensing fails, the same procedure is implemented for a new frequency band,

which is determined according to a previously created band set. As a control mechanism, the receiver checks incorrectly received packets and if number of these packets becomes higher than a threshold, transmission is shifted to the new band.

In real-time implementation, one wireless baby monitor is used as PU and two WARPs are used as SUs. Baby monitor is used in frequency hopping mode with 70 MHz bandwidth from 2.4 to 2.47 GHz. According to test scenario, firstly two WARP nodes operate at Wi-Fi channel 4. When PU starts transmission, transmitter node changes the channel after detecting PU with sensing. In this case, the number of incorrect packets increases and if number of these exceeds the threshold, the receiver also changes the channel. Due to Wi-Fi activity, channel 4 is used initially and channel 9 is used as the secondary channel rather than channels 1, 6 and 11. As shown with figures, the process outputs successful detection results and transmission quality can be kept at a desirable level. As future work, security in CRN testbed is described as an important issue and SDR testbed provides necessary data as tested with some minor experiments. SDR structure is very suitable for CRN based studies as explained in this study. After theoretical modeling, with an SDR testbed, comprehensive analysis could be done and success of such algorithms can be measured in detail.

As mentioned above, there are several studies that target CRN testbed. In the following section, we present an exemplary testbed implementation, that also emphasize wideband deployment in the ISM band along with channelization aspects.

An Energy Detection Based Spectrum Sensing Implementation

Spectrum monitoring is one of the main components of management of the frequency spectrum. With spectrum monitoring, dynamically changing spectrum can be monitored in real-time and vacant areas can be detected. Because of this fact, usage of spectrum monitoring method increases significantly. In the future, it is expected that devices will include this method and will be able to arrange the usage of spectrum by monitoring the spectrum dynamically. In this study, spectrum monitoring method is implemented by using an NI USRP-2921 device and LabVIEW within the scope of energy management system design and implementation for WLAN networks. With this implementation, channel based usage density of the 2.4 GHz ISM band is investigated.

In accordance with the mentioned requirements, LabVIEW software and NI USRP-2921 hardware is used in our implementation. LabVIEW is a visual programming tool that uses visual elements in programming which is a different approach compared to text based tools such as C, C++, Java and MATLAB. Visual programming feature makes LabVIEW easily programmable and understandable. Because of this, LabVIEW is frequently used in SDR based studies.

Programming in LabVIEW is managed in two separate parts, which are the front panel and the block diagram. In the block diagram, programming is done by

using elements from the functions palette and connecting them with virtual cables. Core programming structure in LabVIEW is called virtual instrument (VI). Lab-VIEW contains different function categories like programming, measurement, signal processing and data communication. With the contents of these functions, a communication system can be implemented. LabVIEW has three element types; control, indicator and constant. With control elements, values of some functions are controlled by inserting values in the front panel. Indicator type is responsible for demonstrating results in the front panel. Constant elements have control element properties, but these are managed in block diagrams.

USRP is a hardware that has real-time data transmission ability. USRP devices are used in various communication studies. USRP devices ease the learning, teaching and research processes of communication systems. Due to suitability of the USRP-2921 device to LabVIEW, implementation of SDR systems is quite straightforward. USRP-2921 device operates in 2.4–2.5 GHz and 4.9–5.9 GHz frequency bands. Moreover, it has half-duplex characteristic and has good hardware features, can support instantaneous bandwidth up to 20 MHz and I/Q sampling rate up to 25 MS/s Instruments [10].

Energy Detection Spectrum Monitoring System Model

In this section, the model of the spectrum monitoring method that is used in this study is explained. In our spectrum monitoring system, energy detection method that has a growing usage in the literature, is utilized because of its practical applicability and efficiency. Energy detection is a spectrum detection method that determines the presence of a signal in channel by calculating the signal power of received signal. At the energy detector, power outputs are compared with a threshold that is determined according to noise variance and a decision is made. Energy detection method can be modeled with following hypothesis test Yucek and Arslan [30]. Discrete narrow band signal model can be shown as

$$y[n] = h[n]s[n] + w[n], \tag{2}$$

where $y[n]$, $s[n]$ and $w[n]$ denote received signal, transmitted and desired signal and AWGN component, respectively. Moreover, n represents the sampling index, $h[n]$ is the fading channel coefficient. Test statistic for energy detector can be represented as

$$M = \sum_{n=0}^{N-1} |y[n]|^2, \tag{3}$$

where N denotes length of the observation vector. By comparing determined threshold λ_E and M, that is obtained after test statistics, the usage statistics of the corresponding frequency band can be obtained. This decision process can be modeled with two different hypotheses as

$$H_0 : y[n] = w[n], \tag{4}$$

$$H_1 : y[n] = h[n]s[n] + w[n]. \tag{5}$$

An algorithm for measuring detection performance can consider the two probabilities; P_D, detection probability and P_F, false alarm probability. P_D is the probability that whether a signal is actually at corresponding frequency band or not and high values of this probability are desired in general. This probability can be represented as

$$P_D = Pr(M > \lambda_E | H_1). \tag{6}$$

Furthermore, P_F is the probability of detection of a signal at a frequency band which does not include this signal actually. Opposite to P_D, low values of P_F are desired. P_F can be shown as

$$P_F = Pr(M > \lambda_E | H_0). \tag{7}$$

Threshold λ_E should be defined consistently in order to balance P_D, which should be high and P_F, which should have a low value. However, it is not easy to provide this balance because this process requires the noise knowledge and power properties of detected signals. Signal power is hard to be estimated because of changes in its properties in accordance with transmission characteristics. Therefore, determination of threshold targets proper estimation of false alarm probability Yucek and Arslan [30]. Noise variance is sufficient for the determination of threshold. AWGN component can be modeled as zero mean and $\sigma_w{}^2$ variance, $w[n] \sim N(0, \sigma_w{}^2)$. For easier analysis, signal component can also be modeled as zero mean Gaussian distribution with $\sigma_s{}^2$ variance, $s[n] \sim N(0, \sigma_s{}^2)$. As a result of these models, test statistic, M is obtained as chi-square distribution with $2N$ degree of freedom for $h[n] = 1$. However, channel characteristics cannot be obtained properly in real-life due to dynamic changes. Thus chi-square distribution is not a realistic model for real-time applications. In accordance with these practical limitations, spectrum monitoring system, which is utilized in this study covers calculation of M values in (3) for channels 1, 6 and 11. Moreover, these M values are compared with determined thresholds in order to determine whether channel is vacant or not.

The Structure of the Measurement System

An energy detection method based spectrum monitoring system is implemented by using LabVIEW and an NI USRP-2921 node in order to observe channel based spectrum activity and to understand the usage density of 2.4 GHz ISM band. In this section, details of this implementation are explained with elementary details.

Algorithm for energy detection method based spectrum monitoring implementation is prepared by using LabVIEW. In the code, wideband spectrum of the 2.4–2.48 GHz frequency band is obtained by shifting the carrier frequency. More-
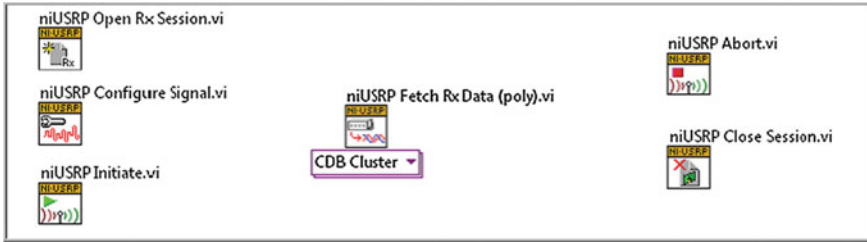
Fig. 1 LabVIEW USRP library receiver VI set

over, the energy in non-overlapping WLAN channels 1, 6 and 11 is calculated in parallel and decisions for channel usages are made by comparing energy results with corresponding thresholds. Each step is detailed below.

In the code, firstly VI sets, which are responsible for management of USRP node, are placed and connected to each other. Because of spectrum monitoring function, USRP node only receives data from channels and transmits nothing, thus only VI set that is used for receiver functionality of USRP is used. This USRP library and corresponding VI set is shown in Fig. 1. Properties of VI set can be summarized as:

niUSRP Open Rx Session: This VI is necessary to start the RF signal reception process of USRP node. Using this, the USRP becomes active.

niUSRP Configuration Property: With this VI, USRP's RF signal reception process can be managed. Various parameters such as IQ sampling rate, carrier frequency, transmit gain, active antenna or channel configurations for usage of multiple USRPs can be configured by using this VI and RF signal that has desired characteristic can be obtained.

niUSRP Initiate: This VI gives feedback to USRP about parameter configurations and informs that samples can be received.

niUSRP Fetch Rx Data: It starts IQ data reception by the USRP node and informs that IQ data can be processed. Without this VI, processing of IQ data cannot be obtained.

niUSRP Abort: This VI gives close session message to USRP and new session with different parameters can be started after this message.

niUSRP Close Session: As the last component of reception set, this VI ends data reception and allocates a memory section for this session.

After the configuration of this VI set, related parameters for wideband spectrum monitoring application are configured with niUSRP Configuration Property component. These parameters are configured as in Fig. 2. Afterwards, the shift of carrier frequency parameter, which is crucial for wideband spectrum analysis, is arranged. USRP-2921 has instantaneous bandwidth up to 20 MHz and USRP cannot operate in bandwidths higher than 20 MHz. However, targeted ISM band has bandwidth of 80 MHz and USRP node cannot support this application without modification. In order to overcome this limitation, carrier frequency can be changed during reception and observation frame can be shifted. To this end, a structure to shift carrier fre-
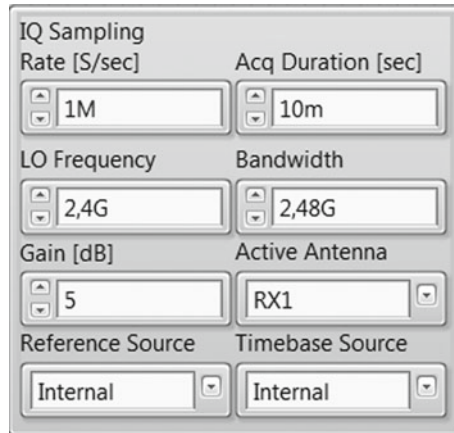
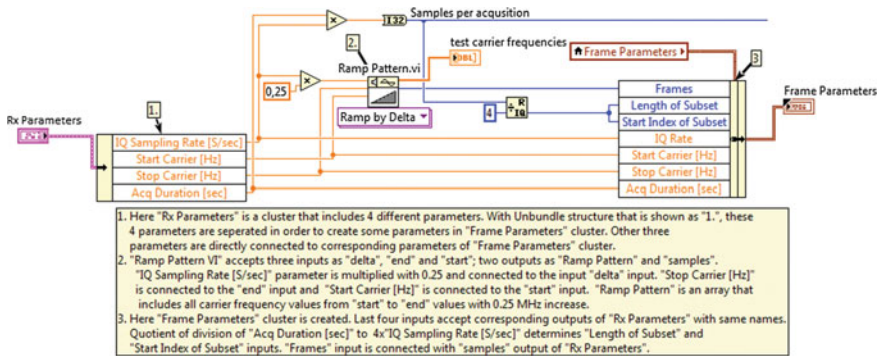**Fig. 2** Parameters utilized in LabVIEW code

Fig. 3 Carrier frequency configuration in LabVIEW code

quency is configured as in Instruments [11] and USRP node becomes able to observe the complete target bandwidth.

As the first step of carrier frequency configuration, initial frequency is set, which is 2.4 GHz as shown in Fig. 2, and this control element is connected to Ramp Pattern VI as input. Ramp Pattern VI accepts a start value, a delta value and an value; creates an array that contains elements, which takes values from the start value to the end value with an increase by a delta value. By setting this delta value as a multiple of "IQ Sampling Rate" parameter with 0.25 and end value as 2.48G, an array that includes all carrier frequency values from 2.4 to 2.48 GHz with 0.25 MHz increments.

Furthermore, a cluster with name of "Frame Parameters" is created as in Fig. 3 to be used in Aggregate Spectrum VI, where wideband spectrum is created by combining smaller spectra. In this cluster, "Frames" value represents total number of carrier frequencies, "Length of Subset" value represents number of frequency bins that are used for taking samples from spectrum interval and "Start Index of Sub-

set" value represents start point of sample in frequency domain. Numerical values of these parameters are demonstrated in Fig. 4.

Later, mentioned parameters that are used as controls are sent to two nested for loop structure. In order to process each carrier frequency separately and to measure each small band consistently, two nested for loops are used. Here, each element of carrier frequency array is processed by exploiting the "Auto-Indexed Tunnel" property of the loop due to the requirement of changing the carrier frequency at each step. With some other values, which will be detailed in following steps, this nested loop structure is created as in Fig. 5. Contents of this structure are the core parts of this code. Reception and processing of data are realized in this structure.

As the second step, RX I/Q SubVI is placed in the nested loop structure. In Lab-VIEW, a SubVI can be defined as narrow-scoped VI, which also has similar properties just as in function structures of text based programming languages. Contents of this SubVI are shown in Fig. 6.

As it can be seen from the figure, some commands are defined in the structure called "Case Structure", which is used for realization of some operation in the case when an input matches the expected value. As long as these values match each other, structure continues to run. As the first component, niUSRP Property Node is placed. As explained earlier, USRP parameters are configured according to controls connected as inputs. "Active Antenna" parameter determines USRP's antenna to be used, the "Number of Samples" parameter determines the number of samples, "Carrier Frequency" parameter determines the carrier frequency and spectrum area to be used for RF signal reception. Afterwards, USRP VI set, which is responsible for signal reception are added and obtained data from USRP, is given as the output of SubVI by using niUSRP Fetch Rx Data component. After this SubVI, Aggregate Spectrum SubVI is prepared, which has the same block diagram view as shown in Fig. 7. Accordingly, the power spectrum is obtained by using input data and Power
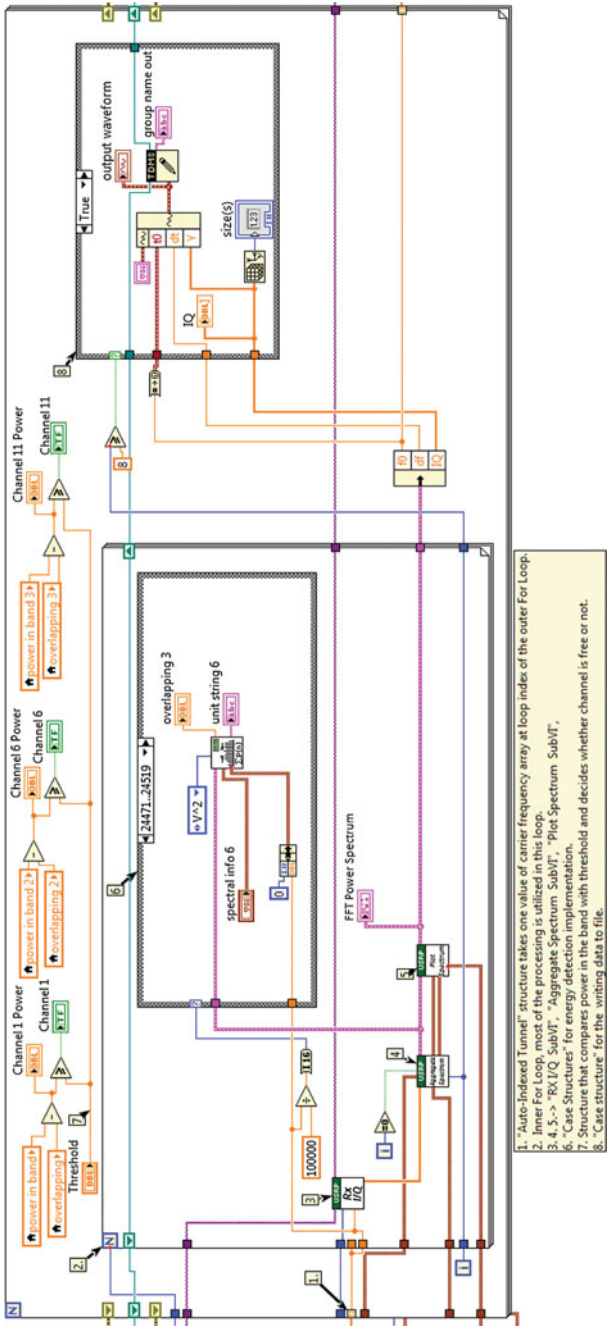
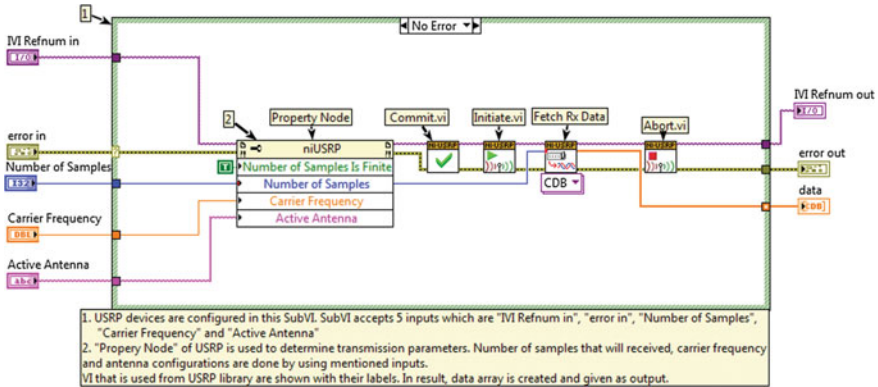**Fig. 5** Nested for loops and correspondent SubVI

1. "Auto-Indexed Tunnel" structure takes one value of carrier frequency array at loop index of the outer For Loop.
2. Inner For Loop, most of the processing is utilized in this loop.
3. 4. 5.-> "RX I/Q SubVI", "Aggregate Spectrum SubVI", "Plot Spectrum SubVI",
6. "Case Structures" for energy detection implementation.
7. Structure that compares power in the band with threshold and decides whether channel is free or not.
8. "Case structure" for the writing data to file.

Fig. 6 Block diagram of RX I/Q SubVI

Spectrum and PSD VI, which is used for the measurement of power spectrum in accordance with the sampling rate. Windowing property of this VI is also configured by the connection of "Window Parameters" control. This control includes various windowing alternatives and windowing configuration is also demonstrated to user by Window Properties VI.

Then, frequency resolution information of the power spectrum, which is obtained with the Power Spectrum and PSD VI, is sent to a SubVI where spectrum details are extracted. If SubVI gets data for the first time, an empty array to keep spectrum measurements is defined and this array is filled iteratively with new spectrum information which is obtained for the corresponding carrier frequency. At the end, combined wideband spectrum is obtained. For the first-time data case, which is valid when data is received for the first time by SubVI, an empty array is created by Initialize Array VI by connection of multiplication of "Frames" and "Length of Subset" controls and 0 to value control. With these connections, the length of array is defined and is totally filled with zeros. Then, by combining this array and frequency resolution information, latest spectrum data is created.

In order to exploit amplitude knowledge obtained from power spectrum, firstly redundant portion is removed with the help of "Start Index Subset" parameter and then the remaining data is inserted into corresponding index of spectrum data array after the removal of previous data which starts from this index. The index value is defined as multiplication of "Iteration" control, which shows the number of iterations of for loops and "Length of Subset" parameter. In this way, spectrum information is managed correctly in accordance with the processed carrier frequency. As the result, obtained power spectrum is combined with previous spectrum data and this process continues similarly whenever this VI is called again (Fig. 7).

After Aggregate Spectrum SubVI, "Case Structure" is generated for energy detection application and band energies of non-overlapping WLAN channels 1, 6 and 11 are calculated in parallel. This structure is composed of different cases, which are defined with some frequency intervals and energy calculation method is run for the
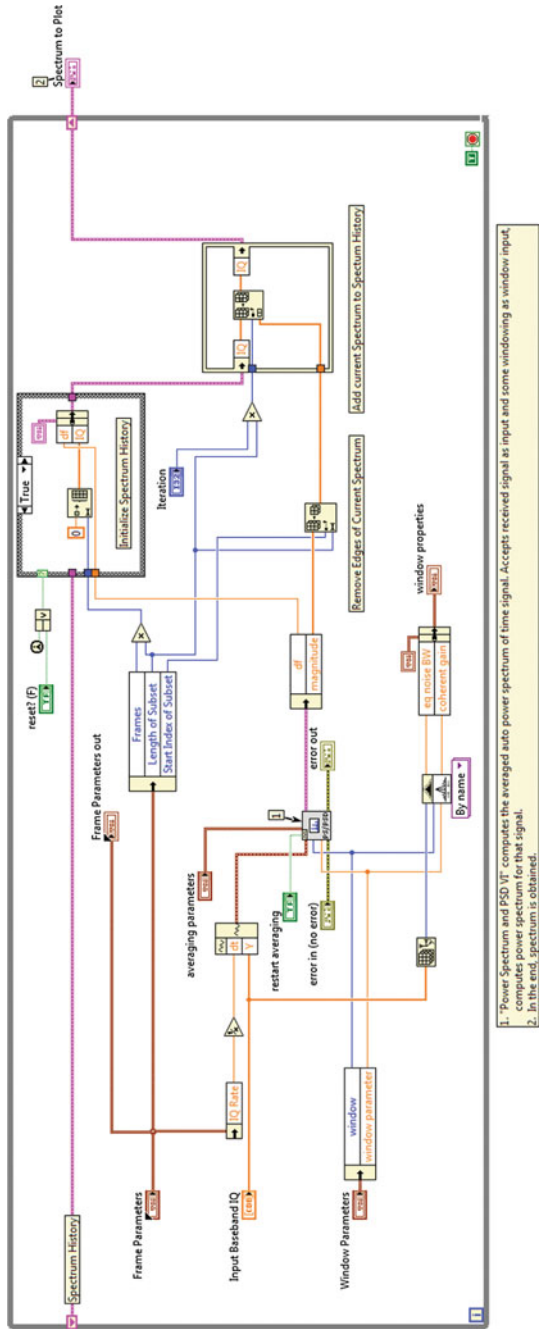
Fig. 7 Block diagram of Aggregate Spectrum SubVI

input carrier frequency value depending on which interval contains this input. 7 different cases are defined in this structure. Three essential cases are created for measurements of channels 1, 6 and 11, where frequency intervals are characterized as 2.402–2.422 GHz, 2.427–2.447 GHz and 2.452–2.472 GHz, respectively. For measurements of other channels, which are located in remaining parts of ISM band, three more cases are also defined. A seventh case, which does not contain energy calculation method, is created in order to manage other possibilities where input value is not included in all six cases. Each main case has the same content and an exemplary block diagram is shown in Fig. 8 for the first case.

Other three cases, which are not located in channels 1, 6 and 11, have similar contents and one more exemplary diagram is shown in Fig. 9 for these cases.

Lastly, the content of case 7, which does not have energy calculation method, is shown in Fig. 10.

In order to plot the obtained spectrum from "Case Structure", Plot Spectrum SubVI is created. This SubVI is connected after Aggregate Spectrum SubVI and spectrum data that is given as output in Aggregate Spectrum SubVI is connected as input to Plot Spectrum SubVI. Moreover, a windowing setting and "Frame Parame-
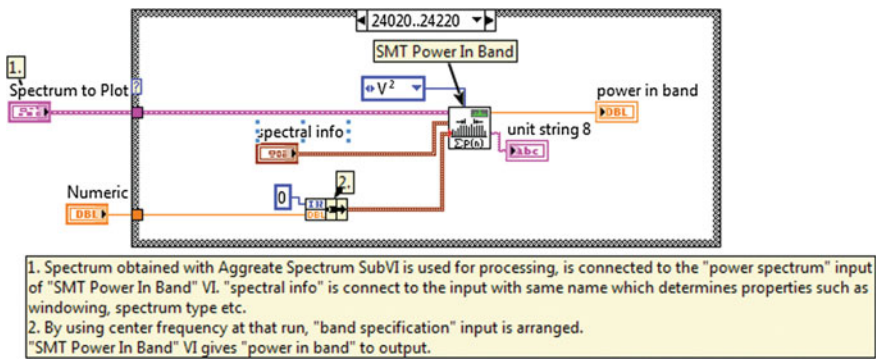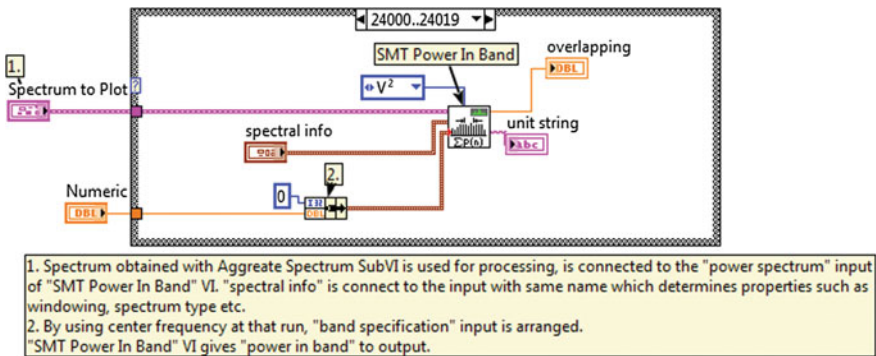

Fig. 8 Case Structure content for three main cases

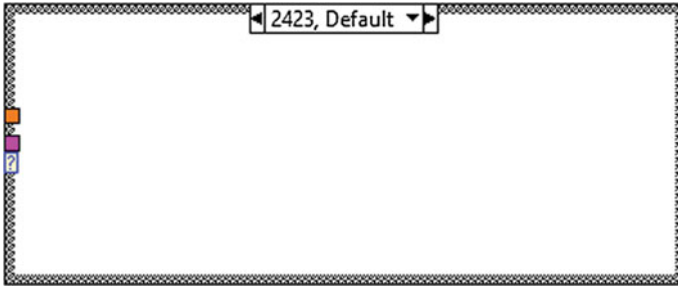**Fig. 9** Case Structure content for overlapping channels
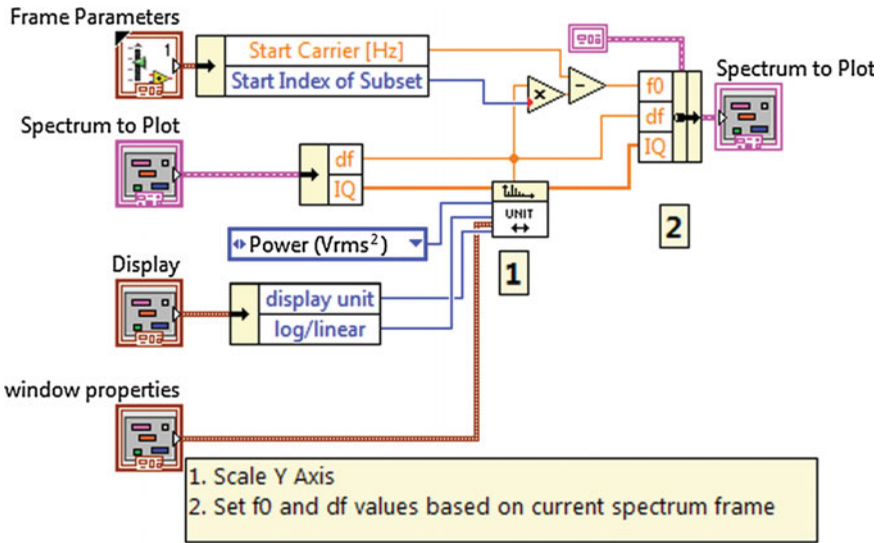
Fig. 10 Block diagram of case 7

Fig. 11 Block diagram of Plot Spectrum SubVI

ters" control are also connected as inputs. Spectrum array includes multiple components and two subarray, which contain frequency resolution. Then amplitude values are extracted. These values are used for start point of the spectrum plot. Similarly, "Start Carrier" parameter, which represents first carrier frequency value, and "Start Index of Subset" parameter are extracted from "Frame Parameters" data set. "Start Index of Subset" parameter is multiplied with the frequency resolution value and subtracted from "Start Carrier" parameter to determine the starting point of the spectrum. Furthermore, Spectrum Unit Conversion VI in the Spectral Measurement Toolkit is added for unit setting of amplitude values in the spectrum. Diagram of this VI is shown in Fig. 11.

Block diagram of LabVIEW code which is created after all previous steps, is demonstrated in Fig. 12.
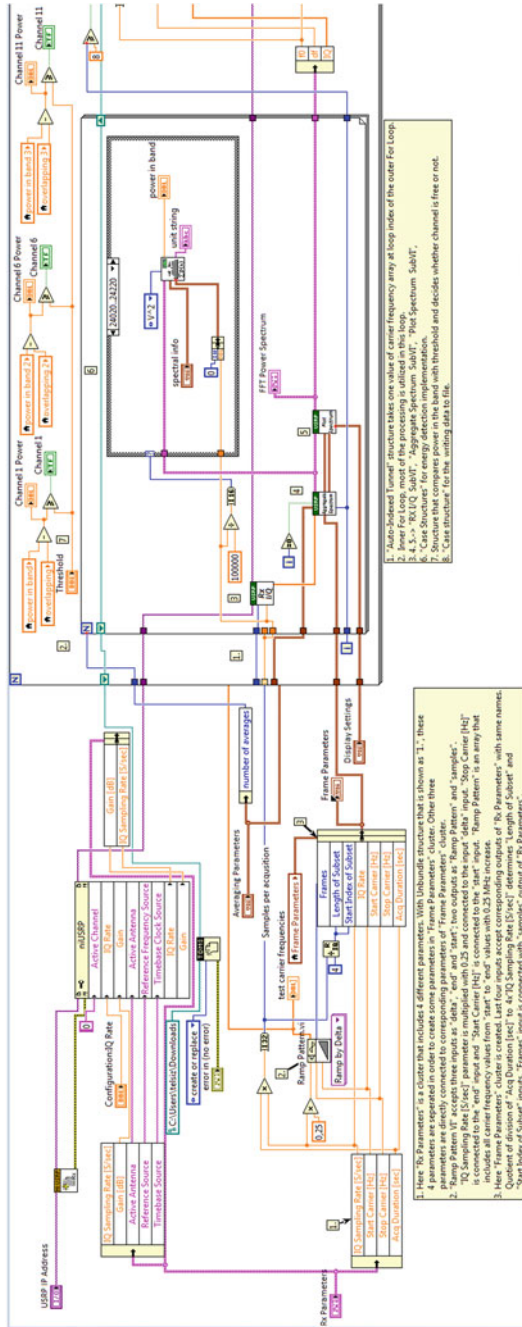
Fig. 12 Block diagram of the spectrum monitoring code

Real-Time Results

In this section, real-time performance measurement results of created energy detection based spectrum monitoring system are shared and explained. The testbed used during these experiments contains one computer with LabVIEW software, one NI USRP-2921 node and one switch for Ethernet connections. A physical view of this testbed is shown in Fig. 13. Before experiments, necessary antenna and cable connections of USRP node are provided for the reception of spectrum information in ISM band and the USRP device is connected to the computer with an Ethernet cable for processing of data. After physical preparation, antenna, carrier frequency and number of samples parameters are configured in LabVIEW. These configured values are shown in Table 1. As explained in system model details, band energy values (M) for each channel are calculated separately. Then, these values are compared with corresponding thresholds (λ_E). Virtual LEDs in LabVIEW are used to visualize the case that M value is higher than the threshold.

After necessary configurations, experiments are executed. USRP device is configured to receive 5000 sample per second and signals from ISM band are received and processed. An exemplary result, which was obtained after one run of the system, is shown in Fig. 14. As it can be observed from the figure, at that moment, channels 1 and 6 are in high usage, but channel 11 is not densely populated. In order to under-

Fig. 13 Physical view of the testbed

Table 1 Configuration parameters of the USRP

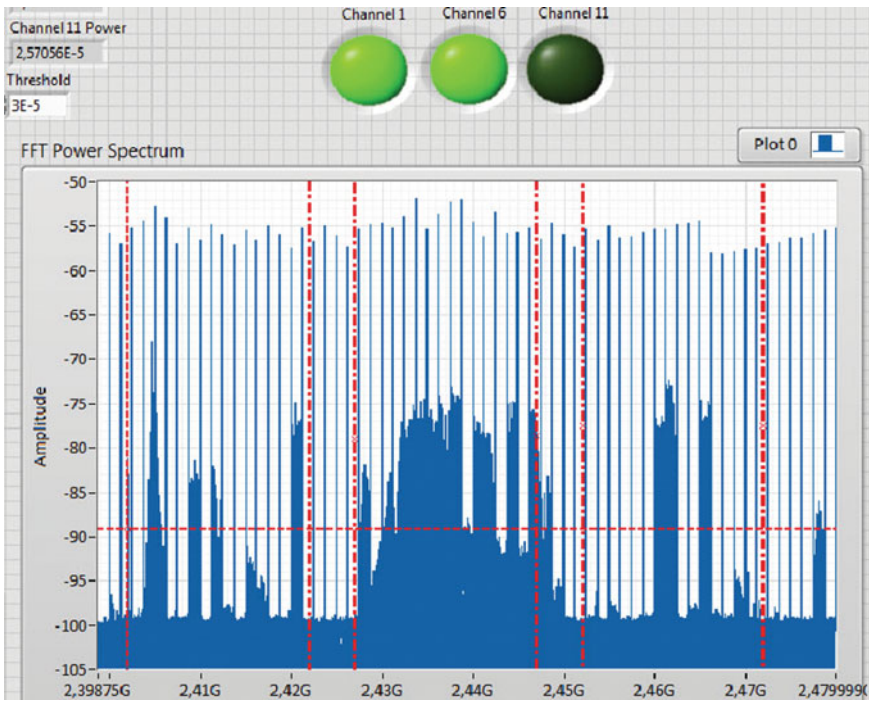| Parameter	Value
First carrier frequency	2.4 GHz
Last carrier frequency	2, 48 GHz
I/Q rate	1 MS/s
Gain	1 dB
Number of carrier frequencies	301

Fig. 14 Exemplary plot obtained after a single experiment

stand the stability of the system, ISM band was observed with Anritsu Spectrum Master MS2711E spectrum analyzer. According to this test, consistent results were obtained, thus we can claim that system has robust performance.

For further consistency, ISM band was observed through long-time experiments. At different moments, system was run for one hour and performance results were observed. According to these experiments, usage density of each channel has decreased compared to previous measurement, and ISM usage has been low in general. As result of these experiments, an exemplary plot is shown in Fig. 15.

As result, real-time and channel based measurements show that energy detection based spectrum monitoring system operates successfully. With these experiments,
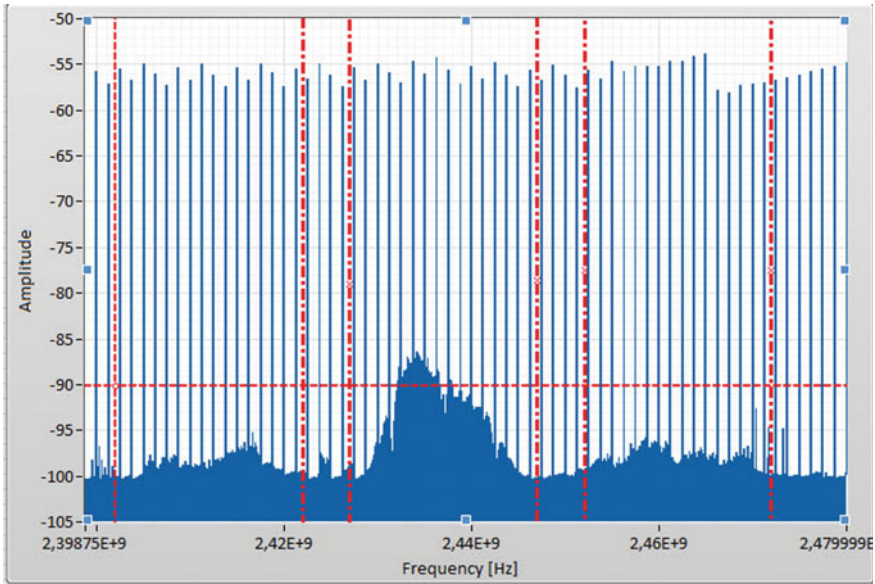
Fig. 15 Average results obtained after one-hour

non-overlapping ISM channels 1, 6 and 11 have been observed and empty channels have been found successfully.

Conclusions

CR technology is very crucial for the future of wireless networks and related algorithms would be very helpful to the spectrum scarcity problem. As supported and explained with various details by this chapter, SDR based studies are very practical and provide comprehensive results. By experimenting with algorithms in real-time with realistic effects of the wireless channel and hardware distortions, more robust CR algorithms can be proposed, which reduce the possibility of unsuccessful deployments that affect PU negatively. In this chapter, by comparing and detailing the important components of spectrum sensing and prominent studies in the literature, we provide crucial insights for the future studies about SDR based implementations of practical CR algorithm. We provide a detailed explanation an practical implementation of the energy based spectrum sensing of channels in the 2.4 GHz ISM band by using SDR tools. As shown, such studies can be implemented in this way without any complexity, and robust algorithms can be proposed by following mentioned structures.

References

1. Ahmed, S., Hossain, M.S., Abdullah, M., Hossain, M.A.: Cooperative spectrum sensing over Rayleigh fading channel in cognitive radio. Int. J. Electron. Comput. Sci. Eng. **1**(4), 2583–2592 (2012)

2. Blad, A., Axell, E., Larsson, E.G.: Spectrum sensing of OFDM signals in the presence of CFO: New algorithms and empirical evaluation using USRP. In: IEEE 13th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), pp. 159–163. doi:10.1109/SPAWC.2012.6292878

3. Bogale, T.E., Vandendorpe, L.: USRP implementation of max-min SNR signal energy based spectrum sensing algorithms for cognitive radio networks. In: IEEE International Conference on Communications (ICC), pp. 1478–1482 (2014). doi:10.1109/ICC.2014.6883530

4. Cabric, D., Mishra, S., Brodersen, R.: Implementation issues in spectrum sensing for cognitive radios. In: Conference Record of the Thirty-Eighth Asilomar Conference on Signals, Systems and Computers., vol. 1, pp. 772–776. doi:10.1109/ACSSC.2004.1399240

5. Chen, Z., Zhang, C., Lin, F., Yu, J., Li, X., Song, Y., Ranganathan, R., Guo, N., Qiu, R.C.: Towards a large-scale cognitive radio network: Testbed, intensive computing, frequency agility and security. In: International Conference on Computing, Networking and Communications (ICNC), pp. 556–562 (2012). doi:10.1109/ICCNC.2012.6167484

6. Denkovski, D., Pavloski, M., Atanasovski, V., Gavrilovska, L.: Parameter settings for 2.4GHz ISM spectrum measurements. In: 3rd International Symposium on Applied Sciences in Biomedical and Communication Technologies (ISABEL), pp. 1–5 (2010). doi:10.1109/ISABEL.2010.5702772

7. Goldsmith, A.: Wireless Communications. Cambridge University Press, Cambridge (2005)

8. Goldsmith, A., Jafar, S.A., Maric, I., Srinivasa, S.: Breaking spectrum gridlock with cognitive radios: an information theoretic perspective. Proc. IEEE **97**(5), 894–914 (2009). doi:10.1109/JPROC.2009.2015717

9. Hickling, R.M.: New technology facilitates true software-defined radio. RF Design Mag. (2005)

10. Instruments, N.: http://www.ni.com (2016)

11. Instruments, N.: Spectrum Monitoring with NI USRP. https://decibel.ni.com/content/docs/DOC-34781 (2016)

12. Jiang, C., Beaulieu, N.C., Zhang, L., Ren, Y., Peng, M., Chen, H.H.: Cognitive radio networks with asynchronous spectrum sensing and access. IEEE Netw. **29**(3), 88–95 (2015). doi:10.1109/MNET.2015.7113231

13. Kishore, R., Ramesha, C.K., Sharma, V., Joshi, R.: Performance evaluation of energy based spectrum sensing in multipath fading channel for cognitive radio system. In: National Conference on Communication, Signal Processing and Networking (NCCSN), pp. 1–6 (2014). doi:10.1109/NCCSN.2014.7001153

14. Kumar, P.V., Sai, M.L.N.: SDR based MIMO link adaptation for cognitive radio application. In: International Conference on Communications and Signal Processing (ICCSP), pp. 577–581 (2014). doi:10.1109/ICCSP.2014.6949907

15. Liang, Y.C., Chen, K.C., Li, G.Y., Mahonen, P.: Cognitive radio networking and communications: an overview. IEEE Trans. Veh. Technol. **60**(7), 3386–3407 (2011). doi:10.1109/TVT.2011.2158673

16. Mitola, J.: The software radio architecture. IEEE Commun. Mag. **33**(5), 26–38 (1995)

17. Najafzadeh, E., George, D., Green, M.P.: Labview-based spectrum occupancy measurements. In: ARSR-SWICOM, The First Conference on Applied Radio Systems Research and Smart Wireless Communications, pp. 1–6 (2012)

18. Nir, V.L., Scheers, B.: Description of a cognitive radio testbed based on USRP platforms and CogWave. In: 9th International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM), pp. 514–519 (2014). doi:10.4108/icst.crowncom.2014.255432

19. Sarijari, M.A., Marwanto, A., Fisal, N., Yusof, S.K.S., Rashid, R.A., Satria, M.H.: Energy detection sensing based on GNU radio and USRP: An analysis study. In: 2009 IEEE 9th Malaysia International Conference on Communications (MICC), pp. 338–342 (2009). doi:10. 1109/MICC.2009.5431525

20. Sarijari, M.A., Rashid, R.A., Fisal, N., Lo, A., Yusof, S., Mahalin, N.: Dynamic spectrum access using cognitive radio utilizing GNU radio and USRP. In: 26th Wireless World Research Forum (WWRF26) (2011)

21. Shahid, H., Yao, Y.D.: Algorithm and experimentation of frequency hopping, band hopping, and transmission band selection using a cognitive radio test bed. In: 23rd Wireless and Optical Communication Conference (WOCC), pp. 1–5 (2014). doi:10.1109/WOCC.2014.6839956

22. Sharma, N., Rawat, D.B., Bista, B.B., Shetty, S.: A testbed using USRP and LabView for dynamic spectrum access in cognitive radio networks. In: 2015 IEEE 29th International Conference on Advanced Information Networking and Applications, pp. 735–740 (2015). doi:10. 1109/AINA.2015.261

23. Shukla, A., Burbidge, E., Usman, I.: Cognitive radios—what are they and why are the military and civil users interested in them. In: EuCAP. The Second European Conference on Antennas and Propagation, pp. 1–10 (2007)

24. Soltani, S., Sagduyu, Y., Shi, Y., Li, J., Feldman, J., Matyjas, J.: Distributed cognitive radio network architecture, SDR implementation and emulation testbed. In: IEEE Military Communications Conference, MILCOM, pp. 438–443 (2015). doi:10.1109/MILCOM.2015.7357482

25. Sun, G., Liu, G., Wang, Y.: SDN architecture for cognitive radio networks. In: 1st International Workshop on Cognitive Cellular Systems (CCS), pp. 1–5 (2014). doi:10.1109/CCS. 2014.6933795

26. Tucker, D.C., Tagliarini, G.A.: Prototyping with GNU radio and the USRP—where to begin. IEEE Southeastcon **2009**, 50–54 (2009). doi:10.1109/SECON.2009.5174048

27. Xu, J., Alam, F.: Adaptive energy detection for cognitive radio: an experimental study. In: 12th International Conference on Computers and Information Technology, ICCIT, pp. 547–551 (2009). doi:10.1109/ICCIT.2009.5407298

28. Yang, J.J., Huang, M., Yu, J., Li, L., Li, L.: USRP: a flexible platform for spectrum monitoring. Appl. Mech. Mater. Trans Tech Publ **610**, 233–240 (2014)

29. Yoshimura, R.S., Mathilde, F.S., Dantas, J.P., de S Jr VA, da Cruz Jr J.H., Bazzo, J.J., Melgarejo, D.C.: A USRP based scheme for cooperative sensing networks. In: Anais do 4 Workshop de Redes de Acesso em Banda Larga—WRA 2014, Brazil, pp. 1–5 (2014)

30. Yucek, T., Arslan, H.: A survey of spectrum sensing algorithms for cognitive radio applications. IEEE Commun. Surv. Tutor. **11**(1), 116–130 (2009). doi:10.1109/SURV.2009.090109

31. Zhang, Z., Zhang, W., Zeadally, S., Wang, Y., Liu, Y.: Cognitive radio spectrum sensing framework based on multi-agent architecture for 5G networks. IEEE Wirel. Commun. **22**(6), 34–39 (2015). doi:10.1109/MWC.2015.7368822

Visible Light Based Throughput Downlink Connectivity for the Cognitive Radio Networks

Bo Xing

Introduction

The Internet of things (IoT) refers to "the networking of physical objects through the use of embedded sensors, actuators, and other devices that can collect or transmit information about the objects" [7]. Gartner reported the population of such smart things could soon reach 21 billion, while Juniper research group predicted that by 2020, there will be 38.5 billion connected devices [48]. In fact, the numbers don't really matter, since when it comes to the IoT era, everything will be connected. Yet, the truth of the matter is how actually are all these things going to work together?

Given the state of IoT today, to fulfil the dream of connecting everything, it might be a bumpy road if certain things aren't settled down. One obstacle is the lack of sufficient spectrum since the applications of IoT will deploy a huge mass of new wireless services in the current available frequency spectrum. Research by CERP (Cluster of European Research Projects) has found that when considering the spectrum of possibilities for IoT, little can be said by now since the technology is still being developed [47]. What's more, the report also pointed out that one possibility is "the device used in the future IoT will employ wireless communications using the frequency spectrum beyond the radio frequency range" [47] to improve current cognitive radio (CR) networks' throughput. This implies that although the current radio spectrum management satisfies IoT's most requirements, some applications will require low-power, low-data rate connectivity but high-speed communications, thus create a critical challenge of limited radio bandwidth.

B. Xing (✉)

Faculty of Engineering and the Built Environment, Computational Intelligence,
Robotics, and Cybernetics for Leveraging E-Future (CIRCLE),
University of Johannesburg, Johannesburg, South Africa
e-mail: bxing2009@gmail.com

To address the above issues, this chapter proposes a visible light (VL) based throughput downlink connectivity scheme for the CR networks (i.e., $CR_{downlink}^{VL}$). The kernel of the proposed architecture is composed of interlinked heterogeneous devices, compartmentalized in various wireless autonomous units (WAUs), built on the CR networks, and downlink connectivity enhanced by visible light communication (VLC) protocol. With this in mind, the remainder of this chapter is organized as follows: first, the relevant background and literature review on the radio spectrum management are described in section "Radio Spectrum Management: An Overview" which is followed by a detailed VLC system in sections "Downlink Connectivity: Visible Light Communication (VLC)" and "A Detailed Overview of VLC System", respectively; second, sections "VLC Protocol—Physical Layer" and "VLC Protocol—Link Layer" elaborate an overview of VLC protocols, i.e., physical layer and link layer; then, the draft progress of applying the proposed method to smart home is summarized in section "VL Based Throughput Downlink Connectivity for the CR Networks: An Exemplary Smart Home System Architecture"; third, the challenges of VLC system are discussed in section "VLC Challenges"; finally, section "Conclusion and Future Work" draws the conclusion and future works of this chapter.

Radio Spectrum Management: An Overview

In general, radio spectrum refers a medium over which all wireless communications take place. Its management requires knowledge of careful planning to maximise its value for all users. In this section, we present a general review of radio spectrum management.

Inception

The rapid development in the radio communication systems can be visualized by a simple comparison between the first broadcast communications and the 300 Mbps already considered the long term evolution (LTE). Traditionally, a fixed spectrum assignment policy called "command-and-control", i.e., assigns a fixed frequency block for a specific license holder, has been widely adopted. Examples include widely used services, such as radio/TV broadcasting, mobile communication, and the daily operation of the emergency services and defence forces. However, according to a survey by the Federal Communications Commission (FCC) [1], under the current regulatory environment, a large portion of the assigned spectrum is used inefficient, ranges from 15 to 85 % with a high variance in time. In addition, as interest in wireless communication has been increasing exponentially over the

last decade, particularly, as the IoT has become the next stage of development for the Internet, a spectrum shortage may be faced in the near future. So the static spectrum policies are unable to manage the spectrum efficiently any longer.

Evolution

In view of this, the FCC are contemplating making the licensed spectrum available to unlicensed users. This leads to a new paradigm pertaining to wireless spectrum allocation, known as cognitive radio (CR), which provides opportunistic access of licensed spectrum to the unlicensed users [27]. The history of the CR dates back to 1999 when Joseph Mitola and Gerald Maguire in [32, 34] presented the analysis of more flexible approach to wireless communication. The idea behind that is the software defined radio (SDR), in which a wireless communication system can be reconfigured by software reprogramming, so that different frequencies can be operated with different protocols [33]. In other words, CR is an intelligent wireless communication system that is awareness of the surrounding environment (i.e., outside world) and have the capability of intelligence (e.g., learning for adaptive tuning of system parameters) [20, 25].

To work as an intelligent wireless communication system, CR had to be able to detect the idle channels of unoccupied licensed spectrum at a particular time and specific geographic location, termed as vacant spectrum bands or "white spaces", and to introduce these spectrum opportunities from licensed primary users to the unlicensed users, called secondary users. More specifically, the CR technology will enable the users to spectrum sensing (e.g., detecting unused spectrum), spectrum decision/management (e.g., capturing the best available spectrum), spectrum sharing (e.g., coordinate access with other users), and spectrum mobility (e.g., vacate the channel when a licensed user is detected) [2].

Downlink Connectivity: Visible Light Communication (VLC)

Within the electromagnetic spectrum, the radio frequency (RF) band has been the most widely used band for wireless communication purposes. However, as IoT pretty much dominated our daily living environment, the throughput downlink connectivity need to be further guaranteed for the CR networks. In this section, we will first list a few obstacles of CR management and then introduce a complementary communication medium, called VLC, as an option to enhance the throughput downlink connectivity for the CR networks.

Motivation of Introducing Additional Technology

Though RF-mediated communication enjoys a tremendous success and the powerfulness of the resultant CR techniques is also widely recognized, the developments of CR are related to many research challenges.

Potential Interference of CR

Technically speaking, one of the most important objectives of CR is to improve spectrum utilization by avoiding interference with licensed or unlicensed users—known in the jargon as spectrum sensing. Several tutorial surveys (e.g., [3, 56]) are convinced that this function is an important backbone of avoiding interference with the primary users. Yet, in practice, to improve the sensing performance is a complicate task and depends on several factors, such as multipath fading, shadowing, and the hidden primary user [46]. Several papers can be found in the literature, dealing with spectrum sensing techniques, such as matched filter [56], energy detection [42], cyclostationary feature detection [16], and covariance-based detection [58]. However, up to now there is no feasible method of detecting all dimensions of radio spectrum apace in CR network due to its requirements of each field of application are different.

In addition, some places in which radio spectrum often falls short to avoid interference. For example, seeking ways to improve the hospital efficiency and in the meantime cutting the hospital environment induced infection rates are always on top of priority list of modern medical community. One promising means is to render doctors a better wireless communication infrastructure which allows them to access and update in-patient's data in a timely manner, say, bedside information retrieval and updating via tablet computers. In another scene, from the critical in-house life support system to the wide spread of wearable devices, conventional wireless communication via the medium of RF cannot always ensure the quality of the transmitted information, say, unaffordable jammed communication caused by surrounding devices' interference. All that is needed is a complemented way of wireless communication which is not susceptible to unclear signal reception affected by other electronics devices.

Security of CR

To fulfil the CR's main objective (i.e., allow unlicensed users to reuse idle frequency which belongs to the licensed user), the CR network had to be able to allow various unknown wireless devices to opportunistically access the spectrum. Consequently, CR networks are prone to serious security threats and are vulnerable to malicious attack [38]. To date, several issues have been pointed out. For example, the authors of [45] made the effort to solve the problem of primary user emulation

attack. In addition, the attack about the overlapping with secondary users is conducted in [30]. More details please refer to [10, 11], in which different types of attacks (e.g., biased utility attack, asynchronous sensing attack, and false feedback attack) and security requirements (e.g., access control, authentication, and availability) are described. Although a lot of topics have been studied, a secure framework for the CR networks is scarce [38].

To solve such attacks, on the one hand, we need to find alternatives to achieve network scalability and stability, such as clustering [54] or routing and channel selection [43]. On the other hand, new technologies that target the security issue are emerging from players such as VLC. For example, to cater for the connectivity demands of the intelligent transport system, VLC can provide a secure link between nodes since it has the remarkable property of being resilient to jamming and Denial of service attacks. In addition, due to its high frequency, visible light cannot penetrate through most objects and walls, thus benefit from an inbuilt secure transmission of information in physical layer [39].

Insufficiency of CR

As the radio spectrum is a finite resource, CR has been proposed in which the dynamic use and reuse of the underutilized spectrum was achieved. However, in the meantime, another limitation in the CR is emerged, i.e., the trade-off between the cost that to achieve throughput of the CR and the level of protection for the primary user [57]. What's more, with the explosive growth of IoT, pressure is mounting for better radio wireless communication technology. Under this circumstance, the radio spectrum used today range from 3 kHz to 300 GHz is outstripping the supply [49]. In other words, the radio spectrum band could not have been used to support the idealized revolutionary IoT evolution.

Given these limitations, a new path is needed to ameliorate the bandwidth saturation problem. For example, under indoor environment, light emitting diodes (LEDs) are commonly used as lighting components, thus becoming appealing for using an indoor VLC system (i.e., with a wavelength between 430THz~790THz) to achieve data transmit, as suggested in [13, 15, 36].

Summary

Overall, the aforementioned disadvantages are thus motivating us to focus on downlink connectivity and to introduce VL into the realm of CR communication. Note that our aim is not to supplant the current RF communication system but to make it more efficient, in particular downlink connectivity. As with RF, light is electromagnetic radiation, but the difference lies in frequency. And, the problem VLC is trying to resolve is not the omnipotence of the radio spectrum communication but to relief the congested downlink connection scenario that data is

transferred via radio spectrum. In other words, the VLC is a compensator instead of a disrupter. For the rest of this section, a brief background of VLC technique will be discussed.

Emergence of VLC

The VLC refers to the communication technology which utilizes the VL source as a signal transmitter. It began life in the mind of Alexander Graham Bell, the brilliant creator of Photophone—the first device that transmitted a voice signal on a beam of sunlight [8]. Since then, dozens of researchers hoped to improve Bell's design by using different types of lighting sources, such as mercury arc lamps [18] and fluorescent lights [23]. In 1999, the concept of using LEDs to provide a dual role, i.e., as illumination and communication medium simultaneously, was proposed [37]. Nowadays, as LEDs are now becoming the lighting source everywhere, applications of using VLC for high speed data transmission are also boomed, e.g., road to vehicular communications [5], in-flight entertainment [40], medical community [12], and in-house GPS tracking system [35].

Building Blocks of VLC

To understand the power of VLC systems, and the things they can do, it is important to introduce three things that add hugely to its success. One is the physical hardware (e.g., LEDs, regular photodiodes, or image sensors). For example, LEDs are semiconductor devices that have the capability of rapid switching with other appropriate electronics. Therefore, the visible light emitted by LEDs can be modulated and encoded with audio/digital information for transmitting, i.e., a kind of data transmission using binary codes. In order to have widespread adoption, the second is the standards. To promote and standardize VLC technology, several initiatives has been broadcasted, such as Visible Light Communication Consortium (VLCC) [52], the HOME Gigabit Access (HOMEGA) project [21], and IEEE 802.15.7 Visible Light Communication Task Group [22]. And, the last one is the ubiquitous infrastructures support. That is, the VLC facilitates the reuse of existing lighting infrastructure (e.g., traffic signals) for the purpose of communication, i.e., with relative lesser efforts and at a lower cost.

Advantages of VLC

In comparison with the incumbent means of wireless communications, using visible light as a communication mediator enjoys the following advantages.

- **Affluent spectrum resources**: Unlike RF radio systems that is a scarce resource and licensed, VLC falls under the category of free-space optical communication, i.e., ten thousand times wider than the RF spectrum. Interested readers please refer to [31] for more detailed description and schematic representation in this regard.
- **Immunity to RF interference and RF band congestion**: Compare with RF, one appealing property of VLC is that the light waves do not create electromagnetic interference to sensitive electronic systems. In other words, it causes no interference to RF-based devices. With this intrinsic characteristic, it dominates, even in RF prohibited areas like hospitals and airplanes.
- **Low deployment cost**: The main technological support that made VLC possible is the low-cost off-the-shelf LEDs. What's more, the infrastructure of VLC is available all around us. All that needs to be done is to combine LEDs with incumbent lighting system to let communication co-exist with the illumination. For example, in vehicular communication, the LED lights already exist in automobiles, while RF-based systems incur additional cost of deploying the equipment.
- **High security**: Due to the nature of the visible light, visible light cannot penetrate through solid objects and walls and thus provides an inherent wireless communication security. For example, since light is visible, it is easy to determine who can listen to (or receive) a message.
- **Energy-efficiency and eco-friendly**: Compared with radio waves which consume a lot of energy, VLC is highly energy efficient since it uses the existing eco-friendly LED lighting system. According to the United States Department of Energy, the average luminous efficacy of best LEDs is around 200 lumens/watt by the year 2020, whereas the incumbent fluorescent bulbs provide 60 lumens/watt [51]. In addition, the lifespan of LEDs is much higher than compact fluorescent. Apart from those advantages, LEDs also have other environmental benefits, such as reduced usage of harmful materials in design and lower heat generation.

A Detailed Overview of VLC System

Transmitter and receiver components are key parts of any communication systems and VLC makes no exceptions. This section offers reader a detailed overview of these elements' composition and how they dedicate themselves to keeping the VLC system running.

Key Components

VLC Transmitter

In VLC communication system, LED luminaire often serve as the transmitter. Since it can be pulsed at very high speeds, it is undetectable to the human eyes. As a complete lighting unit, an LED luminaire comprises an LED lamp, ballast, housing, and other peripherals [39]. Typically, one or more LEDs can be contained in an LED lamp, which is also equipped with a driver circuit for adjusting the current amount passing through the LEDs to realize its controllable brightness. When an LED luminaire is involved for communication, one can modify its driver circuits so that the emitted light can be utilized for modulating the data. For instance, in a simplified modulation scheme, two distinct light intensity levels can be selected for representing and transmitting the data bits "zero" and "one".

In a typical VLC system, a crucial and institutive design prerequisite is the main purpose of the LED luminaries (i.e., illumination) should not be interfered the associated communication utilization. Though how an LED luminaire is designed (say, widely adopted white light and its different generating mechanism) determines the performance of a VLC communication system, the detailed discussion relevant to solid-state light is out of the scope of this chapter. Please refer to [44] for a thorough explanation.

VLC Receiver

As soon as an LED luminaire signalling a piece of modulated information, two kinds of VLC receivers (i.e., photodetector and imaging sensor) can be employed to get that message. The first type of VLC receiver, namely, photodetector or photodiode, is a p-n junction device which can convert the received light into photocurrent; while the latter type, i.e., imaging or camera sensor, could trigger a great interest since most camera sensor embedded mobile devices have the potentialities of being used as ubiquitous VLC receivers. Nevertheless, due to some of imaging sensor's intrinsic limitations (e.g., low frames-per-second capture rate), the throughput of the second type of receiver is rather low in comparison with its counterpart stand-alone photodetector's performance.

VLC Communication Modes

In general, two types of communication mode can be observed in visible light-mediated communication, namely, I2D (infrastructure-to-device) communication and D2D (device-to-device) communication. In a typical indoor environment, the room is flooded with soft illumination provided by LED luminaires from

which data can be sent out to a raft of equipment inhabitant in the room. In such case, if uplink information transmitting from the devices is not possible (e.g., eminent distraction effects imposed by the end-user devices' on-board LEDs), communication channel can be switched back to RF medium for a more effective uplink message conveying. In another similar example, the LEDs used in the existing infrastructure such as street lights and stoplights can also be requisitioned (without making too much modifications) as a gateway to ubiquitous Internet access for both shuttled vehicles and strolling pedestrians. In terms of D2D, the scenes where VLC finds itself a good fit are also legion: smartphone-to-smartphone near-field communication, vehicle-to-vehicle and vehicle-to-road ad-hoc communication network, to name but a few.

VLC Protocol—Physical Layer

Visible Light Propagation Features and Channel Model

LED's Transmitted Power: Luminous Flux

The luminous flux of the transmitter LED (F_T, measured in lumens) can be obtained via Eq. (1) [39] which is often referred to spectral integral method, that is deriving F_T based on the relationship between human eye's luminosity function and an LED's spectral power distribution.

$$\text{Spectral Integral: } F_T = 683 \int_{380}^{750} S_T(\lambda)V(\lambda)d\lambda \qquad (1)$$

where 683 is a constant representing the maximum luminous efficiency (lumens/W), the detailed calculations are out of the scope of this chapter, interested readers please refer to [39] for more details; the upper and lower bounds of integral (i.e., 750 and 380) denote the range of which human eye can see colours; and variables $S_T(\lambda)$ and $V(\lambda)$ stand for luminosity function and the spectral power distribution, respectively.

In addition to the said spectral integral approach, one can also use spatial integral technique, see Eq. (2) [39], built on the spatial emission properties of an LED.

$$\text{Spatial Integral: } F_T = I_0 \int_{0}^{\theta_{\max}} 2\pi \cdot g_t(\theta) \cdot \sin\theta d\theta \qquad (2)$$

where I_0 is called axial intensity and often defined as the luminous intensity (in candelas) at zero degree solid angle. In practice, the value of I_0 is typically

normalized as 987 cd (please also refer to [39] for a colourful illustration); $g_t(\theta)$ denotes the normalized spatial luminous intensity distribution; and when the light intensity reduces to half of the axial intensity, the resulting half beam angle can be represented by θ_{max} (please see [39] for diagrammatic illustration). In terms of this chapter's focal Lambertian source (i.e., LED), the half beam angle is computable via Eq. (3) [39]:

$$\Omega_{max} = 2\pi \cdot (1 - \cos\theta_{max}) \tag{3}$$

where Ω_{max} stands for the entire beam angle.

Luminous Path Loss and Received Optical Power

- **Luminous pass loss**: As the aforementioned equations illustrate, we can further pursue the value of path loss (L_L) based on the obtained luminous flux quantity. In general, L_L can be obtained via Eq. (4) [39]:

$$\text{Luminous Pass Loss (without Lambertial beam distrituion):}$$
$$L_L = \frac{F_R}{F_T} = \frac{g_t(\beta) \cdot A_r \cdot \cos\alpha}{D^2 \cdot \int_0^{\theta_{max}} 2\pi \cdot g_t(\theta) \cdot \sin\theta d\theta} \tag{4}$$

where F_R denotes the receiver flux, computable via Eq. (5) below; D stands for the distance between the receiver and the transmitter; r represents the radius of the receiver aperture; α and β indicate two angels, i.e., the angle between the receiver normal and transmitter-receiver line, and the transmitter viewing angle, respectively; and A_r shows the area of the receiver (please consult [39] about the diagram).

Based on this, we can then calculate F_R (i.e., the receiver flux) through Eq. (5) [39]:

$$F_R = I_0 \cdot g_t(\beta) \cdot \Omega_r \tag{5}$$

where Ω_r describes the receiver solid angle as observed from the transmitter viewpoint, and it can be calculated through Eq. (6) [39]:

$$A_r \cdot \cos\alpha = D^2 \cdot \Omega_r \tag{6}$$

Since majority of LED sources are characterized by Lambertial beam distribution, that is, the spatial luminous intensity distribution can be represented by a cosine function as shown in Eq. (7) [39]:

$$g_t(\theta) = \cos^m(\theta) \tag{7}$$

where the order of Lambertial emission is denoted by m.

With the aid of a set of substitutions, we can further get another luminous pass loss equation (under the Lambertial beam distribution situation) as shown in Eq. (8) [39]:

$$\text{Luminous Pass Loss (with Lambertial beam distriuion):} \\ L_L = \frac{(m+1)A_r}{2\pi D^2} \cdot \ \cos\alpha \cdot (\cos\beta)^m \tag{8}$$

- **Received optical power**: Once the luminous pass loss equation is obtained, the received optical power can also be computed via Eq. (9) [39]:

$$\text{Received Optical Power: } P_{R_O} = \int\limits_{\lambda_{rL}}^{\lambda_{rH}} S_R(\lambda) \cdot R_f(\lambda) d\lambda \tag{9}$$

where the spectral response of the optical filter (typically embedded in the receiving photodector) is denoted by $R_f(\lambda)$; $S_R(\lambda)$ equals to $L_L \cdot S_T(\lambda)$; and the optical filter's lower and upper bound of wavelength cut-off values are represented by λ_{rL} and λ_{rH}, respectively.

As we can see that P_{R_O} mainly depends on three variables, that is, D (distance between the transmitter and the receiver), α (incident angle), and β (irradiation angle). The value of these variables are largely influenced by the receiver's activity and position.

Multipath Propagation with Reflected Paths

In real-world environment, the signals from various LEDs can be simultaneously detected by the receiving photodetector. Under this circumstance, a receiver's P_{R_O} can be computed via Eq. (10) [39]:

$$P_R(total) = \sum_{i=0}^{N} P_R(i) \tag{10}$$

where the total amount of LEDs is denoted by N; and the received optical power from the ith LED (has a line-of-sight (LOS) link and within the receiver's field-of-view (FOV)) is represented by $P_R(i)$.

Table 1 A comparison of different types of noise

Noise type	Typical sources	Reducibility
Ambient light noise	• Solar radiation • Different lamps	High
Shot noise	• Signal • Ambient light	Medium
Thermal noise	• Electrical circuitry	Low

Receiver Noise and Signal-to-Noise Ratio

Typically, there are three main types of noise existing in indoor environment which can interfere visible light optical link. Table 1 shows a brief comparison among these noises.

Since much of the ambient light noise is reducible, the signal-to-noise ratio (SNR) can be acquired via Eq. (11) [39]:

$$\text{Signal} - \text{to} - \text{Noise Ratio: SNR} = \frac{(P_{R_E})^2}{(\sigma_{\text{shot}})^2 + (\sigma_{\text{thermal}})^2} \tag{11}$$

where the standard deviations of shot and thermal noises are denoted by σ_{shot} and $\sigma_{thermal}$, respectively. The calculation of the corresponding variance is out of the scope of this chapter, and interested readers please refer to [39] for a thorough description.

Shadowing

In the indoor circumstances, it is not uncommon that various objects and even humans themselves can sometimes block the VLC link. Considering the exhibited dramatically different propagation features between the newcomer (i.e., VL) and the incumbent RF, one should take the VL shadowing issue very seriously. Unfortunately, the existent studies in this respect are rare and interested readers can use a preliminary attempt (found in [26]) as a guidance.

Modulation Techniques

According to [39, 50], in comparison with RF communication, the most notable feature of VLC communication lies in that encoding data in phase or amplitude of the light signal is impossible. This difference implies that the information need to be encoded by using IM/DD (intensity modulated/direct detection) modulation techniques instead of conventional phase and amplitude modulation approaches. Based on this characteristics, a suitable modulation scheme for VLC should be able to

meet the following two requirements simultaneously, i.e., human being perceivable light's prerequisites, and in the meantime, higher data transfer rate. In terms of the pre-conditions of human's perceivable light, there are two properties often considered in the literature [39].

- **Property 1—Dimming**: With the rapid development of LED driver circuit technology, the illuminance level of an LED can be adjusted to various levels for meeting the application-oriented requirements and also energy-saving purpose. Meanwhile, from biological perspective, in order to be adaptable to varying illumination, human eyes have the capability of altering the diameter of its pupil to regulate the amount of light passing to the retina. According to [41], a non-linear relationship between the measured and the perceived light can be expressed in the form of Eq. (12) below [39]:

$$\text{Perceived light } (\%) = 10 \cdot \sqrt{\text{Measured Light } (\%)} \tag{12}$$

When introducing VLC as a complement to existing CR system, one should bear this property in mind and should not let the dimming level affect the desired communication. Ideally, in VLC scenario, the data ought to be modulated to fit arbitrary dimming level.

- **Property 2—Flickering**: Additionally, any human perceivable light brightness fluctuations should be eliminated, whichever VLC modulation plan is involved. It has been reported in the literature [9], flickering can trigger some serious harmful physiological alteration in humans. Under this circumstance, a recommended light intensity adjustment speed (>200 Hz) can be found in IEEE 802.15.7 Standard [22] or the purpose of rid people of any detrimental effects.

At present, there are four widely-discussed modulation schemes used in VLC, namely, OOK (i.e., on-off keying) modulation, pulse modulation, OFDM (orthogonal frequency division modulation), and CSK (colour shift keying) modulation. The detailed discussion of each of them can be found in [39]. Though each modulation scheme shares different advantages and disadvantages, a quick prediction can be made is that, with the ever-increasing demand of higher data transfer rates, OFDM and CSK are more likely to be the suitable candidates for VLC-based communications.

Multiple Input Multiple Output (MIMO)

In conventional RF-based communications, MIMO systems are widely employed to achieve higher data transfer rates. In the similar way, to achieve higher spectral efficiency in VLC communications, one can deploy multiple LEDs, though it is often hard to convert the VLC's MIMO system into reality. One barrier is, unlike

the traditional RF MIMO systems (which enjoys a wide spatial diversity contributed throughput gains), VLC MIMO systems lack a variety of gains because of a low paths diversity (e.g., indoor environment). Other challenges also include the design of receiver (i.e., photodiode and image sensor) for VLC MIMO systems.

Currently, there are three classes of MIMO techniques proposed for VLC communications, namely, RC (repetition coding), SMP (spatial multiplexing), and SM (spatial modulation). A detailed advantages and disadvantages comparison of all these three VLC MIMO approaches is out of the scope of this chapter and [15] can be used as a reference for further reading.

VLC Protocol—Link Layer

When there are multiple transmitter LEDs existing in the surroundings and a raft of receiver devices hinging on them, it is crucial to manage VLC link layer services such as medium access control (MAC), device association, and device mobility.

Medium Access Control (MAC)

In order to be supported by the MAC protocols, link layer topologies need to be identified for different VLC application scenarios. IEEE 802.15.8 Standard [22] recommends three classes of such topologies for VLC as illustrated in Fig. 3 below.

- **P2P (peer-to-peer)**: In this class, a master (or coordinator) is selected for the link between two pieces of equipment. Since the user has established an uplink to the coordinator device, both machines can talk to each other successfully. Near-field high speed communication is a representative application example found in this category.
- **Star**: When it comes to more than one end user devices connecting with a master device, the star topology will be put into practice. However, given that many bi-directional links co-existing in the same collision area, the MAC design is a demanding tasking.
- **Broadcasting**: Though broadcasting topology shares many similarities with the said star topology, one factor does differentiate them from each other. Since no established uplink involved, the broadcasting topology can make the specific MAC design easy.

In the literature, three main kinds of MAC plans have been suggested for VLC communication, namely, CSMA (carrier sense multiple access), OFDMA (orthogonal frequency division multiple access), and CDMA (code division multiple access). One can refer to [39] for more information. In addition to all this, more advanced approaches like multi-user MIMO [55] are still in their infancy and need to be further devised and nurtured for VLC communication.

Cell Design and Cell Coordination

- **Cell Design**: Unlike the other types of communication networks, managing access to multiple machines in VLC needs to meet specific requirements. Given how illumination is provisioned, the appropriate cell size also varies a lot, ranging from a radius of 10 m (more suitable for multi-devices in public places) to no more than 5 m (mainly for personal usage in private environment).
- **Cell Coordination**: The purpose of facilitating cell coordination is to lower inter-cell interference degree. Some suggestions can be found in IEEE 802.15.7 Standard [22]. One of them is, in an indoor setting, to introduce a central controller entity (connected by all LED transmitters) which can enable the cell operations and device mobility functioning together efficiently. Nevertheless, the practicability of such scheme is greatly hindered by its inherent high deployment cost. More studies are therefore needed to be done towards this direction.

VL Based Throughput Downlink Connectivity for the CR Networks: An Exemplary Smart Home System Architecture

The communications for IoT are noted for its completely autonomous data production, sharing, processing, and actuation in various forms such as peer-to-peer (P2P), peer-to-machine (P2D), and device-to-device (D2D). In scenarios like D2D or P2D, the human intervention is either absent or limited and thus a set of self-x abilities (such as self-organizing, self-configuring, self-managing, self-curing, self-controlling, self-teaching, and self-addressing) are a great necessity. In other words, the landscape of IoT starts with end-users angle, ends with a desired functional output, and takes an interconnected P2P, P2D, and D2D network interlinked with the omnipotent Internet in the middle.

At present, a notable feature of IoT is the incessant improved functionalities (e.g., computing power, accuracy, etc.) accompanied by constant enhanced designs (e.g., cost-effectiveness, reduced energy consumption, miniaturization, etc.). In many IoT communication practice, devices are required to challenge a number of unique conundrums, so that a large scale and multi-dimensional deployment can become a possible. One of the key puzzles include seamless connectivity which greatly influence the total quality-of-service (QoS). Nevertheless, with the amount of "things" growing quickly (ranging from the unbelievable 21 billion by 2020 [48] to the unimaginable myriads), the difficulty of getting all this omnipresent "smart dust" hooked up with Internet so that the useful information can be broadly shared is becoming an evident barrier.

In the literature, there are many methods and techniques that focus on how to improve such connectivity for the general objects which in turn enables them to

share what they see, smell, hear, and feel the surrounding physical world. Among them, CR is arguably the most representative approach that could have profound implications in the forthcoming IoT era. By equipping IoT with a newly added cognition dimension, the level of intelligence and adaptability of the resultant cognitive IoT is largely lifted. Despite the efforts, the current cognizance brought by CR is still limited since only RF is considered as a downlink connection channel. On the one hand the spectrum covered is indeed wide, but the crowded nature is also obvious; whereas on the other hand, the idle visible light spectrum and the associated VLC are advances in wireless communication technology that have yet to lead a dominant trend, but soon might.

This observation stimulates us to present our visions on $CR^{VL}_{downlink}$, and propose an application architecture in the smart home. The most prominent characteristic of a smart home environment lies in that various intelligent objects (e.g., sensors, actuators, gadgetry, etc.) spend most of their time in a constant/periodic working condition (e.g., condition monitoring, decisive event sensing, etc.). Based on this, we first identify the system requirements from various angles. Then, a blueprint of a smart home communication architecture bolstered by VLC and CR techniques is unfolded.

System Requirements

A couple of requirements have to be met for the purpose of implementing a hybrid of VLC and CR wireless communication in smart home environment. We summarize them as follows:

- **Coping with noise and multiple propagation**: Home environment is increasingly characterized by these phenomena and the dedicated communication system must be able to deal with it.
- **Eliminating non-determinism in wireless communication**: the sources of non-determinism should be removed as many as possible by the system.
- **Multiple wireless units co-existence**: A diversity of wireless connection points are demanded to offer necessary connection model, in particular in smart home environment.
- **Backward and forward compatibility**: The great variety of interconnected and Internet-linked objects, differs not only in their very many unlike uses, but also in their operating complication. So the irony of ironies is although innovation, born of fast evolving, is making once dull stuff ever smarter, nobody could actually forecast where the technology will leads. As a result trying to ensure backward and forward compatibility comprehensive enough to embrace all things (e.g., proprietary products) is no doubt a daunting task.
- **Flaw tolerable**: Performance degradation caused by any flaw in a system of systems should be prevented at large. In case of communication errors are detected, the system should perform a switch-over in setting up channels.

- **Service dependable**: The system should guarantee to output a series of reliable communication services.
- **Resources assignment**: It is crucial to understand that there are many challenges facing the smart home proposition (e.g., bandwidth, reaction time, energy consumption, etc.). Most of all the need for strong and active networks at both ends of a communication corridor where the coordinators and end nodes are located respectively. That is why not every scene will likely witness the appearance of wireless communications using VLC techniques challenging the incumbent CR.
- **QoS warrant**: The system have to offer the needed degree of QoS for the provision of services. In other words, to achieve an optimized smart home as a whole, differentiated QoS categories ought to be created so that the best alternatives can be picked and selected.

The Proposed Architecture

In order to meet the aforementioned requisites, the proposed architecture consists of a couple of independent wireless automation units (WAUs), controlled by access point (AP). The communication of the targeted smart home environment, facilitated by CR and VLC, comprises a group of WAUs and an army of switchboard devices (SDs) therein dealing with the specific transmission medium dispatching tasks. Figure below illustrates the proposed $CR_{downlink}^{VL}$ smart home network architecture. The network components are separated into two main parts, namely, the primary and the secondary wireless network modules.

Primary Wireless Network

In the proposed architecture, the existing licensed spectra habitually comprise a primary wireless network where subscribed users gain entry permissions to licensed frequencies. Typical cases in a smart home environment includes mobile terminals connecting to the cellular networks and televisions receiving signals from the affiliated broadcasting networks. Accordingly, a P-BS (i.e., primary based station) serves as the principal component in the primary wireless network.

Secondary Wireless Network

The core spirit of this secondary network module is twofold: First, conventional opportunistic way of awaiting the spectrum holes is still allowed under CR mechanism. To cope with the potential downlink connectivity issue, other spectra

(e.g., visible light) is further introduced as a complement. The APs (i.e., access point) are used as information gateway to the external networks, say, the Internet. In accordance with the need, an S-BS (i.e., secondary base station) is established to take care of communications within its coverage.

The Beneficial Applications

Many applications greatly profit from the added functionalities and optimized performance arising out of the cross between VLC and CR communications. The remaining of this section quotes an in-home instance.

Home Multimedia Share and Share Alike

In a modern home environment, there is always no shortage of mobile phones, personal computers (desktop, laptop, or tablet), televisions, sound systems, and a host of other untold numbers of electronic appliances. With the wide spread of microcomputer embedded intelligent instruments, ordinary people's residences are quickly marching into the period of smart home. In such home communication networks, $CR_{downlink}^{VL}$ is a good application for distributing and sharing multimedia on site, in which the allotment of the radio resources are often a demanding task.

VLC Challenges

The advent of VLC opens up an entirely new quadrant of possibilities. But for all that the VLC is cheap, secure and wide free bandwidth available, sceptics argue that several challenges remain. Some of these issues are addressed in this section, i.e., limited mobility and coverage, underdeveloped hardware, and the uplink channels.

Limited Mobility and Coverage

The most obvious fact is that the VLC only works in places where there are lighting sources, i.e., the communication link relies on the existence of a LOS path between them. This has a drawback that the light can be easily obstructed and cause the signal to be interrupted. Several approaches have been considered. For example, the LOS links can employ narrow FOV transceivers while non-LOS links use wide FOV transceivers. However, both types suffer from interference from ambient light sources [24]. Another alternative is focusing on the mechanism of link switching

and link recovery for movement of the receiver. Yet, this method will increase the collision probability of slot allocation at the initial access [29]. Consequently, at present there is no concrete conclusions as to which solution is the best, and further work is required to develop potential techniques and compare alternatives.

Undeveloped Hardware

A notable advantage of VLC system is that illumination and communication can work simultaneously. However, as the authors of [17] pointed out that the primary purpose of the LED is only for the illumination, which implies that the potential efficacy of VLC technologies is overlooked at this stage. For example, in some recent surveys, the authors of [17] proposed that it needs to consider the dimming control functions of LED in order to provide the necessary power required for VLC to transmit information when the lights are "off", while [28] indicated that the development of new LED materials and devices with better characteristics are needed. In addition, interest in the improvement on light detectors (e.g., photodiode [4, 6] and imaging sensor [19, 53]) grows as well.

Uplink Channels

Apart from above mentioned challenges, how to provide a high speed VLC uplink is difficult as well. Theoretically, VLC can be used for data transmission in either downward or upward direction. Of the two tasks, the former is easier since the VLC channels are naturally downward channels (i.e., broadcast). Yet an effective uplink communication method is more important for the IoT environment due to the sensors need to send their data out [14]. Nowadays, there is a handful of challenges for establishing an effective uplink. For example, using LEDs on end-users devices can cause noticeable disturbance to users [3939]. To cope with those problems, several approaches have been proposed, such as infra-red (IR), RF, and retro-reflective transceivers. But each one has more or less drawbacks and thus further study needs to be done to investigate other solutions and compare the performance of all these options.

Conclusion and Future Work

In this chapter, we promoted the new paradigm of VL based throughput downlink connectivity for CR networks ($CR^{VL}_{downlink}$) communication, by exploiting both CR and VL technologies in IoT communication scene. We first motivated the

promotion of $\text{CR}^{VL}_{downlink}$ from different angles, covering such as the challenge of radio spectrum shortage, the heterogeneity of the involved devices, the weakness of CR, and the promising role of VLC. Then, we introduced $\text{CR}^{VL}_{downlink}$ communication architecture for smart home environment.

The importance of improving and maintaining the connectivity of "everything" is not always fully awared. The proposed architecture is taking an idea (i.e., VLC) with the potential for revolutionary innovation (e.g., higher data transmission speed, lower energy consumption rate, better communication security or privacy, etc.) and utilizing it for something far more excited. This chapter pioneers the use of VLC technology in IoT scenario. The potential yielded answers and momentum are a necessity for further developments, in particular, application-driven solutions. Fortunately more and more researchers are also poring over VLC protocol to make sure that a fair and efficient design blueprint is deliverable. Yet if someone succeeds in bring an operable VLC access network into the existing CR facilitated IoT, reliability and timeliness will be two factors that, among others, need a careful consideration in the subsequent QoS study.

References

1. Akyildiz, I.F., Lee, W.-Y., Vuran, M.C., Mohanty, S.: NeXt generation/dynamic spectrum access/cognitive radio wireless networks: a survey. Comput. Netw. **50**, 2127–2159 (2006)
2. Akyildiz, I.F., Lee, W.Y., Chowdhury, K.R.: CRAHNs: cognitive radio ad hoc networks. Ad Hoc Netw. **7**, 810–836 (2009)
3. Akyildiz, I.F., Lo, B.F., Balakrishnan, R.: Cooperative spectrum sensing in cognitive radio networks: a survey. Phys. Commun. **4**, 40–62 (2011)
4. Almeida, R., Louro, P., Vieira, M.A., Vieira, M.: Visible light communication in traffic links using an a-SiC: H multilayer photodetector. Procedia Technol. **17**, 550–556 (2014)
5. Arnon, S.: Optimised optical wireless car-to-traffic-light communication. Trans. Emerg. Telecommun. Technol. **25**, 660–665 (2014)
6. Arredondo, B., Romero, B., Pena, J.M.S., Fernández-Pacheco, A., Alonso, E., Vergaz, R., Dios, Cd: Visible light communication system using an organic bulk heterojunction photodetector. Sensors **13**, 12266–12276 (2013)
7. Bauer, H., Patel, M., Veira, J.: The Internet of Thins: Sizing Up the Opportunity McKinsey & Company, pp. 1–7 (2014)
8. Bell, A., Adams, W., Preece, W.: Discussion on the photophone and the conversion of radiant energy into sound. J. Soc. Telegraph Engineers **9**, 375–383 (1880)
9. Berman, S.M., Greehouse, D.S., Bailey, I.L., Clear, R.D., Raasch, T.W.: Human electroretinogram responses to video displays, fluorescent lighting, and other high frequency sources. Optom. Vis. Sci. **68**, 645–662 (1991)
10. Chen, R., Park, J.M., Hou, Y.T., Reed, J.H.: Toward secure distributed spectrum sensing in cognitive radio networks. IEEE Commun. Mag., 50–55 (2008)
11. Clancy, T.C., Goergen, N.: Security in cognitive radio networks: threats and mitigation. In: Proceedings of 3rd International IEEE Conference on Cognitive Radio Oriented Wireless Networks and Communications, Singapore, pp. 1–8
12. Ding, W., Yang, F., Yang, H., Wang, J., Wang, X., Zhang, X., Song, J.: A hybrid power line and visible light communication system for indoor hospital applications. Comput. Ind. **68**, 170–178 (2015)

13. Elgala, H., Mesleh, R., Haas, H.: Indoor broadcasting via white LEDs and OFDM. IEEE Trans. Consum. Electron. **55**, 1127–1134 (2009)
14. Ergul, O., Dinc, E., Akan, O.B.: Communicate to illuminate: state-of-the-art and research challenges for visible light communications. Phys. Commun. **17**, 72–85 (2015)
15. Fath, T., Haas, H.: Performance comparison of MIMO techniques for optical wireless communications in indoor environments. IEEE Trans. Commun. **61**, 733–742 (2013)
16. Gardner, W.A.: Exploitation of spectral redudancy in cyclostationary signals. IEEE Signal Process. Mag. **8**, 14–36 (1991)
17. George, J.J., Mustafa, M.H., Osman, N.M., Ahmed, N.H., Hamed, DaM: A survey on visible light communication. Int. J. Eng. Comput. Sci. **3**, 3805–3808 (2014)
18. Groth, M.: Photophones Revisted. Amateur Radio Magazine, pp. 12–17. Wireless Institute of Australia, Melbourne (1987)
19. Haruyama, S., Yamazato, T.: Image sensor based visible light communication. In: Arnon, S. (ed.) Visible light communication. Cambridge University Press, ISBN: 978-1-107-06155-2, University Printing House, Cambridge CB2 8BS, United Kingdom (2015)
20. Haykin, S.: Cognitive radio: brain-empowered wireless communications. IEEE J. Sel. Areas Commun. **23**, 201–220 (2005)
21. Home Gigabit Access (OMEGA) Project
22. Institute of Electrical and Electronics Engineers (IEEE) (2011) 802.15.7 IEEE standard for local and metropolitan area networks–part 15.7: short-range wireless optical communication using visible light
23. Jackson, D., Buffaloe, T., Leeb, S.: Fiat lux: a fluorescent lamp digital transceiver. IEEE Trans. Ind. Appl. **34**, 625–630 (1998)
24. Jivkova, S., Kavehrad, M.: Transceiver design concept for cellular and multispot diffusing regimes of transmission. EURASIP J. Appl. Sig. Process. **1**, 30–38 (2005)
25. Jondral, F.K.: Software-defined radio: basics and evolution to cognitive radio. EURASIP J. Wirel. Commun. Netw., 275–283 (2005)
26. Komine, T., Nakagawa, M.: A study of shadowing on indoor visiblelight wireless communication utilizing plural white LED lighings. In: Proceedings of 1st International Symposium on Wireless Communication System, pp. 36–40 (2004)
27. Kumar, K., Prakash, A., Tripathi, R.: Spectrum handoff in cognitive radio networks: a classification and comprehensive survey. J of Network and Computer Applications **61**, 161–188 (2016)
28. Kumar, N., Lourenço, N.R.: Led-based visible light communication system: a brief survey and investigation. J. Eng. Appl. Sci. **5**, 296–307 (2010)
29. Le, N.-T., Jang, Y.M.: Resource allocation for multichannel broadcasting visible light communication. Opt. Commun. **355**, 451–461 (2015)
30. Mathur, C.N., Subbalakshmi, K.P.: Digital signatures for centralized DSA networks. In: Proceedings of 4th IEEE Conference on Consumer Communications and Networking, pp 1037–1041 (2007)
31. Medina, C., Zambrano, M., Navarro, K.: LED based visible light communication: technology, applications and challenges—a survey. Int. J. Adv. Eng. Technol. **8**, 482–495 (2015)
32. Mitola, J., Maguire, G.Q.: Cognitive radios: making software radios more personal. IEEE Pers. Commun. **6**, 13–18 (1999)
33. Mitola, J.: Software Radio Architecture: Object-Oriented Approaches to Wireless Systems Engineering. Wiley-Interscience (2000)
34. Mitola, J.: Cognitive radio: an integrated agent architecture for software defined radio. KTH Royal Institute of Technology, Stockholm, Sweden (2000)
35. Nakajima, M., Haruyama, S.: New indoor navigation system for visually impaired people using visible light communication. EURASIP J. Wirel. Commun. Netw. **37**, 1–10 (2013)
36. Ntogari, G., Kamalakis, T., Sphicopoulos, T.: Performance analysis of space time block coding techniques for indoor optical wireless systems. IEEE J. Sel. Areas Commun. Strat. **27**, 1545–1552 (2009)

37. Pang, G., Kwan, T., Chan, C.H., Liu, H.: Led traffic light as a communications device. In: IEEE/IEEJ/JSAI International Conference on Intelligent Transportation Systems, Tokyo, Japan, pp 788–793 (1999)
38. Parvin, S., Hussain, F.K., Hussain, O.K., Han, S., Tian, B., Chang, E.: Cognitive radio network security: a survey. J. Netw. Comput. Appl. **35**, 1691–1708 (2012)
39. Pathak, P.H., Feng, X., Hu, P., Mohapatra, P.: Visible light communication, networking, and sensing: a survey, potential and challenges. IEEE Commun. Surv. Tutor. **17**, 2017–2077 (2015)
40. Quintana, C., Guerra, V., Rufo, J., Rabadan, J., Perez-Jimenez, R.: Reading lamp-based visible light communication system for in-flight entertainment. IEEE Trans. Consum. Electron. **59**, 31–37 (2013)
41. Rea, M.S.: The IESNA Lighting Handbook: Reference & Application. Illuminating Engineering Society, North America, New York, USA (2000)
42. Reyes, H., Subramaniam, S., Kaabouch, N., Hu, W.C.: A spectrum sensing technique based on autocorrelation and Euclidean distance and its comparison with energy detection for cognitive radio networks. Comput. Electr. Eng. (in press)
43. Saleem, Y., Salim, F., Rehmani, M.H.: Routing and channel selection from cognitive radio network's perspective: a survey. Comput. Electr. Eng. **42**, 117–134 (2015)
44. Schubert, E.F.: Light-Emitting Diodes, 2nd edn. Cambridge University Press, ISBN 978-0-511-34476-3, The Edinburgh Building, Cambridge CB2 8RU, UK (2006)
45. Sharifi, A.A., Sharifi, M., Niya, M.J.M.: Secure cooperative spectrum sensing under primary user emulation attack in cognitive radio networks: attack-aware threshold selection approach. Int. J. Electron. Commun. (AEÜ) **70**, 95–104 (2016)
46. Soa, J., Kwon, T.: Limited reporting-based cooperative spectrum sensing for multiband cognitive radio networks. Int. J. Electron. Commun. (AEÜ) **70**, 386–397 (2016)
47. Sundmaeker, H., Guillemin, P., Friess, P., Woelfflé, S.: Vision and challenges for realising the Internet of things. European Commission—Information Society and Media DG, Brussels, Belgium (2010)
48. Economist, The: Smart Products, Smart Makers. The Economist **417**, 65 (2015)
49. The Economist: Wireless: the next generation The Economist, pp. 53–54 (2016)
50. Tsonev, D., Videv, S., Haas, H.: Light fidelity (Li-Fi): towards alloptical networking In: Proceedings of SPIE OPTO, Art ID 900702 (2013)
51. United States Department of Energy: Energy savings forecast of solid-state lighting in general illumination applications (2016)
52. Visible Light Communications Consortium (VLCC)
53. Yamazato, T., Takai, I., Okada, H., Fujii, T., Yendo, T., Arai, S., Andoh, M., Harada, T., Yasutomi, K., Kagawa, K., Kawahito, S.: Image-sensor-based visible light communication for automotive applications. IEEE Commun. Mag., 88–97 (2014)
54. Yau, K.-L.A., Ramli, N., Hashim, W., Mohamad, H.: Clustering algorithms for cognitiver radio networks: a survey. J. Netw. Comput. Appl. **45**, 19–95 (2014)
55. Yu, Z., Baxley, R., Zhou, G.: Multi-user MISO broadcasting for indoor visible light communication. In: Proceedings of IEEE ICASSP, May 2013, pp. 4849–4853 (2013)
56. Yücek, T., Arslan, H.: A survey of spectrum sensing algorithms for cognitive radio applications. IEEE Commun. Surv. Tutor. **11**, 116–130 (2009)
57. Zarrin, S.: Spectrum Sensing in Cognitive Radio Networks. Department of Electrical and Computer Engineering. University of Toronto (2011)
58. Zeng, Y., Liang, Y.: Spectrum-sensing algorithms for cognitive radio based on statistical covariances. IEEE Trans. Veh. Technol. **58**, 1804–1815 (2009)

Reconfigurable MIMO Antennas for Cognitive Radios

Rifaqat Hussain and Mohammad S. Sharawi

Introduction

Cognitive radio (CR) is a promising wireless communication technology to be used in next generation cellular and wireless networks. In CR based communication systems, the transceiver can intelligently detect the idle or underutilized resources and instantly move the communication path to those unoccupied channels or underutilized resources. This dynamic spectrum access of underutilized frequency band is used to enhance the spectrum efficiency and thus avoid congestion. In CR platforms, a secondary unlicensed user may be allowed to access the unused spectrum or underutilized spectrum without affecting the primary user activities. The ability to correctly sense the presence of primary user activities is crucial in decision making and allocating the frequency band.

Cognitive Radio and the Need for Reconfigurable Antennas

The concept of CR was first coined by Joseph Mitola III in 1999 as an extension of Software Defined Radio (SDR) [1]. CR can be defined as the situation where radio devices and communication networks are aware of their environment and interact dynamically to adjust the operating parameters according to their needs and learn over a period of time. It basically mimics the human behavior to efficiently manage the frequency spectrum resources with minimum human interaction. CR based appli-

R. Hussain · M.S. Sharawi (✉)
Electrical Engineering Department, King Fahd University of Petroleum
and Minerals (KFUPM), Dhahran 31261, Saudi Arabia
e-mail: msharawi@kfupm.edu.sa

R. Hussain
e-mail: rifaqat@kfupm.edu.sa

cations may have several requirements for efficient spectrum management operation. First, a CR based system requires high level of awareness about the radio environment, users and network needs as well as communication aspects of the radio frequency (RF) spectrum. Secondly, the CR based system is required to adjust the operating parameters autonomously without human intervention. This can be achieved by integrating artificial intelligence features. The third requirement is the switching of the communication channel to a new band [1, 2].

CR systems communicate across multi-channels, operate over multi-bands, follow multi-standard protocols, learn the variants of the RF environment and adapt to it as per user needs. CR systems are capable of detecting and analyzing the radio spectrum and could establish communication with another RF devices across different frequency bands by complying with different standards followed by CR learning. CR learning and adapting new operating parameters capabilities is to ensure seamless real time communication with other systems. At the physical layer of communication networks, CR systems are used to envision cognition as well as reconfigurability to complete the cognition cycle as shown in Fig. 1. CR cognition cycle consist of five main stages [3–5].

1. Spectrum Sensing: The first stage of the cognition cycle, '*Spectrum Sensing*', is to sense or scan the entire RF spectrum to detect the RF activities across various frequency bands and channels. This can be accomplished by developing efficient and optimized techniques for spectrum sensing and detecting across the entire band of interest.
2. Spectrum Analysis: '*Spectrum analysis*' is an important aspect followed by spectrum sensing. It is mainly associated with the following '*Decide*' stage to help in making rational decisions.
3. Decide: The '*Decide*' stage of cognition cycle is based on past observations and experiences of radio spectrum and help in developing cognition engine. This leads to decision making capability based on rationale. This stage actually helps in deciding the current optimal operations of RF devices and networks based on logical reasoning and experiences gained over a period of time.
4. Act: The fourth element of cognition cycle is the reconfigurability stage that is followed by the observed RF environment. A parametric controlled architecture is utilized to adjust and adapt the new set of operating parameters based on users and network needs. The operating parameters reconfigurability may constitute of dynamic switching of multi-band, multi-channel and multi-standard operations.
5. Learn: Learning is an important element of the cognition cycle and is a part of each stage. Learning by observations about RF environment help in anticipating the use of the appropriate communication standards, right protocol and correct mode of operation. Various learning algorithms can be implemented to train the CR devices and networks for right operation such as neural networks and support vector regression.

CR based communication systems are required to perform multitude of functions across several frequency bands or operating bandwidth. Hence it requires an antenna
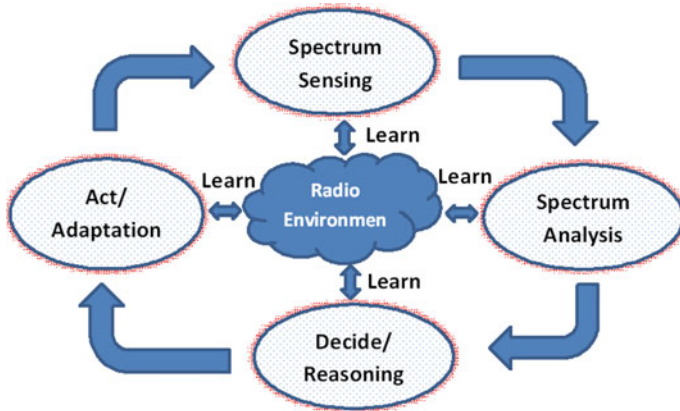
Fig. 1 Cognitive radio cognition cycle

system to scan the wide frequency spectrum via a wide-band sensing antenna and to dynamically switch the operating characteristics across various frequency bands and perform versatile applications by system operating parameters. Thus, reconfigurable antennas become a crucial part of CR based communication systems. Reconfigurable antennas, reconfigurable RF circuits and reconfigurable systems are required to fulfill the system requirements of CR based communication systems, capable of adjusting the operating parameters and behavior of the system to coop the system requirements or RF environmental conditions. The parameters may be to enhance the bandwidth, switching the operating bands, radiation pattern reconfigurability and sense of antenna polarization. Reconfigurable multiple-input-multiple-output (MIMO) antenna system is now widely used in CR based communication systems because of tremendous advantages they offers [6].

Reconfigurable MIMO Advantages

MIMO antenna systems are widely used in current wireless technology, including LTE, WLAN, and mobile WiMAX systems. Such wireless devices are continuously added with numerous features and hence need high system throughput. The high data rate transmission with reliable communication is highly desirable in conditions such as interference, signal fading, and multipath. This can be accomplished by deploying MIMO antenna systems resulting in high system throughput with minimal multipath interference.

The simplified channel capacity for a MIMO antenna system is given by [7]:

$$C = MBlog_2 \left(1 + \frac{N}{M} SNR \right) \tag{1}$$

where *C* is the channel capacity (in bps), *SNR* is the signal to noise ratio of the wireless communication channel, *B* is the bandwidth (in Hz), *M* is the number of antennas at the transmitter side, and *N* is the number of antennas at the receiver side. Channel capacity of MIMO antenna systems depend on the number of antenna elements with least channel coupling. The performance of a MIMO system is significantly degraded with high field coupling between antenna elements.

In CR applications, reconfigurable MIMO antenna systems can be used to combine both the characteristics of MIMO and reconfigurable antennas. Through proper antenna design, channel capacity can be enhanced with its ability to respond to any variation in radio spectrum environment. The objective of reconfigurable MIMO antenna system is to maximize the system throughput with its capability to utilize the available spectrum resources efficiently.

Printed antennas are widely used in mobile terminals and wireless handheld devices for MIMO operation. Printed MIMO antennas have attractive features of low cost, ease of fabrication and ease of integration. The field coupling between closely spaced antenna elements should be reduced to increase the MIMO throughput. Thus the use of multiple antenna elements and characterizing their performance is an inevitable task in modern antenna design. MIMO antenna systems with reconfigurable front ends for efficient bandwidth utilization are becoming popular in recent years [7, 8].

Second Generation CR Features and Outlook

Most of the prospective users of the next generation wireless technology would be mobile terminals and wireless handheld devices. There will be a tremendous demand to access the data resources either for stationary or non-stationary users. Hence it is necessary to standardize the communication protocols and wireless technology to efficiently utilize the available resources that is a fundamental theme of CR technology. This will motivate and lead to the development of CR technology, its standardization and CR network architectures. The integration of CR technology with adaptive CR network architecture is of particular interest to policy makers and standardization agencies for second generation CR platforms [2, 9].

The basic concepts of CR operation is an extension of software defined radio (SDR). It helps in developing a platform that senses the activities on the radio spectrum of interest and reacts to the operating environment. This helps in fully utilizing the idle or underutilized spectrum resources, with minimal external engineering and switching to tune the optimal operating parameters. Some features of second generation CR technology are coping with spectral congestion, dynamic spectral acess (DSA), sensing the current radio frequency spectrum environment, policy and configuration databases, self-configuration, mission-oriented configuration, adaptive algorithms, distributed collaboration, operational state information and distributed resource management [9–12].

Cognitive Radio Antennas

The front end of a CR is embedded with two types of antennas, an ultra wide-band (UWB) sensing antenna and a reconfigurable communication antenna. The UWB antenna is used to scan the entire spectrum of interest and sense the underutilized or idle bands for communication. Reconfigurable antennas are used to switch the band of operation or operating fundamental characteristics according to what was sensed in the first part. These fundamental parameters may include resonance frequency, radiation pattern, polarization and impedance bandwidth [13]. Thus, a CR system can be defined as a radio having a spectrum sensing capability and communicate accordingly.

In modern communication systems, high throughput requirements can be overcome by efficiently utilizing the spectrum resources and using MIMO antenna systems within CR platforms for communication purposes. The hierarchy of the antennas used in CR platforms is shown in Fig. 2 and are categorized as follows.

1. Sensing antenna

   a. Monopole antenna types
   b. Dipole antenna types

2. Communication reconfigurable antenna (using PIN diode, varactor diode, metamaterial (MTM), micro-electromechanical system (MEMS) switches)

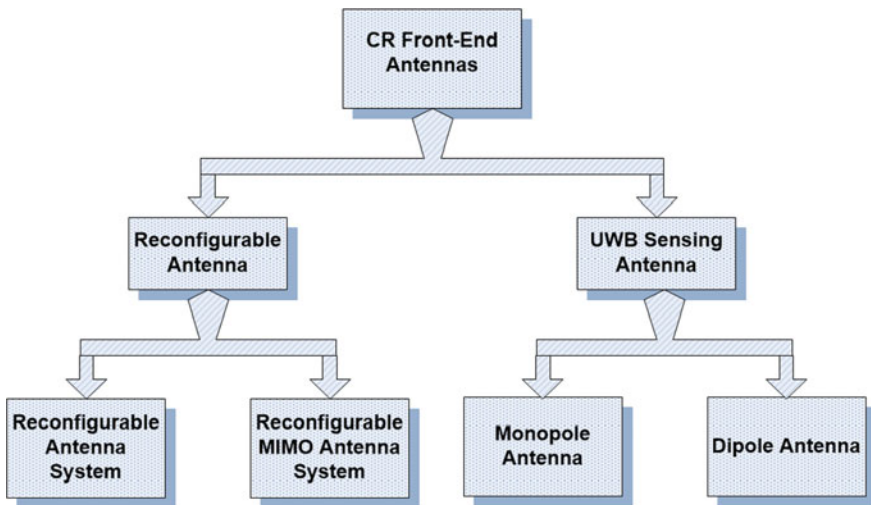   a. Reconfigurable antenna
   b. Reconfigurable MIMO antenna

Fig. 2 Classification of CR antennas

UWB Sensing Antenna

RF spectrum sensing is of primary importance in CR applications for intelligent DSA. UWB sensing antennas are used widely for spectrum sensing. This UWB sensing ensures the continuous reliable spectrum support to licensed primary users and its allocation to secondary users without affecting the primary users activities. It is very critical in DSA for CR application. The frequency allocation to secondary users with minimum effects on the primary users spectrum is the key of CR communication. This can be achieved by the use of UWB antennas covering the whole frequency band of operation and use of UWB devices on the back of CR platforms.

Significance of UWB Sensing

DSA is the backbone of CR communication which can only be realized by UWB spectrum scanning. In practical communication systems, UWB processing, followed by UWB sensing, is the most difficult and challenging task as it requires UWB RF digitizer and other communication equipments to process the data. However, this stage is most crucial and of particular interest. The RF digitizer digitizes the specified ultra-wide bandwidth in real time. The UWB processing becomes more severe in wireless handheld devices and mobile terminals with compact form factor. New architectures have been devised and under continuous investigated over years and new ways are constantly underway to make the spectrum sensing more reliable in CR platforms [12].

UWB Sensing Antenna—Examples

Two types of sensing antennas widely used for UWB sensing in CR applications are printed dipole and monopole antennas. Printed planar antenna structures are easy to work with and are an appropriate choice as sensing antennas in wireless communication systems. The advantages of such structures are low profile structure, offer wide impedance bandwidth, easy to fabricate and most importantly is their omni-directional radiation pattern. However, it is a very challenging to have a compact UWB antenna at low frequency bands below 1 GHz. Examples of both monopole and dipole sensing antennas are provided in this section.

A. **Dipole based Sensing Antennas**: Dipole antennas are used most widely for UWB operation because of their omni-directional radiation pattern and high efficiency. Vivaldi and antipodal vivaldi dipole antennas are printed version and are very popular in CR platforms. The main benefits of vivaldi antennas are their planar structure, easy fabrication, broadband characteristics with compact form factor and can be matched easily to 50 Ω impedance line. A few example of printed dipoles antennas operating as UWB antennas are discussed in this section.

   In [14], a printed vivaldi antenna was presented covering frequency bands from

3.1 GHz to 10.6 GHz The design was realized on a Rogers board with $\varepsilon_r = 3.48$ and height equal to 0.762 mm. The board dimension were $38 \times 60 \, \text{mm}^2$. The antenna design is shown in Fig. 3a. The antenna was optimized to avoid the interference caused by WLAN band. Figure 3b shows the measured VSWR of proposed design with and without the U-shaped slot. The VSWR value remained less than 2:1 in the desired UWB band. The slot produced a sharp band notch centered at 5.45 GHz to avoid WLAN band interference.

A printed UWB antipodal vivaldi antennas with band reject filter based on split ring resonator (SSR) appeared in [15]. The design was compact with board dimensions of $35 \times 32 \, \text{mm}^2$ and loaded with SSR to reject any interference in the frequency band from 5 to 6 GHz. In [16], a dual-purpose, high gain dipole antenna was presented. The antenna was acting as broadband antenna with resonance frequency from 2.4 to 4.02 GHz and also acted as 2×2 MIMO antenna system. The geometry of the proposed design is shown in Fig. 4a. The simulated and measured VSWR curves are given in Fig. 4b. The antenna performance is good in the desired band of 2.4–4.02 GHz. Printed dipole antennas are a good choice to be used as sensing antenna for CR platform. Most of sensing antenna designs covered bands higher than 1.5 GHz and lower frequency UWB antennas were of relatively larger size.

B. **Monopole based Sensing Antenna**: Monopole antennas are also widely used in applications that need high efficiency and high gain. Printed monopole antennas are good candidates as sensing antennas for CR platforms. They can exhibit UWB operation with compact size, easy fabrication process, planar structures, and easy integration with other planer structures. Low profile monopole antennas for UWB applications are designed with different structures, sizes and shapes as per operating band requirements. A number of UWB sensing monopole designs were cited in literature with different shapes such as oval shape,
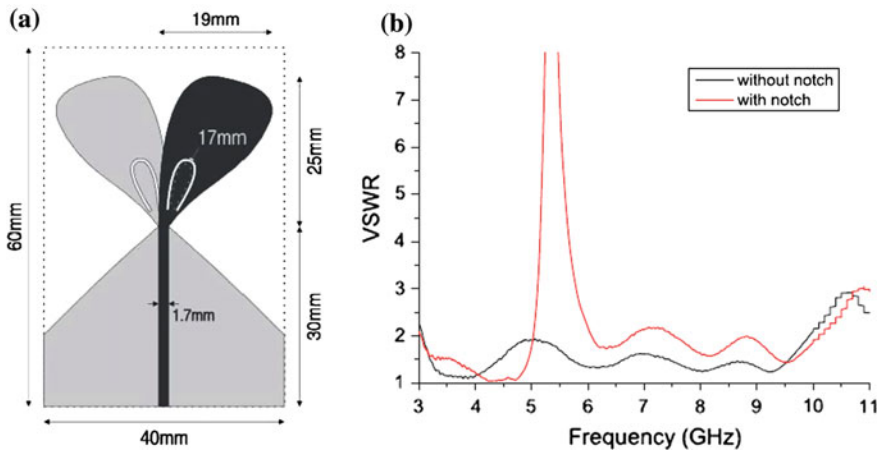

Fig. 3 **a** Geometry of vivaldi antenna with U-shaped slot. **b** VSWR with and without notch [14]
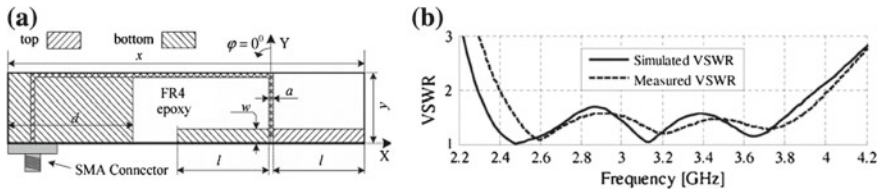
Fig. 4 **a** Geometry of broadband dipole antenna. **b** Measured and simulated VSWR curves [16]

wine-glass shape, fork shape, polygon shape antenna, hour-glass shape and egg shape. The target was UWB unlicensed spectrum bands from 3 to 10 GHz.

In [17], a monopole sensing antenna was presented covering the frequency band 2–5.5 GHz. This was an integrated design of sensing antenna with reconfigurable antennas fabricated on the same structure with a total dimension of 100×47 mm^2 as shown in Fig. 5a. The simualted and measured reflection coefficient curves are shown in Fig. 5b. In [18], an UWB monopole sensing antenna was presented with design volume of $34 \times 31 \times 0.88$ mm^3. This design was unique in a sense that the same UWB band antenna was made reconfigurable using PIN diodes. The switching from UWB to narrow band operation resulted in radiation pattern reconfigurability. This antenna used the same feed line that was used for the reconfigurable part of antenna.

Most of the UWB sensing antenna design were operating above 1.5 GHz while very few designs are available operating below 1.5 GHz. The size of the antenna becomes very large for frequency band below 1 GHz that cannot be accommodated in wireless handheld devices for cellular phone applications. In [19], a printed monopole with board dimensions of $200 \times 40 \times 1.6$ mm^3 was presented as shown in Fig. 6. The frequency bands covered were from 530 to 3000 MHz. A relatively larger dimensions of the design is not suitable for wireless handheld devices and mobile terminals.

Reconfigurable Antennas for CR Platforms

The communication antennas are embedded with numerous features for wide range of applications and service coverage. Thus, the antenna design and its performance is very critical for successful CR implementation. The static characteristics of conventional communication antennas, strict constraints on antenna size and limited system performance resulted in limited operating configurations of antenna system. The problems can be addressed by using the reconfigurable antennas with the existing setup but with reconfigurable configuration using electronic circuitry. The communication antenna can be made frequency reconfigurable, can adjust its radiation characteristics and can change it sense of polarization. Thus, reconfigurable antennas
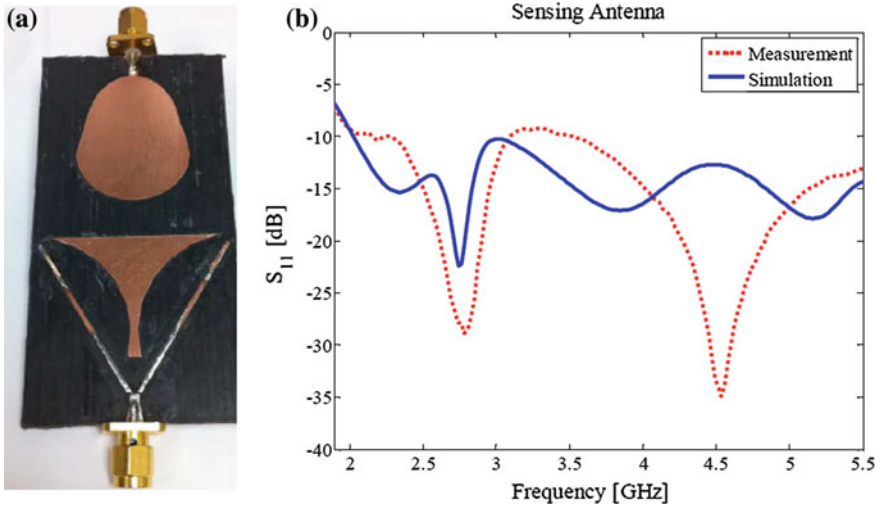
Fig. 5 a Fabricated UWB antenna. **b** S_{11}—simulated and measured [17]
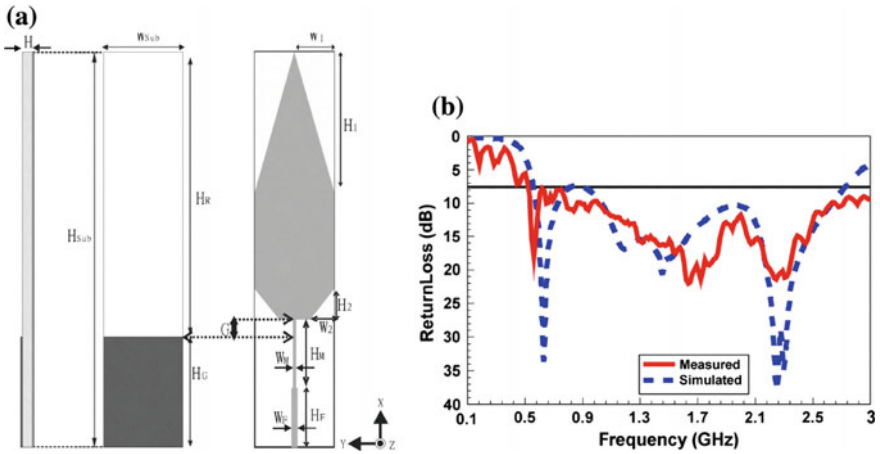

Fig. 6 a Geometry of UWB antenna. **b** Simulated and measured return loss curves [19]

are necessary to dynamically adapt to new parameters as per system requirements or environmental conditions in order to enhance the overall system performance.

Generally, the basic reconfigurable parameters controlled in printed antenna structures in CR application include, frequency reconfigurability, polarization reconfigurability and radiation pattern reconfigurability. The methods or components that are employed for achieving reconfigurability are varactor diodes, PIN diodes, RF MEMS, metamaterial based structures and substrate properties to make the most commonly used. Different types of reconfigurable communication antennas are discussed in the following section.

Frequency Reconfigurable Antennas

In modern wireless communication systems, frequency reconfigurable antennas can play a vital role and have been studied extensively. The same antenna aperture is being utilized to switch its operating bands across various frequency bands resulting in a compact antenna design. The basic theory of frequency reconfigurable operation is to change the effective length of the antenna elements via parasitic loading that enable its operation across different frequency bands.

Several frequency reconfigurable antennas suitable for CR applications have been proposed in literature. In [20–22], varactor diode based frequency reconfigurable communication antennas were presented for CR application for frequency bands from 0.5 to 5 GHz. Varactor diodes are used as variable capacitors and are popular in RF design. The capacitance of the varactor diode can be varied by applying different reverse bias voltages across the diode terminals. Reconfigurability is achieved by varying the reactive loading of the radiating antenna structure by changing the reverse bias voltage. This actually changes the resonance by changing the effecive capacitances of the antenna (recall that an antenna behaves like an RLC resonant circuit) via a continuous range of capacitive loading and hence resulted in smooth variation of the operating frequency bands. The essential DC reverse bias circuitary consists of a combination of an RF choke and a resistance in series with the varactor diode.

In [20], compact frequency reconfigurable F-slot monopole antenna was presented with F-shape slot as shown in Fig. 7a. The varactor diode was accommodated in this F-shape slot. The capacitance was varied from 0.5 to 3 pF with corresponding frequency sweep of 1.70–2.15 GHz was observed as shown in Fig. 7b. The antenna covered wireless bands of DCS, PCS, and UMTS. Antenna was of compact size
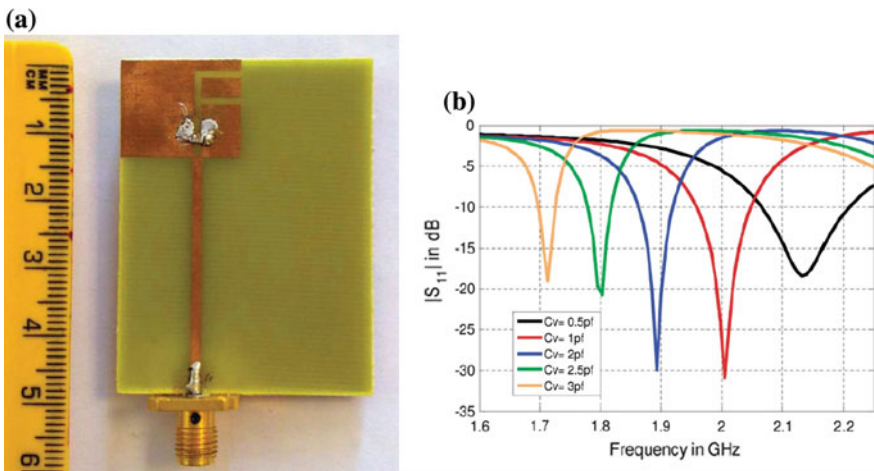

Fig. 7   **a** Fabricated antenna. **b** S_{11}—simulated [20]

with substrate dimensions of $50 \times 37.5\,\text{mm}^2$ suitable for mobile handset applications. A dual band frequency reconfigurable slot antenna was presented in [21]. Varactor diodes based reconfigurable antenna of dimensions $150 \times 150\,\text{mm}^2$, covered frequency band from 1.3 to 2.67 GHZ. In [22], a compact $50 \times 40\,\text{mm}^2$, varactor diode based frequency reconfigurable PIFA was presented. The covered frequency bands were 0.7, 2, 3.5 and 5 GHz.

PIN diode is another choice that could be utilized to change the effective length of the antenna element and hence to obtain reconfigurability. PIN diode can be used as a switch in RF circuit among various antenna parts. It provides conducting path in forward bias condition by connecting antenna elements while it isolates the antenna parts in reverse bias condition. Hence, its two modes of operation can provide different current paths resulting in different resonances and hence achieving reconfigurability. The essential DC biasing circuitry of PIN diode consists of an RF choke and current limiting resistance in series with diode and power supply. PIN diode based frequency reconfigurable antennas were presented in [23–25] for CR applications. In [23], muti-band, frequency agile, inverted-F antenna with dimensions $50 \times 70\,\text{mm}^2$, was presented. The frequency bands covered wrer GSM850 (824–894 MHz), GSM900 (890–960 MHz), DCS (1710–1880 MHz), and PCS (1850–1990 MHz). Similarly, a frequency reconfigurable PIFA was presented in [24] with substrate dimensions of $45 \times 75\,\text{mm}^2$. The quad-band antenna design covered bands GSM900, GSM1800, GSM1900 and UMTS. In [25], dual band C-slot patch antenna was presented with board area of $50 \times 50\,\text{mm}^2$. The frequency band covered were from 5 to 7 GHz.

MEMS are also a popular method for achieving frequency reconfigurability in antenna structure. MEMS are electro-mechanical switches that can be utilized to change the effective length of antenna elements. Some examples of MEMS based frequency agile antennas were presented in [26–28]. In [27], MEMS based antenna design was presented for mobile terminal. The frequency bands achieved were GSM/900 and GSM/1800. The antenna was fabricated on total board dimensions of $40 \times 98\,\text{mm}^2$.

Metamaterial(MTM) based reconfigurable antennas had some attention in recent years because of various features one can get using them. MTM antennas can be made reconfigurable in many ways but varactor diode, PIN diode and MEMS based designs are more popular. In [29], a MTM inspired antenna was investigated for frequency agility. A printed monopole MTM inspired antenna was incorporated with a slot that was fed by a coplanar waveguide (CPW) line as shown in Fig. 8a. The original antenna was exhibiting dual band operation. The lower frequency band was obtained from the incorporated slot and was made reconfigurable using a varactor diode. The capacitance of varactor diode was varied from 0.1 to 0.7 pF with a corresponding smooth transition of resonance frequency from 1.6 to 2.23 GHz as shown in Fig. 8b. The antenna was realized on board dimensions of $30 \times 30\,\text{mm}^2$. Similarly, another MTM based frequency reconfigurable antenna was presented in [30]. A mushroom structure-based zeroth order resonant (ZOR) antenna was used to obtain the frequency reconfigurability with operating band in frequency range from 2 to 2.25 GHz.
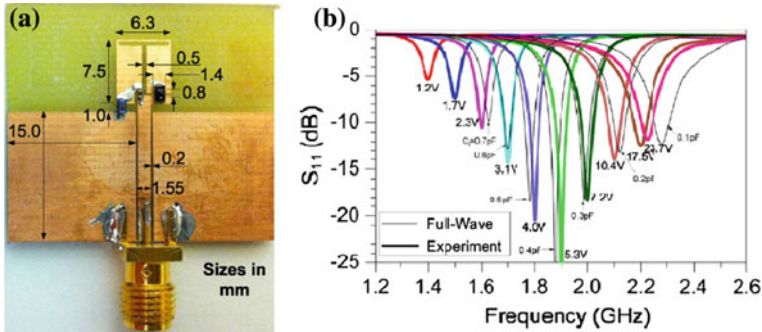
Fig. 8 **a** Geometry of MTM based frequency reconfigurable antenna. **b** S_{11}—Refelection coeffi-cient curves [29]

Radiation Pattern Reconfigurable Antenna

Radiation pattern reconfigurable antennas are gaining wide acceptance in wireless communication technologies. Pattern reconfigurable antennas can be utilized for multitude of functions that would be useful for CR applications. It may include the features like avoiding noisy environments, helpful in anti-jamming techniques, can be utilized for tracking target objects, help in improving the security be dynamically changing the beam and may be helpful in directing the signals toward targeted users [31]. Thus, the pattern reconfigurability would be helpful in improving the system performance of CR platforms.

The radiation pattern and its orientation depends on the distribution of electric/magnetic currents on the antenna surface and hence forming a relationship between source current and the resultant radiation pattern. The relation between the these quantities will determine the pattern reconfigurability. The type of current distribution on the antenna surface would generate a particular pattern that could be made reconfigurable by varying the current distribution. The pattern reconfigurable antenna may suffer from change in frequency response as well as impedance bandwidth mismatch. The mismatch problem can be avoided to some extent by isolating the reconfigurable part from the feeding circuit and thus allowing fixed frequency response with pattern agility [32, 33].

Radiation pattern reconfigurable antennas are either fixed or variable shape beam type. A number of research articles have been cited covering both aspect of pattern reconfigurability. In [34], a beam-steering antenna was presented. The antenna structure was integrated with four RF switches to tilt the beam to four distinct angles, 0°, 90°, 180° and 270°. The antenna was fabricated on Felt substrate with board dimensions of 52×85 mm^2. The reconfigurable antenna is shown in Fig. 9 while Fig. 10 shows the steered beam at four different angles.
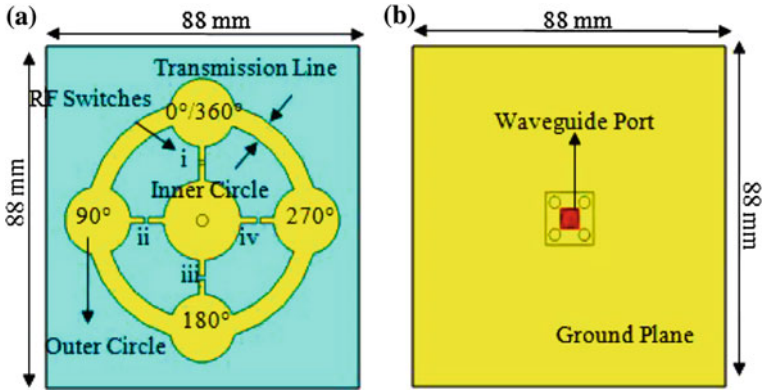
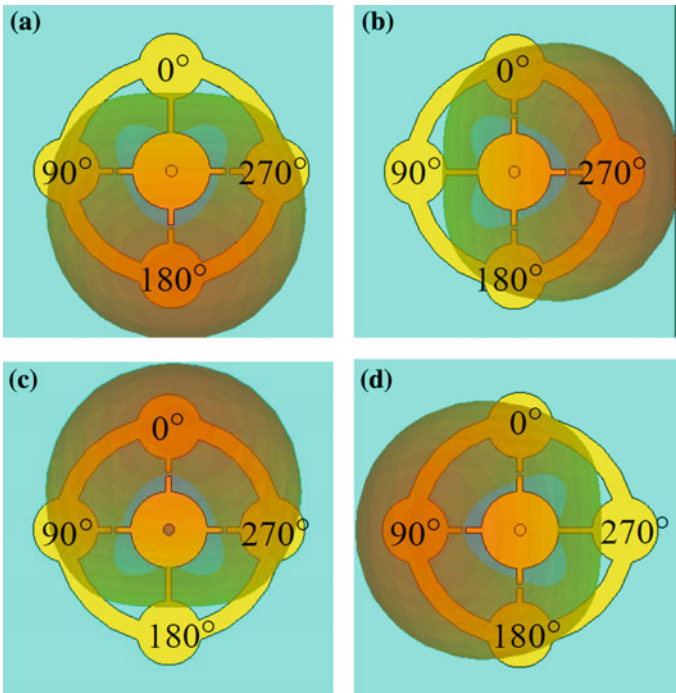**Fig. 9** Geometry of the proposed antenna in [34]

Fig. 10 Simulated 3D pattern of PIN diode configurations. **a** Beam tilted at 180°. **b** Beam tilted at 270°. **c** Beam tilted at 0°. **d** Beam tilted at 90° [34]

Polarization Reconfigurable Antenna

Polarization reconfigurable antennas can also play a major role in CR applications. Polarization reconfigurable antennas can be utilized in radio environment to mitigate the effects of interfering signals as well as it can provide an additional degree of freedom to improve the communication link. The sense of antenna polarization depends on antenna structure, material properties and its feeding network. The main challenge in achieving polarization reconfigurability is at the cost of impedance mismatch and frequency characteristics of antenna elements [33].

In a polarization reconfigurable antenna, the polarization can be controlled to support various polarization modes of operation. Several number of polarization reconfigurable antennas have been reported in literature. In [35], polarization reconfigurable microstrip antenna was presented. The antenna was excited by a microstrip feed line through aperture coupling. Reconfigurability was achieved by a coupling slot and the open stub of the feed line by PIN diodes. PIN diodes were used for switching to achieve vertical and horizontal polarizations. The linearly polarized antenna was converted to circularly-polarized antenna with switchable polarization sense using a perturbation segment. Furthermore, quadri-polarization diversity was also achieved with dual orthogonal linear polarizations and two circular polarizations. The size of proposed design was 100×100 mm^2 and is shown in Fig. 11a while the simulated and measured axial ratio are given in Fig. 11b.

Reconfigurable MIMO Antennas for CR Platforms

MIMO technology is a disruptive wireless communication one that could be utilized to increase the channel capacity and hence meet the high data rate requirements by employing multiple antennas on the communication link. This could be accomplished without any extra power or bandwidth. The antenna design with MIMO con-
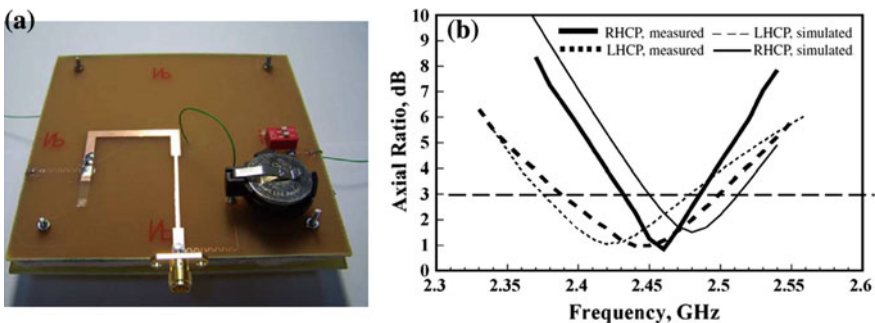

Fig. 11 **a** Fabricated antenna protype. **b** Measured and simulated results of axial ratio for two circularly polarized modes [35]

figuration providing high data rates and, that is capable of efficient utilization of the frequency spectrum, are highly desirable for CR applications especially in wireless handheld devices and mobile terminals. In this section, three basic types of reconfigurability are discussed for printed antennas in MIMO configuration for CR platforms.

Frequency Reconfigurable MIMO Antenna Systems

For next generation CR application, high data rate requirements and dynamic configuration of antenna parameters resulted in investigating reconfigurable MIMO antenna systems. A number of research works can be found in literature for CR applications. Some of the frequency reconfigurable MIMO antennas for CR applications are discussed in this section.

A 2-element dual-band varactor diode based modified inverted F-shaped MIMO antenna system was presented in [36]. The fabricated models of reconfigurable PIFA are shown in Fig. 12. The antenna elements were mounted on an elevated board at height 5.8 mm above the main substrate board. The dimensions of the single element were 12×30 mm^2. The single element modified PIFA consists of meander lines with a folded patch to bring down resonance frequency. The detailed structure of single antenna element is shown in Fig. 13.

A rectangular slot on the top layer was created to accommodate the varactor diode for reactive antenna loading. The varactor diodes were used to vary the resonating bands from 750 to 1160 MHz in the lower band and 1540–1900 MHz in the higher band in simulation. The simulated and measured reflection coefficients curves are shown in Fig. 14a, b. The MIMO antennas covered several well-known frequency bands including GSM-750, GSM-870, LTE-900, LTE-1800, with several other bands as well. The simulated 3-D gain patterns for antenna elements A-1 and A-2 are shown in Fig. 14c, d, respectively. The tilt in the radiation patterns help in lowering the field coupling and enhanced the MIMO performance. The design was realized on FR4 substrate with typical mobile phone back-plane area of 120×65 mm^2. The reconfigurable antenna was evaluated for MIMO parameters and showed good MIMO performance metrics.

In [37], 4-elements meandered line F shape frequency reconfigurable MIMO antenna was presented with enhanced isolation as CR MIMO communication antenna. Antenna was fabricated on Rogers 4003 substrate with $\varepsilon_r = 3.55$. The fabricated prototype is shown in Fig. 15 with top view in Fig. 15a and bottom view in Fig. 15c while the detailed dimensions and biasing circuitry of are given antenna are shown in Fig. 15b. The antenna was develop on smart phone back plane dimensions of $65 \times 120 \times 1.56$ mm^3 with single element dimension was a summation of 7.9×37 mm^2 and 17.7×19.6 mm^2. The antenna was suitable to be used in wireless handheld devices and mobile terminals.

The antenna had a unique combination of PIN and varactor diodes to tune the antenna at various distinct bands as well as smooth variation of tuned band over a larger bandwidth. The antenna covered several frequency band in the frequency
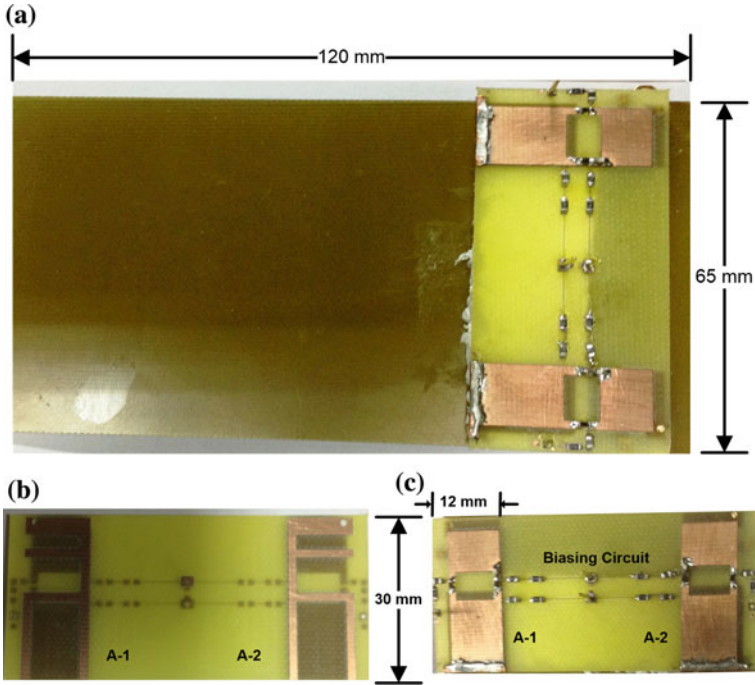
Fig. 12 **a** Fabricated antenna model. **b** *Bottom* view of elevated F-shaped antenna. **c** *Top* view elevated F-shaped antenna [36]
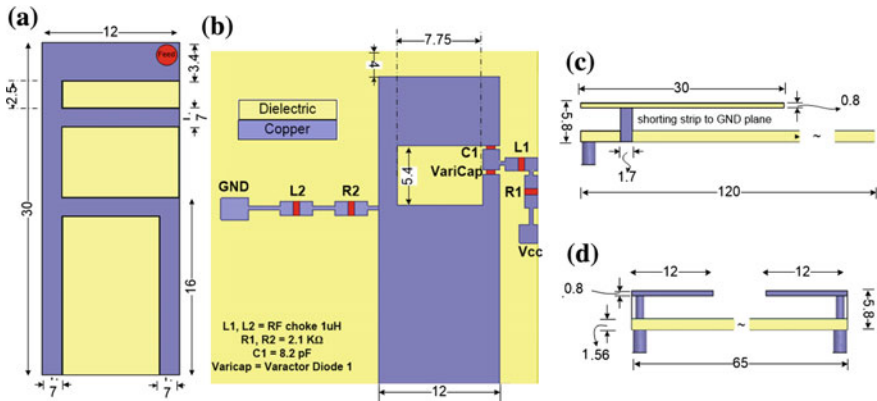

Fig. 13 Detailed schematic of reconfigurable MIMO antenna. **a** *Top* view. **b** *Bottom* view. **c** *Side* view. **d** *Front* view [36]
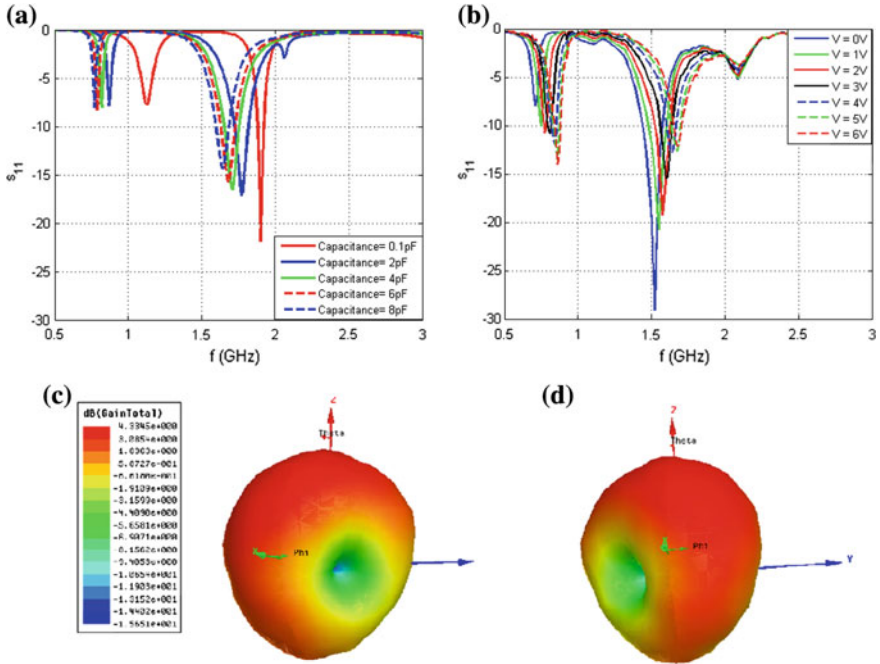
Fig. 14 **a** Simulation reflection coefficient curves. **b** Measured reflection coefficient curves. **c** Simulated gain at 1700 MHz for antenna-1. **d** Simulated gain at 1700 MHz for antenna-2 [36]
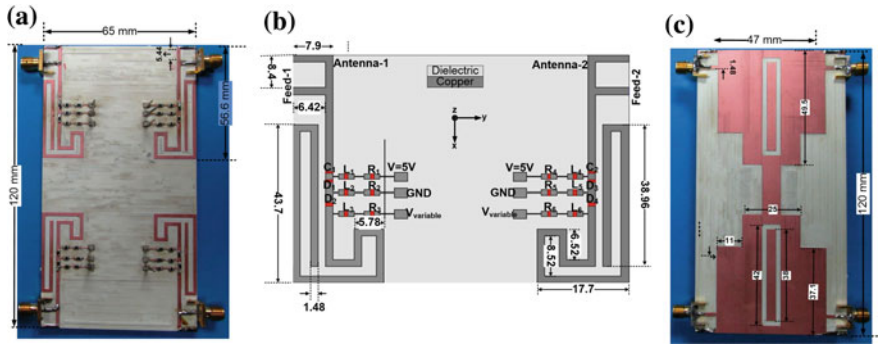
Fig. 15 Proposed four elements MIMO antennas system for CR platform. **a** *Top* view. **b** Detailed view of antenna elements. **c** *Bottom* view—all dimensions are in millimeter (mm) [37]

range from 0.7 to 3 GHz and suitable for CR platforms deployed in cellular networks. The PIN diodes were utilized for band switching while varactor diodes were used to tune the antenna over a wide-band. The MIMO antenna was operating in two modes. In mode-1, the PIN diodes were switched OFF while the dynamic reactive loading of the antenna had negligible effects on the operating frequency. Mode-1 had a dual

Fig. 16 S_{11} curves. **a** Mode-1: simulated and measured. **b** Mode-2: simulated [37]

Table 1 Comparison of frequency reconfigurable MIMO antenna systems

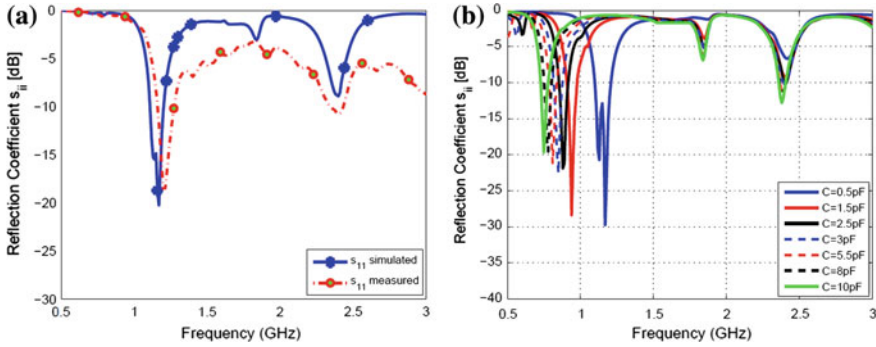| S.no	Antenna type	Reconfigurability mean	No of elements	Size (mm²)	Frequency bands (GHz)
[38]	PIFA	PIN	2	60×120	0.9, 0.95, 1.25, 1.55, 1.8,1.84, 2.45
[39]	Monopole	Varactor	4	60×120	0.78–2.4
[40]	Dipole	Varactor	2	118×40	0.646–0.848, 1.648–2.074
[41]	Dipole	Varactor	2	66×38	2.3–3.5
[42]	Monopole	PIN	2	80×40	2.4, 5.2 and 3.5
[43]	Pixel	MEMS	13×13	50.8×50.8	4–7

band operation at 1170 and 2420 MHz with a −6 dB operating bandwidth of at-
least 100 MHz in both bands. In Mode-2, the PIN diodes were switched ON and
varactor diodes loading had a significant effects on the lower frequency band below
1 GHz. In mode-2, first frequency band achieved was from 743 to 1240 MHz and
second band at 2400 MHz with minimum −6 dB operating bandwidth of 60 MHz,
and 120 MHz, respectively. The reflection coefficients curves for both these bands are
shown in Fig. 16a, b. Table 1 compares the features of the reconfigurable antenna that
apperared in [38–43] in terms of antenna type, reconfigurability mean, number of
MIMO antenna elements, board size and frequency bands covered by these antennas.

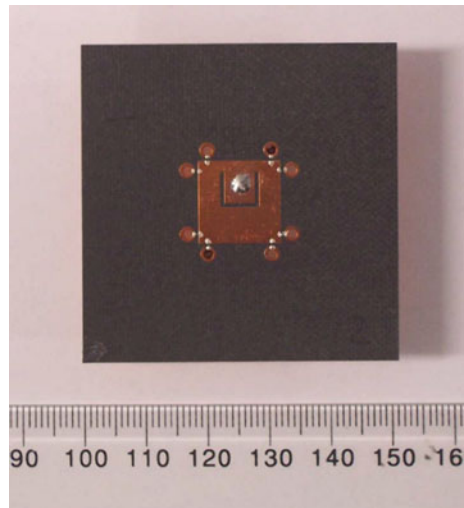## Pattern Reconfigurable MIMO Antennas

A pattern reconfigurable MIMO antenna system integrate the features of pattern
diversity antenna along with enhancing system performance with MIMO configura-

tion. Thus, pattern reconfigurable MIMO antennas can be utilized for efficient signal transmission, power saving by directing signal toward target, anti-jamming capabilities and secure communication along with high system throughput. These are the requirements of second generation CR platform. Several work have been reported in literature and could be utilized for next generation CR application. One of the pattern reconfigurable U-slot patch antenna was presented in [44] for MIMO applications. The patch antenna was short circuited with ground plane using PIN diodes. The different states of PIN diodes helped in determining the pattern. The antenna was acting as monopolar patch or normal patch and hence resulted in pattern diversity was achieved by switching the pattern between conical and boresight patterns electrically. The antenna was operating at frequency band of 5.32 GHz. U-slot antenna with shorting post is shown in Fig. 17 with antenna dimension of 12×12 mm^2.

In [45], pattern reconfigurable MIMO antenna was presented. The antenna utilized the parasitic elements characteristic modes to achieve pattern diversity. Figure 18a shows the fabricated design of the given pattern reconfigurable antenna. The analysis were performed on 30×90 mm^2 board dimension at 2.28 GHz. The antenna showed pattern diversity by switching to different pattern orientation as shown in Fig. 18b, c.

Table 2 summarizes the features of the pattern reconfigurable antenna that appeared in [46–50]. The antennas features included number of elements, size, frequency of operation and number of switchable patterns.

Fig. 17 Fabricated model of pattern reconfigurable U-slot antenna [44]

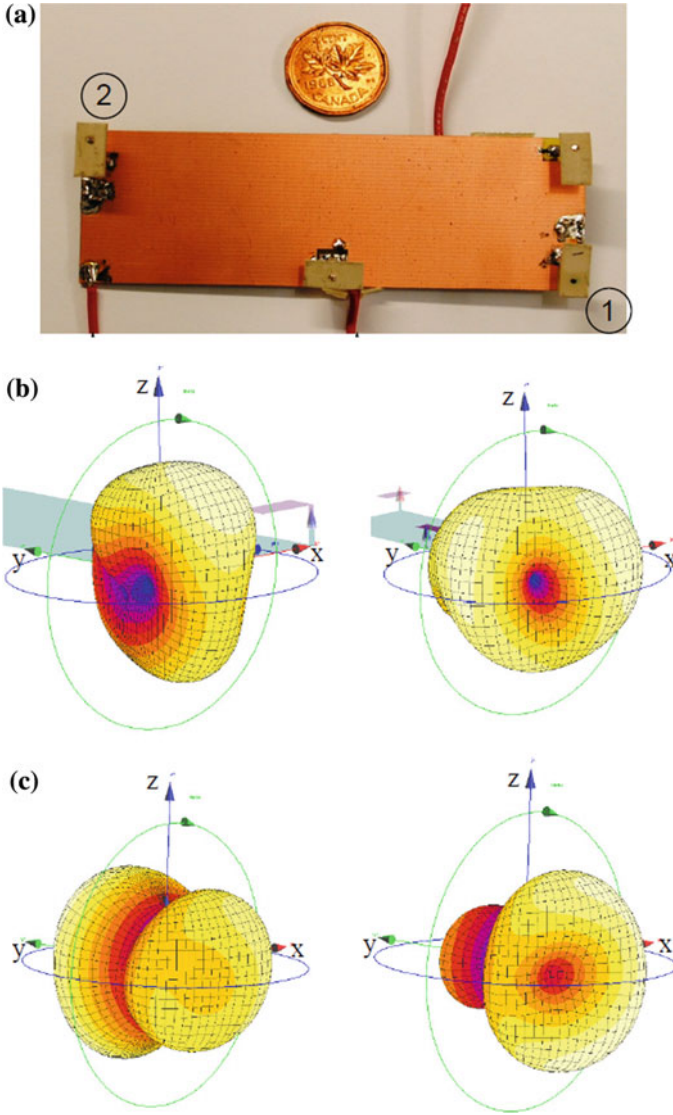**Fig. 18** **a** Fabricated mode. **b** Radiation patterns from ports 1 and 2 in state 1. **c** Radiation patterns from ports 1 and 2 in state 2 [45]

Polarization Reconfigurable MIMO Antennas

Polarization reconfigurable MIMO antennas are the prospective CR communication antennas with potential to improve the system performance in dynamic environment. The high system throughput requirements can be met with switching capability of

Table 2 Comparison of pattern reconfigurable MIMO antenna systems

S.no	Number of elements	Frequency (GHz)	Size (mm²)	Number of switchable patterns
[46]	2	2.3	–	2
[47]	4	2.455	16π	2
[48]	4	2.45	20	4
[49]	4	2.5–2.8	55×55	4
[50]	4	8.2	30×30	4

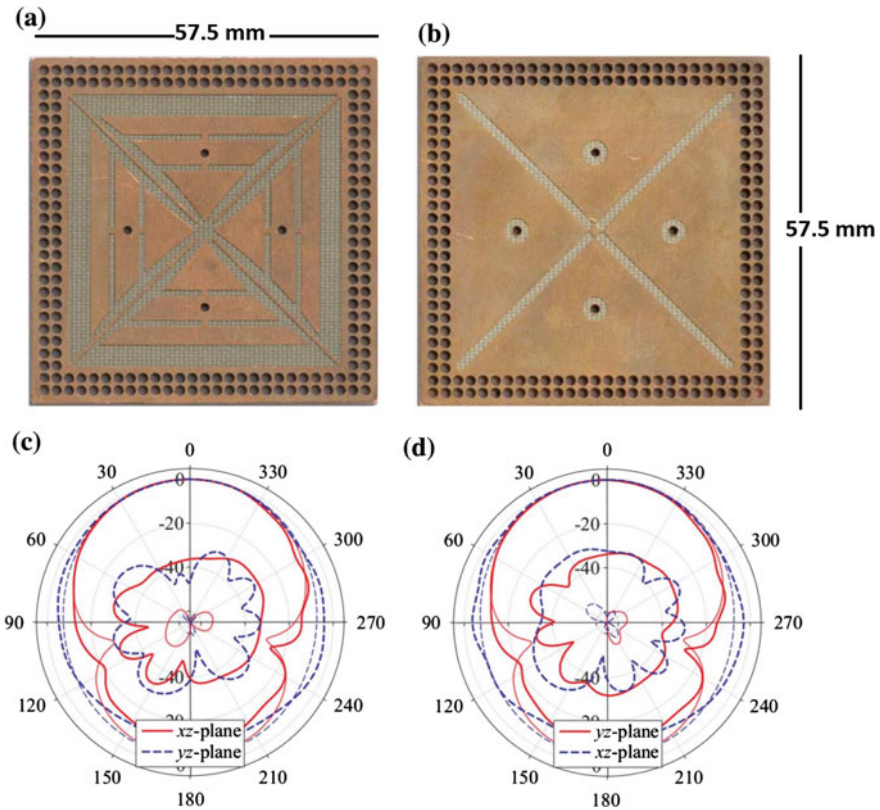**Fig. 19** Dual-polarized MIMO antenna. **a** *Top* view. **b** *Bottom* view. **c** Simulated and measured radiation patterns at 2.6 GHz—vertical polarization. **d** Simulated and measured radiation patterns at 2.6 GHz—horizontal polarization [51]

polarization in reconfigurable MIMO antennas. Several polarization reconfigurable MIMO antennas were reported in literature and few of these antennas are discussed in this section.

In [51], a dual-polarized slot based MIMO antenna system was presented as shown in Fig. 19a as top view while Fig. 19b shows the bottom view of the antenna. The rectangular slot is used to fed by two pairs of triangular patches and resultant radiation pattern exhibited orthogonal polarizations. The vertical and horizontal polarization wavers were achieved by feeding port 1 and port 2. The additional slots were utilized to control the length of the current paths in patch antenna elements. Figure 19c, d shows the simulated and measured radiation patterns at 2.6 GHz for both vertical and horizontal polarization, respectively. The antenna structure was optimized to achieve high cross-polarization discrimination (XPD) in all directions.

In [52], polarization diversity MIMO patch antenna system was presented. The sense of polarization switching was achieved by changing the current path using PIN diodes. The antenna was dual-purpose, exhibiting both polarization and pattern diversity antenna. The antenna design was realized on board size of 51×51 mm^2. In [53], a dual-feed PIFA MIMO antenna was presented for WLAN and LTE applications. The two feeding plates were placed perpendicular to each other to obtain the polarization and pattern diversity. The design was fabricate on FR4 with board dimensions of 40×100 mm^2. A dual-polarised monopole-slot MIMO antenna was presented in [54]. The antenna was resonating at 2.4 GHz and could change its polarization between vertical and horizontal orientation. The design was compact and was occupying a volume of $50 \times 16 \times 16$ mm^3.

Integrated Sensing and MIMO Antennas for CR

Over the past decade, reconfigurable MIMO antennas system were investigated extensively for CR applications. Most of the front-end antennas for CR application were either UWB antennas or single reconfigurable communication antenna with sensing antenna. Very few MIMO reconfigurable communication antennas were integrated with UWB sensing antenna on the same substrate for CR applications. MIMO antenna systems integrated with UWB sensing antenna complying with 3G/4G wireless standards are highly desirable for next generation CR communication. In this section, such as integrated antenna systems are discussed for CR platforms.

Since CR front ends requires spectrum sensing antenna as well as narrow band reconfigurable antenna operation, an integrated solution for both antennas on single platform is required. A number of single sensing and reconfigurable antennas for CR application are provided in [55–58] and a comparison between them in terms of size and covered frequency bands for both sensing and reconfigurable are provide in Table 3.

The sensing and MIMO communication antennas integration are of vital importance and is discussed in coming paragraphs. In [60, 61], 2-element frequency reconfigurable MIMO antennas integrated with UWB sensing antenna were presented. The communication MIMO antennas covered several frequency bands between 0.7 and 3 GHz while the sensing antenna was supposed to cover this entire, 0.7–3 GHz, UWB range.

Table 3 Comparison of single integrated sensing and reconfigurable antenna systems

S.no	Size (mm^2)	Sensing antenna band (GHz)	Reconfigurable antenna bands (GHz)
[55]	33.54 × 49	2.65–10.3	3.55–5.18, 5.12–6.59, 7.10–8.01
[56]	58 × 65.5	3.1–11	5.3–9.15, 3.4–4.85
[59]	50 × 50	3–11	3.88, 4.56, 4.84, 5.44, 6.36, 7.48, 8, 8.88
[57]	70 × 50	2.1–10	2.1–3, 3–3.4, 3.4–5.56, 5.4–6.2, 6.3–10
[58]	60 × 50	3 × 10	5.5, 2.5, 9.5
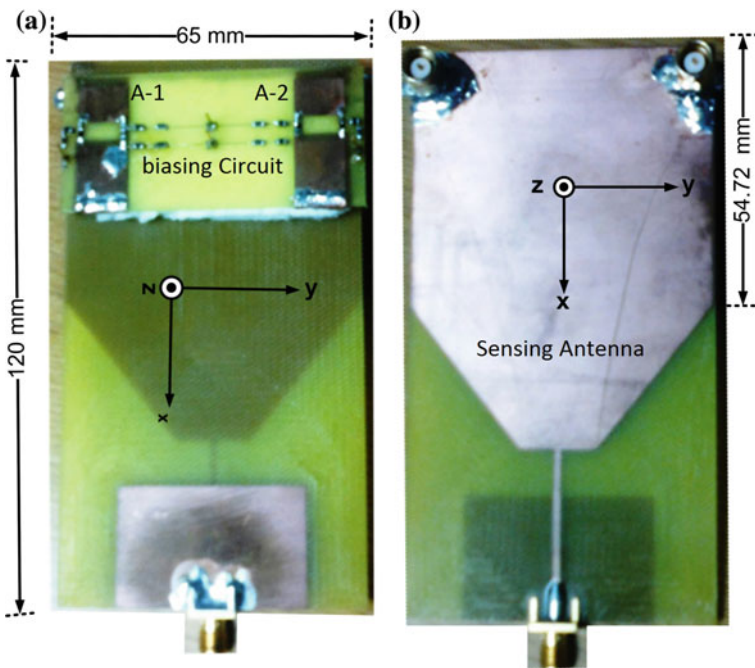
Fig. 20 Integrated CR antenna system. **a** *Top* view. **b** *Bottom* view [60]

In [60], modified reconfigurable PIFA-like MIMO antenna was presented along with sensing antenna for CR applications as shown in Fig. 20a. The GND plane of the MIMO antenna was utilized as UWB antenna as shown in Fig. 20b. The given UWB sensing antenna is basically monopole structure covering a frequency range from 710 to 3600 MHz. The sensing antenna exhibited a omnidirectional radiation pattern, which is the basic requirement of CR sensing antenna. The reflection coefficient and omni-directional radiation pattern of sensing antenna are shown in Fig. 21a, b, respectively.
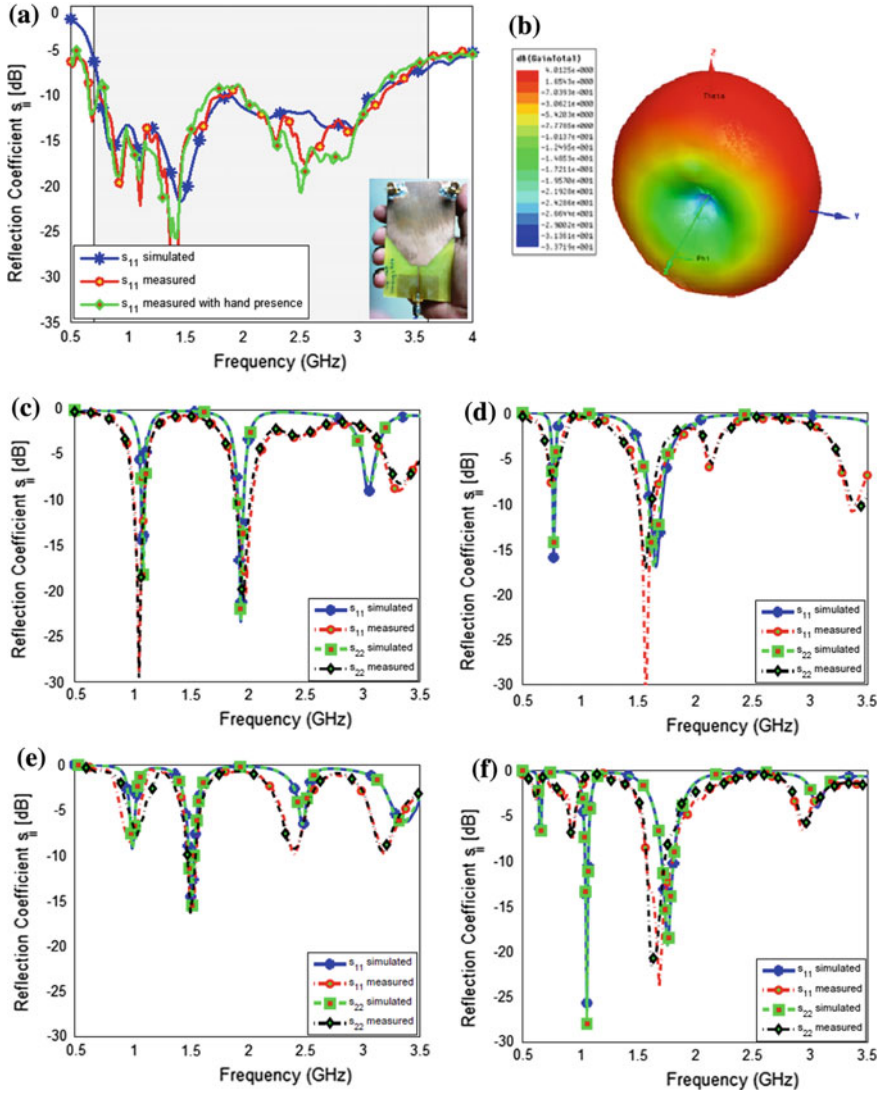
Fig. 21 **a** S_{11} of sensing antenna. **b** Gain pattern at 1500 MHz. **c** S_{11} mode-1. **d** S_{11} mode-2. **e** S_{11} mode-3. **f** S_{11} mode-4 [60]

This 2-element MIMO antenna with PIN diode based reconfigurability covered several frequency bands from 755 MHz and up to 3450 MHz. The reconfigurable MIMO antenna had four operating modes. The reflection coefficient curves of all 4-modes are given in Fig. 21c–f, respectively. The 4-mode operation of reconfigurable antenna resulted in frequency operation of 770, 1060, 1508, 1660, 1760, 1945, 2450, 3050 and 3350 MHz. The MIMO antennas were evaluated for MIMO performance
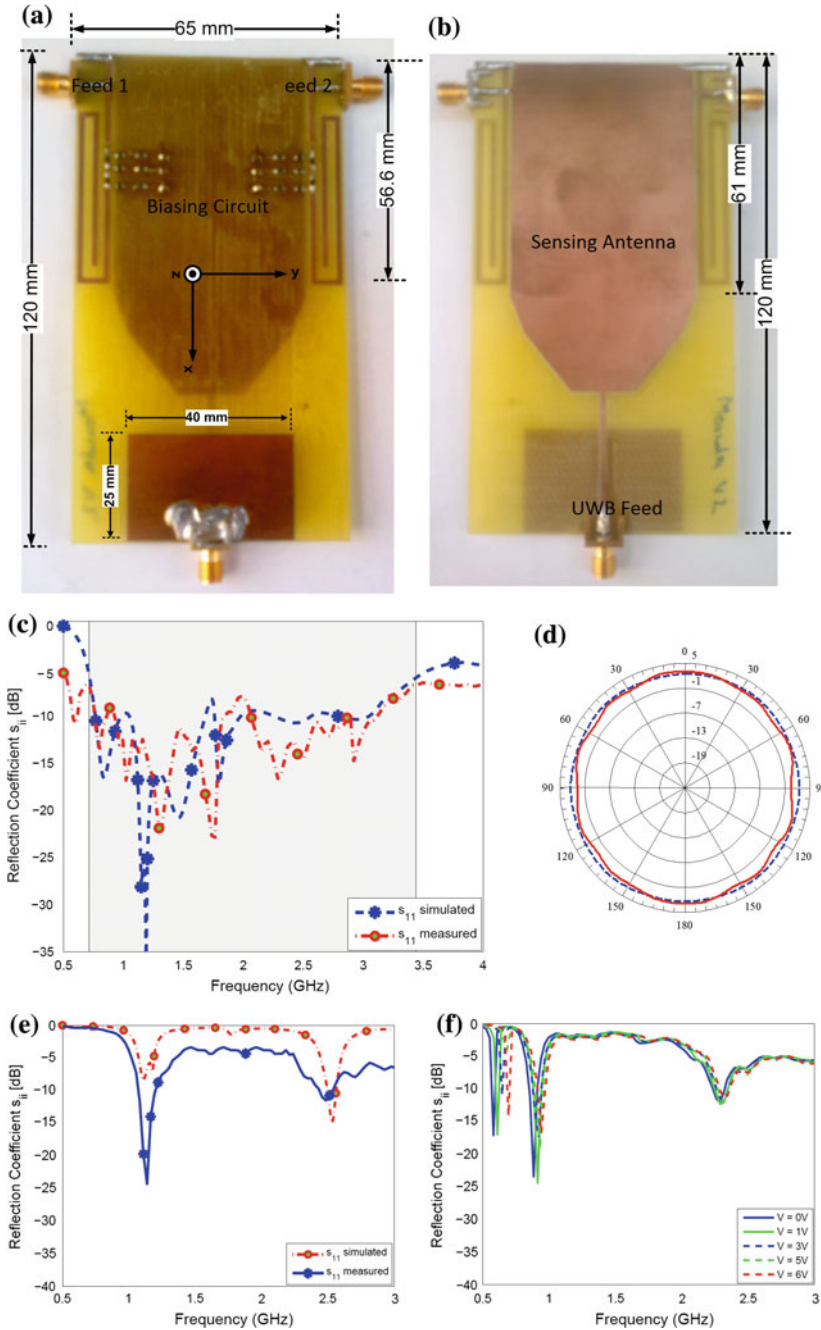
Fig. 22 **a** Fabricated antenna. **b** Fabricated sensing antenna. **c** S_{11}—simulated and measured for sensing antenna. **d** 2-D pattern of sensing atenna—yz plane. **e** S_{11} mode-1: simulated and measured. **f** S_{11} mode-2: measured [61]
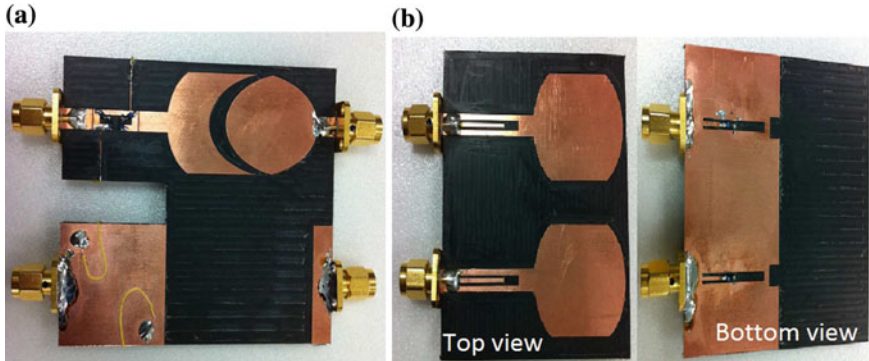
Fig. 23 **a** Interweave MIMO CR antenna system. **b** Underlay MIMO CR antenna system [62]

metrics and good results were obtained. The dimensions of the single antenna element were $30 \times 12 \times 0.8$ mm^3 and was fabricated on FR-4 substrate with $\varepsilon_r = 4.4$.

Similarly, a planar reconfigurable MIMO antenna embedded with an UWB sensing antenna for CR platform was presented in [61]. The complete integrated solution is well suitable for wireless handheld devices and mobile complying the wireless standards in CR applications. The fabricated model of the proposed design is shown in Fig. 22. The MIMO antennas are shown in Fig. 22a while UWB sensing antenna is shown in Fig. 22b. The sensing antenna was printed monopole antenna with omni-directional radiation pattern. The sensing antenna covered the target frequency bands of 700–3000 MHz. The reflection coefficient curves and 2-D gain pattern of the UWB sensing antenna are shown in Fig. 22c, d, respectively.

The reconfigurable MIMO antenna was modified monopole antenna. The frequency agility was achieved using a unique combination of PIN and varactor diode to tune the antenna over a wide band. 2-modes of operation were achieved using PIN either in ON or OFF state. The frequency bands covered were the lower frequency bands ranged from 573 to 680 MHz, 834–1120 MHz along with other distinct bands of 1100 and 2480 MHz as well for these 2-mode of operation. The reflection coefficient curves for 2-modes of operation are given in figure these antennas are given in Fig. 22e, f, respectively. The complete front-end CR antennas including sensing and communication were developed on single substrate dimensions of 65×120 mm^2 with single element size of 7.9×56.6 mm^2.

A reconfigurable MIMO filter-antenna (filtenna) was presented in [62] to cover the design aspects of two types of CR antenna, spectrum interweave and spectrum underlay systems. In spectrum interweave, the idle spectrum in sensed and utilized for communication while in spectrum underlay system, secondary users are allowed to use the primary user spectrum within the threshold power level. The fabricated model of MIMO filtenna design for two types of CR platforms are shown in Fig. 23a, b. The MIMO antenna was targeting CR applications along with sensing antenna to cover both interweave and underlay CR systems. The reconfigurable antenna structure was integrated with band-pass filter for interweave system and

band-reject filter for underlay system. The antenna system was realized on substrate area of 70×80 mm^2 for interweave system and 65×70 mm^2 for underlay system. Both the sensing and the reconfigurable antennas were designed for frequency range 3–6 GHz.

Future Prospects and Requirements of CR

One of the future disruptive technology in wireless communication is CR. It will enable new directions for research and will develop new set of communication standards. CR devices are fully programmable that sense the radio environment and respond accordingly. The CR dynamic response include switching its mode of operation, dynamic spectrum access and may adapt to new set of networking protocols for optimal performance. Future directions for CR network were extensively reported in [9, 63]. Some of the future research trends and challenges in implementing CR networks are discussed in this section.

- **Spectrum Sensing Algorithms**: Spectrum sensing is one of the key elements in CR communications to sense the surrounding of spectral environment, interferences and physical layer parameters. Spectrum sensing technology requires UWB operations. Spectrum sensing is a challenging task to obtain the signatures from multiple devices and the signal processing of obtained features. Effective spectrum sensing is important area for researchers in near future. The methods involve in effective spectrum sensing requires to improve the dynamic performance of system, reduce complexity of algorithms that can be implemented on real-world cognitive radio platforms. The spectrum sensing method still need to include the following:

1. Independent UWB energy detectors of unknown signals
2. Matched filter detection of known signals
3. Cyclo-stationary detectors to enhance the signal detection
4. Collaborative sensing by deploying multiple radios to get an improved signal estimate

- **Seamless spectrum handovers**: For next generation CR applications, seamless spectrum handovers are highly desirable to maintain high system throughput even during switch over to other spectrum settings. This is required for multimedia streams and real-time traffic applications. Spectrum handover scenarios are discussed in details in [2]. In [2], several open research topics regarding spectrum handovers are suggested. This will require high speed digital electronics on backend to process the signals during minimum time to ensure the quality of service (QoS) of CR platforms.
- **Proactive spectrum selection and interference avoidance**: Intelligent spectrum access and decisions to switch the operating parameters are necessary for CR operation. In CR operation, primary users are supposed to be unaware of the secondary

users while sharing the same spectrum. In CR application, the secondary users operate with maximum performance without interfering with the primary users activities. Thus, it suggests and recommended to integrate the learning algorithms based on the observations and past experiences. The learning may include the interaction with the environment with different actions and responses. Proactive spectrum selection and intelligent decision for CR settings may require extensive research to mature the CR communication.

- **DSA and new trends**: Dynamic spectrum access (DSA) still needs significant research to mature the technology. DSA is an autonomous radio spectrum observation method to continuously adapt to select frequency band to avoid interference. Researchers are continuously suggesting new schemes to improve the device agility performance [16, 17, 28]. It has been shown that systematic coordination techniques are much more effective to be used instead of random DSA. However, it is still needed to work on spectrum coordination protocols for improved coordination. The biggest challenge in DSA methods are to minimize the risk of interfering between primary and secondary users within acceptable limits.

- **Propagation path loss model of frequency agile characteristics of radio signals**: The propagation characteristics and path loss of frequency agile radio signals should be investigated in CR scenarios. In [12], propagation characteristics of the frequency was discussed for CR networks. However, there is much more space for research in this particular segment for CR network.

- **Alternatives to the common channel**: The existing proposed CR schemes used common control channel to coordinate all the activities. However, this method has limited scalability options, more prone to jamming attacks and less room for overall system performance. Hence, it is needed more efficient and robust methods to be investigated [36]. To improve overall CR system performance with scalability and robust communication.

- **Energy efficiency**: Spectrum sensing is one of the key elements of CR communication and the most power consuming process. It is highly desirable to reduce the power consumption in CR by reducing the number of sensing channels. The whole scanning of spectrum for arbitrarily channel selection is not an appropriate choice for CR communication. Learning based approach was recommended to minimize the spectrum sensing. Thus, CR devices utilize the prior learning experiences to optimize the energy usage instead of sensing. This would be hot topic of research in future till we reached to matured level of CR communication.

- **Validation of CR protocols**: Currently, most of CR protocols are validated either analytically or based on simulation results. It is not convincing and need real validation of these communication protocols. Prototyping and practical demonstration is necessary as proof of concept and is highly desirable to work on and has a research potential.

- **Cognitive algorithms for adaptation and resource management**: It is highly recommended to enhance the CR adaptation algorithm with one minimal processing and efficiently manage all the CR devices and network resources. Some direction for future research may include the following:

1. Dynamic adaptation algorithms with compliance with CR communication protocols
2. Robust adaptation algorithms and mode switching
3. Use artificial intelligence and machine learning algorithms for adaptation and resource management
4. Develop decentralized adaptive CR networks

- **Network security**: CR network security has an important new dimension of research and development in future wireless communication [43]. A CR platform with multiple authorized and unauthorized users are prone to data security threat. Hence, it necessary to develop trusted software and hardware CR architectures to minimize the threat to CR network security.

Conclusions

Future generation of wireless communication technology with tremendous system throughput requirement lead to empathize new ways to meet the ever growing demands. Once such option is to use CR (an extension of SDR) based innovative solutions to meet the high data rate requirements. CR is a method of efficient spectrum utilization, available resources and bandwidth that is able to perform multitude of functions across various frequency bands and wireless standards. The two front-end antennas, sensing and reconfigurable communication, for CR application are very crucial and it requires compact and efficient antenna design. The function of such antennas are self-awareness of radio spectrum environment by complete spectrum scanning along with communication by reconfigurable antennas over several bands. The reconfigurable communication antennas are required to improve the data rate and efficiently utilize the available bandwidth. Hence, it is recommended to be used in MIMO configurations for enhanced system throughput and improved performance. Intelligent spectrum sensing may provide additional capability of efficient spectrum utilization and be a part of a learning cycle. MIMO reconfigurable antenna design for CR applications has received considerable attention among the researchers over past decade. In the near future, the prospect usage of CR platforms integration within wireless handheld devices and mobile phone is very promising.

In this chapter, the two basic front-end antennas for CR platform were discussed. The antennas are the UWB sensing antenna and the reconfigurable communicating antenna. The need of reconfigurable antennas and reconfigurable MIMO antennas are highlighted with emphasis on the advantages of MIMO antennas for CR applications. Different types of sensing and communication antennas are discussed with numerous examples. The emphasis was on the printed planar structure for future generation CR applications. The chapter concluded with the future directions and challenges of CR antennas for wireless communication standards.

References

1. Mitola, J. III: Cognitive radio for flexible mobile multimedia communications. In: IEEE International Workshop on Mobile Multimedia Communication, pp. 3–10 (1999)
2. Mitola, J.: Cognitive radio architecture evolution. Proc. IEEE **97**(4), 626–641 (2009)
3. Jayaweera, S., Christodoulou, C.: Radiobots: architecture, algorithms and realtime reconfigurable antenna designs for autonomous, self-learning future cognitive radios. The University of New Mexico, Technical Report EECE-TR-11-0001 (2011)
4. Thomas, R.W., Friend, D.H., Dasilva, L.A., Mackenzie, A.B.: Cognitive networks: adaptation and learning to achieve end-to-end performance objectives. IEEE Commun. Mag. **44**(12), 51–57 (2006)
5. Tawk, Y., Costantine, J., Christodoulou, C.G.: Cognitive-radio and antenna functionalities: a tutorial [wireless corner]. IEEE Antennas Propag. Mag. **56**(1), 231–243 (2014)
6. Zhang, R., Liang, Y.-C.: Exploiting multi-antennas for opportunistic spectrum sharing in cognitive radio networks. IEEE J. Sel. Topics Signal Process. **2**(1), 88–102 (2008)
7. Sharawi, M.S.: Printed Mimo Antenna Engineering. Artech House (2014)
8. Piazza, D., Kirsch, N.J., Forenza, A., Heath, R.W., Dandekar, K.R.: Design and evaluation of a reconfigurable antenna array for mimo systems. IEEE Trans. Antennas Propag. **56**(3), 869–881 (2008)
9. Steenkiste, P., Sicker, D., Minden, G., Raychaudhuri, D.: Future directions in cognitive radio network research. NSF workshop report, vol. 4, issue no. 1, pp. 1–2 (2009)
10. Force, F.S.P.T.: Report of the spectrum efficiency working group (2002)
11. Kim, S.W., Pan, M., Joshi, G.P., Gazi, O., He, J., Coupechoux, M.: Applications of cognitive radio networks: recent advances and future directions. Int. J. Distrib. Sensor Netw. (2016)
12. Sadhu, B., Harjani, R.: Cognitive Radio Receiver Front-Ends: RF/Analog Circuit Techniques. Springer (2014)
13. Aboufoul, T., Alomainy, A., Parini, C.: Reconfiguring uwb monopole antenna for cognitive radio applications using gaas fet switches. IEEE Antennas Wirel. Propag. Lett. **11**, 392–394 (2012)
14. John, M., Ammann, M.J., McEvoy, P.: Uwb vivaldi antenna based on a spline geometry with frequency band-notch. In: IEEE International Symposium on Antennas and Propagation Society, pp. 1–4 (2008)
15. Sarkar, D., Srivastava, K.V.: Srr-loaded antipodal vivaldi antenna for uwb applications with tunable notch function. In: International Symposium on Electromagnetic Theory (EMTS), pp. 466–469. IEEE (2013)
16. Le, D.T., Karasawa, Y.: A simple broadband antenna for mimo applications in cognitive radio. IEICE Trans. Commun. **95**(1), 18–26 (2012)
17. Tawk, Y., Christodoulou, C., Costantine, J., Barbin, S.: A frequency and radiation pattern reconfigurable antenna system with sensing capabilities for cognitive radio. In: IEEE International Symposium on Antennas and Propagation Society, pp. 1–2 (2012)
18. Aboufoul, T., Ali, K., Alomainy, A., Parini, C.: Combined pattern and frequency reconfiguration of single-element ultra-wideband monopole antenna for cognitive radio devices. In: IEEE 7th European Conference on Antennas and Propagation (EuCAP), pp. 932–936 (2013)
19. Chen, W.-S., Chen, Y.-T., Chen, H.-T., Kuo, J.-S.: Wideband printed monopole antenna for wireless applications. In: IEEE International Symposium Antennas and Propagation Society, APSURSI'09
20. Elfergani, I., Abd-Alhameed, R., See, C., Sadeghpour, T., Noras, J., Jones, S.: Small size tuneable printed f-slot antenna for mobile handset applications. Microw. Opt. Technol. Lett. **54**(3), 794–802 (2012)
21. Behdad, N., Sarabandi, K.: Dual-band reconfigurable antenna with a very wide tunability range. IEEE Trans. Antennas Propag. **54**(2), 409–416 (2006)
22. AbuTarboush, H.F., Nilavalan, R., Peter, T.: Pifa based reconfigurable multiband antenna for wireless applications. In: IEEE International Conference on Electromagnetics in Advanced Applications (ICEAA), pp. 232–235 (2010)

23. Li, Y., Zhang, Z., Chen, W., Feng, Z., Iskander, M.F.: A quadband antenna with reconfigurable feedings. IEEE Antennas Wirel. Propag. Lett. **8**, 1069–1071 (2009)
24. Park, Y., Sung, Y.: A reconfigurable antenna for quad-band mobile handset applications. IEEE Trans. Antennas Propag. **60**(6), 3003–3006 (2012)
25. Abutarboush, H.F., Nilavalan, R., Cheung, S., Nasr, K.M., Peter, T., Budimir, D., Al-Raweshidy, H.: A reconfigurable wideband and multiband antenna using dual-patch elements for compact wireless devices. IEEE Trans. Antennas Propag. **60**(1), 36–43 (2012)
26. Sorrentino, R., Gatti, R.V., Marcaccioli, L., Mencagli, B.: Electronic steerable mems antennas. In: IEEE First European Conference on Antennas and Propagation, EuCAP, pp. 1–8 (2006)
27. Valkonen, R., Luxey, C., Holopainen, J., Icheln, C., Vainikainen, P.: Frequency-reconfigurable mobile terminal antenna with mems switches. In: IEEE Proceedings of the Fourth European Conference on Antennas and Propagation (EuCAP), pp. 1–5 (2010)
28. Mirzajani, H., Nasiri, M., Ghavifekr, H.B.: A new design of mems-based wideband frequency reconfigurable microstrip patch antenna. In: 8th International Symposium on Mechatronics and its Applications (ISMA), pp. 1–6. IEEE (2012)
29. Mirzaei, H., Eleftheriades, G.V.: A compact frequency-reconfigurable metamaterial-inspired antenna. IEEE Antennas Wirel. Propag. Lett. **10**, 1154–1157 (2011)
30. Choi, J., Lim, S.: Frequency reconfigurable metamaterial resonant antenna. In: IEEE Asia Pacific Microwave Conference (APMC), pp. 798–801 (2009)
31. Zhang, S., Huff, G., Feng, J., Bernhard, J.: A pattern reconfigurable microstrip parasitic array. IEEE Trans. Antennas Propag. **52**(10), 2773–2776 (2004)
32. Yang, S.-L., Luk, K.-M.: Design of a wide-band l-probe patch antenna for pattern reconfiguration or diversity applications. IEEE Trans. Antennas Propag. **54**(2), 433–438 (2006)
33. Bernhard, J.T.: Reconfigurable antennas. Synth. Lect. Antennas **2**(1), 1–66 (2007)
34. Jais, M. Jamlos, F.: 2.45 ghz beam-steering textile antenna for wban application. In: IEEE International Symposium Antennas and Propagation Society, pp. 394–395 (2013)
35. Chen, R.-H., Row, J.-S.: Single-fed microstrip patch antenna with switchable polarization. IEEE Trans. Antennas Propag. **56**(4), 922–926 (2008)
36. Hussain, R., Sharawi, M.S.: A reconfigurable dual-band mimo antenna system for mobile terminals. In: Loughborough Antennas and Propagation Conference (LAPC), pp. 110–113 (2014)
37. Hussain, R., Sharawi, M.S.: Planar meandered-f-shaped 4-element reconfigurable multiple-input-multiple-output antenna system with isolation enhancement for cognitive radio platforms. IET Microw Antennas Propag. **10**(1), 45–52 (2016)
38. Hussain, R., Sharawi, M.S.: A compact reconfigurable multi-band mimo antenna system. In: IEEE International Symposium on Antennas and Propagation Society (APSURSI), pp. 480–481 (2014)
39. Hussain, R., Sharawi, M.S.: Planar four-element frequency agile mimo antenna system with chassis mode reconfigurability. Microw. Opt. Technol. Lett. **57**(8), 1933–1938 (2015)
40. Hu, Z., Hall, P., Gardner, P.: Reconfigurable dipole-chassis antennas for small terminal mimo applications. Electron. Lett. **47**(17), 953–955 (2011)
41. Ghasemi, A., Ghahvehchian, N., Mallahzadeh, A., Sheikholvaezin, S.: A reconfigurable printed monopole antenna for mimo application. In: 6th European Conference on Antennas and Propagation (EUCAP), pp. 1–4. IEEE (2012)
42. Jin, Z.-J., Lim, J.-H., Yun, T.-Y.: Frequency reconfigurable multiple-input multiple-output antenna with high isolation. IET Microw. Antennas Propag. **6**(10), 1095–1101 (2012)
43. Grau, A., Lee, M.J., Romeu, J., Jafarkhani, H., Jofre, L., De Flaviis, F.: A multifunctional mems-reconfigurable pixel antenna for narrowband mimo communications. In: IEEE International Symposium on Antennas and Propagation Society, pp. 489–492. IEEE (2007)
44. Qin, P.-Y., Guo, Y.J., Weily, A.R., Liang, C.-H.: A pattern reconfigurable u-slot antenna and its applications in mimo systems. IEEE Trans. Antennas Propag. **60**(2), 516–528 (2012)
45. Kumar, K.K.: A pattern reconfigurable chassis-mode mimo antenna. IEEE Trans. Antennas Propag. **62**(6), 3290–3298 (2014)
46. Lim, I., Lim, S.: Pattern reconfigurable antenna for adaptive multi-input multi-output switching applications. In: IEEE International Symposium on Antennas and Propagation Society, pp. 1–2. IEEE (2012)

47. Piazza, D., Mookiah, P., D'Amico, M., Dandekar, K.R.: Pattern and polarization reconfigurable circular patch for mimo systems. In: Proceedings of the Third European Conference on Antennas and Propagation (EuCAP), pp. 1047–1051. IEEE (2009)

48. Lai, M.I., Jeng, S.-K.: Compact pattern reconfigurable antenna array based on l-shaped slots and pin diodes for adaptive mimo systems. In: IEEE International Symposium on Antennas and Propagation Society, pp. 1–4. IEEE (2008)

49. Nguyen, V.A., Ook, P.S. : Compact switched and reconfigurable 4-ports beam antenna array for mimo applications. In: IEEE MTT-S International Microwave Workshop Intelligent Radio for Future Personal Terminals (IMWS-IRFPT), pp. 1–3. IEEE (2011)

50. Mookiah, P., Piazza, D., Dandekar, K.R.: Reconfigurable spiral antenna array for pattern diversity in wideband mimo communication systems. In: IEEE International Symposium on Antennas and Propagation Society, pp. 1–4. IEEE (2008)

51. Oh, T., Lim, Y.-G., Chae, C.-B., Lee, Y.: Dual-polarization slot antenna with high cross-polarization discrimination for indoor small-cell mimo systems. IEEE Antennas Wirel. Propag. Lett. **14**, 374–377 (2015)

52. Raj, J.S.K., Bonney, J., Herrero, P., Schoebel, J.: A reconfigurable antenna for mimo application. In: Antennas & Propagation Conference: LAPC 2009, Loughborough, pp. 269–272. IEEE (2009)

53. Chattha, H.T., Huang, Y., Boyes, S.J., Zhu, X.: Polarization and pattern diversity-based dual-feed planar inverted-f antenna. IEEE Trans. Antennas Propag. **60**(3), 1532–1539 (2012)

54. Li, Y., Zhang, Z., Zheng, J., Feng, Z.: Dual-polarised monopole-slot co-located mimo antenna for small-volume terminals. Electron. Lett. **47**(23), 1259–1260 (2011)

55. Jin, G., Zhang, D., Li, R.: Optically controlled reconfigurable antenna for cognitive radio applications. Electron. Lett. **47**(17), 948–950 (2011)

56. Tawk, Y., Christodoulou, C.: A new reconfigurable antenna design for cognitive radio. IEEE Antennas Wirel. Propag. Lett. **8**, 1378–1381 (2009)

57. Tawk, Y., Costantine, J., Avery, K., Christodoulou, C.: Implementation of a cognitive radio front-end using rotatable controlled reconfigurable antennas. IEEE Trans. Antennas Propag. **59**(5), 1773–1778 (2011)

58. Ghanem, F., Hall, P., Kelly, J.: Two port frequency reconfigurable antenna for cognitive radios. Electron. Lett. **45**(11), 534–536 (2009)

59. Al-Husseini, M., El-Hajj, A., Tawk, Y., Kabalan, K.Y., Christodoulou, C.G.: A simple dual-port antenna system for cognitive radio applications. In: International Conference on High Performance Computing and Simulation (HPCS), pp. 549–552. IEEE (2010)

60. Hussain, R., Sharawi, M.S.: Integrated reconfigurable multiple-input-multiple-output antenna system with an ultra-wideband sensing antenna for cognitive radio platforms. IET Microw. Antennas Propag. **9**(9), 940–947 (2015)

61. Hussain, R., Sharawi, M.S.: A cognitive radio reconfigurable mimo and sensing antenna system. IEEE Antennas Wirel. Propag. Lett. **14**, 257–260 (2015)

62. Tawk, Y., Costantine, J., Christodoulou, C.: Reconfigurable filtennas and mimo in cognitive radio applications. IEEE Trans. Antennas Propag. **62**(3), 1074–1084 (2014)

63. Akyildiz, I.F., Lee, W.-Y., Vuran, M.C., Mohanty, S.: Next generation/dynamic spectrum access/cognitive radio wireless networks: a survey. Comput. Netw. **50**(13), 2127–2159 (2006)

Cognitive Radio and the New Spectrum Paradigm for 5G

Hua Mu and Teck Hu

Dynamic Spectrum Access and Cognitive Radio

Introduction

In traditional spectrum allocation policy, frequency bands are allocated to fixed licensed users for exclusive access. However, this exclusive spectrum allocation policy could be inefficient when the traffic pattern of licensed users is sparse resulting in unused spectrum resources. During the recent years spectrum demands are soaring due to the increase of wireless devices and applications. Therefore the wireless industry is eager to see solutions to increase spectrum efficiency. Cognitive radio is a spectrum sharing concept which allows unlicensed users to access licensed spectrum in a non-intrusive manner. An illustration of this cognitive spectrum access concept is shown in Fig. 1. In cognitive radio networks, unlicensed users learn the spectrum usage information from the radio environment and try to fill in those spectrum white spaces (frequency bands not occupied by licensed users). This is considered as a dynamic spectrum access paradigm for users can now access unlicensed spectrum dynamically according to real time spectrum usage of licensed users. The success of cognitive radio depends on both technical advancement and flexible spectrum regulation. In this chapter, we provide an extensive review on the current state of cognitive radio techniques as well as possible applications of cognitive radio in 5G wireless communications.

H. Mu (✉)
ECE Department, Florida Institute of Technology, Melbourne, FL, USA
e-mail: hmu@fit.edu

T. Hu (✉)
Cortona Research, Melbourne, FL, USA
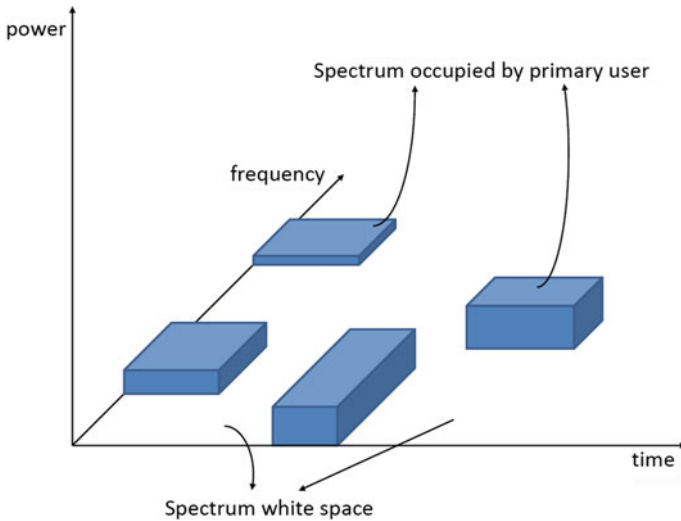e-mail: teck.hu@gmail.com

Fig. 1 Cognitive radio spectrum access concept

To enable cognitive radio techniques, the first step is to find the spectrum white space, which is the spectrum not utilized by any licensed user at certain time within certain geo-location. Also, the unlicensed users must vacate the band whenever a licensed user becomes active. In another word, the cognitive users need to learn the licensed users' spectrum usage information and be constantly aware of the spectrum usage. This can be done either through active spectrum sensing or by consulting with geo-location base database.

Spectrum Sensing

Cognitive users may learn from the radio environment and identify spectrum white space through spectrum sensing. Different spectrum sensing methods are based on different levels of knowledge of licensed users' signal. Energy detector (ED) [1] is the most popular method due to its simplicity. The unlicensed users who perform spectrum sensing don't need to know anything about licensed users' signal characteristics and the implementation of ED is simple. By comparing the output of ED with a predefined threshold, the sensing node makes a decision on whether a licensed user is present in the frequency band. Consider an ED based on n received sample values. Let y_i be the i-th sample value, then the energy detector can be formulated into the following hypothesis test:

$$\frac{1}{n}\sum_{i=1}^{n}|y_i|^2 \underset{H_0}{\overset{H_1}{\gtrless}} T,$$

where H_1 denotes the hypothesis that licensed signal is present while H_0 represents the alternative. While ED is a low-complexity spectrum sensing approach, its performance degrades when the noise variance is unknown or there is interference in the frequency band of interest. In order to improve sensing accuracy, other methods which rely on advanced signal processing such as feature detection may be adopted exploiting the statistics of licensed user's signal such as cyclostationary signatures.

Figure 2 shows a comparison of several spectrum sensing techniques in terms of its accuracy and complexity [2]. As we can see, energy detector has the lowest complexity. However, its sensing accuracy is much lower than other advanced sensing schemes. Waveform-based sensing provides the highest accuracy with a medium complexity. It uses the known pattern such as preamble and/or midamble in licensed users' signal to detect the existence of licensed users.

Recently, it is proposed in some studies that spectrum sensing and spectrum access can be combined together to improve overall spectrum efficiency. Instead of making a hard decision on whether a licensed user's signal is present or not in a certain band, the soft output from ED or other types of spectrum sensing block can be used directly in deciding the probability to access this band as well as the transmit power. Of course, this joint spectrum sensing and access is modeled into an optimization problem that incurs high computational complexity compared with traditional hard-decision spectrum sensing schemes [3–5].

Within a cognitive radio network, spectrum sensing can be performed by individual cognitive devices or jointly by a group of cognitive users. The latter case is usually referred to as cooperative sensing and how to combine sensing results from multiple cognitive users is of great importance. Several common combining techniques are selection combining, majority vote, soft combining, etc.
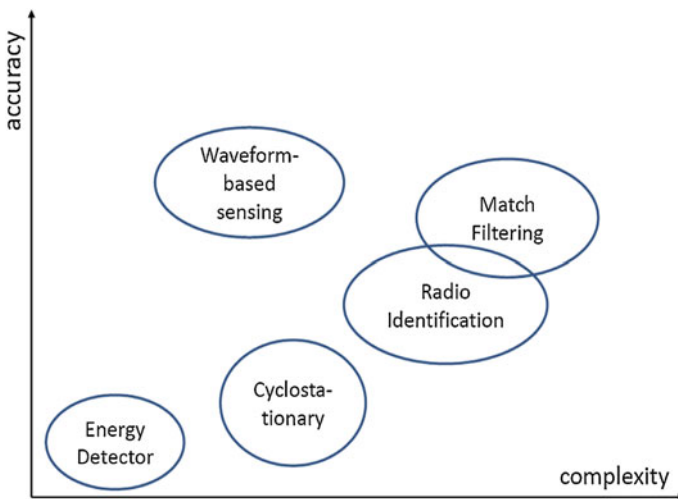

Fig. 2 Comparison of different spectrum sensing schemes

Geo-Location Database

An alternative to spectrum sensing is to let licensed user report their spectrum usage to a geo-location based database (GDB) and let unlicensed users consult with this database to find spectrum white space. This database provides so-called Radio Environment Maps (REM) containing information of spectrum policies and parameters of the network to unlicensed users. This approach has been adopted by TV white space devices [6]. Before any unlicensed devices can access TV channels, they need to consult with this geo-location database to find the parameters they can use, such as channel number and transmit power. This geo-location database contains the spectrum usage of all licensed users (also called incumbents). Therefore, conflicts between licensed and unlicensed users can be avoided.

Dynamic Spectrum Access (DSA)

Once a spectrum white space is found, unlicensed user needs to choose proper base band signal characteristics. Although licensed and unlicensed users will occupy different frequency bands to avoid interference, the out-of-band (OOB) leakage of unlicensed users' transmitted signals is always present and will interfere with licensed users. One way to mitigate this out of band interference involves spectral shaping and is closely related to the underlying air-interface waveform. In the following section, we will cover spectral shaping in a multicarrier setting. Also, from a networking perspective, how to manage dynamic access of multiple users sharing the same spectrum white space has a significant impact on system level performance such as the overall throughput. Therefore we will discuss spectrum access management for cognitive radio networks as well.

Spectral Shaping

After finding the spectrum holes through reliable spectrum sensing or database inquiry, the unlicensed users can then access without conflicting with licensed users' transmission. However, OOB power leakage of unlicensed users' signal must be strictly suppressed to avoid adjacent channel interference to licensed users (and also among unlicensed users themselves). For example, for TV whiter space (TVWS) applications, cognitive devices are only allowed to transmit with adjacent channel leakage ratio (ACLR) no larger than –72.8 dB.

   It is well known that orthogonal frequency division multiplexing (OFDM) is robust to multi-path fading and its implementation is considered low-complexity due to the existing of fast Fourier transform (FFT) algorithm. OFDM has been widely used in modern wireless communication systems including WiFi and LTE. When OFDM is used as the underlying waveform in cognitive radio networks, this

multi-carrier scheme provides much flexibility to use available spectral holes. More importantly, it enables the usage of non-contiguous spectrum.

However, the traditional rectangular-shaped OFDM signals exhibit high magnitude sidelobes, causing considerable interference to adjacent subcarriers. This is due to the use of rectangular pulse shape, whose Fourier transform is a Sinc function with high sidelobes. Therefore, effective sidelobe suppression is important in designing OFDM-based cognitive radio systems. Several sidelobe suppression techniques have been proposed recently. Schaich and Wild [7] presents a modification to traditional OFDM, namely filtered-OFDM, in which a different pulse shape with smooth edges is used to replace rectangular pulse shape. A common choice of this new pulse shape is raised-cosine pulse. This filtering technique can either be subcarrier based or subcarrier group based. Either way, it can bring down the sidelobes and therefore reduces interference among different users sharing the spectrum by using different parts of the OFDM subcarriers. However, filtering usually results in higher peak-to-average power ratio (PAPR), which makes the OFDM symbol subject to non-linear distortion when passing through power amplifiers.

Another approach to mitigate out-of band interference is to use filter bank multicarrier (FBMC) [7] instead of OFDM. It is also referred to as OFDM/OQAM because it uses offset QAM modulation. By carefully designing the underlying prototype filter, FBMC may be able to remove inter-symbol and inter-carrier interference. However, FBMC may introduce even larger PAPR values. Also, unlike filter-OFDM which is complaint to current LTE, FBMC, due to its use of OQAM which is not supported in LTE, would need a redesign if applied in LTE complaint systems.

While all spectral shaping techniques for OFDM have their advantages and disadvantages, at least one spectral shaping should be applied to traditional OFDM to satisfy the ACLR requirement on cognitive radio systems.

Dynamic Spectrum Access Policy

When the available spectrum holes are detected, cognitive radio network needs to decide adaptively on a spectrum access policy, deciding which unlicensed users may share the spectrum white space available for cognitive access. This adaptive spectrum access policy includes various transmission parameters, decided based on environmental conditions and capabilities of the transmitters. Such parameters include the selection of frequency bands, access probabilities, transmit power, etc. Cognitive devices should be equipped with a cognitive engine which has the ability to make a decision on DSA policy and adjust transmit parameters accordingly. There are several ways to determine the DSA policy:

- **Joint Optimization**: DSA policy can be obtained by solving an optimization problem. Such problem can be formulated to maximize the system throughput with some constraints on transmit power, interference level, etc. The parameters

being optimized over may include which frequency band to access; transmit power over each band, etc. This kind of optimization problem is similar to resource allocation problems extensively researched in the literatures and can usually be solved by dual method, together with Lagrangian decomposition. Some examples can be found in [5, 8].

- **Game Theory**: Game theory is another powerful tool in solving complex optimization problems especially when multiple objectives are involved. When multiple unlicensed users try to access a given set of spectrum holes, the DSA problem can be formulated into a potential game, as in [9]. Some recent researches show interests in modelling the spectrum access problem into a congestion game. In congestion game, each unlicensed user is allowed to choose one idle channel for accessing to maximize its own utility [10–12].
- **Machine Learning**: Reinforcement learning can also be applied to DSA problems in cognitive networks. For example, the spectrum usage by primary network can be modeled as a Markov process and Q-learning is used to determine the state transition probabilities.

Cognitive Radio Application and Standardization

The commercial use of cognitive radio technology is mainly focused on TVWS. Cognitive access of TWVS was prototyped in 2007–2008 by companies like Microsoft, Motorola, Philips, Adaptrum and I2R. In Nov. 2008, Federal communications commission (FCC) in the US approved the unlicensed use of TWVS. The FCC regulation requires that unlicensed devices in TVWS consult with geo-location based database which contains the unused TV channels and allowed transmitting power before accessing TV spectrum. This is due to the concern that the existing spectrum sensing techniques are not mature or reliable enough to enable the implementation of cognitive radio networks and won't be able to satisfy the ACLR requirement set by FCC. However, as the spectrum sensing researches going forward, in the future, unlicensed devices may be allowed to perform spectrum sensing and use their own sensing results to facilitate the access to TVWS, which may enable more flexible and dynamic cognitive network design. Also, spectrum sensing can be incorporated by cognitive radio networks as a secondary method in addition to geo-location based database inquiries to monitor the quality of potential channels and enhance spectrum sharing among multiple secondary users.

In UK, the regulatory organization, the office of communications (Ofcom) also proposed to allow unlicensed use of TV white space for cognitive devices. Its Digital Dividend Review Statement, released in Dec. 2007 promotes interleaved spectrum usage for cognitive devices [13]. In 2009, Ofcom suggests a sensitivity level for sensing-only devices of −120 dBm for digital TV and −126 dBm for wireless microphones. Ofcom is also more focused on GDB approach due to the concern that low-cost sensing hardware may not provide enough accuracy [13].

The European Communications Committee (ECC) started project SE43 on white space and cognitive issues in May, 2009. This project addressed the protections of broadcast services as well as the definition of the requirements for GDB approach and assessment of the potential available spectrum for unlicensed devices utilizing spectrum white space. In report ECC159 and subsequent reports ECC185 and ECC186, it is pointed out that GDB is the favorable approach compared to spectrum sensing.

Several major standardization efforts on DSA for cognitive radio are outlined in the following:

- **IEEE Dynamic Spectrum Access Networks Standards Committee (DySPAN-SC)** focuses on DSA to improve spectrum efficiency, including the management of radio transmission interference and coordination of wireless technologies including network management and information sharing amongst networks deploying different wireless technologies. DySpan-SC also includes an ad hoc group on DSA in vehicular environment.
- **The European telecommunications standards institute (ETSI) technical committee** on reconfigurable radio systems works on system level solutions related to software defined radio (SDR) and cognitive radio. Their four working groups address the issues of system aspects, radio equipment architecture, cognitive management and control and public safety. The aim is to exploit the capabilities of reconfigurable radio and networks for self-adaption to a dynamic, ever-changing environment to improve supply chain, equipment and spectrum utilization.
- **IEEE 802.22** is developed for wireless regional area networks (WRANs) and is the first wireless standard based on cognitive radios. It is designed to provide wireless broadband access in large areas with radius up to 100 km. This standard is based on time division duplex (TDD) OFDM with bandwidth options of 6, 7 and 8 MHz for different TV channel bandwidths in different countries. Channel aggregation can also be applied to enhance the throughput by combining multiple TV channels to support the same user.
- Besides IEEE802.22, **IEEE 80.211af** task group aims to modify the PHY and MAC layer of 802.11 to standardize a WiFi technology using TVWS with coverage of a few km. As a wireless local area network (LAN) standard, the communication range is up to 1 km. Its PHY is based on OFDM and MAC protocol is based on CSMA/CA. IEEE802.11af is also referred to as WhiteFi or Super WiFi. Some early deployments have been conducted by companies and universities such as Microsoft, Google and Rice University mainly use GDBs to obtain available TV bands.
- **IEEE 802.19** is the wireless coexistence technical advisory group within the IEEE 802 LAN/MAN committee. It focuses on the coexistence of wireless networks accessing unlicensed spectrum. While multiple wireless devices accessing the same unlicensed spectrum in the same location, they cause interference to each other and measures should be taken to ensure that the coexisting networks are all functional.

- **ECMA-392** is a standard released in June 2012. It specified a physical layer and MAC layer for cognitive devices operating in TVWS, such as high speed video streaming and internet access on personal portable electronics, home electronics equipment and computers and peripherals. The PHY of ECMA-392 is also based on OFDM while its MAC layer supports prioritized contention access (PCA) and channel reservation access (CRA).

Cognitive Radio in Beyond 4G Cellular Communications

Although the current standardization and application of cognitive radio have focused on TV bands, it is believed that cognitive radio can be incorporated into 5G to improve spectrum efficiency in order to meet the challenging requirement in 5G such as huge capacity, massive connectivity high reliability and low latency. CR concept can be used in two ways. First, CR can be used to enable the coexistence of multiple networks using licensed shared access (LSA) or dynamic spectrum leasing. Second, with the deployment of massive MIMO and cell densification, cognitive radio concept can be used to mitigate interference by dynamically adjusting transmission parameters in time, frequency and spatial domains.

Spectrum and Channel Model

Introduction

The millimeter Wave (mmWave) by definition spans 30–300 GHz, with wavelength of 1 cm to 1 mm. At these frequencies, it is expected that much larger carrier bandwidth is available to enable the support of peak rates ranging in the range of 10-20 Gbps which is much higher than those that are supported in the current 4G wireless systems. The identification of the specific mmWave bands is the responsibility of the international regulatory body ITU-R through its World Radio Conference (WRC). Within ITU, ITU-R WP5D defines radio requirements and scenarios for evaluations needed to certify a new radio access as part of the IMT-family. Together, these two processes have the important implication that any IMT band can be used to deploy any new radio access that has been certified as part of the IMT (International Mobile Telecommunications) family. Hence, the acceptance of new mmWave bands by ITU-R has global consequences on the radio access that can be developed and eventually deployed in these new bands.

Due to its shorter range, mmWave is well suited for indoor short range high speed transmission as an alternative to HDM. For outdoor applications, wireless local loop enabling the last hop connection to homes and offices provides a much cheaper options to network operators compared to high cost of fiber connection installations or in some instances, existing final connections are monopolized by the incumbent cable providers.

Another promising outdoor application for mmWave is for cell densification, which is a key technique to improve spectrum efficiency due to its spectrum reuse. The benefit of using mmWave for cell densification is two-fold: first it reduces the cost of cell deployment. Smaller cells or relays can be plug in whenever needed and removed later. Secondly, due to the directional signal propagation property of the mmWave, interference can be significantly reduced.

Regulatory and Standardization

At present, the existing spectrum for mobile communications is all below 6 GHz. International Mobile Telecommunications (IMT) spectrum below 6 GHz has been approved in the recently WRC-15 meeting and it consists of 8 subbands, shown in Table 1 below, with bandwidth ranging from 20 MHz up to 315 MHz. Approximately a total of 1200 MHz of spectrum is allocated and identified for IMT Mobile use, comprises of both TDD and FDD duplexing modes. It should be noted that in these bands, more spectrum is allocated as FDD than for TDD, on the region and countries.

Among these bands, 92 MHz bandwidth from 698 to 790 MHz, 1427–1518 MHz, and 3400–3600 MHz are new global allocations. From Table 1, the spectrum below as allocated by WRC-15 can be seen to be rather fragmented.

As mentioned in the previous section, due to potential of availability of wider bandwidths, the mm-wave bands above 30 GHz hold promise for providing high peak data rates in specific areas where traffic demands are very high. Various regulatory administrations have started investigation and consideration of potential new bands for 5G. Given the need for more bandwidth, these investigations generally have been directed towards opportunities in the 6–100 GHz frequency range. In the same WRC-15 meeting, some frequency bands have been agreed to be studied for WRC-19 considerations. They are between 24.25 GHz and 86 GHz. These frequency bands are listed below in Table 2.

Table 1 Spectrum allocation after WRC-15

Band (MHz)	Spectrum amount (MHz)
450–470	20
698–960	262
1427–1518	91
1710–2025	315
2110–2200	90
2300–2400	100
2500–2690	190
3400–3600	200
Total	1268

Table 2 Candidate bands for WRC-19

| Band (GHz)	Spectrum amount (GHz)
24.25–27.5	3.25
37–40.5	3.5
42.5–43.5	1
45.547–47.047	1.5
50.4–52.6	2.2
66–76	10
81–86	5

Even in the presence of potentially large spectrum at these mmWave bands, there is still a clear need for tools that would help address the ever increasing demand for spectrum. In general, spectrum sharing is needed when the refarming or re-purposing of a specific spectrum is not possible. Other sharing scenarios would be the sharing needed within the unlicensed bands. Mobile operators are increasingly looking at unlicensed spectrum as an important spectrum resource. For non mmWave bands, large bandwidth up to 200 MHz of unlicensed spectrum is available in the 5150–5350 MHz band in many countries around the world. For mmWave, wide bands of unlicensed spectrum are available around the 60 GHz spectrum.

Channel Model

Channel model is crucial to any air interface design and more so for the high frequency bands at mmWave since the deployments of narrow beamforming and large bandwidth in these bands are especially vulnerable to dynamic blockage, spatial dependency, multi-path resolvability, and severe path and penetration loss. There are various efforts worldwide targeting 5G channel measurements and modeling such as the METIS202 and in 3GPP, to name just two among the many. Various existing channel models are applicable only for below 6 GHz and there have been effort to extend these models to higher frequencies. In 3GPP, studies on channel model for frequency above 6 GHz have produced various channel models. Key features and requirements of channel models at mmWave is the additional flexibility and characteristics to account for the large bandwidth and penetration loss and blockage, in addition to providing spatial consistency in the delay and angular domain.

To illustrate the spatial consistency aspect, Fig. 3 below shows the measurement results on spatial consistency in the 73 GHz band [14]. A strong correlation is observed in both delay and angular domain when moving along the predefined routes. Such characteristic is crucial towards the deployments of advanced techniques such as beam forming. Hence, the channel models therefore need to also provide additional flexibility for the elevation dimension, thereby allowing the considerations of two dimensional antenna systems.

Fig. 3 Measurements on spatial consistency in 73 GHz band
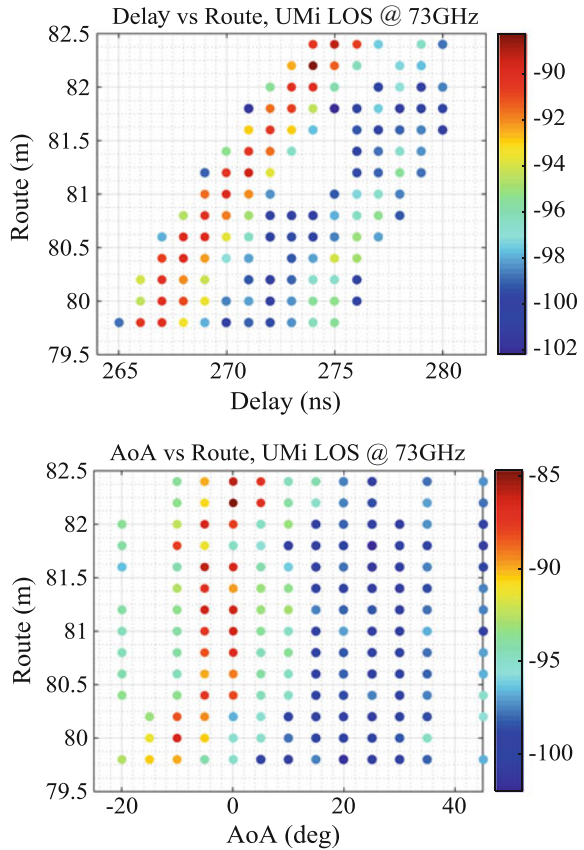
To briefly summarize, the new channel model supporting frequency bands in the mmWave should have the following capabilities:

- Ability to incorporate 2D antenna arrays with azimuth and elevation angles of departure and arrival of the multipath components.
- Support of carrier frequency up to 100 GHz and the support of channel bandwidth up 2 GHz.
- Support of different array type such as linear, planar, cylindrical and spherical arrays, with arbitrary polarization.

An example of such channel model is the Clustered Delay Line (CDL) model that is reported in 3GPP Technical Report 3GPP TR 38.900. The CDL model is applicable to frequency ranging from 0.5 GHz to 100 GHz with a maximum bandwidth of 2 GHz. The rms delay spread values of the CDL model are

normalized to 1 s, and the strongest cluster is normalized to 0 dB. Before using the models, the delays should be scaled according to the desired delay spread.

$$\tau_{n,\,scaled} = \tau_{n,\,norm} DS_{desired}$$

where

$\tau_{n,\,norm}$ is the original normalized delay value of the nth cluster (or nth tap in the TDL)

$\tau_{n,\,scaled}$ is the new delay value of the nth cluster/tap

$DS_{desired}$ is the wanted delay spread

An example of how the scaling parameters are selected is provided in Table 3 where the values have been chosen such that the RMS delay spreads span the range observed in measurements corresponding to the typical deployment and evaluation scenarios

Different CDL models have been reported for different channel profiles such as LOS and NLOS scenarios. Table 4 below reports one such CDL model for NLOS that is denoted as CDL-A in TR 38.900. The CDL models the channel by clusters where each cluster is parameterized by its normalized delay values and its corresponding power, Azimuth angle of Departure (AoD), Azimuth angle of Arrival (AoA), Zenith angle of Departure (ZoD) and Zenith angle of Arrival (ZoA). Furthermore, for these clusters its corresponding angular spreads (ASA, ASD, ZSA, ZSD) are provided together with the cross polarization power ratios (XPR) for each ray of each cluster. Greater details on how this channel model is generated can be found in the same technical report (TR 38.900).

Another important aspect of high frequency operations is the understanding and modelling of the RF impairments, which are highly dependent on the carrier frequency. Measured statistics of phase noise [11] in four different bands are given in Table 5, to illustrate the significant variations and differences of the delay spread and Angle of Arrival (AoA) as a function of the measurement route. In addition to phase noise, the non-linearity in the PA also varies with the carrier frequency and the RF system fabrication materials [15]. PA non-linearity impacts the required power back-off and the respective energy efficiency as well as the spectrum leakage and block error rate in the mmWave bands.

Model	$DS_{desired}$ (ns)
Very short delay spread	10
Short delay spread	30
Nominal delay spread	100
Long delay spread	300
Very long delay spread	1000

Table 3 Examples of scaling parameters

Table 4 CDL Parameters (CDL-A as reported in 3GPP TR 38.900)

Clusters						
Cluster	Delay	Power	AoD	AoA	ZoD	ZoA
#	s	dB	°	°	°	°
1	0.0000	−13.4	−178.1	51.3	50.2	125.4
2	0.3819	0	−4.2	−152.7	93.2	91.3
3	0.4025	−2.2	−4.2	−152.7	93.2	91.3
4	0.5868	−4	−4.2	−152.7	93.2	91.3
5	0.4610	−6	90.2	76.6	122	94
6	0.5375	−8.2	90.2	76.6	122	94
7	0.6708	−9.9	90.2	76.6	122	94
8	0.5750	−10.5	121.5	−1.8	150.2	47.1
9	0.7618	−7.5	−81.7	−41.9	55.2	56
10	1.5375	−15.9	158.4	94.2	26.4	30.1
11	1.8978	−6.6	−83	51.9	126.4	58.8
12	2.2242	−16.7	134.8	−115.9	171.6	26
13	2.1718	−12.4	−153	26.6	151.4	49.2
14	2.4942	−15.2	−172	76.6	157.2	143.1
15	2.5119	−10.8	−129.9	−7	47.2	117.4
16	3.0582	−11.3	−136	−23	40.4	122.7
17	4.0810	−12.7	165.4	−47.2	43.3	123.2
18	4.4579	−16.2	148.4	110.4	161.8	32.6
19	4.5695	−18.3	132.7	144.5	10.8	27.2
20	4.7966	−18.9	−118.6	155.3	16.7	15.2
21	5.0066	−16.6	−154.1	102	171.7	146
22	5.3043	−19	126.5	−151.8	22.7	150.7
23	9.6586	−33.1	−56.2	55.2	144.9	156.1
Per-cluster parameters						
Parameter		ASD	ASA	ZSD	ZSA	XPR
Unit		°	°	°	°	dB
Value		5	11	3	3	10

Table 5 Measured phase noise in different bands (dBc/Hz)

Frequency (GHz)	Material	100 Hz	100 kHz	1 MHz	10 MHz
6	GaAs	−67	−103	−123	−124
30	GaAs	−55	−83	−112	−116
62	Si CMOS	−44	−68	−97	−113
72.6	GaAs	−34	−89	−114	−128

Enabling Technologies

At the mmWave bands, due to the high path-loss and its susceptibility to blockage, coverage and mobility support are the main challenges. On the other hand, the large amount of bandwidth available provides the possibility for very high data rate transmission. As an example, up to 10 GHz bandwidth is available around 70 GHz, achieving a trial peak data rate of 70 Gbps [16].

One the most important enabling technology for mmWave is the application of Massive Multiple Input and Multiple Output (m-MIMO). m-MIMO is the common terminology to refer to the use of large number of antenna elements where the term massive typically referred to as the use of number of transmit and receive antennas that are greater than 8. m-MIMO under the certain spatially diverse channels have to deploy to provide high spectral efficiency transmission, while in some deployment conditions can be used to increase the reliability by transmitting the same data stream over multiple antenna elements to exploit the diversity gain. With beamforming, narrow beam is formed to reduce the interference and hence increase the received signal to noise and interference ratio. This is important from the coverage consideration since at these mmWave bands, coverage is limited due to the high path loss. In summary, it is essential for mmWave bands to apply proper beamforming scheme(s) to combat large path-loss and track users in mobility.

Even with m-MIMO, the challenges with coverage and mobility persists at these mmWave bands. One solution is the combined use of low frequency bands and high frequency bands to enable coverage and mobility support as well as supporting the very high data rate transmission. The lower frequencies are used to carry the control signalling where coverage is crucial and the data channels are carried in the higher bands that also provide the additional bandwidth for high peak rates support.

At these mmWave bands, to control the implementation complexity, the design on MIMO and beamforming needs to be adapted from the full digital pre-coding as currently being used. Instead of full digital precoding, adaptive analog beamforming need to be used in conjunction with digital precoding. An implementation of combined adaptive analog beamforming with digital beamforming, also commonly referred to as *Hybrid Beamforming*, is significantly less complex than digital pre-coding as it primarily relies on simple phase shifters.

Different array structures have been proposed and studied extensively in the literature and even verified in some trials [16]. An example of a hybrid beamforming using phased array is illustrated in Fig. 4 where arrays of phase shifters are used for the analog part of the beamforming [14].

To fully utilize the spatial multiplexing gains and the array gains of massive MIMO, knowledge of channel state information at the transmitter is essential. In TDD systems, the channel state information (CSI) can be obtained by exploiting the channel reciprocity using sounding. In FDD system, the CSI has to be obtained through UE measurement and reporting. Compared with traditional MIMO systems, the issue of channel acquisition is much more challenging in massive MIMO
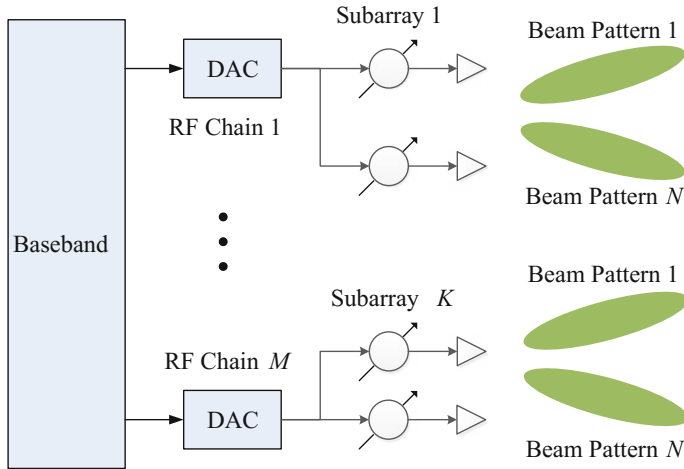
Fig. 4 Hybrid beamforming with phased array

systems due to the higher channel dimension. The accuracy of the CSI therefore significantly influences how much benefits can be really achieved from large number of antennas.

As the beam is much narrower with m-MIMO through its beamforming operation, the multi-point coordination of a distributed m-MIMO becomes feasible through its beam-directed coordination. By such transmission, UEs in anywhere of the cell, even for those cell edge UEs, can be served by multiple directed beams from different transmission points (i.e. Base stations).

License Shared Access

License shared access (LSA) is a concept for spectrum sharing under an exclusive license regime. It is a spectrum sharing approach serving mid to long-term industrial needs and is hopefully a transition stage towards complete dynamic spectrum sharing. In traditional communication systems, fixed spectrum allocation assigns spectrum to licensees. In 5G, in order to handle the spectrum scarcity problem, new spectrum is typically made available through spectrum refarming, such as the refarming of TV broadband. It has also been widely proposed that mmWave can be used as a whole new spectrum for 5G. While the spectrum allocation in mmWave band could incorporate many new concepts, the traditional wide area cell applications still need spectrum resources below 6 GHz. LSA could be a solution to utilize below 6 GHz spectrum in a more efficient way. It allows commercial access to spectrum that cannot be cleared from incumbent users at least in the near-term. Those incumbent users are typically government or other commercial systems with

sparse spectrum usage. LSA introduces spectrum sharing without removing any licensed users through reframing and grants QoS guarantee for both licensed and unlicensed users.

LSA was first proposed by European Telecommunications Standards Institute Reconfigurable Radio Systems (ETSI RRS) Technical Standardization Committee to promote spectrum sharing in 5G [15, 17]. Its spectrum allocation period will be in the order of multiple years. As the standard evolves, the allocation period will shorten and eventually fit the dynamic spectrum access model defined by IEEE DySPAN.

A two-tier LSA framework includes an incumbent (spectrum owner) and an LSA licensee (cellular operator, for example). The difference between LSA framework and cognitive systems such as TVWS is that LSA provides QoS guarantees for LSA licensees while in cognitive radio networks the secondary users access the spectrum completely opportunistically and are served without any QoS guarantee. A three-tie LSA framework has also been proposed to include general authorized access (GAA) users, which has the lowest priority in accessing the spectrum and receives no QoS guarantee.

Use Cases

LSA is especially beneficial for Mobile Network Operator (MNO) to increase load capacity during peak hours in certain area. MNOs can off load part of the traffic to some other spectrum using LSA license. Also, as the emerging of Internet of Things (IoT) devices, the burst traffic pattern from such devices can also be served by LSA schemes. Some of the possible use cases of LSA are presented as follows:

MNOs: Although MNOs may have sufficient capacity most of the time using their licensed spectrum, during peak hours congestion may occur and MNOs can utilize LSA scheme on a secondary basis to add additional capacity. By accessing an incumbent's spectrum, MNOs can offload some of its traffic while still providing QoS guarantee for their users. The MNOs' access to incumbent's spectrum is exclusive and guaranteed with certain QoS within an agreed geographic area, time duration and frequency range. Also, some of the internet traffics carried by MNOs are more suitable for LSA transmission. Some non real-time traffic, such as advertising information can wait until some spectrum holes are found in incumbents' spectrum since the delivery of such data is not sensitive to delays. Since MNOs get LSA licensed for a relative long time (several years), they can develop a business model with revenue estimation and infrastructure investment analysis.

On the other hand, MNOs may only need to access a portion of incumbent's spectrum resources during peak times and within a particular area of higher than normal user density. Such overflowed traffic can be well supported by spectrum holes within incumbents' licensed spectrum using LSA architecture.

Integrating LSA into LTE is relatively easy. Because LTE's resource management is based on resource blocks (RBs) which contains 12 subcarriers and each subcarrier has a bandwidth of 15 kHz. Even if the available LSA spectrum is not continuous, they can be divided into RBs and assigned to one or more LTE users. This is especially useful when LTE network experiences a sudden increase in user number but each user requires a relatively small bandwidth. Even when users request bandwidth consuming services such video-streaming, carrier aggregation can be used to combing LTE bands with LSA bands to support them. Therefore, incorporating LSA into LTE only requires minor modification to the existing LTE system.

WiFi: WiFi networks can also be provided with an LSA license to utilize incumbent's spectrum within a small area. With the growth of WLAN devices and traffic, some WiFi networks may experience overflow occasionally. By providing WiFi networks with LSA license, WiFi traffic can be off-loaded to some spectrum white space to better serve the WiFi devices and provide them with more consistent connection quality.

Internet of Things: IoT is an emerging trend to enable the smart city and smart planet concept. Many wearable devices, industrial equipments and vehicles are all connected wirelessly. These wireless devices generate short, burst, low power traffic which can be well served by spectrum holes in incumbents' spectrum. Therefore, LSA may be a good solution for the connection of IoT devices.

2-Tie LSA Architecture

The 2-tie LSA structure proposed by ETSI is shown in Fig. 5. The target frequency is the 2.3-2.4 GHz band in Europe, currently used by cordless cameras, SAP/SAB portable video links, telemetry, radiolocation and defense systems.

This is a two-tie structure consisting of incumbents, LSA licensee, a repository, an LSA controller and a Network Operations Administration and Maintenance (QA&M) block. The roles and functions of the blocks in this LSA architecture are listed as follows:

- **LSA Repository**: This is a database containing the relevant information on spectrum usage by the incumbents in the spatial, frequency and time domains. The incumbent users will provide updated spectrum resource usage information, as well as power level information to this repository. To ensure the confidentiality of incumbent users' data, this LSA repository will be country-specific and under the purview of the National Regulation Authority (NRA). Also, it is recommended that LSA repository be directly managed by a trusted third parity to protect incumbent users' privacy and guarantee fairness in spectrum management.
- **LSA Controller**: The LSA controller computes LSA spectrum availability in the spatial, frequency and time domains based on LSA rules and spectrum usage
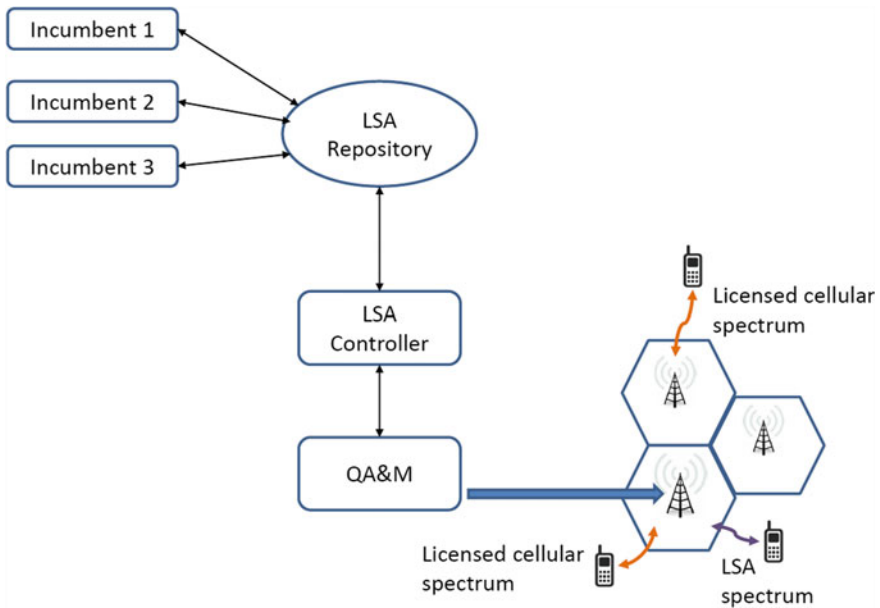
Fig. 5 LSA architecture from ETSI [15]

information from LSA repository. The repository will notify the controller of changes to the sharing agreement or the creation of such agreement. An LSA controller may be interfacing with more than one LSA repository and more than on LSA licensee's network. It is also recommended that LSA controller to be managed a trusted third party.

- **OA&M**: This block is in control of the operation of LSA licensee network. It actually manages the LSA licensed spectrum. It uses spectrum availability information obtained from LSA controller and gives commend to LSA licensee network. It provides LSA licensees with transmission parameters and makes sure they comply with the sharing agreement. Use the information from network QA&M, LSA licensee base station (BS) enables user equipment (UE) to access certain LSA spectrum subject to QoS requirements, data rates or data plans.

Now let's discuss how an LSA licensee, e.g. mobile BS, access or vacate the LSA spectrum. A mobile network will access LSA spectrum at the appropriate time indicated by LSA controller. The mobile network's QA&M block will instruct the relevant BS to enable the transmission in LSA frequency band. Then existing load-balancing algorithm in the mobile network will make use of the newly available spectrum resources and allow some devices to access the new band if needed. Several techniques are available to transfer a device to LSA spectrum. These techniques can also be combined.

- **Reselection procedure**: User equipment in idle mode migrates into the coverage area of LSA spectrum may utilize this spectrum.
- **Inter-frequency handover procedures**: The mobile network initiates handover to transfer users operating in underlying band to LSA band.
- **Carrier aggregation procedures**: To enable user equipment to operate on both underlying band and LSA band through carrier aggregation.

When the LSA access time period granted by the LSA controller expires or when an incumbent switches on and requires it spectrum back, the existing load-balancing algorithm in mobile network needs to vacant the LSA spectrum and transfer users back to underlying bands, through one or more of the following techniques:

- **Reselection procedure**: Users in idle mode will start vacating the LSA spectrum by reselection of frequency bands.
- **Inter-frequency handover procedure**: Handover is initialized for users operating in LSA band back to underlying band.
- **Carrier Aggregation procedure**: The mobile network will reconfigure appropriate users to stop using LSA bands and use underlying bands only.

3-Tie LSA Architecture

While the 2-tie model is originally proposed in Europe, US PCAST report and FCC's spectrum sharing plan for small cells in 3.5 GHz band promotes a 3-tie model consisting of incumbent, LSA licensee and general authorized access licensee. This 3-tie model is also referred to as priority access. The main focus in US for this model is 3.5 GHz band and potentially 1.7/2.1 GHz bands as well.

The get-all-around (GAA) devices in this 3-tie model usually operate at lower power as unlicensed or "licensed by rule" devices and they have the lowest priority compared to incumbent and LSA licensee. Different from incumbent and LSA use, GAA use may not be within a controlled network and mahy operate similar to WiFi access point. GAA licensees may consult with the central database in order to select frequency bands opportunistically without conflicting with incumbents and LSA licensees and decide transmit parameters accordingly. However, GAA users receive no QoS guarantee.

Current State of LSA

Currently LSA architecture is mainly proposed and advocated in US and EU. When introducing LSA concept to current networks, a main concern is the compatibility,

i.e. current wireless devices can use LSA spectrum with minor modification. Therefore, the most sought after LSA spectrum should be what is currently supported by existing UE models. For example, in some regions, only a subset of LTE channels are implemented and licensed. Then those unused LTE channels become target LSA spectrum.

US is currently planning to implement LSA/PA within 3.5 GHz band (between 3550 to 3650 MHz) according to FCC [18]. This band is currently occupied by radiolocation and aeronautical radio navigation services under the rules G110 and US245 [19]. These are currently government exclusive allocations and only secondary non-government radiolocation services are additionally permitted. The use of this 3.5 GHz band is limited under current condition. Turning it into LSA spectrum for LTE networks would only incur minimal regulatory and commercial effort. The EU is currently looking at deploying LSA using 2.3 GHz band, according to ECC 205 [20]. Both 3.5 GHz and 2.3 GHz bands coincide very well with LTE bands and could be the primary choice of LSA spectrum to augment existing LTE spectrum portfolios.

LSA to Dynamic Spectrum Access

Currently proposed LSA schemes are mainly for mid to long-term spectrum regulation. The incumbents own the spectrum while LSA licensees may access without causing harmful interference to incumbents. As the evolution to 5G, incumbents may lease their spectrum to several LSA systems for smaller time period, in smaller regions and may divide their spectrum into more bands and lease to different LSA networks. Therefore, current mid to long-term spectrum leasing will evolve to short term leasing and eventually turn into dynamic spectrum access scheme for multiple LSA networks. The evolution of LSA may be summarized into four stages:

- Macro-cell over long-term LSA
- Small-cell over long/mid-term LSA
- Dynamic and pro-active LSA
- LSA for everything

In summary, LSA is a new complementary regulatory framework and provides flexible spectrum allocation to secondary networks based on location, frequency and time. Unlike cognitive radio systems, the LSA licensees in LSA framework also receive QoS guarantees, making it suitable for applications and devices of cellular systems. The 3-tie model supports additional GAA users who access opportunistically and receive no QoS guarantee.

Conclusions

In this chapter, we discussed several aspects of new spectrum paradigm for 5G wireless systems. The aim is to increase spectrum efficiency to support the soaring demand from increasing numbers of wireless devices and applications. The key ideas include improve spectrum efficiency by allowing spectrum reuse and introducing new spectrum. The cognitive radio concept and LSA architecture fall into the first category while the mmWave spectrum is a new member of 5G wireless spectrum portfolio.

Despite the advancing in various technologies, the operators and regulatory bodied play dominant roles for commercializing cognitive radio and LSA networks. The business model and economic impact should be carefully analyzed. The success of mmWave largely depends on the modeling and understanding the physical channels and building transceivers that can function in this high frequency band.

References

1. Tandra, R., Sahai, A.: Fundamental limits on detection in low SNR under noise uncertainty. In: Proceedings of the International Conference on Wireless Networks, Communications and Mobile Computing, pp. 464–469 (2005)
2. Yucek, T., Arslan, H.: A survey of spectrum sensing algorithms for cognitive radio applications. IEEE Commun. Surv. Tutor. **11**(1), 116–130 (2009)
3. Asghari, V., Aissa, S.: Adaptive rate and power transmission in spectrum-sensing system. IEEE Trans. Wirel. Commun. **9**(10), 3272–3280 (2010)
4. Barbarossa, S., Sardellitti, S., Scutari, G.: Joint optimization of detection thresholds and power allocation for opportunistic access in multicarrier cognitive radio networks. In: Proceedings of the 3rd IEEE International Workshop Computational Advances in Multi-sensor Adaptation Processing (CAMSAP), Aruba, Dutch Antilles (2009)
5. Hua, Mu, Tugnait, Jitendra K.: Joint Soft-decision cooperative spectrum sensing and power control in multiband cognitive radios. IEEE Trans. Signal Process. **60**(10), 5334–5346 (2012)
6. Fitch, M., Nekovee, M., Kawade, S., Briggs, K., MacKenzie, R.: wireless service provision in TV white space with cognitive radio technology: a telecom operator's perspective and experience. IEEE Commun. Mag. **49**(3), 64–73 (2011)
7. Schaich, F., Wild, T.: Waveform contenders for 5G: OFDM vs. FBMC vs. UFMC. In: 2014 6th International Symposium Communications, Control and Signal Processing (ISCCSP), pp. 457–460 (2014)
8. Choi, K., Hossain, E., Kim, D.: Downlink subchannel and power allocation in multi-cell OFDMA cognitive radio networks. IEEE Trans. Wirel. Commun. **10**(7), 2259–2271 (2011)
9. Nie, N., Comaniciu, C.: Adaptive channel allocation spectrum etiquette for cognitive radio networks. In: Proceedings of the 1st IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, (DySPAN 2005), pp. 269–278 (2005)
10. Liu, M., Wu, Y.: Spectrum sharing as congestion games. In: Proceedings of the IEEE Allerton (2008)
11. Cheng, N., Zhang, N., Lu, N., Shen, X., Mark, J.: Opportunistic spectrum access for cr-vanets: a game theoretic approach. IEEE Trans. Veh. Technol. **63**(1), 237–251 (2013)
12. Law, L., Huang, J., Liu, M.: Price of anarchy for congestion games in cognitive radio networks. IEEE Trans. Wirel. Commun. **11**, 3778–3787 (2012)

13. Ofcom: "Digital Dividend Review": a statement on our approach towarding the digital dividend (2007)
14. GPP, R1-162170, Discussions on the frequency bands above 6 GHz. RAN1#84bis (2016)
15. ETSI: Electromagnetic compatibility and Radio spectrum Matters (ERM); Mobile broadband services in the 2300 MHz –2400 MHz frequency band under Licensed Shared Access regime —System Reference Document (2012)
16. Deutsche Telekom: Deutsche Telekom and Huawei show record-breaking over the air speed (2016). https://www.telekom.com/media/company/301824
17. European Telecommunications Standards Institute: ETSI TS 103 154: Reconfigurable Radio Systems (RRS); System requirements for operation of Mobile Broadband Systems in the 2300–2400 MHz band under Licensed Shared Access (LSA) regime (2014)
18. FCC: Amendment of the Commission's Rules with Regard to Commercial Operations in the 3550–3650 MHz Band (2014)
19. FCC: Online Table of Frequency Allocations, 47 C.F.R. § 2.106 (2016)
20. ECC 205: Licensed Shared Access (LSA), CEPT Working Group Frequency Management (2014)