Signals and Communication Technology

Maria-Gabriella Di Benedetto
Faouzi Bader *Editors*

Cognitive Communication and Cooperative HetNet Coexistence

Selected Advances on Spectrum Sensing, Learning, and Security Approaches

Springer

Signals and Communication Technology

For further volumes:
http://www.springer.com/series/4748

Maria-Gabriella Di Benedetto
Faouzi Bader
Editors

Cognitive Communication and Cooperative HetNet Coexistence

Selected Advances on Spectrum Sensing, Learning, and Security Approaches

Springer

Editors
Maria-Gabriella Di Benedetto
Rome
Italy

Faouzi Bader
Rennes
France

ISSN 1860-4862 ISSN 1860-4870 (electronic)
ISBN 978-3-319-01401-2 ISBN 978-3-319-01402-9 (eBook)
DOI 10.1007/978-3-319-01402-9
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013950672

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The aim of this book, the first of two volumes, is to present selected research that has been undertaken under COST Action IC0902 "Cognitive Radio and Networking for Cooperative Coexistence of Heterogeneous Wireless Networks" (http://newyork.ing.uniroma1.it/IC0902/). COST (European Cooperation in Science and Technology) is one of the longest-running European frameworks supporting cooperation among scientists and researchers across Europe.

Written by leading researchers from both academia and major industrial research laboratories, this book will provide electrical, R&D electronic and communication engineers, as well as researchers, undergraduate, graduate, and postgraduate students with a unique and comprehensive overview of recent advances in cognitive radio (CR) and networks.

The book will allow the reader to have access to avant-garde results, in the CR and cooperative heterogeneous coexistence field, in a thorough and detailed way.

Topics covered by the book include newly developed techniques on spectrum sensing algorithms for cognitive white-space systems using either single or multiple antennas, and novel blind free-bands detectors exploiting cyclic autocorrelation function sparsity. The reader will also explore learning and optimization techniques and mechanisms based on biomimetic approaches for Self-Organization in Macro–Femtocell Coexistence; these methods are based on the recently developed "Docition" concept for cognitive networks. The book further includes extensive discussions regarding issues related to the implementation of the different steps of the cognitive cycle, and, in particular, on reasoning. Application of such concepts to cognitive networks allows defining a potential architecture for their integration into current TCP/IP networks. Indeed, using the cognitive paradigm represents a way toward (i) addressing the multiple timescales of operation of a network and (ii) gaining additional information on the cause–effect relationships between network configuration and performance. A very interesting chapter that describes how cognitive networking can be implemented to support green network operation, proposing a test case and demonstrating its potential in a 3G cellular context, is also included in the book. A complete description of the latest groundbreaking field trials is also available to the reader along with the current and future market requirements and so-called killer applications. Finally, in the last chapter, main conclusions and recommendations regarding the available test beds are reported.

A brief description of each of the chapters is as follows.

Chapter 1, by A. Mariani et al., focuses on "Recent Advances on Wideband Spectrum Sensing for Cognitive Radio". Spectrum sensing plays a fundamental role in cognitive radio networks allowing the discovery of spectrum opportunities and enabling primary user protection. Many spectrum sensing techniques have been proposed in the literature, from the most popular algorithms such as energy detection and feature detectors, to most advanced techniques such as eigenvalue-based strategies and cooperative approaches. Most of these techniques have been conceived to assess the occupancy of primary users within a single frequency band. Improved knowledge of the surrounding radio environment can be reached exploiting wideband spectrum sensing, which consists in a joint observation of multiple frequency bands and joint detection on the occupancy of each sub-band. In this chapter, a novel methodology for wideband spectrum sensing based on an information theoretic approach is provided. This technique does not require the setting of a decision threshold, which is a problem for many spectrum sensing algorithms due to the dependence on unknown parameters or difficulties in the statistical description of the decision metrics. A general formulation of the problem is also provided in this chapter, which is valid for any kind of spectral representation, and then specifies the analysis for the case in which simple DFT is used. Performance analysis is provided to the reader and is based both on numerical simulations and laboratory tests using SDR implementation.

Complexity of spectrum activity, and its effect on opportunistic access, is the topic addressed by I. Macaluso et. al., in Chap. 2. The authors analyse the relationship between the number of observed channels, the duty cycle (DC), the complexity of each channel activity, and the performance of a learning-enhanced technique for opportunistic spectrum access. The findings show that the probability of finding a free channel among a group of observed ones strongly depends on the DC and the complexity of the channel activity. Moreover, it is shown that if a cognitive radio is able to select the channels with best characteristics, reducing the number of observed channels has little effect on the performance of the learning algorithm. Hence, a pre-processing phase that a cognitive radio can use to focus on a subset of channels is introduced, which results in a more effective spectrum exploitation. In particular, the authors show that a cognitive radio system can use the DC and the Lempel–Ziv complexity to characterize spectrum usage and select a subset of channels yielding to best performance of the learning technique. The reader will notice that the introduction of the pre-processing stage, which is akin to the orientation phase of the cognitive cycle reduces the number of observed channels, thus, impacting the amount of resources devoted to the sensing stage.

Chapter 3, by S. Dikmese et al., explores the effect of deviating from the simplistic signal model commonly utilized in modelling energy detection in spectrum sensing. The first part of the chapter briefly introduces basic energy detection models, as well as the concept of wideband, multi-mode spectrum sensing using Fast Fourier Transform (FFT) and analysis filter bank for spectrum analysis. The second part of the chapter examines the effects of various forms of frequency dependency in energy detection. First, the effect of a non-ideal

frequency response of the sensing receiver is analyzed and related modifications to optimize energy detection performance are developed. Second, the effects of various forms of frequency dependency in energy detection are analyzed. Next, the authors address the case where the sensing window in the time-frequency plane includes both zones where the PU signal is present and zones where it is absent, leading to mixed decision statistics. The case of non-flat primary user spectrum is addressed in the last part of this chapter, and includes the effect of non-flat transmitted spectra and channel frequency selectivity.

Chapter 4, by D. Riviello et al., focuses on spectrum sensing algorithms and techniques for secondary users operating in digital TV white-spaces. Here, the primary signal conforms to the ETSI DVB-T or DVB-T2 standard, and consists in a continuous sequence of OFDM symbols. Since no further assumptions are made on primary signal characteristics, results obtained for DVB-T(2) can be straightforwardly extended to any standard using OFDM modulation. Note that both single-antenna and multi-antenna techniques are considered in this chapter; performance is assessed and compared against simple sensing techniques such as energy detection.

Z. Khalaf and J. Palicot propose in Chap. 5 a new blind free-band detector exploiting the cyclic autocorrelation function (CAF) sparsity. They first show that the CAF of a linearly modulated signal is a sparse function in the cyclic frequency domain. Using this property, the authors propose a new CAF estimator that uses compressed sensing with Orthogonal Matching Pursuit (OMP). This new proposed estimator outperforms classic estimators operating under similar conditions. Several cases are analysed with and without using filtering at the transmission and reception sides. The impact of a propagation channel is also considered in the analysis. Using this new CAF estimator, in the second part of this chapter, the performance of two blind free-bands detectors are analysed and compared. The first is a soft version of the proposed CAF, where it is assumed that two estimated CAF of two successive packets of samples have close cyclic frequencies. The second estimator uses the symmetry property of the Second Order Cyclic Autocorrelation.

Chapter 6, by A. Georgakopoulos et al., addresses novel concepts on intersystem coexistence and cooperation through cognitive control channels. Cognitive management systems have been proposed as an extension of the ETSI/RRS functional architecture in order to enable the coordination of the network elements with the operators infrastructure. In the functional architecture, two management systems are presented, namely: (i) the Cognitive Management System for the Coordination of the Infrastructure (CSCI), which is responsible for the detection of situations where an Opportunistic Network (ON) would be useful (prior to the formation of the ON) and (ii) the Cognitive System for the Management of the Opportunistic Network (CMON), which is responsible for the creation, maintenance and termination of a given ON based on the context and policy information provided by the CSCI. Both systems are separate functional blocks of the functional architecture and interact with other components via pre-specified interfaces. For the cooperation of CSCIs and CMONs, specific mechanisms need to be defined in order to increase the accuracy of obtained knowledge on the context of

the operational environment. Therefore, this chapter focuses on the definition of required interfaces that are introduced in order to enable communication between the cognitive management systems and also to the related groups of information that is needed to be conveyed between these systems in order to ensure proper interaction. Furthermore, high-level evaluation of the load associated with the aforementioned groups of information is provided based on analytical models.

Chapter 7, by S. Nallgonda et al., covers cooperative spectrum sensing with censoring of cognitive radios, focusing on the case of Rayleigh fading under majority logic fusion. The chapter starts on a comparative analysis of the performance of cooperative spectrum sensing (CSS) under different hard decision fusion rules such as AND logic, OR logic and Majority logic. Also soft information fusion such as maximal ratio combining (MRC)-based fusion is considered, if the sensing channel is subject to fading and shadowing, including Rayleigh/Rician fading and Lognormal shadowing. The impact of fading and shadowing on the performance of CSS is analyzed in terms of missed detection, false detection and total error probabilities. Complementary receiver operating characteristics (RoC) are also discussed under several scenarios of fading and data fusion. The second part of the chapter focuses on the performance of a CSS scheme with censoring of CRs, based on the quality of the radio link in the R-channel. In this chapter, two censoring schemes are investigated: (i) rank-based, where a subset of SUs, associated with best estimated channel coefficients, are selected out of all available SUs and (ii) threshold-based, where an SU is selected if its R-channel amplitude is above some pre-selected threshold level. The performance of these two censoring-based CSS schemes is studied in the presence of Rayleigh and Nakagami-m fading in the R-channels, considering majority logic or MRC fusion rules at FC. Note that the studies presented in this chapter are useful in designing energy-constrained cognitive radio networks.

Chapter 8, by M.-G. Di Benedetto et al., focuses on the medium access control (MAC) for distributed cognitive ultra wide band (UWB) networks. It is claimed that the MAC should be specifically conceived around the impulsive characteristics of the UWB radio signal, and as such foresee and eventually optimize strategies for power sharing and management. MAC functions that can benefit from specific UWB features are discussed, and sensing in the presence of potential narrowband coexisting networks is analyzed. Interference models for impulsive communication systems are also taken into consideration and reviewed, with the aim of defining packet error rate. In particular, the pulse collision model is compared against traditional interference models based on the standard Gaussian approximation. Performance evaluation using the proposed models incorporate theoretical and simulation analyses of MAC properties that are peculiar to impulsive communications.

The notion of Radio Environmental Maps (REMs) has become very prominent in the cognitive radio and dynamic spectrum access field. REMs can be seen as databases or knowledge bases storing different kinds of radio environmental information, such as locations of transmitters and receivers, models of the propagation environment and various spatial/temporal spectrum usage measurements.

L. Gavrilovska et al., address this topic in Chap. 9. An extended study on the "Integration of Heterogeneous Spectrum Sensing Devices Towards Accurate REM Construction" is reported. The reader is introduced to a recently developed generic REM construction architecture capable of integrating heterogeneous spectrum sensing devices by combining the spectrum sensing and the database approach for accurate radio environmental mapping with a specific focus on the device calibration procedure as a quintessential part of the integration process. Theoretical and practical aspects related to the realization of a REM prototype are discussed, as well as on-the-field experimental results obtained with different hardware.

Chapter 10, by A. Galindo-Serrano et al., further addresses the Radio Environment Map topic. The main issue here is to introduce an automatic and remote self-optimization process based on geo-location information exploitation for cellular coverage optimization. Specifically, the REM is used for cellular network coverage hole detection purposes. This coverage hole detection approach drastically reduces the required drive tests and enhances the network with self-responsive capabilities to handle key obstacles toward cellular networks autonomy. In this chapter, the REM is handled in a more general Cognitive Radio (CR) context than TV white-spaces, and it is considered as a mean to represent spatio-temporal characteristics of the radio environment by using concepts and tools from spatial statistics, like point processes, spatial random fields, pair correlation functions, point interaction models, spatial interpolation techniques, etc. Coverage may in fact be the most important and highest-priority target for cellular operators.

Chapter 11, by A. Imran and L. Giupponi, presents the use of several bio inspired approaches, called biomimetics, for the design of Self-Organization (SO) in heterogeneous network scenarios, and in particular the use of learning, game theory and optimization as Biomimetic approaches for SO in Macro–Femtocell Coexistence. Mainly, these approaches are further categorized in indirect and direct biomimetics. Under the concept of indirect biomimetics, the authors discuss in detail the emerging paradigms in learning theory that have been recently shown to have strong potential for designing SO solutions in heterogeneous networks such as Start-up Docition, IQ-Driven Docition, Performance-Driven Docition and Perfect Docition. The authors further investigate a rather under explored paradigm of direct biomimetic. Building on case studies of self-organizing systems in nature, the authors extracte the generic SO design principles, that can be used as a direct biomimetic approach for designing distributed, scalable and agile solutions, to many problems in complex heterogamous networks.

Chapter 12, by H. Bogucka and M. Parzy, focuses on a practical perspective of cooperation and competition for spectrum sharing in cognitive radio networks. In this chapter, practical issues of cooperation among cognitive radio nodes competing for available resources in decentralized network are considered. It is pondered how the theory of competition and cooperation (game theory) meet the practice, by discussing the quantitative metrics of the cost of avoiding cooperation (the Price of Anarchy—PoA), of having limited knowledge of the competitors (the Price of Ignorance—PoI), and of limited time for learning the network environment (the Price of Impatience—PoIm) in dynamically changing radio channels.

This chapter provides the description of practical approaches to spectrum sharing and allocation, which make use of limited, i.e., intentionally reduced information, that the CR nodes have to exchange.

Chapter 13, by L. Abarca et al., consists in a synopsis on security threats in cognitive radio networks (CRNs), specially focusing on the Primary User Emulation (PUE) attack. The chapter includes details on how the location of a transmission source can be a valuable tool to detect PUE attacks, whenever the position of true primary transmitters is known, as it is the case for TV towers in IEEE 802.22 networks. Due to its wireless nature, CRNs inherit most of the threats already reported in the literature in the context of wireless networks. However, the flexibility and reconfigurability capabilities of these networks not only may make conventional attacks easier but also expose them to new risks. In the first part of this chapter, the reader has an outline on main threats to cognitive radio networks, providing a classification of the current threats and detailing the new specific security challenges. The second part describes the behaviour of the PUE attack, with its different implementations and variants. The third part describes a novel method able to detect PUE attacks, based on a cooperative location technique.

Chapter 14, by F. Granelli et al., addresses "Cognition as a Tool for Green Next Generation Networks", with a focus on reasoning. Network performance is a multifaceted concept, including simple measures such as throughput as well as user-level QoS, and a recently added parameter to the equation, i.e., the power consumption. The need for identifying suitable methodologies to optimize performance from the above viewpoints, including energy saving, is driving interest of research toward the emergence of the "green networks". Green networking represents an appropriate scenario where cognition and adaptation are required. How cognitive networking can be implemented to support green network operation is discussed in this chapter, also based on a test case that demonstrates its potentials in 3G cellular contexts.

Chapter 15, by K. Katzis et al., addresses the topic of no less importance "Testbeds and Implementation Issues". Efficient design of CR engines requires the capability of experimentally verify the proposed solutions, and the identification of engine components and of corresponding implementation choices is a fundamental step toward this direction. Within this context, this chapter aims at presenting testbeds and related implementation issues, including CR engine architecture regarding its software and hardware components and available technologies, available platforms and finally implementation issues of CR engines related to standardization.

We would like to express our sincere thanks to all the contributors for their contribution and enthusiasm in participating to the creation of this book. The encouragement and technical support provided by Springer have been crucial to the realization of this book project. The support of COST throughout the IC0902 Action is gratefully acknowledged.

We hope that the book will provide the reader with comprehensive treatises of salient aspects in the field of "Cognitive Communications and Cooperative HetNet Coexistence". We also hope that the book will motivate the research community, especially in young researchers, toward solving the several issues that are left open for future research.

Sapienza University of Rome, Italy — Maria-Gabriella Di Benedetto
SUPELEC, France — Faouzi Bader

Acknowledgments

We would like to place on record our sense of gratitude to the European Cooperation in Science and Technology (COST) framework allowing the coordination of nationally funded research on a European level for their unconditional support and encouragement during the realization of this book on "Cognitive Communications and Cooperative HetNet Coexistence".

We would also like to thank the leaders of the working groups and the special interest groups of COST Action IC902 for their help in the careful selection of the different topics forming this book.

In producing this book, we were fortunate to receive the assistance of many colleagues, students and friends who were extremely generous with their time. We would like to express our deep appreciation to all of them, and to all who helped in careful reviewing of the chapter proofs, a hard task highly requiring time and patience.

We are thankful to all members of COST Action IC902 for their enthusiasm in participating in all training events and organized schools. The technical content of this book is in part thanks to their efforts in participating in these events.

Contents

Chapter 1
Recent Advances on Wideband Spectrum Sensing for Cognitive Radio

Andrea Mariani, Andrea Giorgetti and Marco Chiani

Abstract Spectrum sensing plays a fundamental role in cognitive radio (CR) networks allowing to discover spectrum opportunities and enabling primary user (PU) protection. However, it represents also one of its most challenging aspects due to the requirement of performing radio environment analysis in a short observation time and the fact that its performance can be strongly affected by harsh channel conditions and lack of knowledge about the PU characteristics. In literature, many techniques have been proposed, starting from the most popular algorithms, such as energy detection, to the most advanced, such as, e.g., eigenvalue based detection and cooperative approaches. Most of these techniques have been conceived to assess the occupancy of PUs within a single frequency band. A better knowledge of the surrounding radio environment can be reached exploiting wideband spectrum sensing, that consists in a joint observation of multiple bands and joint detection on the occupancy of each sub-band. Recently, different wideband approaches have been proposed, mainly derived from advanced spectral analysis techniques such as multitaper methods and compressive sensing. In this chapter, we propose a novel methodology for wideband spectrum sensing based on the computation of a frequency domain representation of the received samples and the use of information theoretic criteria (ITC) to identify which frequency components contain PU signals. This technique does not require the setting of a decision threshold, a problem for many spectrum sensing algorithms due to dependence on unknown parameters or difficulties in the statistical description of the decision metrics. We provide a general formulation of the problem, valid for any kind of spectral representation and then focus on the case

A. Mariani (✉) · A. Giorgetti · M. Chiani
Department of Electrical, Electronic and Information Engineering "Guglielmo Marconi" - DEI, Alma Mater Studiorum University of Bologna, Via Venezia 52, 47521 Cesena, FC, Italy
e-mail: a.mariani@unibo.it

A. Giorgetti
e-mail: andrea.giorgetti@unibo.it

M. Chiani
e-mail: marco.chiani@unibo.it

M.-G. Di Benedetto and F. Bader (eds.), *Cognitive Communication and Cooperative HetNet Coexistence*, Signals and Communication Technology,
DOI: 10.1007/978-3-319-01402-9_1,

in which discrete Fourier transform (DFT) is used. This choice is motivated by the simplicity of implementation and the fact that DFT blocks are already available in many wireless systems, such as OFDM receivers. This wideband spectrum sensing approach can be adopted by a single CR node in a standalone manner or within a cooperative sensing scheme. Numerical results show that the algorithm derived for DFT can be also applied as an approximated approach when more accurate frequency representations, such as multitaper method (MTM) spectrum estimates, are adopted. Wideband ITC based sensing can be applied in scenarios in which approaches that require a high level of sparsity of the received signal (such as compressive sensing) can not be adopted.

1.1 Introduction

In the last ten years opportunistic spectrum radio assess strategies have gained an increasing interest both in the academia and industries. This fact has been driven by two aspects: the so called *spectrum scarcity* problem and the attempt to reach a more *efficient utilization* of the spectrum resources. Indeed, in spite of the nominal absence of available spectrum, measurements of the radio frequency occupation indicate that large portions of the licensed bands are not used for significant periods of time [1]. Thus, a more efficient utilization of the spectrum can be reached through the adoption of flexible devices, able to analyze the surrounding radio environment, discover unused spectrum resources and use them without interfering higher priority users, called PUs. These actions describe the essential characteristics of the opportunistic spectrum access (OSA), where users with a lower priority, named secondary users (SUs), "adopt dynamic spectrum access (DSA) techniques to exploit spectral opportunities"[1] [3]. The expression "spectral opportunities" can be generally used to indicate situations in which the SUs have some occasion to transmit. In this work, as in most of the CR literature, a spectral opportunity indicates the presence of a portion of spectrum that is temporarily or locally unused. These unoccupied bands are often referred as *spectrum holes* or *white spaces*.[2]

The OSA techniques have been studied in particular in the context of CR. Recently International Telecommunication Union (ITU) defined a CR system as "a radio system employing technology that allows the system to obtain knowledge of its operational and geographical environment, established policies and its internal state; to dynamically and autonomously adjust its operational parameters and protocols according to its obtained knowledge in order to achieve predefined objectives; and to learn from the results obtained" [5]. Thus we can identify three main key characteristics of CR systems [6]:

[1] The SUs are unlicensed or light-licensed users; in the former case the expression "opportunistic unlicensed access" is often used [2].

[2] The expression "white space" is mainly used with reference to digital television (DTV) bands. It is however accepted as a general term [4].

- capability to obtain knowledge;
- capability to dynamically and autonomously adjust its operational parameters and protocols;
- capability to learn.

Therefore the first step in CR/OSA systems is to implement strategies to acquire information from the radio environment in order to identify ongoing licensed transmissions and preserve them [7]. The main task of this stage consists in identifying which channels are available for opportunistic transmissions, that is equivalent to a PU detection problem. Secondly, it can be useful to acquire some additional information, such as some characteristics of the identified signals, interference measurements, etc.. In literature mainly three solutions have been proposed [2]:

- Geolocation databases;
- Beacon signals;
- Spectrum sensing.

The geolocation database solution is based on the consultation by the SU network of a database that stores the information on the spectral occupancy in the nodes locations and additional information, such as the maximum permitted equivalent isotropic radiated power in each bands. The advantages of this approach are that it is virtually error free and is not affected by radio channel characteristics. However, it is a quite expensive solution. Indeed the secondary nodes require to incorporate some localization technique (e.g. GPS) and Internet connection in order to access the database information. Moreover, additional costs are related to the design, implementation, maintenance and administration of the database, and the costs for gathering the PU occupancy information [8].

The beacon based approach consists in the adoption of a beacon signal that is broadcasted to the secondary nodes providing the PU occupancy information. This solution has a very high infrastructural cost, also requiring some modifications of the current licensed systems. However, some CR networks implementations foresee the adoption of a cognitive pilot channel (CPC) to support cognitive operations such as spectrum allocation [9]. This dedicated channel could be also adopted to convey sensing information towards the SU nodes.

Spectrum sensing (SS) is defined by IEEE as “the act of measuring information indicative of spectrum occupancy” [4]. It consists therefore in the implementation of an autonomous process of the SUs, that on the basis of the received signals analyze the spectrum. It offers the advantage of no infrastructural costs nor modifications in the licensed systems. Moreover, SS makes the SU network completely autonomous and capable of a reactive behaviour. The SU nodes implementation costs depend on the algorithms adopted. The main disadvantages of SS are that its behaviour is generally related to the tradeoff between performance (e.g. detection rate) and observation time and the fact that it can suffer adverse radio channel characteristics, leading to the hidden node phenomenon [7, 10].

The choice of the proper technique to be adopted depends on the particular OSA problem under investigation. In particular, the characteristics of the PUs (such as

their temporal dynamics, bandwidth and power) are the most important features to be considered for choosing the proper strategy. Indeed, the adoption of geolocation database fits particularly with highly predictable PUs such as TV signals, that are continuous transmissions broadcasted from known locations. In this case, to evaluate the presence of a TV signal, it could be sufficient, for example, to enquire a database on a daily basis. SS instead is the most promising solution for unpredictable signals that transmit from unknown locations such as programme making and special events (PMSE) signals (like wireless microphones).

The adoption of CR systems is not limited to the licensed spectrum, but they can also operate in unlicensed bands, in which different networks with the same right to access the spectrum are present [11]. Here the main objective is the coexistence of the CR networks that must share the spectrum resources available in an efficient way. In this context SS plays a fundamental role in supporting high level cognitive functionalities such as interference management.

1.1.1 Sensing in the TV White Spaces

In the context the TV white spaces, Federal Communications Commission (FCC) recently decided to remove the requirement that white space devices (WSDs) should implement SS [12]. This decision came after some studies on the minimum sensitivity required at the secondary nodes to ensure DTV and PMSE signal protection in the worst propagation conditions. These analyses showed that common sensing approaches do not guarantee the detection performance required, leaving the implementation of sensing algorithms as an optional feature [12]. However, this decision seems to be moved more by the willing to come up with a regulation on WSDs in a short time, enabling companies to access the white space market, rather than a definitive mistrust in SS strategies. Indeed, the FCC states that [12]:

> Specifically, we are taking the following actions: While we are eliminating the sensing requirement for TV Bands Devices (TVBDs), we are encouraging continued development of this capability because we believe it holds promise to further improvements in spectrum efficiency in the TV spectrum in the future and will be a vital tool for providing opportunistic access to other spectrum bands.

Then, while eliminating SS as a mandatory function, FCC strongly encourages research activities to make possible a sensing based WSDs future generation. In Europe, on the basis of single user sensing algorithms, the European Communications Committee (ECC) came to the same conclusion also suggesting the potential benefit in using a combination of sensing and geolocation database to provide adequate protection to digital TV receivers [13, 14]. It is worthy to note that most advanced sensing techniques have not been considered in the drawing up of these rules. For instance, it is emblematic the case of the ECC report 159 in which, while assessing the benefits of cooperative sensing strategies, the conclusions are drawn considering single node sensing only [14].

In conclusion, in the next few years we expect the born of the first generation of WSDs that probably will be based on geolocation implementation, being the adoption of SS algorithms not mandatory. In spite of this fact, the research community is motivated anyway to continue the investigation in new sensing techniques with the first aim to propose new algorithms with higher detection performances. The most promising approaches consist in advanced techniques based on cooperation among SUs and multiple antennas WSDs, that currently has not been deeply analyzed by regulatory bodies and that deserve more attention in the definition of future rules. More generally, further efforts in SS research must be motivated by the fact that a geolocation database based secondary system can adopt OSA strategies, but cannot be properly considered a cognitive system, due to the lack in autonomy and reactivity to the environment that characterize the original Mitola's proposal, and that only SS can provide. In addition, beyond PU protection, the sensing task has an important role in supporting higher level cognitive functionalities such as resource allocation and spectrum efficiency [15, 16].

1.2 Overview of Spectrum Sensing Algorithms

In this section we present an overview of the main algorithms proposed for SS. It is not simple to provide a unique classification of the sensing techniques, especially because there are lots of possible approaches and many algorithms can be included in more than one class. Here we choose to adopt a classification based on the detectors' practical requirements, defining the following four groups [17]:

- Fundamental detectors
  We include in this class the basic detectors, typically proposed for the observation of a single band by a single antenna receiver.
- Diversity based sensing
  These detectors require some kind of diversity to be implemented, such as multiple antennas or oversampling. We include in this class the eigenvalue based detection algorithms.
- Cooperative sensing[3]
  These algorithms are based on the adoption of multiple CR nodes.
- Wideband sensing
  We include in this group algorithms that are suited for the analysis of multiple bands observations.

[3] Note that cooperative sensing schemes could be included in the class of the diversity based algorithms, because the adoption of several sensing nodes is essentially a technique for exploiting spatial diversity. However, we separate the class of cooperative algorithms because they have some peculiar characteristics that are not common to other diversity based techniques, such as the selection of the fusion strategy to be adopted, presence of error prone reporting channels, unbalances in the average received power, etc..

In the following we review the main characteristics of the first three classes, while wideband techniques are discussed in the next section.

1.2.1 Fundamental Detectors

Since five years ago, most of the papers on CR introduced sensing asserting that "sensing algorithms can be classified in energy detector, matched filter and cyclostationarity detector". These techniques, indeed, are the very basic strategies that can be adopted in simple sensing problems in presence of single antenna receivers that operate on a single frequency band.

- Energy based detection
  The energy detector (ED) is the most simple and popular algorithm for signal detection. Its implementation consists in an estimate of the received power followed by a comparison with a decision threshold. Theoretically the ED is derived as the generalized likelihood ratio test (GLRT) for the detection of a deterministic unknown signal in additive white Gaussian noise (AWGN) or as a sufficient likelihood ratio (LR) statistic when the signal to be detected is described as a zero mean Gaussian process. Its statistic has been widely studied in literature (see e.g. [7, 18, 19]) and due to its simplicity of implementation and analysis, is currently the standard sensing algorithm adopted, for example, in studies on higher level CR functionalities and by regulatory bodies [13, 20]. Frequency domain EDs have also been proposed [21]. The main impairment of the ED is the fact that its statistic depends on the noise power level, which is required for setting the decision threshold according to the Neyman-Pearson (NP) approach [18]. In practice, noise uncertainty can cause performance losses due to an inaccurate threshold setting and in some cases the presence of the so called SNR wall, which is a minimum SNR level under which it is impossible to rich the desired probability of detection (P_D) and probability of false alarm (P_{FA}) [22, 23]. It has been demonstrated that in practical systems proper design of the noise power estimator allows to counteract the noise uncertainty problem [24]. In particular, the conditions for the avoidance of the SNR wall are related to the statistical properties of the noise power estimator [25].

- Feature based detection
  When some additional knowledge on the signal to be detected is available, it can be adopted signal detection. In particular, the most common algorithms in this class are:

  – Autocorrelation based detectors
    These algorithms can be adopted when the autocorrelation of the signal to be detected presents some peculiar peaks. The most popular autocorrelation based algorithms are the cyclic prefix based algorithms for the detection of orthogonal frequency-division multiplexing (OFDM) signals [26–28].

– Waveform based detector
  If some portions of the primary signal is known, we can build a detector that exploits this knowledge, usually correlating the known feature with the received signal sequence. This can be the case, for example, of signals with known preambles or with some known pilot patterns [29]. The extreme case is the matched filter (MF) detector, that requires the knowledge of the complete signal sequence. Even if the MF is often mentioned in SS algorithm surveys, the assumption of perfect knowledge of the PU sample sequence is unrealistic in practical CR implementations [26, 28].
– Cyclostationarity based detection
  When a signal presents some periodicity in the autocorrelation function, this corresponds to the presence of some correlation in the frequency domain, called cyclostationary feature [30]. As for the autocorrelation features in the time domain, this property can be adopted for detecting PU signals. Many cyclostationary detection algorithms have been proposed in literature, usually based on the estimation of the cyclic autocorrelation function or the cyclic spectrum [30–32].

Feature based algorithms are generally used for detecting the presence of specific PU signals. Being suited for particular communication standards, they are unable to evaluate the presence of different transmissions. The main impairment of these tecniques is the susceptibility to synchronization errors and frequency offsets, that implies the adoption of a synchronization stage [33]. We can consider therefore the feature based detector architecture as a simplified PU signal receiver chain that aims at detecting the presence of the PU transmission, without the need of extracting the information symbols. It is also possible to built general purpose feature detectors, capable, for example, of identifying any possible autocorrelation peak or cyclostationary feature. However these algorithms are very expensive from a computational point of view, time consuming and suffer of synchronization errors [31, 33].

1.2.2 Diversity Based Sensing

In this section we present some algorithms that can be adopted in presence of some diversity reception mechanisms. We refer in particular to multiple antennas systems, that have been widely studied in literature [34]. The same techniques can be also adopted with oversampled signals. In these situations, from the original sample sequence we can extract a set of subsequences which number corresponds to the oversampling factor and use them as they were collected at different antennas [35]. The same algorithms can be also adopted in cooperative sensing systems.[4] In all these cases it is possible to compute the sample covariance matrix (SCM) of the received samples and derive decision tests based on its functions. These algorithms

[4] See footnote 3.

are generally called "eigenvalue based algorithms". Alternative approaches are based on ITC.

- Eigenvalue based detectors
  The eigenvalue based algorithms are binary tests in which the decision metric is a functions of the eigenvalues of the SCM. They have attracted a lot of attention providing good performance results without requiring the knowledge of the noise power nor any prior information on the PU signals [36–40]. Considering the most general scenario, with possible multiple PUs, the GLRT is the so called sphericity test, well known in statistics literature and recently re-proposed for SS with the name of arithmetic-geometric mean ratio test (AGM) [36, 40]. Alternatively, in situations in which we expect to have a single PU, the GLRT is the ratio of maximum eigenvalue to the trace (MET) [39]. Others metrics have been also proposed, such as the maximum to minimum eigenvalues ratio (MME) [37], also addressing the case in which multiple antennas are uncalibrated [40, 41].

- ITC based detectors
  A different approach for the detection of PU signals is to estimate the dimension of the observed sample set. If we receive only noise, the eigenvalues of the covariance matrix of the observed samples are all equal to the noise power σ^2. Otherwise, if some signals are present, some of them are greater than σ^2. Estimating the number of PU signals is thus a model order selection problem, in which the order of the model is the number of eigenvalues of the covariance matrix, estimated by the SCM. The selection problem can be solved by means of ITC [38, 42]. If the estimated model order is greater than zero, it means that at least one PU has been detected [43]. Mainly Akaike information criterion (AIC) and minimum description length (MDL) have been adopted [43]. This approach allows to implement detectors that do not need to set a decision threshold. Note that this implies that we cannot control the tradeoff between false alarm and detection probabilities.

1.2.3 Cooperative Sensing

A very promising solution for improving the sensing performance of the SU networks is to exploit cooperation among secondary nodes. In particular, exploiting the SUs spatial diversity, cooperative strategies can be adopted to counteract channel effects, such as multipath and shadowing, that cause the hidden node problem [7, 10]. Cooperative SS has reached an increasing attention in the last few years, and many different schemes have been proposed. We refer to [44] and the references therein for an extended overview on cooperative techniques and their principal issues. The main requirement in cooperative sensing is related to the availability of channels for signaling among the SUs, that in most of the literature studies consist in fixed control channels.

Cooperative algorithms can be classified on the basis of how SUs share their sensing data and in which point of the network the final decision is taken. We have

mainly two approaches, the centralized and the distributed.[5] Mixed strategies can be also adopted.

- Centralized cooperative sensing
  In centralized cooperative strategies the sensing information from all the SUs is reported to a central identity, called *fusion center*, that takes the global decision. This information is then provided to the cognitive manager of the network that will use it for supporting resource allocation strategies. In some cases the global decision must be sent back to the SUs by means, e.g., of broadcasting [7, 44].
- Distributed cooperative sensing
  Distributed schemes differ from centralized ones for the absence of a specific fusion center. In this case, indeed, the SUs communicate among themselves and converge to a unified decision taken by each SU on the basis of a common policy [44].
- Mixed strategies
  Besides the centralized and distributed approaches, some mixed strategies can be adopted. For example, a relay assisted cooperative scheme can be used in situations in which some SUs experience a weak report channel and the remainders can be used for forwarding their sensing results to the fusion center [44]. Another solution is the clustered sensing scheme, in which cluster-heads act as second level fusion centers, collecting the sensing results from the SUs within their cluster. Then this data can be shared among other cluster-heads or can be forwarded to a global fusion center. An example of cluster based cooperative sensing can be found in [20].

With respect to the information that is shared among the SUs, cooperative strategies can be divided in hard fusion and soft fusion schemes:

- Hard fusion schemes
  When the SUs share their local binary decisions on the presence of PUs, we talk about hard fusion schemes. Locally the SUs can adopt any of the single node sensing techniques described previously. These schemes are convenient for the minimum amount of data that must be exchanged among the secondary nodes. In this case the fusion strategies are typically linear fusion rules such as AND, OR, and majority rules. Also Bayesian approaches can be adopted, such as the Chair-Varshney optimal rule [7].
- Soft fusion schemes
  In place of the local binary decisions, the SUs can share a richer information, such as their likelihood ratios, in order to improve the sensing result. Therefore these schemes generally require a larger amount of data to be shared, mainly depending on the metric chosen and its representation. It has been demonstrated that in many practical situations representing the sensing information with few bits is sufficient for reaching a detection performance equivalent to the unquantized case [45, 46]. If the amount of data to be exchanged is not a problem, algorithms that imply the

[5] Note that in some works the term "distributed" is used as a synonym of cooperative, and expressions such as "non-centralized" are adopted.

transmission of all the SUs' observations to the fusion center have been proposed. In this case eigenvalue based algorithms can be adopted also in the cooperative case [37, 40, 46].

1.3 Wideband Spectrum Sensing: A Review

Most of the SS techniques proposed in literature have been conceived to assess the presence of transmissions within a single frequency band [7]. A better knowledge of the surrounding radio environment can be reached exploiting wideband spectrum sensing, that consists in a joint observation of multiple bands and joint decision on the occupancy of each sub-band. The aim of wideband SS is to distinguish which frequency components contain PU signals from which contain only noise.

Wideband spectrum sensing strategies are generally based on the adoption of some frequency domain representations of the received samples, and thus are related to spectral analysis techniques. The scope of spectral analysis is to provide a reliable estimate of the energy distribution in the frequency domain, and therefore it has a big impact of the environment awareness of the SUs. In CR contexts non parametric techniques are the more suitable strategies because they do not require any assumptions on the received signal.[6] Wideband sensing algorithms are generally constituted by a spectrum estimation stage followed by the adoption of some metric to evaluate the occupancy of each sub-band. The starting point of these techniques is the classical non parametric spectrum estimation theory, based on the periodogram and its derivatives, such as the Welch's method. The most advanced spectrum estimation approach in this context is the multitaper method [47–49]. If the SUs know the power spectral density (PSD) profile of the signal to be detected, the optimum detector in low signal-to-noise ratio (SNR) regimes assumes the structure of an estimator-correlator [50].

The application of wideband SS is primarily related to hardware front-end requirements such as the linearity of analog components and analog-to-digital converters characteristics [51, 52]. To get around such constraints some wideband techniques are based on sequential sensing on multiple bands, frequency sweeping or filter-banks approaches [53–55]. Strategies to reduce hardware complexity have been proposed in the context of compressed sensing, which is a special signal processing technique that can be applied to signals with a sparse representation [56, 57]. In the context of CR, it can be adopted in particular in situations in which the PU signal occupancy is sparse in the frequency domain. The main advantage of this technique is that it allows to analyze a large portion of spectrum without requiring a high sampling rate [57].

Wideband sensing has been also studied in the context of the so called multiband joint detection, that is based on the maximisation of the aggregate opportunistic throughput, a metric that takes into account the trade-off between sensing time and

[6] Generally, the unique assumption is that the received signal samples are taken from a stationary random process.

transmission time in CR systems [58]. In [59, 60] wideband SS has been formalized as a generalized likelihood ratio (GLR) detector, assuming the presence of a given amount of unoccupied spectrum. Alternative wideband approaches are based on ITC, e.g. in [61], where such tools are proposed in a channelized sub-Nyquist scheme. In [62] standard ITC have been adopted to detect the presence of occupied sub-bands using an ED in each sub-band, and a similar approach has been applied to multiband OFDM in [63].

1.4 Wideband Sensing by Model Order Selection

In this section we formulate wideband SS as a model order selection problem solved using ITC. Most of the sensing algorithms proposed in literature are based on the adoption of decision thresholds, which setting is a difficult task in practice due to the dependence on unknown parameters. In particular, considering energy based techniques, including frequency domain analysis, threshold setting depends on the noise power level that must be properly estimated in real implementations [25]. The proposed wideband approach is blind instead, since it does not require the knowledge of the noise power nor any a priori information about the number and the characteristics of the signals present in the observed frequency band.

The proposed wideband sensing technique is based on N observations of a frequency domain vector $\mathbf{x}_i = \left(x_{i,1} \dots x_{i,q} \dots x_{i,N_\mathrm{b}}\right)^\mathrm{T}$, where $i = 1, \dots, N$, and N_b is the number of frequency components considered. We will refer to the elements of $\mathbf{x}_i$ as frequency bins.[7] The problem can be formulated considering a very general approach, in which the vector $\mathbf{x}_i$ can be any kind of frequency domain representation. For instance, it can be a PSD estimate, the output of a filter bank, a compressed sampling reconstruction of the spectrum or, simply, the result of a N_b points DFT.

If PU signals are present in the observed frequency band, we assume that they occupy k^* frequency bins, while the remaining $N_\mathrm{b} - k^*$ contain only noise. Our objective is to identify the occupied k^* bins. In order to accomplish this goal, we formulate wideband SS as a model order selection problem in which k^* is the order of the model [17]. The proposed algorithm estimates k^* and also identifies the occupied bins.

Assuming the radio environment is stationary during the overall sensing period, we collect the N vectors $\mathbf{x}_i$ in the observation matrix

$$\mathbf{Y} = (\mathbf{x}_1 | \dots | \mathbf{x}_i | \dots | \mathbf{x}_N) \,. \tag{1.1}$$

Let us sort all vectors $\mathbf{x}_i$ such that the power levels of the frequency bins are now arranged in decreasing order. We denote with $\widetilde{\mathbf{x}}_i$ and $\widetilde{\mathbf{Y}}$ the ordered vectors and the corresponding ordered matrix, respectively. Thanks to ordering, once the model order is estimated, the frequency bins containing PU signals are the first k^* bins of the vectors $\widetilde{\mathbf{x}}_i$. Thus, after recovering the order of the model, we identify the bins that

[7] This is in accordance to the DFT based scenario studied in the following.

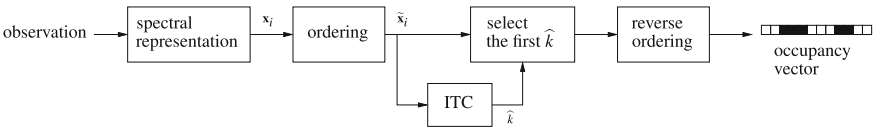

Fig. 1.1 Block diagram of the proposed wideband SS strategy [17]

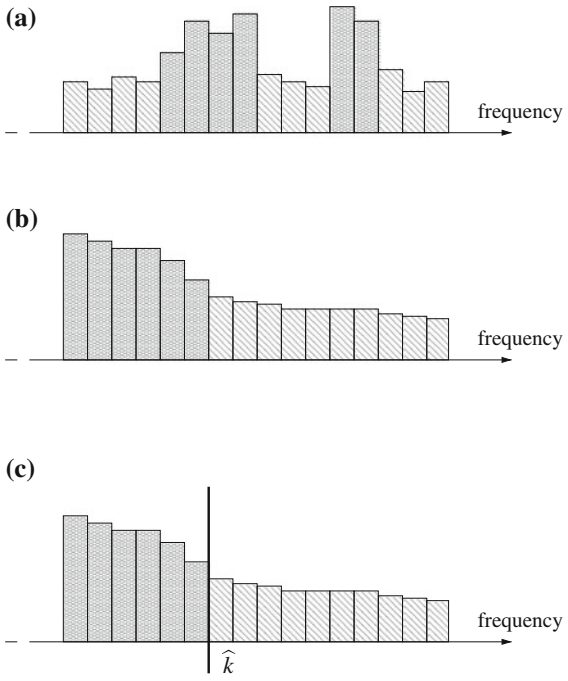

Fig. 1.2 ITC based wideband sensing process [17]. First a frequency representation vector $\mathbf{x}_i$ is collected (**a**) and it is ordered (**b**). Using ITC we obtain $\widehat{k}$ that estimates k^*. Thus the bins that contain signal components are identified as the first $\widehat{k}$ bins of the ordered vector (**c**). **a** Frequency bins vector. **b** Ordering. **c** Selection

contain signal components, and, thanks to a reverse ordering operation, we obtain the occupancy vector, which is a N_b length binary vector in which the q-th element is one if the q-th bin is declared occupied. This wideband sensing process can be represented by the block diagram in Fig. 1.1; the first three steps are depicted in detail in Fig. 1.2. Note that in practical implementations ordering is based on the estimated received power in each frequency bin.

For solving the model order selection problem (i.e. estimating k^*) we adopt ITC, a typical approach used in statistics for choosing the model that better fits data among a family of possible models [64]. In our problem we have N_b possible models, where

the k-th corresponds to the case in which we assume that only $k \in \{0, \ldots, N_b - 1\}$ bins are occupied.[8]

To adopt ITC, we start from the analysis of the log-likelihood of the received observation matrix, $\ln f\left(\widetilde{\mathbf{Y}}|\boldsymbol{\Theta}^*\right)$, where $\boldsymbol{\Theta}^*$ is the vector that contains the unknown parameters of the model, which number depends on k^*.[9] According to ITC, the best choice for estimating k^* is given by

$$\widehat{k} = \arg\min_k \left\{-2 \ln f\left(\widetilde{\mathbf{Y}}|\widehat{\boldsymbol{\Theta}}^{(k)}\right) + \mathscr{P}(k)\right\} \tag{1.2}$$

where $\widehat{\boldsymbol{\Theta}}^{(k)}$ is the vector of the estimated parameters in the k-th hypothesis and $\mathscr{P}(k)$ is the ITC penalty term.[10] Different choices of the penalty term lead to different criteria, each one characterized by different performance and complexity. In the next section we review the most common and simple techniques, adopted throughout this chapter.

The advantage of using model order selection is that it leads to a blind algorithm which does not require any a priori knowledge of parameters, such as the noise power or PU characteristics. In addition, it does not require the setting of thresholds, avoiding problems such as deriving the exact threshold selection rule. The unique assumption of the proposed strategy is that at least one frequency bin contains only noise. This "minimum sparsity requirement" make this method appealing for scenarios in which wideband algorithms that require a high level of sparsity of the received signal (such as compressive sensing) can not be adopted.

1.4.1 Information Theoretic Criteria

In [65] Akaike first proposed an information theoretic criterion for statistical model identification based on the observation of N independent, identically distributed (i.i.d.) samples of the N_b dimensional random variable (r.v.) $\mathbf{X}$, generated by the "true" distribution $f\left(\mathbf{X}|\boldsymbol{\Theta}^*\right)$. The model selection problem consists in identifying the model that better fits data among a set of possible models

$$\left\{f\left(\mathbf{X}|\boldsymbol{\Theta}^{(k)}\right)\right\}_{k \in \mathscr{K}} \tag{1.3}$$

characterized by the model order k. $\mathscr{K}$ is the set of the possible values assumed by k. Akaike proposed to select the model that minimizes the Kullback-Leibler (K-

[8] We will refer to the k-th model also as the k-th hypothesis.

[9] Varying the number of occupied frequency bins we have a different set of parameters that describe the model [64].

[10] Using the notation $\mathscr{P}(k)$ we emphasize that the penalty depends on k through the vector $\widehat{\boldsymbol{\Theta}}^{(k)}$. Note that in general $\mathscr{P}(k)$ could also depend on other parameters, e.g. N_b, N and other functions of the observation.

L) distance from $f\left(\mathbf{X}|\boldsymbol{\Theta}^*\right)$, i.e.

$$\widehat{k} = \arg\min_k \mathbb{E}\left\{\ln \frac{f\left(\mathbf{X}|\boldsymbol{\Theta}^{(k)}\right)}{f\left(\mathbf{X}|\boldsymbol{\Theta}^*\right)}\right\}. \tag{1.4}$$

This criterion is equivalent to minimize the cross entropy

$$-\int f\left(\mathbf{X}|\boldsymbol{\Theta}^*\right) \ln f\left(\mathbf{X}|\boldsymbol{\Theta}^{(k)}\right) d\mathbf{X} \tag{1.5}$$

for which a natural estimate, under the k-th hypothesis, is given by the average log likelihood

$$\frac{1}{N}\sum_{i=1}^{N} \ln f\left(\mathbf{x}_i|\widehat{\boldsymbol{\Theta}}^{(k)}\right). \tag{1.6}$$

Akaike noted that the average log likelihood is a biased estimate of the cross entropy, and added a penalty term that asymptotically, for large N, compensates the estimation error. Exploiting the asymptotical chi squared distribution of the log likelihood, Akaike derived the AIC, in which the penalty term is

$$\mathscr{P}_{\text{AIC}}(k) = 2\,\phi(k) \tag{1.7}$$

where $\phi(k)$ is the number of degrees of freedom in the k-th hypothesis. Alternative ITC can be derived adopting the Bayesian approach, which chooses the model that maximizes the posterior probability $\mathbb{P}\left\{\boldsymbol{\Theta}^{(k)}|\mathbf{X}\right\}$ [66]. In this context, the most popular and simple criterion is the Bayesian information criterion (BIC) with penalty term [66]

$$\mathscr{P}_{\text{BIC}}(k) = \phi(k)\,\log N. \tag{1.8}$$

For large enough samples BIC coincides with the MDL criterion, which attempts to construct a model which permits the shortest description of the data [67]. The AIC and BIC approaches are the most popular ITC adopted in many statistical and engineering problems [38, 42, 62, 68]. Although the AIC metric provides an unbiased estimate of the K-L divergence, in many situations it tends to overestimate the true order of the model, even asymptotically [69]. In some cases, consistency can be reached by properly modelling the penalty term [42, 70]. In particular, when the penalty is in the form $\mathscr{P}(k) = \phi(k) \cdot c$, it can be demonstrated that it is required, for N that goes to infinity, that $c/N \to 0$ to avoid underestimation and $c/\log\log N \to +\infty$ to avoid overestimation [71]. Further conditions can be derived in order to solve specific selection problems [72]. Here we consider three consistent criteria, defined by the penalty terms

$$\mathscr{P}_{\mathrm{CAIC1}}(k) = \phi(k)\ (\log N + 1) \tag{1.9}$$

$$\mathscr{P}_{\mathrm{CAIC2}}(k) = 2\,\phi(k)\ \log N \tag{1.10}$$

$$\mathscr{P}_{\mathrm{CAIC3}}(k) = 3\,\phi(k)\ \log N. \tag{1.11}$$

Note that CAIC1 has been proposed by [69] and CAIC2 has been adopted in [68].

An alternative criterion based on the large sample distribution of maximum likelihood (ML) estimators is the consistent AIC with Fisher information (CAICF), which penalty term is [69]

$$\mathscr{P}_{\mathrm{CAICF}}(k) = \phi(k)\ (\log N + 2) + \log\left|\mathbf{J}\left(\widehat{\boldsymbol{\Theta}}^{(k)}\right)\right| \tag{1.12}$$

where $\mathbf{J}\left(\widehat{\boldsymbol{\Theta}}^{(k)}\right)$ is the estimate of the Fisher information matrix (FIM) of the observation and $|\cdot|$ is the determinant operator.

Note that the formulation of the ITC as in (1.2) supports the interpretation of these techniques as extensions of the ML principle in the form of penalized likelihood. The penalty term is introduced as a cost for the increased complexity of the model, related to the presence of unknown parameters that must be estimated [65, 73]. Thus ITC extend the ML approach in the sense that they take into account both the estimation (of the unknown parameters) and the decision (among the possible models) processes. Note that the ML approach performs poorly in model selection problems, always leading to the choice of the maximum number of parameters considered [66].

1.4.2 DFT Based Wideband Algorithms

In this section we apply the ITC based wideband sensing strategy described in Sect. 1.4 to the case in which simple DFT is used as spectral representation of the received signal. We adopt DFT motivated by its simplicity and by the fact that its implementation can be already available in many systems, such as OFDM receivers. In particular, we consider two practical situations with uncorrelated and correlated frequency bins. The first case exploits only the received energy, while the latter jointly exploits the energy level and spectral correlation to discern PU signals from noise.

At the i-th time instant, the output of the DFT can be expressed as

$$\mathbf{x}_i = \mathbf{s}_i + \mathbf{n}_i \tag{1.13}$$

where $\mathbf{n}_i$ represents the AWGN and $\mathbf{s}_i$ is the aggregation of the PUs signals.[11] We assume that the time domain received sample vector is modeled as zero mean complex Gaussian, that is a common assumption in communications literature.[12] Thanks to

[11] Including the channel effects.

[12] This is a proper assumption for many practical problems, such as the case of OFDM signals, that are widely adopted in recent communication systems.

the linearity of the DFT operation, $\mathbf{x}_i$ is a vector of zero mean complex Gaussian r.v.s with covariance matrix $\boldsymbol{\Sigma}_{\mathbf{x}} = \mathbb{E}\left\{\mathbf{x}_i \mathbf{x}_i{}^{\mathrm{H}}\right\}$. After the collection of N DFT outputs we order of the vector $\mathbf{x}_i$ according to the received power in each bin, i.e., according to the vector $\left(\nu_1, \ldots, \nu_q, \ldots, \nu_{N_\mathrm{b}}\right)$, where $\nu_q = (1/N)\sum_{i=1}^{N} \left|x_{i,q}\right|^2$. Thus we obtain a new vector $\widetilde{\mathbf{x}}_i$ with power in each frequency bin in descending order. Note that the vector $\widetilde{\mathbf{x}}_i$ is zero mean with covariance matrix $\boldsymbol{\Sigma}_{\widetilde{\mathbf{x}}} = \mathbb{E}\left\{\widetilde{\mathbf{x}}_i \widetilde{\mathbf{x}}_i^{\mathrm{H}}\right\}$. If the number of frequency bins containing signals is k, $\boldsymbol{\Sigma}_{\widetilde{\mathbf{x}}}$ can be expressed as

$$\boldsymbol{\Sigma}_{\widetilde{\mathbf{x}}} = \boldsymbol{\Sigma}_{(k)} \bigoplus \sigma^2 \mathbf{I}_{N_\mathrm{b}-k} \tag{1.14}$$

where $\boldsymbol{\Sigma}_{(k)}$ is a $k \times k$ submatrix, $\mathbf{I}_p$ is a $p \times p$ identity matrix, σ^2 is the unknown noise power at each frequency bin, and $\bigoplus$ is the direct sum operator [74]. Note in particular that $\boldsymbol{\Sigma}_{(k)} = \mathbb{E}\left\{\widetilde{\mathbf{x}}_{(k)}\, \widetilde{\mathbf{x}}_{(k)}^{\mathrm{H}}\right\}$, with $\widetilde{\mathbf{x}}_i^{\mathrm{T}} = \left[\widetilde{\mathbf{x}}_{(k),i}^{\mathrm{T}}\ \widetilde{\mathbf{n}}_{(k),i}^{\mathrm{T}}\right]$. Then the log-likelihood function of $\widetilde{\mathbf{Y}}$ can be expressed as

$$\begin{aligned}\ln f\left(\widetilde{\mathbf{Y}}|\boldsymbol{\Theta}^{(k)}\right) = &- N_\mathrm{b} N \ln \pi - N \ln |\boldsymbol{\Sigma}_{(k)}| - N\left(N_\mathrm{b} - k\right) \ln \sigma^2 \\ &- N \mathrm{tr}\left\{\boldsymbol{\Sigma}_{(k)}{}^{-1}\, \mathbf{S}_{(k)}\right\} - \frac{N}{\sigma^2} \mathrm{tr}\left\{\mathbf{N}_{(k)}\right\}\end{aligned} \tag{1.15}$$

where $\mathbf{S}_{(k)} = (1/N)\sum_{i=1}^{N} \widetilde{\mathbf{x}}_{(k),i}\, \widetilde{\mathbf{x}}_{(k),i}^{\mathrm{H}}$ and $\mathbf{N}_{(k)} = (1/N)\sum_{i=1}^{N} \widetilde{\mathbf{n}}_{(k),i}\, \widetilde{\mathbf{n}}_{(k),i}^{\mathrm{H}}$.

1.4.2.1 Independent Frequency Bins

In the case in which the frequency bins are independent, $\boldsymbol{\Sigma}_{(k)}$ is diagonal, and the log-likelihood reduces to

$$\begin{aligned}\ln f\left(\widetilde{\mathbf{Y}}|\boldsymbol{\Theta}^{(k)}\right) = &- N_\mathrm{b} N \ln \pi - N \sum_{q=1}^{k} \ln \sigma_q^2 - N\left(N_\mathrm{b} - k\right) \ln \sigma^2 \\ &- N \sum_{q=1}^{k} \frac{\widehat{\sigma}_q^2}{\sigma_q^2} - \frac{N}{\sigma^2} \mathrm{tr}\{\mathbf{N}_{(k)}\}\end{aligned}$$

where $(\sigma_1^2, \ldots, \sigma_k^2) = \mathrm{diag}\{\boldsymbol{\Sigma}_{(k)}\}$ and $(\widehat{\sigma}_1^2, \ldots, \widehat{\sigma}_k^2) = \mathrm{diag}\{\mathbf{S}_{(k)}\}$. In this case the parameter vector is given by $\boldsymbol{\Theta}^{(k)} = \left(\sigma_1^2, \ldots, \sigma_k^2, \sigma^2\right)$, that can be estimated as $\widehat{\boldsymbol{\Theta}}^{(k)} = \left(\widehat{\sigma}_1^2, \ldots, \widehat{\sigma}_k^2, \widehat{\sigma}^2\right)$, where

$$\widehat{\sigma}^2 = \frac{\mathrm{tr}\{\mathbf{N}_{(k)}\}}{\left(N_\mathrm{b} - k\right)}. \tag{1.16}$$

Then, removing the terms that do not depend on k, the log-likelihood can be expressed as [17]

$$\ln f\left(\widetilde{\mathbf{Y}}|\widehat{\boldsymbol{\Theta}}^{(k)}\right) = -N\sum_{q=1}^{k}\ln\widehat{\sigma}_q^2 - N\left(N_\mathrm{b}-k\right)\ln\widehat{\sigma}^2. \tag{1.17}$$

Note that (1.17) corresponds to the result derived in [62]. For the independent frequency components case the number of degrees of freedom corresponds to the length of $\boldsymbol{\Theta}^{(k)}$, i.e. $\phi(k) = k+1$.

1.4.2.2 Correlated Frequency Bins

In some practical applications, the signals collected present a non negligible spectral correlation (see [75] for some examples). Thus in this section we remove the assumption that the frequency bins are uncorrelated, and we study the most general case, assuming no particular structures for the correlation matrix. In this case the number of degrees of freedom of the model is given by $\phi(k) = k^2+1$, that accounts for the $k\times k$ Hermitian matrix $\boldsymbol{\Sigma}_{(k)}$ and the noise power. Adopting the ML estimates $\widehat{\sigma}^2$ and $\widehat{\boldsymbol{\Sigma}}_{(k)} = (1/N)\sum_{i=1}^{N}\widetilde{\mathbf{x}}_{(k),i}\,\widetilde{\mathbf{x}}_{(k),i}^{\mathrm{H}}$, and removing the terms that do not depend on k, from (1.15) we obtain [17, 76]

$$\ln f\left(\widetilde{\mathbf{Y}}|\widehat{\boldsymbol{\Theta}}^{(k)}\right) = -N\sum_{q=1}^{k}\ln\widehat{\alpha}_q - N\left(N_\mathrm{b}-k\right)\ln\widehat{\sigma}^2 \tag{1.18}$$

where $\widehat{\alpha}_q$ is the q-th eigenvalue of the sample covariance matrix $\mathbf{S}_{(k)}$. In this case the vector of the unknown parameters is given by $\widehat{\boldsymbol{\Theta}}^{(k)} = \left(\widehat{\alpha}_1,\ldots,\widehat{\alpha}_k,\widehat{\sigma}^2\right)$.

1.4.2.3 Performance Metrics

The performance of the wideband approach can be evaluated in terms of probability to correctly detect k^*, $P_\mathrm{k} \triangleq \mathbb{P}\{\widehat{k}=k^*\}$.[13] The probability of incorrect detection can be evaluated in terms of probability of overestimation, $P_\mathrm{over} \triangleq \mathbb{P}\{\widehat{k}>k^*\}$, and the probability of underestimation, $P_\mathrm{under} \triangleq \mathbb{P}\{\widehat{k}<k^*\}$. Note that these performance metrics are very severe metrics; for example, the cases in which $\widehat{k}=k^*+1$ and $\widehat{k}=k^*+10$ are both considered overestimation events, irrespective of the actual distance from k^*.[14]

[13] Numerical simulations show that the difference between P_k and the probability of correctly identifying the set of occupied frequency bins is very small, which means that when the algorithms correctly estimate k^* they generally correctly estimate also the occupied set. See [17] for some numerical examples.

[14] Note that in some practical applications the adoption of algorithms that tend to overestimate k^* may be used by means of including a protection margin to preserve low SNR PU transmissions.

For some practical cases we will also analyze the probability of detection related to the q-th bin, P_{D}^{q}.

1.5 Cooperative Wideband Spectrum Sensing

The wideband sensing strategy described in the previous section can be also implemented in a cooperative context in which several CR nodes share their occupancy vectors to reach a global decision. We denote as

$$\mathbf{d}^{(j)} = \left(d_1^{(j)}, \ldots, d_q^{(j)}, \ldots, d_{N_{\mathrm{b}}}^{(j)}\right)^{\mathrm{T}} \tag{1.19}$$

the occupancy vector of the j-th SU. We assume a centralized approach in which a K out of M rule is applied to assess the presence of PUs in each frequency bin. To focus on the performance evaluation of the proposed sensing strategies, we assume that an error free separate reporting channel is used for sending the local CR decisions to the fusion center (FC). The q-th element of global occupancy vector, $\mathbf{d}^{\mathrm{F}}$, is given by

$$d_q^{\mathrm{F}} = \begin{cases} 1, & \sum_{j=1}^{M} d_q^{(j)} \geq K \\ 0, & \sum_{j=1}^{M} d_q^{(j)} < K. \end{cases} \tag{1.20}$$

The choice of the parameter K determines the specific voting rule. Choosing $K = 1$ we implement the OR strategy, which in general allows higher probability of detection. This approach is the most protective toward PUs, but leads to higher false alarm probabilities. The approach that minimize the number of false alarm events is the AND rule, that can be obtained with $K = M$. However, the AND rule allows the secondary network to declare a band occupied only when all the nodes agree upon the presence of PUs, and thus it performs poorly in presence of harsh channel conditions. In order to reach a good trade-off between false alarm and detection probabilities, intermediate values of K can be chosen, such as $K = M/2$ that leads to the so called majority rule, which in some contexts minimize the total error probability [77].

The probability that the q-th each bin is declared occupied is given by

$$P_{\mathrm{D}}^{q,\mathrm{F}} = \mathbb{P}\left\{\sum_{j=1}^{M} d_q^{(j)} \geq K\right\} = \sum_{h=K}^{M} \mathbb{P}\left\{\sum_{j=1}^{M} d_q^{(j)} = h\right\} \tag{1.21}$$

and can be derived from the single node probabilities of detection P_{D}^{q}, with $q = 1, \ldots, N_{\mathrm{b}}$, and the distribution of a Poisson binomial r.v..

1.6 Numerical Results and Discussion

In this section we show some numerical examples to assess the performance of the proposed wideband sensing technique. We compare the ITC algorithms to simple energy based approaches in which the estimated received power of each frequency bin is compared to the decision threshold. We indicate with ED the ideal ED, which assumes that the noise power is known exactly, and with ED_{k} the estimated noise power (ENP)-ED described in [25] in which the $N_{\mathrm{b}} - k^*$ bins with lower received power are used for estimating the noise power.[15] The threshold is set according to the Neyman-Pearson criterion considering a probability of false alarm $P_{\mathrm{FA}} = 0.01$.

1.6.1 Independent Frequency Bins

In this section we consider the case of independent frequency bins, that is the most interesting case in practice due to the fact that most of communication signals are generated by white data sequences, which give spectral uncorrelated transmissions [78].

In Fig. 1.3 we consider the proposed wideband SS strategy using DFT with $N_{\mathrm{b}} = 128$, in presence of a single white Gaussian signal, that occupy exactly 64 bins, and AWGN. The number of DFT outputs considered is $N = 1000$. We can see that, increasing the SNR, all the consistent ITC present a step wise behaviour, assessing the correct detection probability to a fixed value for high SNR. The AIC instead confirms its non consistent behaviour. The corresponding probabilities of incorrect estimation are shown in Fig. 1.4. We can see that at high SNR P_{under} goes to zero and an incorrect detection always consists in a false alarm event. Note that this property is very important in CR scenarios, because it implies that ITC never misdetect the presence of PUs if the SNR is sufficiently high. Considering the ED based approaches we can see that they perform quite poorly providing almost 50 % of overestimations for high SNR levels. In Fig. 1.5 we perform the same analysis of Fig. 1.3 in a Rayleigh fading scenario. Here we consider frequency-flat fading on all frequency bins. With respect to the AWGN case we can see that fading has a big impact on the performance of the wideband algorithms, increasing the SNR value at which they reach a target probability of correct detection.

From the previous analysis it emerges that CAICF, CAIC2, and CAIC3 are the ITC algorithms that allows a better sensing performance allowing almost 100 % probabilities of correct detection of the occupied bins set. Compared to simple ED strategies, the proposed wideband ITC algorithms allow a more accurate identification of the occupied bands.

[15] For simplicity, here we do not use the exact distribution of the ordered vector. Thus the ENP-ED approach adopted can be considered as an approximated strategy valid for large samples use cases.

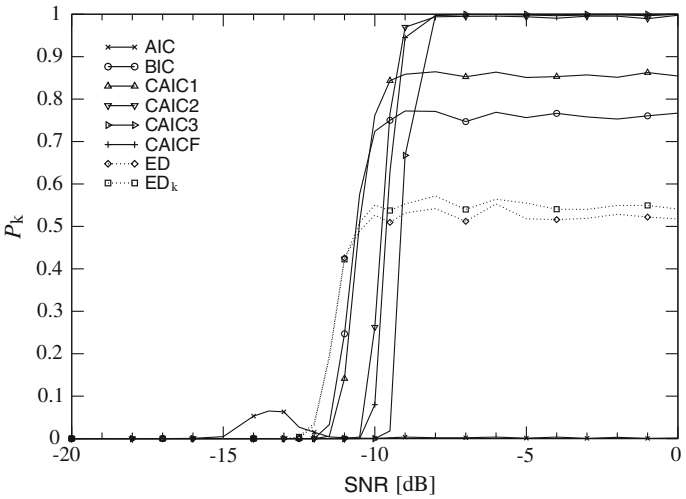

Fig. 1.3 Probability to estimate the correct number of occupied bins as function of the SNR. The number of occupied bins is 64, $N_b = 128$ and $N = 1000$

1.6.2 Correlated Frequency Bins

In presence of frequency correlated observations the wideband ITC approach described in Sect. 1.4.2.2 can be adopted. For simplicity we consider only AIC and BIC, and use the notation $\mathsf{AIC}^{\mathrm{i}}$, $\mathsf{BIC}^{\mathrm{i}}$ to denote the adoption of the independence based algorithm and $\mathsf{AIC}^{\mathrm{c}}$, $\mathsf{BIC}^{\mathrm{c}}$ for the correlated case.

To study the performance of the algorithms we adopt a set of Gaussian samples, generated as an autoregressive sequence, in which consecutive samples have a correlation coefficient ρ. In Fig. 1.6 we show P_{ok} assuming $\rho = 0.8$, $k^* = 64$, $N_b = 128$ and $N = 1000$. Note that in this case the AIC is the algorithm that provides the better performance, reaching $P_{ok} \approx 1$ at around $\mathsf{SNR} = -3$ dB. Further numerical results assessing the performance of the wideband ITC based technique in the correlated frequency case are provided in [76].

1.6.3 Multiband Sensing

In this section we analyze a multiband scenario in which three OFDM like signals are present in the observed band. The PSD of the spectrum of the three signals is depicted in Fig. 1.7. In the following we indicate with SNR the SNR of the two lower frequency signals. Note that the higher frequency signal has a SNR drop of -3 dB. In Fig. 1.8 we show P_D^q when the wideband algorithm proposed in Sect. 1.4.2.1 is adopted. It is interesting to note that for very low SNR, such as $\mathsf{SNR} = -20$ dB,

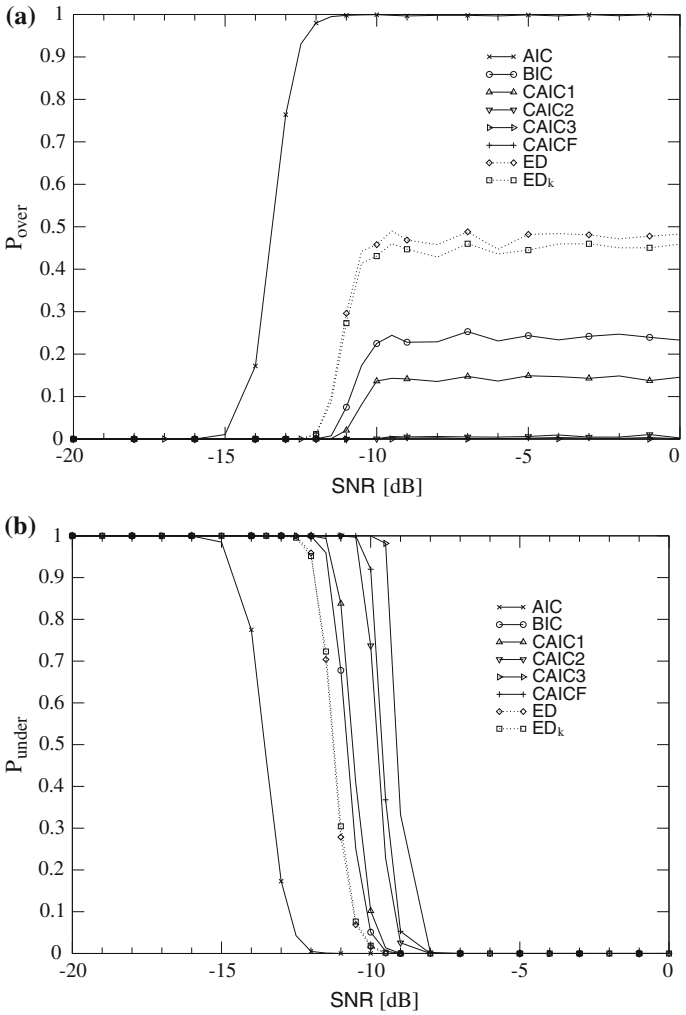

Fig. 1.4 Probabilities to overestimate and underestimate the correct number of occupied bins as function of the SNR. The number of occupied bins is 64, $N_b = 128$ and $N = 1000$. **a** Probability of overestimation. **b** Probability of underestimation

the algorithm that performs better is the AIC. In Sect. 1.6.1 we noted that generally AIC tends to overestimate the number of bins occupied; this property turns to be an advantage at low SNR levels allowing a better probability to detect the presence of signals. On the other hand AIC always leads to a non negligible number of false alarms in unoccupied bins. From Fig. 1.8b we can see that in this case study all ITC performs well at $\mathsf{SNR} = -10\,\mathrm{dB}$.

ITC are conceived for being statistical approaches that choose the model that best approximates data among a family of models. It is interesting therefore to analyze

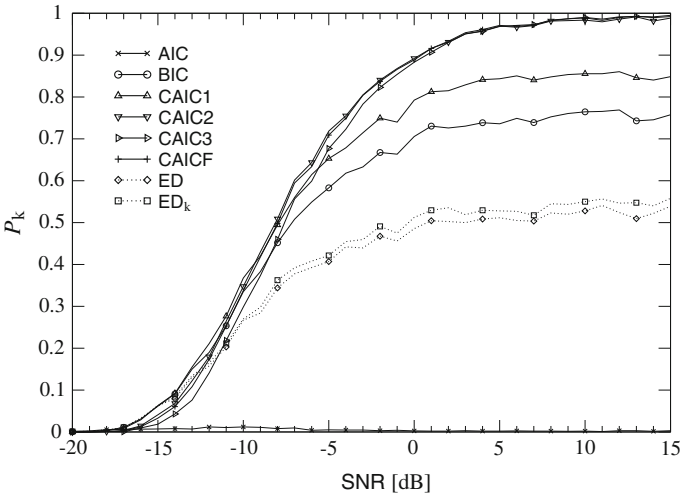

Fig. 1.5 Probability to estimate the correct number of occupied bins as function of the SNR in Rayleigh fading. The number of occupied bins is 64, $N_b = 128$ and $N = 1000$

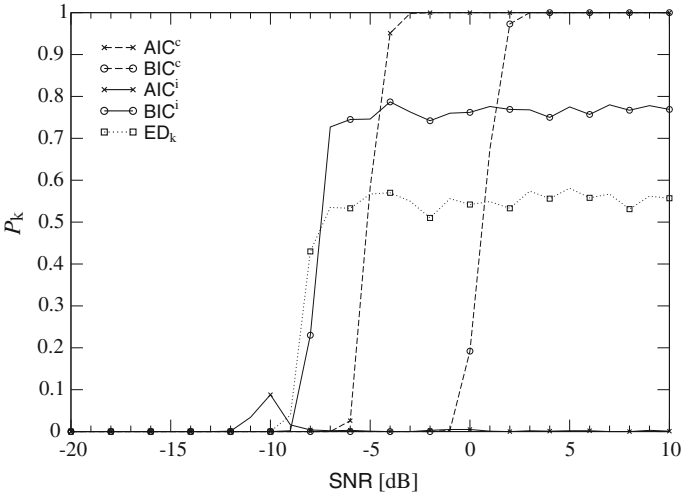

Fig. 1.6 Probability to estimate the correct number of occupied bins as function of the SNR in presence of a frequency correlated signal. The number of occupied bins is 64, $N_b = 128$ and $N = 1000$

if these algorithms provide a good detection performance also when the true model that underlie the generation of the observation is not in the considered model set. This case has also an important impact on practical situations in which the exact statistical description of the collected data is not known or it is too complex to apply ITC in a rigorous way, and thus algorithms derived for simpler models are adopted.

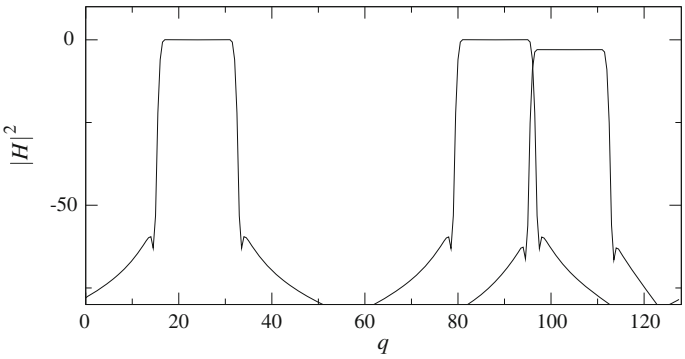

Fig. 1.7 Squared magnitude of the signals adopted in the multiband scenario normalized to 0 dB

Here we consider the case in which the spectral representation used is a spectrum estimate derived using MTM or Welch periodogram, both with N_b points in frequency domain, and we apply the wideband ITC derived for the DFT analysis in Sect. 1.4.2. This can be considered an approximated strategy in which these spectrum estimates, that in general are chi squared distributed, are approximated to Gaussian r.v.s[16] [79]. In Figs. 1.9 and 1.10 we can see that the MTM and Welch strategies provide a very good detection performance that outperforms the DFT based approach for low SNR levels. Then we can benefit from better spectrum estimates (DFT has a non negligible spectral leakage) and apply the wideband approach proposed in Sect. 1.4.2.

1.6.4 Cooperative Wideband Sensing

In Sect. 1.5 we introduced a cooperative sensing strategy that extends single user wideband SS. Here we apply this cooperative approach to the multi band scenario described in the previous section. In Fig. 1.11 we compare the single user P_D^q with the corresponding cooperative performance with different choices of K when AIC is adopted. We consider the presence of $M = 6$ SUs in the AWGN scenario used in Fig. 1.8. When $K = 1$ we implement an OR fusion strategy that provide a very high probability of detection in the occupied bins, but also a very high number false alarms. When $K = M$ we implement the AND rule that allows a very low probability of false alarm, at the expense of a low probability of detection. The performance of the majority rule, with $K = M/2$, is more balanced providing a small number of detection errors in both occupied and unoccupied bins.

[16] Note that this approximation is valid when the chi squared distribution has a high number of degrees of freedom.

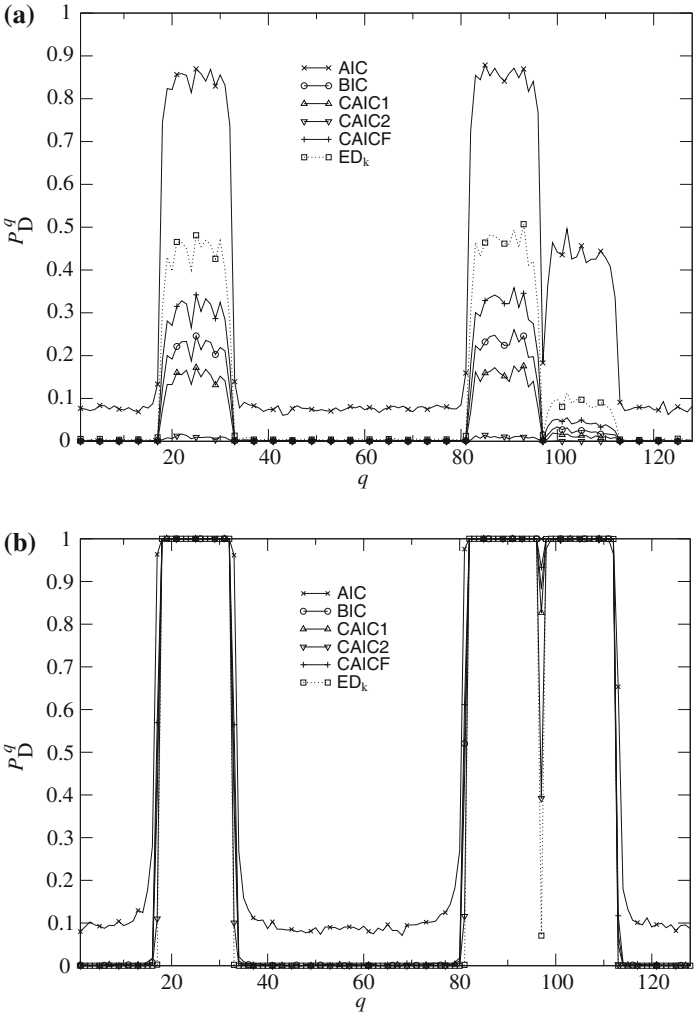

Fig. 1.8 Probability of detection for each frequency bin when the DFT is adopted for the multi band scenario depicted in Fig. 1.7. $N_b = 128$ and $N = 1000$. **a** SNR $= -20$ dB. **b** SNR $= -10$ dB

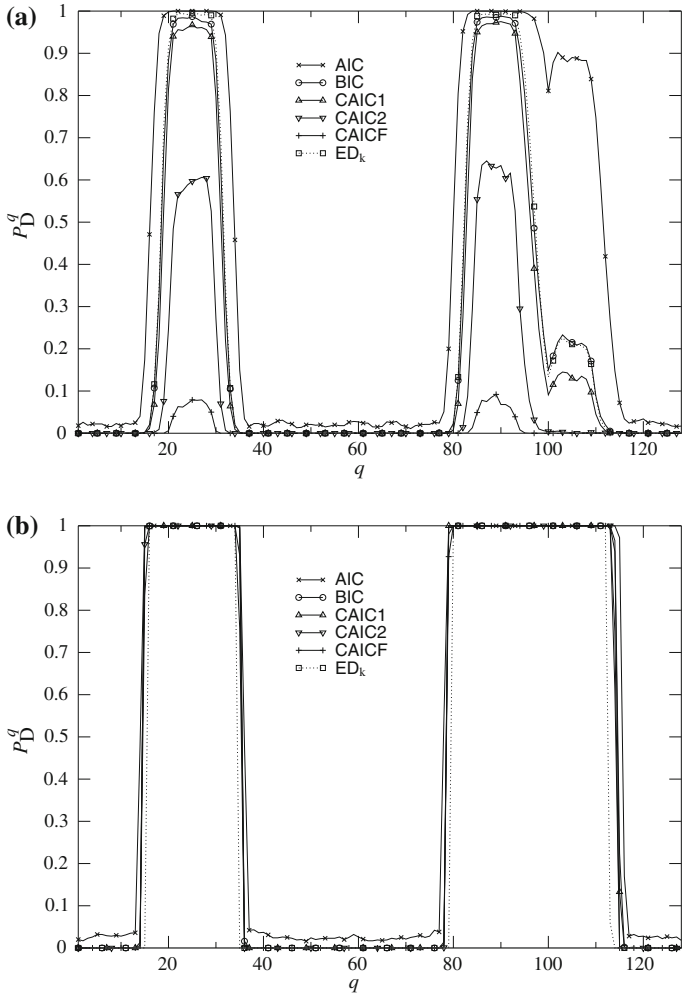

Fig. 1.9 Probability of detection for each frequency bin when a 128 points MTM spectrum estimate is adopted for the multi band scenario depicted in Fig. 1.7. $N = 1000$. **a** SNR $= -20$ dB. **b** SNR $= -10$ dB

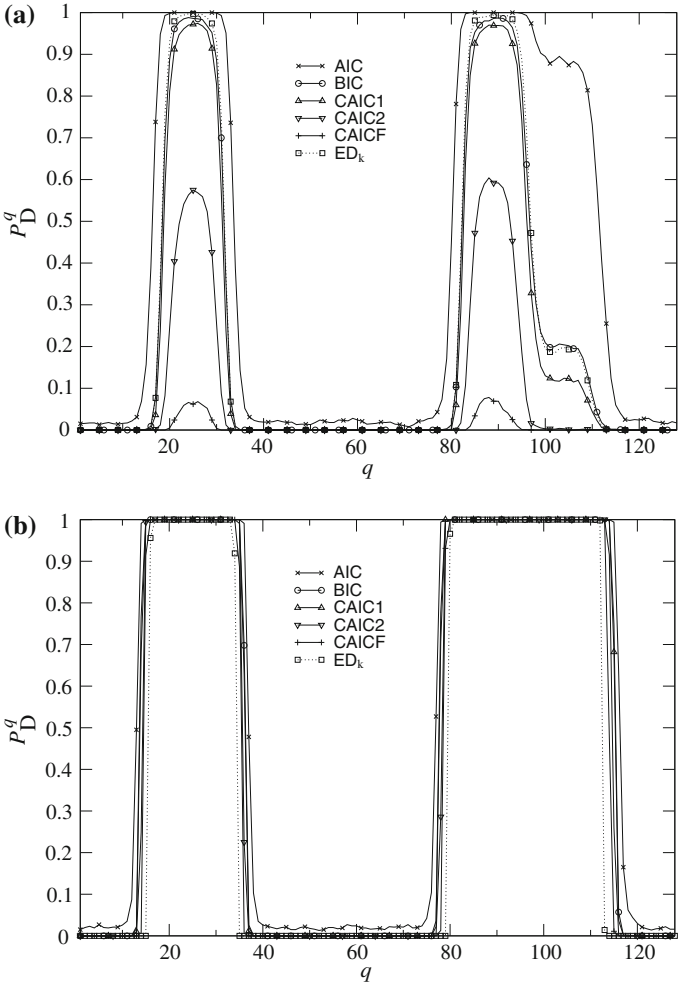

Fig. 1.10 Probability of detection for each frequency bin when a 128 points Welch spectrum estimate is adopted for the multi band scenario depicted in Fig. 1.7. $N = 1000$. **a** SNR $= -20$ dB. **b** SNR $= -10$ dB

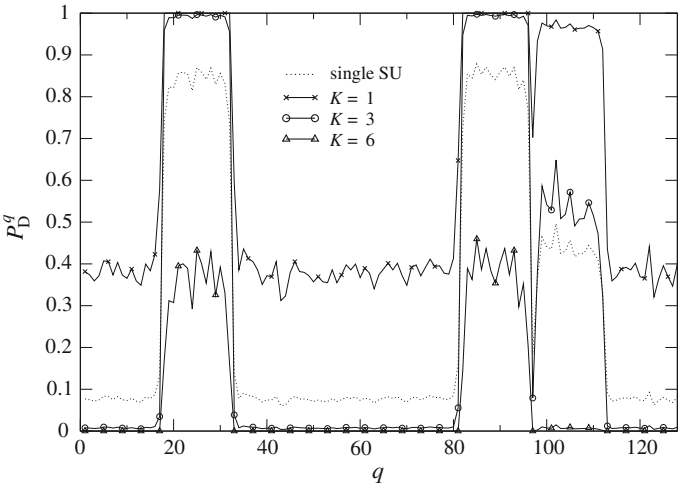

Fig. 1.11 Probability of detection for each frequency bin in a cooperative sensing scheme with six SUs. The fusion rule adopted is the K out of M hard combining. $N_b = 128$ and $N = 1000$

1.7 Conclusions

In this chapter we proposed a wideband spectrum sensing technique based on ITC. We described a general approach that can be applied to any spectral representation and then focused on the simple DFT case. The proposed technique is completely blind since it does not require any knowledge about the noise power and characteristics of the signals present in the observed band. In particular we showed that consistent ITC can reach an almost one probability to correctly identify the number of occupied bins, outperforming simple ED based approaches. Numerical results revealed that this wideband approach can be applied both with independent and correlated frequency components. In particular, the derived DFT based algorithm can be applied as an approximated approach in situations in which the exact distribution of the observation is unknown or too complex, such as when advanced techniques like MTM spectrum estimation are adopted. Wideband ITC based sensing can be applied in scenarios in which approaches that require a high level of sparsity of the received signal (such as compressive sensing) can not be adopted.

References

1. FCC: Spectrum Policy Task Force Report, ET Docket 02–135, Nov (2002)
2. Ghasemi, A., Sousa, E.S.: Spectrum sensing in cognitive radio networks: requirements, challenges and design trade-offs. IEEE Commun. Mag. **46**(4), 32–39 (2008)
3. IEEE Standard Definitions and Concepts for Dynamic Spectrum Access: Terminology Relating to Emerging Wireless Networks, System Functionality, and Spectrum Management (2008)

4. IEEE Standard Definitions and Concepts for Dynamic Spectrum Access: Terminology Relating to Emerging Wireless Networks, System Functionality, and Spectrum Management. Amendment 1: Addition of New Terms and Associated Definitions (2013)
5. ITU-R: Definitions of Software Defined Radios (SDR) and Cognitive Radio Systems, Sept (2009). [Online]. Available: http://www.itu.int/pub/R-REP-SM.2152
6. Filin, S., Harada, H., Murakami, H., Ishizu, K.: International standardization of cognitive radio systems. IEEE Commun. Mag. **49**(3), 82–89 (2011)
7. Kandeepan, S., Giorgetti, A.: Cognitive Radios and Enabling Techniques. Artech House Publishers, Boston (2012)
8. Goncalves, V., Pollin, S.: The value of sensing for TV white spaces. In: Proceedings of IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN 2011) (2011)
9. Tonelli, O., Berardinelli, G., Cattoni, A.F., Srensen, T.B., Mogensen, P.E.: Software architecture design for a dynamic spectrum allocation-enabled cognitive radio testbed. In: Proceedings of European Signal Processing Conference (EUSIPCO 2011), Barcelona, Spain (2011)
10. Ghasemi, A., Sousa, E.S.: Collaborative spectrum sensing for opportunistic access in fading environments. In: Proceedings of IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN 2005) (2005)
11. Akyildiz, I., 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)
12. FCC: Second memorandum opinion and order 10–147, Sept (2011)
13. ECC: Report 159 Technical and Operational Requirements for the Possible Operation of Cognitive Radio Systems in the 'White Spaces' of the Frequency Band 470–790 MHz, Jan (2011)
14. ECC: http://www.erodocdb.dk/Docs/doc98/official/pdf/ECCREP185.PDF. Report 185 Complementary Report to ECC Report 159 Further Definition of Technical and Operational Requirements for the Possible Operation of White Spaces Devices in the Band 470–790 MHz, Jan (2013)
15. El-Sherif, A.A., Liu, K.J.R.: Joint design of spectrum sensing and channel access in cognitive radio networks. IEEE Trans. Wireless Commun. **10**(6), 1743–1753 (2011)
16. Lee, W.Y., Akyildiz, I.F.: Optimal spectrum sensing framework for cognitive radio networks. IEEE Trans. Wireless Commun. **7**(10), 3845–3857 (2008)
17. Mariani, A.: Spectrum sensing algorithms for cognitive radio applications. Ph.D. thesis, Philosophy Doctoral Program in Electronics, Computer Science and Telecommunications, Alma Mater Studiorum University of Bologna, Cesena, Italy (2013)
18. Urkowitz, H.: Energy detection of unknown deterministic signals. Proc. IEEE **55**(4), 523–531 (1967)
19. Mariani, A., Giorgetti, A., Chiani, M.: Energy detector design for cognitive radio applications. In: Proceedings of IEEE International Waveform Diversity & Design Conference (WDD2010), pp. 53–57. Niagara, Canada (2010). Invited Paper
20. Saad, W., Han, Z., Basar, T., Debbah, M., Hjorungnes, A.: Coalition formation games for collaborative spectrum sensing. IEEE Trans. Veh. Technol. **60**(1), 276–297 (2011)
21. Mustonen, M., Matinmikko, M., Mammela, A.: Cooperative spectrum sensing using quantized soft decision combining. In: Proceedings of IEEE International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM 2009) (2009)
22. Sonnenschein, A., Fishman, P.M.: Radiometric detection of spread-spectrum signals in noise of uncertain power. IEEE Trans. Aerosp. Electron. Syst. **28**(3), 654–660 (1992)
23. Tandra, R., Sahai, A.: SNR walls for signal detection. IEEE J. Sel. Top. Signal Proc. **2**, 4–17 (2008)
24. Mariani, A., Giorgetti, A., Chiani, M.: SNR wall for energy detection with noise power estimation. In: Proceedings of IEEE International Conference on Communications (ICC2011). Kyoto, Japan (2011)
25. Mariani, A., Giorgetti, A., Chiani, M.: Effects of noise power estimation on energy detection for cognitive radio applications. IEEE Trans. Commun. **59**(12), 3410–3420 (2011)
26. Chaudhari, S., Koivunen, V., Poor, H.: Autocorrelation-based decentralized sequential detection of OFDM signals in cognitive radios. IEEE Trans. Signal Process. **57**(7), 2690–2700 (2009)

27. Axell, E., Larsson, E.: Optimal and near-optimal spectrum sensing of OFDM signals in AWGN channels. In: Proceedings of IEEE International Workshop on Cognitive Information Processing (CIP 2010), pp. 128–133 (2010)
28. Danev, D., Axell, E., Larsson, E.: Spectrum sensing methods for detection of DVB-T signals n AWGN and fading channels. In: Proceedings of IEEE Conference on Personal, Indoor and Mobile Radio Communications (PIMRC 2010) (2010)
29. Cabric, D., Tkachenko, A., Brodersen, R.: Spectrum sensing measurements of pilot, energy, and collaborative detection. In: Proceedings of IEEE Military Communications Conference (MILCOM 2006) (2006)
30. Gardner, W.A.: Introduction to Random Processes with Applications to Signals and Systems, 2nd edn. McGraw-Hill, New York (1990)
31. Mariani, A.: Detection algorithms based on cyclic spectrum analysis for cognitive radio. Second Faculty of Engineering, Alma Mater Studiorum University of Bologna, Cesena, Italy, Master's thesis (2009)
32. Chiani, M., Giorgetti, A., Mariani, A., Montanari, M.: Cognitive radio for defense scenarios Technical challenges and spectrum sensing issues. In: Proceedings of SDR Italy'09 Workshop, Form Software Defined Radio to Cognitive Networks (2009)
33. Cabric, D.: Addressing feasibility of cognitive radios. IEEE Signal Process. Mag. **25**(6), 85–93 (2008)
34. Bizaki, H.K.: MIMO Systems Theory and Applications. InTech, Rijeka (2011)
35. Zheng, Y., Liang, Y.C.: Maximum-minimum eigenvalue detection for cognitive radio. In: Proceedings of IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2007) (2007)
36. Lim, T.J., Zhang, R., Liang, Y.C., Zeng, Y.: GLRT-based spectrum sensing for cognitive radio. In: Proceedings of IEEE Global Communications Conference (GLOBECOM 2008) (2008)
37. Penna, F., Garello, R., Spirito, M.A.: Cooperative spectrum sensing based on the limiting eigenvalue ratio distribution in Wishart matrices. IEEE Commun. Lett. **13**(7) (2009)
38. Chiani, M., Win, M.Z.: Estimating the number of signals observed by multiple sensors. In: Proceedings of IEEE International Workshop on Cognitive Information Processing (CIP), pp. 156–161. Elba Island, Italy (2010)
39. Wang, P., Fang, J., Han, N., Li, H.: Multiantenna-assisted spectrum sensing for cognitive radio. IEEE Trans. Veh. Technol. **59**(4), 1791–1800 (2010)
40. Mariani, A., Giorgetti, A., Chiani, M.: Test of independence for cooperative spectrum sensing with uncalibrated receivers. In: Proceedings of IEEE Global Communications Conference (GLOBECOM 2012). Anaheim, CA, USA (2012)
41. Leshem, A., van der Veen, A.J.: Multichannel detection of Gaussian signals with uncalibrated receivers. IEEE Signal Process. Lett. **8**(4), 120–122 (2001)
42. Wax, M., Kailath, T.: Detection of signals by information theoretic criteria. IEEE Trans. Acoust. Speech Signal Process. **33**, 387–392 (1985)
43. Wang, R., Tao, M.: Blind spectrum sensing by information theoretic criteria for cognitive radios. IEEE Trans. Veh. Technol. **59**(8), 3806–3817 (2010)
44. Akyildiz, I., Lo, B., Balakrishnan, R.: Cooperative spectrum sensing in cognitive radio networks: A survey. Phys. Commun. **4**(1), 40–62 (2011)
45. Ma, J., Zhao, G., Li, Y.: Soft combination and detection for cooperative spectrum sensing in cognitive radio networks. IEEE Trans. Wireless Commun. **7**(11), 4502–4507 (2008)
46. Mariani, A., Giorgetti, A., Paolini, E., D' Angelo, A., Tieri, C., Schillaci, S., Chiani, M.: Implementation issues in cooperative spectrum sensing with soft fusion. In: Proceedings of IEEE Military Communications and Information Systems Conference (MCC 2013). Saint Malo, France (2013)
47. Thomson, D.: Spectrum estimation and harmonic analysis. Proc. IEEE **70**(9), 1055–1096 (1982)
48. Erpek, T., Leu, A., Mark, B.: Spectrum sensing performance in tv bands using the multitaper method. In: Proceedings of IEEE Signal Processing and Communications Applications Conference (SIU 2007) (2007)

49. Zhang, Q.T.: Theoretical performance and thresholds of the multitaper method for spectrum sensing. IEEE Trans. Veh. Technol. **60**(5), 2128–2138 (2011)
50. Quan, Z., Zhang, W., Shellhammer, S.J., Sayed, A.H.: Optimal spectral feature detection for spectrum sensing at very low SNR. IEEE Trans. Commun. **59**(1), 201–212 (2011)
51. Cabric, D., Mishra, S., Brodersen, R.: Implementation issues in spectrum sensing for cognitive radios. In: Proceedings of Asilomar IEEE Conference on Signals, Systems and Computers, vol. 1, pp. 772–776 (2004)
52. 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)
53. Krasner, N.: Efficient search methods using energy detectors-maximum probability of detection. IEEE J. Sel. Areas Commun. **4**(2), 273–279 (1986)
54. Kandeepan, S., Piesiewicz, R., Aysal, T.C., Biswas, A.R., Chlamtac, I.: Spectrum sensing for cognitive radios with transmission statistics: Considering linear frequency sweeping. EURASIP J. Wirel. Commun. Netw. (2010)
55. Farhang-Boroujeny, B.: Filter bank spectrum sensing for cognitive radios. IEEE Trans. Signal Process. **56**(05), 1801–1811 (2008)
56. Sun, H., Chiu, W.Y., Jiang, J., Nallanathan, A., Poor, H.V.: Wideband spectrum sensing with sub-Nyquist sampling in cognitive radios. IEEE Trans. Signal Process. **60**(11), 6068–6073 (2012)
57. Tian, Z., Giannakis, G.B.: Compressed sensing for wideband cognitive radios. In: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2007), vol. 4 (2007)
58. Quan, Z., Cui, S., Sayed, A., Poor, V.: Optimal multiband joint detection for spectrum sensing in cognitive radio networks. IEEE Trans. Signal Process. **57**(3), 1128–1140 (2009)
59. Taherpour, A., Gazor, S., Nasiri-Kenari, M.: Invariant wideband spectrum sensing under unknown variances. IEEE Trans. Wireless Commun. **8**(5), 2182–2186 (2009)
60. Coulson, A.: Blind detection of wideband interference for cognitive radio applications. EURASIP J. Adv. Signal Process. **2009**, 1–13 (2009)
61. Rashidi, M., Haghighi, K., Owrang, A., Viberg, M.: A wideband spectrum sensing method for cognitive radio using sub-nyquist sampling. In: Proceedings of IEEE International Digital Signal Processing and Signal Processing Education Workshop (DSP/SPE 2011) (2011)
62. Liu, S., Shen, J., Zhang, R., Zhang, Z., Liu, Y.: Information theoretic criterion-based spectrum sensing for cognitive radio. IET Commun. **2**(6), 753–762 (2008)
63. Fujii, M., Watanabe, Y.: A study on interference detection scheme using AIC for UWB MB-OFDM systems. In: Proceedings of International Workshop on Multi-Carrier Systems Solutions (MC-SS), 2011 (2011)
64. Stoica, P., Selen, Y.: Model-order selection: a review of information criterion rules. IEEE Signal Process. Mag. **21**(4), 36–47 (2004)
65. Akaike, H.: Information theory and an extention of the maximum likelihood principle. In: Proceedings of International Symposium on Information Theory, pp. 267–281 (1972)
66. Schwarz, G.: Estimating the dimension of a model. Ann. Stat. **6**(2), 461–464 (1978)
67. Rissanen, J.: An introduction to the MDL principle (2004). http://www.mdl-research.org
68. Giorgetti, A., Chiani, M.: Time-of-arrival estimation based on information theoretic criteria. IEEE Trans. Signal Process. **61**(8), 1869–1879 (2013)
69. Bozdogan, H.: Akaike's information criterion (AIC): The general theory and its analytical extensions. Psychometrica **52**(3), 345–370 (1987)
70. Akaike, H.: On newer statistical approaches to parameter estimation and structure determination. Int. Fed. Autom. Control **3**, 1877–1884 (1978)
71. Nishii, R.: Maximum likelihood principle and model selection when the true model is unspecified. J. Multivar. Anal. **27**(2), 392–403 (1988)
72. Zhao, L.C., Krishnaiah, P.R., Bai, Z.D.: On detection of the number of signals when the noise covariance matrix is arbitrary. J. Multivar. Anal. **20**(1), 26–49 (1986)
73. Hansen, M.H., Yu, B.: Model selection and the principle of minimum description length. J. Am. Stat. Assoc. **96**(454), 746–774 (2001)

74. Horn, R.A.: Johnson. Matrix analysis. Cambridge University Press, Cambridge (1990)
75. Coulson, A.J.: Do wireless data signals exhibit spectral autocorrelation? In: Proceedings of IEEE Australian Communications Theory Workshop (AusCTW 2008) (2008)
76. Giorgetti, A., Mariani, A., Chiani, M.: Spectrum holes detection by information theoretic criteria. In: Proceedings of International Conference on Cognitive Radio and Advanced Spectrum Management (COGART 2011). Barcelona, Spain (2011). Invited Paper
77. Zhang, W., Mallik, R.K., Letaief, K.B.: Optimization of cooperative spectrum sensing with energy detection in cognitive radio networks. IEEE Trans. Wireless Commun. **8**(12), 5761–5766 (2009)
78. Proakis, J.G.: Digital Communications, 4th edn. McGraw-Hill, New York (2001)
79. Percival, D.B.: Walden. Spectral Analysis for Physical Applications. Cambridge University Press, Cambridge (1993)

Chapter 2
Channel Usage Patterns and Their Impact on the Effectiveness of Machine Learning for Dynamic Channel Selection

Irene Macaluso, Hamed Ahmadi, Luiz A. DaSilva and Linda Doyle

Abstract The diverse behavior of different primary users (PU) in various spectrum bands impacts a cognitive radio's ability to exploit spectrum holes. This chapter summarizes the results of our previous studies on the impact of the complexity of primary users' behavior on the performance of learning algorithms applied to dynamic channel selection. In particular, we characterize the observable spectrum utilization with respect to the duty cycle of the channels and to the complexity of the primary user's activity. We use the term complexity to refer to the unpredictability associated with the primary user's wireless resource usage, which we quantitatively characterize using Lempel-Ziv complexity. We evaluate the effectiveness of two learning-based dynamic channel selection algorithms by testing them with real spectrum occupancy data collected in the GSM, ISM, and DECT bands. Our results show that learning performance is highly correlated with the level of PU activity, estimated by the duty cycle, and the amount of structure in the use of spectrum, estimated by the Lempel-Ziv complexity.

2.1 Introduction

Opportunistic spectrum access may rely on a combination of geolocation databases and spectrum sensing to detect spectrum holes. In their search for spectrum holes, secondary users (SUs) can use learning methods to predict the next channel state from

I. Macaluso (✉) · H. Ahmadi · L. A. DaSilva · L. Doyle
CTVR Telecommunications Research Center, Trinity College, Dublin, Ireland
e-mail: irene.macaluso@gmail.com

H. Ahmadi
e-mail: ahmadih@tcd.ie

L. A. DaSilva
Department of Electrical and Computer Engineering, Virginia Tech, Arlington, VA, USA
e-mail: dasilval@tcd.ie

L. Doyle
e-mail: linda.doyle@tcd.ie

M.-G. Di Benedetto and F. Bader (eds.), *Cognitive Communication and Cooperative HetNet Coexistence*, Signals and Communication Technology,
DOI: 10.1007/978-3-319-01402-9_2,

their past observations on each channel. However, because of the diverse behavior of primary users in different spectrum bands, spectrum holes exhibit different characteristics, which in turn affect the performance of a learning algorithm [1, 2].

This chapter summarizes the results of our previous studies on the impact of the complexity of PUs' behavior on the performance of learning algorithms applied to dynamic channel selection (DCS) [2, 3]. In particular, we consider two learning approaches. In [2], we presented a reinforcement learning-based method for spectrum opportunity prediction; in [3], we presented a Markov process-based learning. We analyzed the performance of the two approaches with respect to the duty cycle (DC) of the channels and to the complexity of the PU's activity, relying on actual spectrum measurement data.

In both cases we characterized the PU activity by using a measure of complexity proposed by Lempel and Ziv [4]. Our results showed that the amount of structure in the PU activity, estimated by the Lempel-Ziv (LZ) complexity, has a significant impact on the performance of the learning-based DCS approaches. In particular, our studies showed that the LZ complexity of the PU's behavior can account for up to a 30 and 20 % difference in the probability of success of reinforcement learning (RL) and Markov process-based learning respectively.

We begin with a short discussion on the literature of learning algorithms applied to DCS in Sect. 2.2. Section 2.3 details our use of Lempel-Ziv complexity to quantify the amount of structure in the usage of the bands by the PU. Section 2.4 describes the two learning-based dynamic channel selection approaches. Section 2.5 presents the relationship between the effectiveness of learning-based DCS and the amount of structure in the usage of the bands by the PU, relying on spectrum measurements conducted at RWTH Aachen and by us at Trinity College Dublin to determine, for any given time slot and set of channels, whether there is PU activity. We summarize our conclusions in Sect. 2.6.

2.2 Existing Works

In the literature, several learning algorithms have been proposed and applied to predict the channel state. All these approaches need to learn the channel occupancy model by observing the PU activity for a certain number of time slots. This phase is called training or learning period. The learning algorithms are able to make predictions on the channel state after they are trained. Learning algorithms with higher accuracy, lower complexity and shorter training time are preferred. In this section we briefly review some of the existing works.

In [5], Clancy claims that a hidden Markov model (HMM) can be a suitable method to model the channel occupancy as a function of time. In [6], the authors model the channel state occupancy of a PU on each channel as a Poisson distribution and use an HMM to predict the availability of a channel. The HMM is trained with the Baum-Welsh algorithm (BWA) [7], predicting the presence of PUs to avoid transmission collisions. An SU will occupy an idle channel until a PU becomes active in that

channel, then it will switch to another predicted idle channel. Simulation results show that the probability of collision is reduced compared to a random selection of channels to be sensed by the SU. An artificial neural network is proposed in [8] to predict the channel state for the next time slot. In [9], the authors evaluate the performance of their proposed neural network when the statistics of the channel are changing. The authors of [10] formulate this problem of efficiently using the spectrum as a Markov decision process and propose a solution strategy based on reinforcement learning techniques. In [11], we propose a modified, less complex HMM. We compare the prediction accuracy of our proposed method with that of conventional HMM and show that it achieves the same prediction accuracy with much less computational complexity.

The aforementioned works test their proposed algorithms on data which are generated based on some assumptions on the probability distribution of the PU activity (synthetic data), and none of them applies their algorithm on data that are collected from actual sensing (real data).

There are some recent works which use real data in their studies. In [12], authors conduct spectrum measurements in Guangzhou city, and then approximate the prediction error with the beta distribution. Kone et al. propose frequency bundling in [13], where secondary devices build reliable channels by combining multiple unreliable frequencies into virtual frequency bundles. Their experiments on real data show that bundling random channels together can provide sustained periods of reliable transmission.

The works above only study the performance of their proposed algorithms and compare them with other learning algorithms. On the other hand, few works [2, 3] investigated the predictability of the behavior of the PUs and its impact on the learning performance, which is the main focus of the rest of this chapter.

2.3 Complexity of Primary User Activity

The term complexity is used with different meanings in the literature [14]. In some domains, e.g. dynamic systems and statistical mechanics, both completely ordered or disordered sequences are associated with low complexity [15]. In other words, complexity does not increase monotonically with disorder. In this chapter our goal is to investigate the effectiveness of learning algorithms in exploiting the regularities of channel utilization; hence we use the term complexity to refer to the unpredictability, or uncertainty, associated with the PU's wireless resource usage. Accordingly, we adopt a measure of complexity that associates high values of complexity to completely disordered, i.e. unpredictable, spectrum occupancy sequences.

We quantitatively characterize the structure of a spectrum occupancy sequence by making use of a measure of complexity proposed by Lempel and Ziv [4]. In particular, we adopt the normalized Lempel-Ziv complexity, which measures the rate of production of new patterns in a sequence. The complexity coefficient c is computed by scanning the sequence and incrementing c every time a new substring

of consecutive symbols is found. Then c is normalized via the asymptotic limit $n/\log_2(n)$, where n is the length of the sequence [16]. Lempel-Ziv complexity is a property of individual sequences and it can be computed without making any assumptions about the underlying process that generated the data. This feature is of the utmost importance when one is dealing with real data (in our case, sensed channel status in a variety of frequency bands). Furthermore, LZ complexity is strongly related to the source entropy. In fact, if the source is ergodic, the normalized LZ complexity has been proven [17] to be equal to the source entropy almost surely.

It is interesting to note that the LZ complexity is closely related to the concept of Kolmogorov complexity [18], which measures the complexity of a binary string s as the bit length of the shortest program that produces s and halts afterwards. Whereas Kolmogorov complexity refers to the shortest program among all the possible classes of programs, LZ complexity makes use of one class of programs that can only perform copy and paste operations [16]. Although Kolmogorov complexity is known to be algorithmically uncomputable, it should be noted that in the case of ergodic sources Brudno's theorem states that the entropy rate of the source is equal to the Kolmogorov complexity per symbol of almost all emitted strings ([19], as discussed in [20]). Lempel-Ziv complexity and Kolmogorov complexity are deterministic complexity measures: by looking for the shortest description that allows to exactly reproduce the data, they inevitably include the noise in such description.

2.4 Learning and DCS

We consider the problem of an SU searching for a channel to occupy opportunistically while the PU is inactive in that channel. We do not assume the SU to have any a priori knowledge of the pattern of activity of the PU or the long-term probability that each channel is occupied. We explore two alternative learning strategies to decide which channel to sense prior to each transmission slot: (i) the SU applies reinforcement learning; (ii) the SU applies Markov-based learning. Our goal is to study the effect of both the levels of PU activity and the complexity of the PU behavior as defined in Sect. 2.3 on the effectiveness of learning.

Our model considers a single SU that can use one of N equal-bandwidth frequency channels opportunistically. Time is slotted and alternates between a sensing phase and a transmission phase. The SU is allowed to transmit in the time slot if the selected channel in the sensing phase is still free. The SU's choice of which channel to attempt transmission in will therefore affect its performance: the more successful the SU is in predicting which channel is the least likely to contain PU activity in the next time slot, the greater its likelihood to opportunistically utillize the channels.

In our study, we rely on spectrum measurements conducted at RWTH Aachen and by us at Trinity College Dublin to determine, for any given time slot and set of channels, whether there is PU activity. For each band that we investigate, we evaluate the adaptation techniques described in Sects. 2.4.1 and 2.4.2 for different combinations of N channels to be explored by the SU.

2.4.1 Reinforcement Learning

A secondary user selects one among N channels according to the policy determined by an RL algorithm. Under the assumption of full observability of the state of the channels, Q-learning [21] is the most natural candidate. Moreover, as Q-learning does not require a model of the agent's environment, it is suitable to deal with real spectrum occupancy data.

The goal of Q-learning is to find an optimal policy, i.e. the sequence of actions that maximizes the expected sum of discounted rewards. The idea is to reward an SU if it selects a free channel, while also including a cost of switching channels to discourage too frequent channel changes. The SU state at time t is given by $\mathbf{s_t} = [X_{1,t}, \ldots, X_{N,t}, c_t]$, where $X_{i,t} \in \{0, 1\}$ indicates whether the ith channel is free (0) or occupied (1) and $c_t \in \{1, \ldots, N\}$ is the index of the channel the SU is accessing at time t. At time t the SU performs an action $a_t \in \{1, \ldots, N\}$, i.e. it selects a channel c_{t+1}. At time $t + 1$ it receives a reward $r_{(t+1)}(\mathbf{s_t}, a_t)$:

$$r_{(t+1)}(\mathbf{s_t}, a_t) = (1 - X_{a_t,t+1}) - e(1_{a_t,c_t}) \tag{2.1}$$

where $1_{a_t,c_t}$ is 0 if $a_t = c_t$ and 1 otherwise, and $e \in [0, 0.5]$.

Based on the received reward, the SU updates the Q-values according to [21]:

$$Q(\mathbf{s_t}, a_t) := Q(\mathbf{s_t}, a_t) + \alpha r_{(t+1)} + \alpha(\gamma \max_{a_{t+1}} Q(\mathbf{s_{t+1}}, a_{t+1}) - Q(\mathbf{s_t}, a_t)) \tag{2.2}$$

where $0 \leq \gamma < 1$ is the discount factor and α is the learning rate. If $\gamma > 0$, the agent takes into account not only the immediate reward but also the delayed reward when it chooses which action to take.

In a stationary environment Q-learning is proven to converge to the optimal policy if $\alpha \to 0$ and all the state-action pairs are visited an infinite number of times. During the learning stage, an exploration strategy is required to allow the agent to visit all the state-action pairs. A randomized strategy is commonly adopted: the agent selects a random action with a probability ε and the best estimated action with probability $1 - \varepsilon$. At the beginning the algorithm starts with a large value of ε, which decreases as the Q-learning converges.

As the stationary condition is generally not satisfied for real spectrum occupancy data, we fix the learning factor to 0.1 to allow the agent to adapt to the changes in the environment. Moreover, we set the ε-value to 0.01 to allow the agent to perform exploratory actions from time to time in order to discover changes in the environment.

2.4.2 Markov-Based learning

As mentioned earlier, the SU needs to select a channel (from the set of channels that it can access) for transmission which is the least likely to be occupied by the

PU. Other than reinforcement learning, we can use Markov process-based learning algorithms to learn the channels' occupancy model and predict the availability of the channels for the next time slot. Here, we study a low complexity Markov process-based learning algorithm which can accurately predict the channel status of the next time slot.

Our Markov process-based learning algorithm has K states and M possible observations, where $\mathscr{S}$ and $\mathscr{O}$ represent the sets of possible states and observations, respectively.

We denote sequences of states by $\mathbf{x}$, and we use $\mathbf{y}$ to indicate the sequence of observations. Each element of $\mathbf{x}$, denoted by $x(t) \in \mathscr{S}$, $\forall t$, is the state at time t; each element of $\mathbf{y}$, denoted by $y(t) \in \mathscr{O}$, $\forall t$, is the observation at time t.

The transition probabilities between the states are stored in a $K \times K$ matrix ($\mathbf{A}$). The distribution of the observation outcomes at each state is described by the respective column vector of the $K \times M$ emission matrix $\mathbf{B}$. We represent this Markov process by $\lambda = \{\mathbf{A}, \mathbf{B}, \pi\}$, where π is the initial state distribution.

The algorithm is trained off-line over a training sequence. In this Markov process-based learning model, the number of states (i.e. K) grows dynamically as learning proceeds. Here, we have two possible observations ($M = 2$). We observe a zero when we sense a free channel, and we observe a one by sensing a busy channel.

The transitions between states depend on the length of the string of consecutive zeros or ones observed. This means that in our system each state represents a number of observed consecutive zeros or ones. Positive states represent the number of observed consecutive ones, and non-positive (negative and zero) states represent the number of observed consecutive zeros.

During the training phase, we create the Markov chain using the training data set and based on the number of consecutive zeros and ones. Then, it is possible to count the number of times each particular transition or output observation is applied in a set of training data. As proven in [7], counting functions for the output observations provide maximum likelihood estimates for the desired model parameters. Suppose that the maximum number of negative and positive states in the Markov chain, after the training, are q and p, respectively. The set of states is $S = \{s_{q-}, \ldots, s_0, \ldots, s_p\}$ which has the cardinality of $N = q + p + 1$. The elements of transition and emission matrices will be computed by:

$$a_{i,j} = \frac{f_{i,j}(\mathbf{x})}{\sum_{k=-q}^{p} f_{i,k}(\mathbf{x})} \quad \forall i, j \in \{-q, \ldots, p\}, \tag{2.3}$$

$$b_{i,j} = \frac{g_{i,j}(\mathbf{x}, \mathbf{y})}{\sum_{m=1}^{M} g_{i,m}(\mathbf{x}, \mathbf{y})} \quad \forall i \in \{-q, \ldots, p\}, j \in \{1, \ldots, M\}, \tag{2.4}$$

where $a_{i,j}$ is the transition probability from s_i to s_j, and $b_{i,j}$ is the probability of o_j at s_i. The counting functions $f_{i,j}$ and $g_{i,j}$ simply count the number of transitions from state s_i to state s_j and the number of observations o_j at state s_i, respectively.

During the training process, the transition and emission probabilities over the observations can be easily calculated by (2.3) and (2.4). As a result, we can predict the observation by:

$$\check{y}(t) = \begin{cases} 0 & p(x(t), 0|\lambda) \geq p(x(t), 1|\lambda) \\ 1 & otherwise \end{cases}, \tag{2.5}$$

where $\check{y}(t)$ indicates the predicted observation for time t.

In case of an inaccurate prediction, the system will notice the prediction error after observing $y(t)$. Since $x(t+1)$ only depends on the observation outcome rather than on the predicted result, the system will move to the correct state and errors will not propagate. Retraining the system is only needed when the statistics of the behavior of PUs on the channel are changing. To account for this, the system can be retrained after a certain number of time slots or whenever the prediction accuracy drops below a certain threshold.

2.5 Results

In this section we analyze the performance of the two learning solutions presented in Sect. 2.4. The results presented in this section refer to sequences of spectrum occupancy over 12 h (from 11:00 to 23:00) and duty cycle $DC \in [0.3, 0.8]$. We examined a number of frequency bands: the 2.4 GHz ISM band, the DECT band, and the GSM900 and GSM1800 bands. For each band, we considered all the possible combinations of $N = 3$ and $N = 4$ channels to evaluate the RL approach and the Markov-based learning approach respectively. Each combination corresponds to one instance of the DCS problem. One of the fundamental issues common to all RL approaches is the convergence time when the dimension of the state-action space is large. For this reason, we consider combinations of $N = 3$ channels in the case of RL. The convergence time of the Markov process-based learning algorithm increases linearly with N, thus allowing us to consider combinations of 4 channels. However, the number of channels that an SU can observe (N) strongly depends on the hardware and sensing capabilities of the SU.

Our empirical findings show that the performance benefits of the two learning algorithms are highly correlated with the level of PU activity observed and the amount of structure in these observations, estimated by the LZ complexity.

2.5.1 Reinforcement Learning

Each combination of channels corresponds to one instance of the RL problem. For each instance, we run 10^3 independent simulations. For each simulation, first we compute the optimal policy using the Q-learning algorithm and considering only

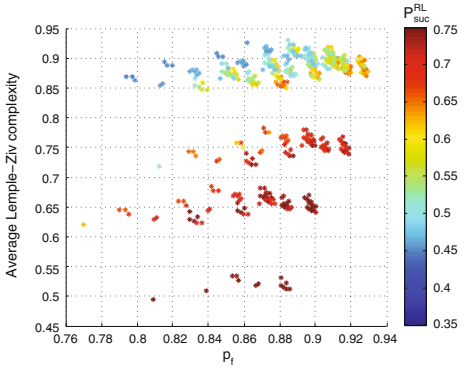

Fig. 2.1 ISM band (outdoor), RWTH Aachen dataset. Probability of success of Q-learning as a function of the average LZ complexity and the probability of at least one free channel existing. Each point represents a particular instance of Q-learning applied to $N = 3$ channels. The total number of possible combinations which we analyzed is $\binom{15}{3} = 455$, where 15 is the number of channels with $DC \in [0.3, 0.8]$

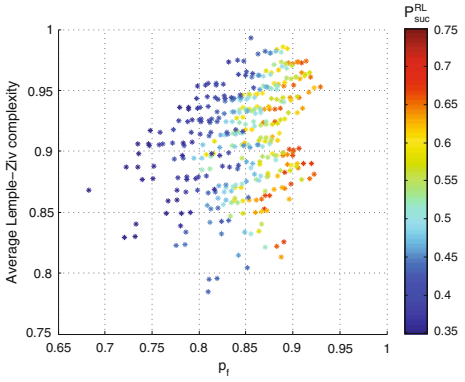

Fig. 2.2 GSM1800 band (outdoor), RWTH Aachen dataset. Probability of success of Q-learning as a function of the average LZ complexity and the probability of at least one free channel existing. Each point represents a particular instance of Q-learning applied to $N = 3$ channels. The total number of possible combinations which we analyzed is $\binom{15}{3} = 455$, where 15 is the number of channels with $DC \in [0.3, 0.8]$

the first hour of the sequences of spectrum occupancy. Then, the resulting policy is evaluated over the remaining 11 h. For each simulation the probability of success is computed according to the number of times that a free channel was selected over the length of the spectrum occupancy sequences. The probability of success of each RL instance is the average over the 10^3 simulations.

Figures 2.1, 2.2 show the probability of success of Q-learning for the ISM band and the GSM1800 band respectively, as a function of the average LZ complexity and the probability of at least one free channel existing $p_f = 1 - \prod_{i=1}^{N} \delta_{1,i}$, where

$\delta_{1,i}$ is the duty cycle of the i-th channel. The LZ complexity of each channel is computed using the algorithm described in [16], while the average value is used as an approximation of the complexity of each combination of channels.

As expected, the probability of success increases with p_f. However, it can be observed that the performance of RL is also strongly dependent on the complexity of the PU behavior. For example, for the ISM band (see Fig. 2.1), the values for the average LZ complexity span a considerable range and we can easily observe that, when p_f remains constant, the RL performance decreases when LZ complexity increases.

Both the average LZ complexity and p_f do not cover the same range for each band. As it can be observed in Figs. 2.1 and 2.2 the DC values, and therefore the p_f values, span almost the same range of values for both the ISM band and the GSM1800 band. However, the GSM1800 channels exhibit a less structured, i.e. more complex, activity on average than the ISM channels. Accordingly, the performance in the ISM band is on average better.

The data collected by us at TCD exhibits a significantly lower LZ complexity than the measurements taken in Aachen (see Fig. 2.3). Accordingly, the performance of RL in the ISM band in TCD is significantly better than the performance in the ISM band for the Aachen dataset.

In any case, all the frequency bands examined by us exhibit the same kind of relationship between average LZ complexity, p_f and the performance of RL, confirming that the LZ complexity is a valid metric for the analysis of learning performance.

2.5.2 Markov-Based Learning

We test the Markov process-based learning algorithm on exactly the same sets of channels that we have used for the RL algorithm. Unlike the RL algorithm, to measure the prediction accuracy of the Markov process-based learning algorithm for each combination of channels we need to run the algorithm only once. In other words, averaging the prediction accuracy results over a number of independent simulations is not required because the prediction accuracy of the Markov process-based learning algorithm over the same test set having the same training set will always be the same.

The data set is recorded over 12 h and it consists of 25000 samples. We train the Markov process-based algorithm over the first 1000 data samples (∼first 30 min) and test it over the remaining 24000 samples. At each time slot the algorithm either selects the channel with the highest predicted probability of being free, or predicts that all the channels are busy.

Figures 2.4, 2.5 show the probability of success as a function of average LZ complexity and the probability of at least one free channel existing. We use the same data sets that we have used for measuring the performance of the Q-learning algorithm, and we compute the probability of success over all possible combinations of existing channels in the dataset for $N = 4$.

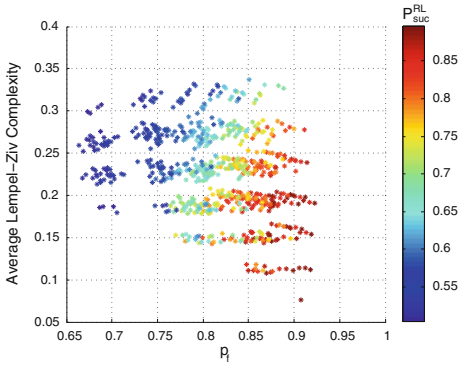

Fig. 2.3 ISM band (indoor), TCD dataset. Probability of success of Q-learning as a function of the average LZ complexity and the probability of at least one free channel existing. Each point represents a particular instance of Q-learning applied to $N = 3$ channels. The total number of possible combinations which we analyzed is $\binom{17}{3} = 680$, where 17 is the number of channels with $DC \in [0.3, 0.8]$

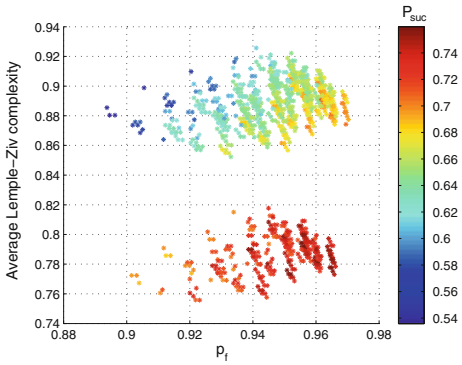

Fig. 2.4 ISM band (outdoor), RWTH Aachen dataset. Probability of success of Markov process-based learning algorithm as a function of the average LZ complexity and the probability of at least one free channel existing. Each point represents a particular instance of the algorithm applied to $N = 4$ channels. The total number of possible combinations which we analyzed is $\binom{15}{4} = 1365$, where 15 is the number of channels with $DC \in [0.3, 0.8]$

Similar to the Q-learning algorithm results, in Figs. 2.4, 2.5, we see that the probability of success using Markov process-based learning algorithm increases with p_f and it also strongly depends on the complexity of the PU behavior. In these two figures we again observe the same phenomenon. For any given p_f, the probability of success has an inverse relation with the average LZ complexity, i.e., when the PU behavior is more complex the probability of success reduces and vice versa.

In Fig. 2.6, we apply the Markov process-based learning algorithm to the data collected from DECT band. For the DECT band, the average LZ complexity is always greater than 0.93. Accordingly, the probability of success is only moderate.

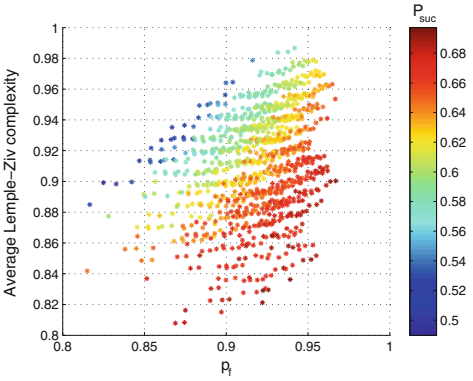

Fig. 2.5 GSM1800 band (outdoor), RWTH Aachen dataset. Probability of success of Markov process-based learning algorithm as a function of the average LZ complexity and the probability of at least one free channel existing. Each point represents a particular instance of Markov process-based learning algorithm applied to $N = 4$ channels. The total number of possible combinations which we analyzed is $\binom{15}{4} = 1365$, where 15 is the number of channels with $DC \in [0.3, 0.8]$

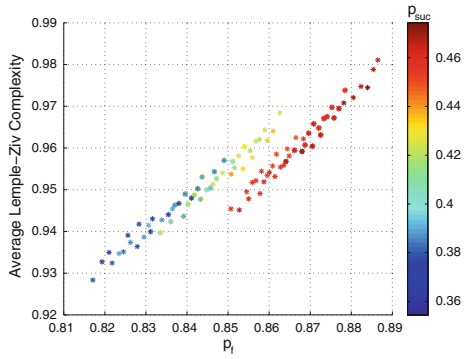

Fig. 2.6 DECT band (indoor), RWTH Aachen dataset. Probability of success of Markov process-based learning algorithm as a function of the average LZ complexity and the probability of at least one free channel existing. Each point represents a particular instance of Markov process-based learning algorithm applied to $N = 4$ channels. The total number of possible combinations which we analyzed is $\binom{10}{4} = 210$, where 10 is the number of channels with $DC \in [0.3, 0.8]$

From the above results, we can conclude that, if possible, it is more beneficial to observe channels (and frequency bands) with higher probability of being free and lower complexity for opportunistic access.

2.6 Conclusions

In this chapter, we presented two learning algorithms that are used to predict the availability of a channel. Moreover, we analyzed the probability of success in finding an unoccupied channel for these algorithms. Our findings show that the performance of both algorithms strongly depends on the behavior of PUs. This means that, although the duty cycle of a channel is the metric usually taken into the account in previous works on learning algorithms applied to DCS, the complexity of PU behavior plays an equally important role.

We also showed that both learning algorithms performed similarly. Both algorithms performed the same for the combinations of channels with high probability of success, while the Markov process-based learning algorithm performed better on the channel combinations with low probability of success. The additional observed channel on Markov process-based scenarios might be the reason of its better performance on low quality channel combinations.

Acknowledgments This material is based upon works supported by the Science Foundation Ireland under Grants No. 10/CE/I1853 and 10/IN.1/I3007.

References

1. Macaluso, I., Forde, T., DaSilva, L., Doyle, L.: Recognition and informed exploitation of grey spectrum opportunities. IEEE Veh. Tech. Mag. **7**(2), 85–90 (2012)
2. Macaluso, I., Finn, D., Ozgul, B., DaSilva, L.: Complexity of spectrum activity and benefits of reinforcement learning for dynamic channel selection. IEEE J. Sel. Areas Commun. (Cogn. Radio Ser.) **31**(11), 2237–2248 (2013)
3. Ahmadi, H., Macaluso, I., DaSilva, L.: The effect of the spectrum opportunities diversity on opportunistic access. In: IEEE ICC'13 (2013)
4. Lempel, A., Ziv, J.: On the complexity of finite sequences. IEEE Trans. Inform. Theory **22**(1), 75–81 (1976)
5. Clancy, T., Walker, B.: Predictive dynamic spectrum access. In: SDR Forum's Technical Conference (2006)
6. Akbar, I., Tranter, W.: Dynamic spectrum allocation in cognitive radio using Hidden Markov Models: Poisson distributed case. In: IEEE SoutheastCon'.07 (2007)
7. Candy, J.: Bayesian Signal Processing: Classical, Modern, and Particle Filtering Methods. Wiley-Interscience, New York (2009)
8. Tumuluru, V., Wang, P., Niyato, D.: A neural network based spectrum prediction scheme for cognitive radio. In: IEEE ICC'10 (2010)
9. Tumuluru, V., Wang, P., Niyato, D.: Channel status prediction for cognitive radio networks. Wireless Commun. Mobile Comput. (2010)
10. Berthold, U., Fu, F., Van der Schaar, M., Jondral, F.: Detection of spectral resources in cognitive radios using reinforcement learning. In: New Frontiers in Dynamic Spectrum Access Networks, 2008. DySPAN 2008. 3rd IEEE Symposium on (2008)
11. Ahmadi, H., Chew, Y.H., Tang, P.K., Nijsure, Y.A.: Predictive opportunistic spectrum access using learning based hidden markov models. In: IEEE PIMRC'11 (2011)
12. Song, C., Zhang, Q.: Intelligent dynamic spectrum access assisted by channel usage prediction. In: IEEE INFOCOM'10 (2010)

13. Kone, V., Yang, L., Yang, X., Zhao, B., Zheng, H.: On the feasibility of effective opportunistic spectrum access. In: Proceedings of the 10th Annual Conference on Internet Measurement. ACM (2010)
14. Lindgren, K., Nordahl, M.: Complexity measures and cellular automata. Complex Syst. **2**(4), 409–440 (1988)
15. Grassberger, P.: Toward a quantitative theory of self-generated complexity. Int. J. Theor. Phys. **25**(9), 907–938 (1986)
16. Kaspar, F., Schuster, H.: Easily calculable measure for the complexity of spatiotemporal patterns. Phys. Rev. A **36**(2), 842–848 (1987)
17. Ziv, J.: Coding theorems for individual sequences. IEEE Trans. Inform. Theory **24**(4), 405–412 (1978)
18. Kolmogorov, A.: Three approaches to the definition of the concept quantity of information. Problemy Peredachi Informatsii **1**(1), 3–11 (1965)
19. Brudno, A.: Entropy and the complexity of the trajectories of a dynamical system. Trans. Moscow Math. Soc. **44**(2), 127–151 (1983)
20. Boffetta, G., Cencini, M., Falcioni, M., Vulpiani, A.: Predictability: a way to characterize complexity. Phys. Rep. **356**(6), 367–474 (2002)
21. Watkins, C., Dayan, P.: Q-learning. Mach. Learn. **8**(3), 279–292 (1992)

Chapter 3
Analysis and Optimization of Energy Detection for Non-Flat Spectral Characteristics

Sener Dikmese, Tero Ihalainen and Markku Renfors

3.1 Introduction

Spectrum sensing is one of the key elements of cognitive radio (CR) systems. Among various alternative spectrum sensing algorithms energy detection, also known as radiometer, is the most widely considered one. In spite of its shortcomings, notably its sensitivity to the uncertainty of the sensing receiver noise level, it is widely considered as a useful tool in cognitive radio and various other applications of radio scene analysis. Most of the studies on energy detection utilize a simplistic signal model, where the whole frequency band under sensing contains either noise only or noise plus a primary user (PU) signal, both with constant power spectral density (PSD). In practice, the sensing receiver has non-ideal frequency response, the transmitted PSD is not flat, the frequency-selective multipath channel has an effect on the received PU PSD, and the sensing window does not necessarily coincide with the frequency channel and time period of the PU transmission. The purpose of this chapter is to investigate the effects of these deviations from the basic model on the energy detector performance.

Energy detection is commonly formulated as a Neyman-Pearson type binary hypothesis testing problem, which is governed by probability distributions of the chi-square or gamma type [1, 2]. This model has also been extended to Nakagami and Rayleigh fading channels in [3]. Elaborating this model in the intended direction leads to elaborate analytical formulations, employing generalized forms of the

S. Dikmese (✉) · M. Renfors
Department of Electronics and Communications Engineering, Tampere University of Technology, P.O. Box 553, FI-33101 Tampere, Finland
e-mail: sener.dikmese@tut.fi

M. Renfors
e-mail: markku.renfors@tut.fi

T. Ihalainen
CTO Nokia Research Center, Nokia Corporation, Visiokatu 3, 33720 Tampere, Finland
e-mail: tero.ihalainen@nokia.com

M.-G. Di Benedetto and F. Bader (eds.), *Cognitive Communication and Cooperative HetNet Coexistence*, Signals and Communication Technology,
DOI: 10.1007/978-3-319-01402-9_3,

chi-square distribution. Instead of trying this, we utilize the widely used approach of Gaussian approximation of the actual probability distribution functions (PDFs) of the decision statistic [4]. This will be seen to lead to relatively simple and intuitive models for the considered energy detection problems. The results are validated through computer simulation based experimental results, showing clearly better agreement with the developed models than with the simplistic models where the frequency dependency is ignored. Similar methods have been recently used, for example, for analyzing linear cooperative spectrum sensing [5–7].

This chapter is organized as follows. Section 3.2 introduces briefly the basic energy detection model, as well as the idea of wideband, multi-mode spectrum sensing utilizing fast Fourier transform (FFT) or analysis filter bank for spectrum analysis. Then in Sect. 3.3, the effects of various forms of frequency dependency in energy detection are analyzed. First, the effect of non-ideal frequency response of the sensing receiver is analyzed and a simple metric for characterizing the effect is presented. Next, we address the case where the sensing window in time-frequency plane includes both zones where the PU signal is present and zones where it is absent, leading to mixed decision statistic. This model has applications in the detection of a reappearing PU during secondary transmission, as well as in spectrum sensing in situations where the frequency range of the PU signal is unknown. Moving-average processing is proposed as an effective way of handling these situations. Finally, the case of non-flat primary user spectrum is addressed, considering both the effects of non-flat transmitted PU spectrum and channel frequency selectivity.

3.2 Basic Energy Detection Schemes

3.2.1 Analytical Model

We start with the very basic spectrum sensing problem setting, which can be formulated as a binary hypothesis testing problem [1, 2]:

$$\begin{aligned} \mathscr{H}_0 &: \; y[n] = w[n] \\ \mathscr{H}_1 &: \; y[n] = s_{PU}[n] + w[n]. \end{aligned} \tag{3.1}$$

Here $y[n]$ is the signal observed by the sensing receiver. Under hypothesis $\mathscr{H}_0$, i.e., in the absence of the primary user, it consists only of zero-mean, complex, circularly symmetric, wide-sense stationary white (uncorrelated) Gaussian noise $w[n]$. Under hypothesis $\mathscr{H}_1$ the PU signal $s_{PU}[n]$ is present, along with the channel noise, and also the PU signal is modeled as as zero-mean, complex, circularly symmetric, wide-sense stationary white Gaussian sequence. In energy detection, the test statistic is obtained as

$$T(y) = \frac{1}{N} \sum_{n=0}^{N-1} \left| y[n] \right|^2, \tag{3.2}$$

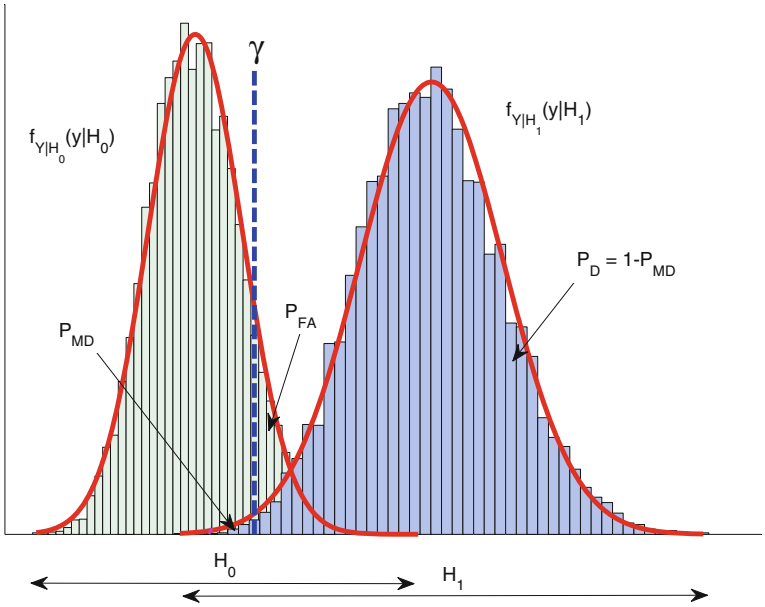

Fig. 3.1 Illustration of the binary hypothesis testing problem assuming Gaussian approximation of the test statistic. Also examples of experimental probability distributions are shown

where N is the length of the observation sequence, referred to as sample complexity. For tractability of the later developments, we adopt the Gaussian approximation for the probability distributions [4]:

$$\begin{aligned} T(y)|_{\mathscr{H}_0} &\sim \mathscr{N}\left(\sigma_n^2, \tfrac{1}{N}\sigma_n^4\right) \\ T(y)|_{\mathscr{H}_1} &\sim \mathscr{N}\left(\sigma_{PU}^2 + \sigma_n^2, \tfrac{1}{N}(\sigma_{PU}^2 + \sigma_n^2)^2\right). \end{aligned} \tag{3.3}$$

Here σ_{PU}^2 and σ_n^2 are the variances of the PU signal and noise sequences, respectively. The hypothesis testing problem is illustrated in Fig. 3.1. The false alarm probability P_{FA} and detection probability P_D, can be expressed as:

$$P_{FA} = \Pr(T(y) > \lambda|\mathscr{H}_0) = Q\left(\frac{\lambda - \sigma_n^2}{\sqrt{\sigma_n^4/N}}\right) \tag{3.4}$$

$$P_D = \Pr(T(y) > \lambda|\mathscr{H}_1) = Q\left(\frac{\lambda - (\sigma_n^2 + \sigma_{PU}^2)}{\sqrt{(\sigma_n^2 + \sigma_{PU}^2)^2/N}}\right), \tag{3.5}$$

where $Q(\cdot)$ is the standard Gaussian complementary cumulative distribution function. The PU signal variance cannot be assumed to be known due to the unknown

channel characteristics. Hence, the threshold value, λ, is commonly calculated based on the assumed noise variance and targeted false alarm probability as follows:

$$\lambda = Q^{-1}(P_{FA})\sqrt{\sigma_n^4/N} + \sigma_n^2. \tag{3.6}$$

The sample complexity N is chosen in such a way that the missed detection probability, $P_{MD} = 1 - P_D$, becomes small enough for the minimum PU power level expected to be detectable.

We make use of the equations in [4], that relate the sample complexity to P_{FA}, P_{MD}, and the operating $\text{SNR} = \sigma_{PU}^2/\sigma_n^2$. In case there is no uncertainty and the noise variance is completely known, the required sensing time in samples, $\hat{N}$, to achieve target P_{FA} and P_{MD} writes:

$$\hat{N} = \frac{\left[Q^{-1}(P_{FA}) - Q^{-1}(1 - P_{MD})(1 + SNR)\right]^2}{SNR^2}. \tag{3.7}$$

On the other hand, when the energy detector is assumed to operate under a noise level uncertainty of $x = \pm 10log_{10}\rho$ dB, the sample complexity can be approximated [4] according to

$$\hat{N} \approx \frac{\left[Q^{-1}(P_{FA}) - Q^{-1}(1 - P_{MD})\right]^2}{\left[SNR - (\rho - \frac{1}{\rho})\right]^2}. \tag{3.8}$$

This introduces the so-called SNR wall: For example, with 0.1 dB uncertainty of the noise variance, the sensing time grows without limits when the SNR approaches -13.3 dB. An experimental study of the effects of noise uncertainty can be found in [8]. The noise power estimation problem is addressed in [9], indicating ways to mitigate the SNR wall effects.

3.2.2 FFT and Filter Bank Based Flexible Schemes for Multiband Sensing

Most of the latest broadband wireless systems are based on multicarrier modulation. Especially, cyclic prefix based orthogonal frequency division multiplexing (CP-OFDM) techniques are characterized by simplicity and robustness of the receiver signal processing functions. However, alternative multicarrier techniques have been considered increasingly in the recent developments. Especially, filter bank multicarrier (FBMC) techniques have been shown to have various potential benefits in the cognitive radio context [10, 11]. When considering multicarrier techniques for CR, we can make use of the existing signal processing blocks (the FFT of an OFDM receiver or the analysis filter bank (AFB) of an FBMC receiver) for spectrum sensing

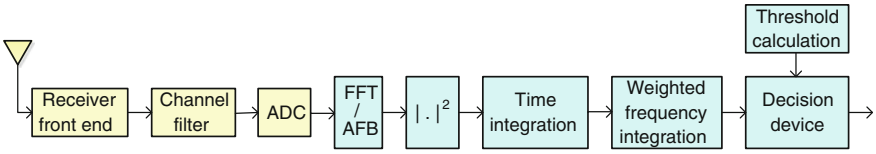

Fig. 3.2 Block diagram for energy detection with FFT/AFB based spectrum analysis

purposes. A block diagram for FFT or AFB based energy detection algorithms with weighting process (to be discussed in Sect. 3.3) is shown in Fig. 3.2.

The main benefits of FBMC waveforms over OFDM include good spectral localization of the transmitted signal spectrum, as well as high frequency selectivity of receiver signal processing. These characteristics help to improve the interference control especially in scenarios where the users are not precisely synchronized with each other. From the spectrum sensing point of view, the benefits of AFB are prominent in high-dynamic-range scenarios, i.e., when performing spectrum sensing in the presence of strong transmissions at nearby frequencies. In these cases, the spectrum leakage of basic FFT processing degrades the sensing performance significantly [10–12]. However, in the considerations of this chapter the differences between FFT and AFB based methods are not essential.

We focus here on spectrum sensing in a wideband multichannel receiver, where energy detection is carried out at subband level at the output of an FFT or AFB, which is used for splitting the received signal into relatively narrow signal bands. The output of these blocks is expressed as $y_k[m]$ where $k = 0, \ldots, K-1$ is the subband index and m is the subband sample index. Typically, the sampling rate of the subbands is equal to the analog-to-digital converter (ADC) sampling rate divided by the number of subbands in the filter bank, K, in which case the system is critically-sampled. In case of FBMC/OQAM (offset quadrature amplitude modulation) [13, 14], 2x oversampling of the subband signals is also commonly considered. In the spectrum sensing context, the subband signals can be expressed as

$$y_k[m] = w_k[m] \qquad \mathscr{H}_0 \tag{3.9a}$$

$$y_k[m] = s_k[m] + w_k[m] \qquad \mathscr{H}_1. \tag{3.9b}$$

Here $s_k[m]$ is the transmitted signal by a primary user as it is present at the mth FFT or AFB output sample in subband k, and $w_k[m]$ is the corresponding noise sample. When additive white Gaussian noise (AWGN) only is present, the noise in each subband is modeled as a zero-mean Gaussian random variable with variance σ_n^2, i.e., $w_k[m] \sim \mathscr{N}(0, \sigma_n^2)$. The PU signal can also be modeled as a zero-mean Gaussian variable $s_k[m] \sim \mathscr{N}(0, \sigma_k^2)$ where, σ_k^2 is the PU signal variance (power) in subband k.

The subband-wise test statistics can be obtained in the same way as in Eq. (3.2). However, it is possible to calculate the test statistic over multiple subbands and over a certain subband sample interval:

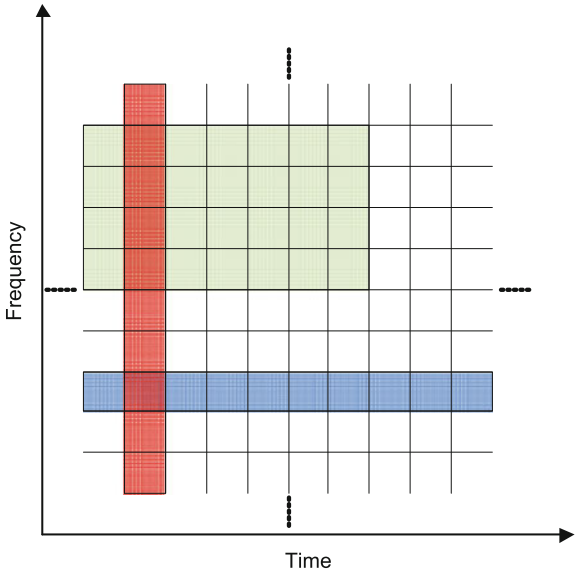

Fig. 3.3 Illustration of different integration zones in time-frequency plane

$$T(y_{m_0,k_0}) = \frac{1}{N_t N_f} \sum_{k=k_0-\lfloor N_f/2 \rfloor}^{k_0+\lceil N_f/2 \rceil-1} \sum_{m=m_0-N_t+1}^{m_0} |y_k[m]|^2. \tag{3.10}$$

Here N_f and N_t are the averaging filter lengths in frequency and time, respectively. It follows from Parseval's theorem that for a spectral component fully covered by the subband integration range, FFT based subband integration and full-band time-domain integration over the same time interval give the same decision statistic. In this case $N = N_f N_t$. The same applies also to critically-sampled orthogonal filter banks. Assuming that the PU spectrum is flat over the sensing band, the probability distribution of the decision statistic can be modeled as follows:

$$T(y_{m_0,k_0})|_{\mathscr{H}_0} \sim \mathscr{N}\left(\sigma_n^2, \frac{1}{N_t N_f}\sigma_n^4\right) \tag{3.11a}$$

$$T(y_{m_0,k_0})|_{\mathscr{H}_1} \sim \mathscr{N}\left(\sigma_{PU}^2 + \sigma_n^2, \frac{1}{N_t N_f}(\sigma_{PU}^2 + \sigma_n^2)^2\right). \tag{3.11b}$$

The threshold value λ can be obtained from Eq. (3.6) with $N = N_f N_t$ and the false alarm probability can be obtained from Eq. (3.4).

FFT/AFB processing makes it possible to tune the sensing frequency band to the expected band of the PU signal, as well as sensing multiple PU bands simultaneously. Figure 3.3 illustrates the tradeoffs in choosing the integration range in the time-frequency plane.

In a cognitive radio system, spectrum sensing is needed both for initial sensing for a transmission opportunity, as well as for monitoring of the spectrum during SU system operation, in order to detect rapidly a possible reappearing PU. Multicarrier transmission techniques have a high commonality of receiver processing and spectrum sensing as selected time-frequency zones can be left unused for spectral monitoring purposes, while data transmission continues on the others [15].

3.3 Energy Detection in the Presence of Frequency Variability

Our target in this section is to develop tractable models for energy detection in cases where the PU signal and/or noise are not white within the sensing frequency band. The spectrum of the PU signal part is determined by the spectrum of the transmitted waveform, channel frequency response and the sensing receiver filter, whereas the spectrum of the channel noise depends on the receiver filter frequency response only. First, a generic analytic model for the decision statistic is developed. This model is then applied for different scenarios in later sections.

3.3.1 Analytical Modeling

We start with a basic setup where a complex, zero-mean, wide-sense stationary, non-white Gaussian distributed signal $y[n]$ with variance σ_y^2 is entering an energy detector. The goal is to derive the mean and variance of the test statistic calculated according to Eq. (3.2). Our approach is based on splitting signal $y[n]$ to K approximately frequency-flat subbands using FFT or orthogonal AFB. Then the test statistics calculated directly for $y[n]$ and from the subband samples $y_k[m], k = 0, \ldots, K-1$, $m = m_0, \ldots, m_0 + N/K - 1$ (assuming that N is an integer multiple of K) are the same.

Now the subband sample sequences can be assumed to be independent and each of them is wide-sense stationary, zero-mean, and can be assumed to be white with variance $\sigma_{y_k}^2$. The overall test statistic is obtained as a combination of the subband-wise test statistics. If we use Gaussian approximation for the subband-wise test statistics, then the overall test statistic can be interpreted as a linear combination of Gaussian variables. Its mean and variance can be expressed as:

Table 3.1 Theoretical versus experimental β-values

	Theoretical β	Experimental β		
		SNR = 0 dB	SNR = −6 dB	SNR = −99 dB
FIR1	1.24	1.15	1.22	1.24
FIR2	1.57	1.48	1.56	1.57
ITU-R Veh A channel 1	1.31	1.14	1.28	1.31
ITU-R Veh A channel 2	2.38	2.18	2.33	2.36

$$\mu_{T(y)} = \frac{1}{K}\sum_{k=0}^{K-1}\mu_{T(y_k)} = \frac{1}{K}\sum_{k=0}^{K-1}\sigma_{y_k}^2 = \sigma_y^2 \tag{3.12a}$$

$$\begin{aligned}\sigma_{T(y)}^2 &= \frac{1}{K^2}\sum_{k=0}^{K-1}\sigma_{T(y_k)}^2 = \frac{1}{K^2}\sum_{k=0}^{K-1}\frac{1}{N/K}\sigma_{y_k}^4 = \frac{1}{NK}\sum_{k=0}^{K-1}\sigma_{y_k}^4 \\ &= \frac{1}{N}\sigma_y^4 \cdot \frac{1}{K}\sum_{k=0}^{K-1}\left(\frac{\sigma_{y_k}}{\sigma_y}\right)^4.\end{aligned} \tag{3.12b}$$

Based on the Central Limit Theorem [1], the true distribution approaches the Gaussian approximation when the number of subband samples grows.

We can express the distribution of the test statistic as

$$T(y) \sim \mathcal{N}\left(\sigma_y^2, \frac{\beta}{N}\sigma_y^4\right) \tag{3.13}$$

where

$$\beta = \frac{1}{K}\sum_{k=0}^{K-1}\left(\frac{\sigma_{y_k}}{\sigma_y}\right)^4. \tag{3.14}$$

From this formulation, we can see that the value of β characterizes the effect of non-flat signal spectrum on the required sample complexity for reaching certain level of test statistic variance. Clearly, the minimum value of β is one, so the required sample complexity is minimized when the signal is white.

Table 3.1 shows a comparison of the theoretical β-values versus experimental β-values in example cases, including finite impulse response (FIR) filters and frequency selective channel responses. The experimental β-values are obtained as the ratio of the simulated test statistic variance to the theoretical test statistic variance without frequency dependency. The simulated variances are calculated from 100000 simulation instances with the same filter or channel frequency response but independent 16-QAM-modulated data sequences and noise sequences, using 1k FFT length and sample complexity of 1000 samples.

The FIR filters are two cases of the decimation-by-2 filters considered in Sect. 3.3.2, one with order 10 and $\beta = 1.57$, the other with order 19 and $\beta = 1.24$.

The used channel model is ITU-R Vehicular A model with 30.72 MHz sampling rate and 78 taps. Two different channel instances are shown in Table 3.1. Checking 100000 independent channel instances, the range of theoretical β-values was [1.13, 2.47] and in 90 % of the cases it is in the range [1.47, 1.99]. A general observation from Table 3.1 is that for white Gaussian input signal (the SNR $= -99$ dB case), the developed theoretical model applies quite well in the considered cases. The PU signal is modeled as 16-QAM signal, which does not follow the Gaussian distribution. For the 0 dB SNR, the experimental β-value is significantly smaller than the theoretical one, i.e., the theoretical model is pessimistic in terms of sensing performance. With SNR $= -6$ dB, the match is already rather good.

3.3.2 Effects of the Frequency Response of Sensing Receiver

Here we focus on analyzing the effects of non-ideal sensing receiver frequency response on the performance of energy detection. Any practical receiver would include a chain of multiple filtering stages, but for clarity of discussion, we assume a single stage model, referred to as sensing filter. Depending on the relative bandwidths of the PU signal and sensing filter, different scenarios need to be considered. Throughout this section, we assume an idealized model where the channel noise is white and PU signals have ideally band-limited spectra, i.e., constant power spectrum in the passband and stepwise transition bands.

3.3.2.1 Partial Band Sensing

The most simple situation to analyze is the case where both the noise and PU signal are white in the considered sensing band. This would be applicable, for example, for the case where a relatively narrow frequency slot between active SUs is used for spectrum monitoring. We assume that the signal power level is scaled to be constant under different choices of the sensing filter.

Now the sensing filter frequency response affects the observed signal in the same way under $\mathscr{H}_0$ and $\mathscr{H}_1$. More precisely, the variances of the test statistics are affected by the same factor β, while the means of the test statistic are not affected. Then the sample complexity needs to be increased by the factor of β to compensate the effects of the sensing filter.

Let us assume that after down-conversion by the receiver front-end, the sensing frequency band is located in the angular frequency range $[-\omega_c, \omega_c]$. Then the ideal solution would be an ideal lowpass filter with cut-off frequency of ω_c. FFT or AFB processing can be used for approximating well such an ideal sensing filter. On the other hand, we can use the subband model as a tool for analyzing and optimizing basic time-domain sensing filter designs. Using the developed model, we can evaluate the impact of practical filter design on the required sample complexity, trying to find a good compromise between the filter complexity and sensing performance. Let us

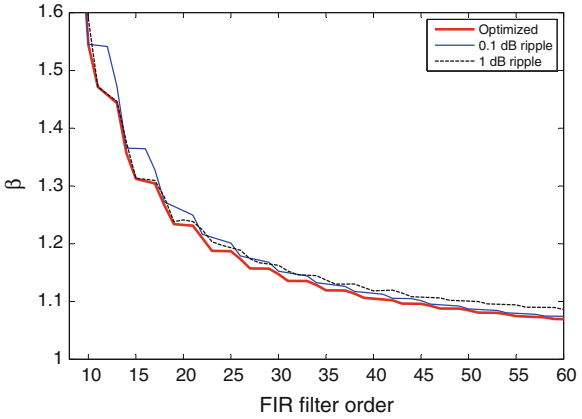

Fig. 3.4 β as a function of the sensing filter order when using linear-phase minimax-optimized FIR filters

assume that $F_k, k = 0, \ldots, K-1$ are uniformly spaced samples of the amplitude response of the sensing filter in the range $[-\omega_c, \omega_c]$. To normalize the signal variance, we assume that

$$\sum_{k=0}^{K-1} F_k^2 = K. \tag{3.15}$$

Then we obtain:

$$\beta = \frac{1}{K} \sum_{k=0}^{K-1} F_k^4. \tag{3.16}$$

As an example, we consider how the β-value depends on the filter order and frequency response parameters when using linear-phase FIR filters designed using the Parks-McClellan algorithm, i.e., equiripple passband and stopband. We assume that the sensing bandwidth is half of the sampling rate, i.e., $\omega_c = \pi/2$ and that 40 dB stopband attenuation is required to suppress the spectral components outside the sensing band to a level which does not effect the sensing performance. Figure 3.4 shows the results for three cases: (i) ± 0.1 dB passband ripple, (ii) ± 1 dB passband ripple, (iii) optimized case. The optimized case is found by exhaustive search covering different combinations of passband edge frequency and passband ripple. We can see that for low filter orders, the sensing performance is rather sensitive to the filter design.

As another example, we consider subband-wise sensing using an AFB. In [15] the idea of using narrow sensing subbands interlaced with data subcarriers was proposed for spectrum monitoring purposes. In FBMC/OQAM with typical spectrally well-contained prototype filter design, a group of three unused subcarriers provides an empty gap of two subcarrier spacings between active groups of secondary users'

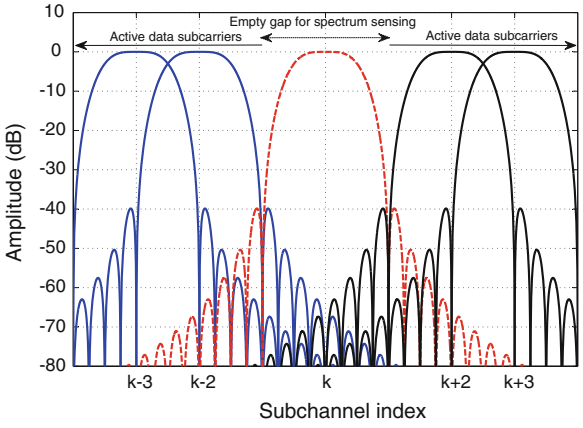

Fig. 3.5 Spectrum monitoring scenario using a single-subband sensing filter in a two-subband-wide spectrum gap

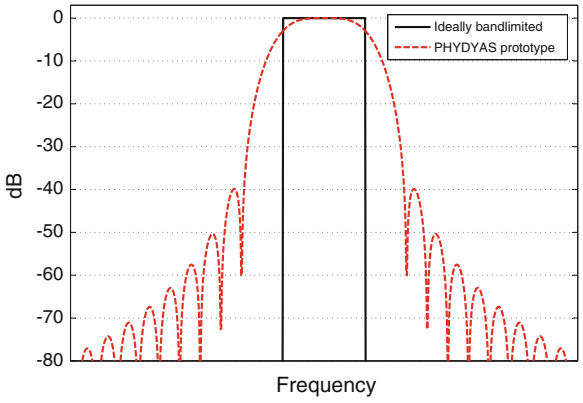

Fig. 3.6 AFB subband frequency response in relation with an ideal subband filter with the same subband spacing

data carriers, as shown in Fig. 3.5. Then the center subband can be used for spectrum sensing during data reception without interference from adjacent groups of active data subcarriers. Figure 3.6 shows the frequency response of the PHYDYAS filter bank prototype filter [13, 16] in relation to the response of an ideal lowpass filter with the same subchannel spacing. This prototype response is the baseband equivalent of all subband filters of the FBMC receiver. Since the filter response is far from the ideal one, it is important to examine its effect on the spectrum sensing performance.

In the FBMC/OQAM receiver, subchannel processing (e.g., fractionally-spaced equalization / frequency domain fine synchronization) is carried out at 2x oversampled rate [13, 16]. In other words, complex-valued subchannel samples are processed

instead of the real-valued ones, which are finally sufficient for data reception at the end of the subchannel processing chain. In the spectrum sensing context, it is interesting to evaluate the possible benefit from using oversampled subband signals instead of critically-sampled ones.

We assume that both the noise and PU signal power spectra are flat withing the sensing gap. Let's start with the case of ideally lowpass filtered and critically-sampled subband signals. In this case we obtain N_t independent uncorrelated observations per subband. The same applies also to the critically-sampled AFB subband observations, as the subband filter is of the root-raised cosine (RRC) type.

In case of 2x oversampled subband signals, and ideal sensing filter would have double bandwidth comparing with the critically-sampled case and the sample complexity would be doubled as well. In order to analyze the effect of non-ideal subband filter, we can apply the model developed above. By sampling the subband frequency response at K uniformly spaced frequencies, we can calculate using Eq. (3.16) that for the subband filter $\beta = 1.646$. This means that the required sample complexity using the oversampled subband sequence is about 1.65 times higher than when using an ideal filter with 2x oversampling. On the other hand, the required sample complexity with oversampled subband signal is 0.82 times the required sample complexity with critically-sampled subband signal [17]. It is clear that the use of oversampled subband sequences provides significant benefit only in single subband wide sensing, and possibly also in edge subcarriers when the integration range includes a few subcarriers.

This analysis has been validated experimentally, through simulation based complementary receiver operation characteristic (CROC) plots shown in Fig. 3.7. A CROC plot represents the detection probability, P_D, as a function of P_{FA}. We can see that the experimental data, obtained from an FBMC test-bed (with PHYDYAS prototype filter) exploiting sensing subband based energy detection, matches quite well with the analytical model ($\beta = 0.82$), much better than with the simplistic basic model with $\beta = 1$.

3.3.2.2 Band Edge Detection and Transmission Burst Detection

Typically, it is assumed that the center frequency and the bandwidth of the PU signal are known. Further, it is commonly assumed that the PU is either absent or active during the whole sensing interval used for constructing the test statistic. In practice, it may happen that the PU becomes active during the measurement period (reappearing PU) [18], or the sensing frequency band does not match the frequency band of the PU signal. In such cases, only some fraction of the integration window matches the time-frequency zone of the PU activity. Such a transient phase, in time or frequency direction, and the related test statistic distributions are illustrated in Fig. 3.8 [15].

The distribution of the transient phase test statistic, $T(y)\big|_{TR}$, can be derived as follows. The integration window can virtually be split into two distinct sub-windows such that one of them contains only those observation samples before PU becomes active ($N - N_1$ samples; see Fig. 3.8) while the other one contains the rest N_1

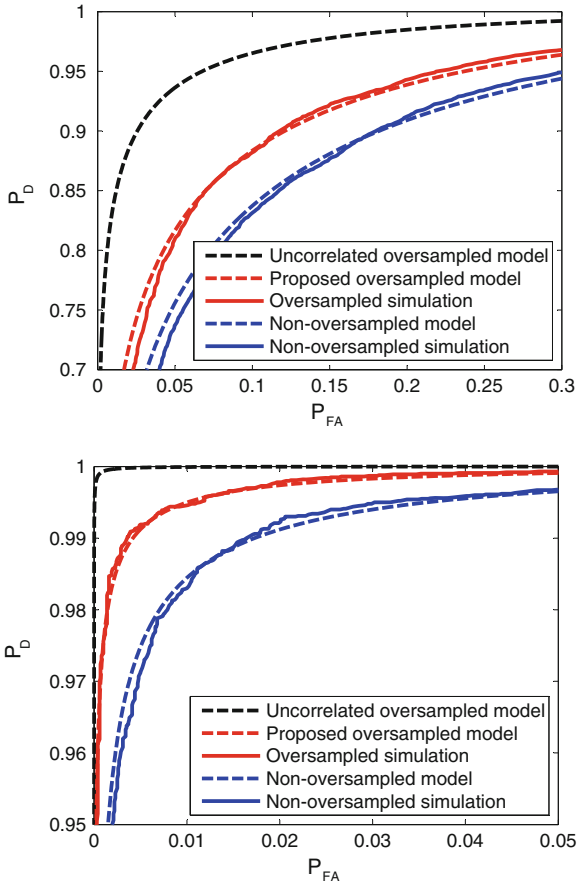

Fig. 3.7 Verification of the analytical model of sensing filter effects in single subband sensing using CROC plots for PU $SNR|_{H1} = -6\,\text{dB}$ and 10000 independent trials. Critical sampling and 2x oversampling with respect to the subband spacing. Sample complexities 100 (*upper*) and 400 (*lower*) critically-spaced samples

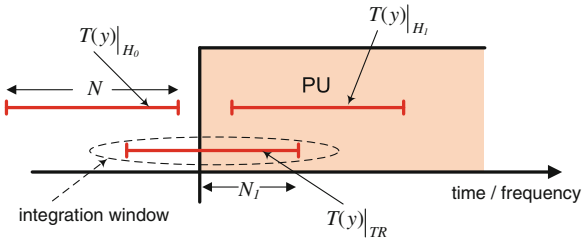

Fig. 3.8 A model for the distribution analysis of the transient phase test statistic

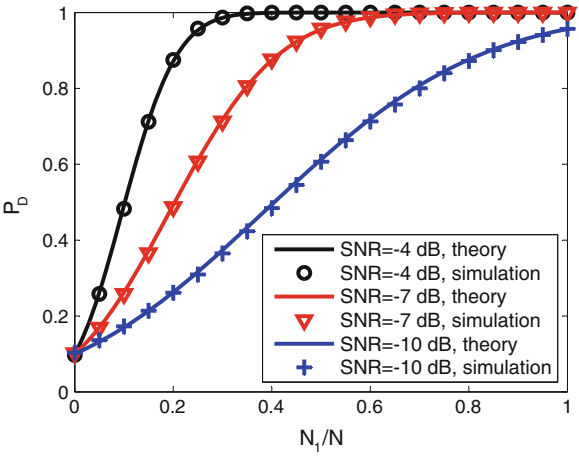

Fig. 3.9 Comparison between the analytic and experimental transient phase detection performance. The case with $P_{FA} = 0.1$, sample complexity of $N = 1000$ and 10000 independent trials

samples. The distributions corresponding to the sample subsets within these virtual sub-windows write $\mathcal{N}\left(\sigma_n^2, \frac{1}{N-N_1}\sigma_n^4\right)$ and $\mathcal{N}\left(\sigma_{PU}^2+\sigma_n^2, \frac{1}{N_1}(\sigma_{PU}^2+\sigma_n^2)^2\right)$, respectively. The overall sequence of N samples can be interpreted as a linear combination of these independent normal random variables using relative weights of $\frac{N-N_1}{N}$ and $\frac{N_1}{N}$, respectively. Next, by making use of the following property of the normal distribution: given

$$\begin{cases} X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2) \\ X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2) \\ \quad \Rightarrow \quad aX_1 + bX_2 \sim \mathcal{N}(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2) \end{cases} \tag{3.17}$$

we obtain an expression for the mixture-distribution as

$$T(y)\big|_{TR} \sim \mathcal{N}\left(\sigma_n^2 + \frac{N_1}{N}\sigma_{PU}^2, \frac{1}{N}\left((1-\frac{N_1}{N})\sigma_n^4 + \frac{N_1}{N}(\sigma_{PU}^2+\sigma_n^2)^2\right)\right). \tag{3.18}$$

Figure 3.9 shows that the developed model matches quite well with experimental transient phase energy detection performance. One effective way to handle such transient situations (i.e., unknown PU channel band edges and/or transmission burst timing) is moving average filtering. In FFT or AFB based energy detection, moving average processing can be carried out in time and/or frequency direction to locate the PU transmission in time or frequency.

3.3.3 Effects of Non-Flat Primary User Spectrum

3.3.3.1 Generic Analysis

In the flexible multiband sensing scheme of Fig. 3.2, after the receiver front-end and ADC, FFT or AFB is used to split the signal into relatively narrow subbands. Depending on the bandwidth of the candidate PU signal, a number of consecutive subbands is combined in the sensing process, possibly after weighting to optimize the energy detection performance for non-flat PU signals. Two different ideas are applied for the weighting process. (i) Constant weights are optimal for a PU signal with flat power spectrum, and they may provide a good approximation also for a non-flat spectrum, if the number of subbands is selected properly. This is also a natural choice if there is no prior information about the PU spectrum. (ii) Optimized weights can be found if there is prior information about the PU power spectrum. Naturally, a frequency selective channel has an effect on the spectrum of the received signal, and the channel cannot be assumed to be known. If the channel can be assumed to be flat fading, then the power spectrum of the received signal has the same shape as that of the transmitted signal, and weighting process can be used. Even in case of frequency selective channel, weights derived from the transmitted spectrum give some gain in average performance according to the simulation results.

Under both $\mathscr{H}_0$ and $\mathscr{H}_1$ the decision statistic is approximated as a sum of independent Gaussian variables with different variances, and the resulting distributions can again be assumed to be Gaussian with zero mean. Hence the probability distribution of the decision statistic T_κ for center frequency κ can be written for arbitrary weighting coefficients as

$$\begin{aligned} f(T_\kappa)|_{\mathscr{H}_0} &\sim \mathscr{N}\left(\sum_{k=\kappa-\overline{N_f}}^{\kappa+\overline{N_f}} w_k^2 \sigma_n^2,\ \frac{1}{N} \sum_{k=\kappa-\overline{N_f}}^{\kappa+\overline{N_f}} w_k^4 \sigma_n^4 \right) \\ f(T_\kappa)|_{\mathscr{H}_1} &\sim \mathscr{N}\left(\sum_{k=\kappa-\overline{N_f}}^{\kappa+\overline{N_f}} w_k^2 (\sigma_k^2 + \sigma_n^2),\ \frac{1}{N} \sum_{k=\kappa-\overline{N_f}}^{\kappa+\overline{N_f}} w_k^4 (\sigma_k^2 + \sigma_n^2)^2 \right). \end{aligned} \tag{3.19}$$

Here $N = N_t N_f$ is the overall sample complexity, as in Sect. 3.2.2. To simplify the notation, we assume that the window size in frequency direction is odd, $N_f = 2\overline{N_f} + 1$. The integration in frequency direction takes the weighted average of the time filter outputs, with the weight value w_k for subband k. The PU signal power in subband k is denoted by σ_k^2. Next we address the problem of optimizing the subband weights.

For arbitrary weight values, the false alarm probability P_{FA} is formulated as [12]

$$P_{FA} = Q\left(\frac{\lambda - \sum_{k=\kappa-\overline{N_f}}^{\kappa+\overline{N_f}} w_k^2 \sigma_n^2}{\sqrt{\frac{1}{N}\sum_{k=\kappa-\overline{N_f}}^{\kappa+\overline{N_f}} w_k^4 \sigma_n^4}}\right). \tag{3.20}$$

When the signal is present, probability of detection P_D with arbitrary weight values can be expressed as

$$P_D = Q\left(\frac{\lambda - \sum_{k=\kappa-\overline{N_f}}^{\kappa+\overline{N_f}} w_k^2 (\sigma_n^2 + \sigma_k^2)}{\sqrt{\frac{1}{N}\sum_{k=\kappa-\overline{N_f}}^{\kappa+\overline{N_f}} w_k^4 \left(\sigma_n^2 + \sigma_k^2\right)^2}}\right). \tag{3.21}$$

The threshold value λ can be calculated as

$$\lambda = Q^{-1}\left(P_D\right)\sqrt{\frac{1}{N}\sum_{k=\kappa-\overline{N_f}}^{\kappa+\overline{N_f}} w_k^4 \left(\sigma_n^2\left(1 + SNR_k\right)\right)^2} + \sum_{k=\kappa-\overline{N_f}}^{\kappa+\overline{N_f}} w_k^2 \left(\sigma_n^2\left(1+SNR_k\right)\right). \tag{3.22}$$

Here $SNR_k = \sigma_k^2/\sigma_n^2$. To obtain the optimum threshold value λ, it is required to have the knowledge of the noise power and subband powers of the primary signal. Then the required sample complexity for given detection probability can be calculated as

$$\hat{N} = \frac{\left[Q^{-1}\left(P_{FA}\right)\sqrt{\sum_{k=\kappa-\overline{N_f}}^{\kappa+\overline{N_f}} w_k^4 \sigma_n^4} - Q^{-1}\left(P_D\right)\sqrt{\sum_{k=\kappa-\overline{N_f}}^{\kappa+\overline{N_f}} w_k^4 \left(\sigma_n^2\left(1 + SNR_k\right)\right)^2}\right]^2}{\left[\sum_{k=\kappa-\overline{N_f}}^{\kappa+\overline{N_f}} w_k^2 \sigma_n^2 SNR_k\right]^2}. \tag{3.23}$$

In practice, the required time record length N_t is determined based on the targeted minimum detectable PU power level. When using optimum weighting coefficients, the frequency block length N_f should be chosen to include all subbands which essentially contribute to the decision statistics. Including extra subbands adds complexity

but should not harm the sensing performance, because the corresponding weights become small.

Assuming that the weights are normalized for constant noise power level (i.e., $\sum w_i^2 = 1$), and taking partial derivatives of Eq. (3.23) with respect to the weights, the optimum weight values can be obtained as:

$$w_k{}^2 = \frac{1}{\sum\limits_{k=\kappa-\overline{N_f}}^{\kappa+\overline{N_f}} \sigma_k^2} \sigma_k{}^2 = \frac{\sigma_k^2}{\sigma_{PU}^2}. \tag{3.24}$$

where σ_{PU}^2 is the total power of the PU signal.

It is not a surprise that this solution can be seen as an approximation of matched filtering [19] in the sense that squared magnitude response of the sensing filter is approximately the same as the transmitted power spectrum [8]. Alternatively, this can be seen as maximum ratio combining of statistically independent variables. Both models result in the maximization of the SNR in the presence of PU.

It is interesting to consider the gain of using optimum weights instead of constant weights, which would be a conceptually simpler approach, even though the difference in implementation complexity is minor. In case of constant weights, it is important to choose optimum number of subbands. Subbands with small PU power level contribute to the noise variance under $\mathscr{H}_0$, but might actually reduce the SNR under $\mathscr{H}_1$, thus reducing the sensing performance. Given the PSD of the PU, the optimum number of subbands can be found numerically based on Eq. (3.23) for target values of P_{FA}, P_D, and PU SNR.

3.3.3.2 Bluetooth Case

Frequency selective channel has to be assumed in most of the practical wireless communication scenarios, and in those cases, knowledge of the received PU signal power spectrum cannot be assumed. However, for certain short-range and/or narrowband cases, flat-fading channel model is commonly assumed. Bluetooth is one important example of such systems [20].

The ISM band around 2.4 GHz is an unlicensed frequency band which is utilized by various applications, including 802.15 based Bluetooth (BT) devices, cordless phones, WLAN signals, and even microwave ovens introduce interferences on certain part of this band. The frequency hopped frequency shift keying (FH-FSK) -based 802.15 Bluetooth signal has 79 different frequency channels at center frequencies starting from 2.402 GHz and ending at 2.480 GHz, with 1 MHz spacing. The nominal bandwidth of BT signal is 1 MHz and the hopping rate is 1600 hops/sec. We consider first a simplified scheme with continuous BT signal at the 33rd channel. The corresponding BT signal spectrum can be seen in Fig. 3.10.

In the following, we consider the performance of the multiband spectrum sensing scheme with weighting process to sense a BT signal and spectral holes. Assuming

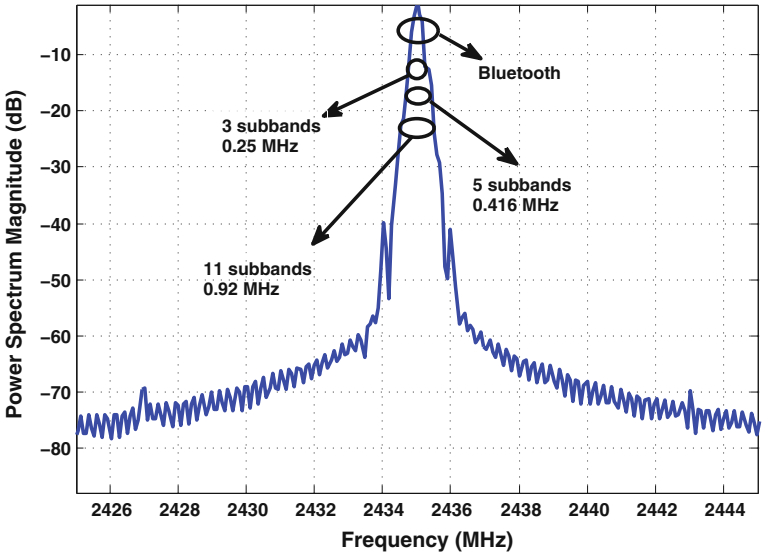

Fig. 3.10 Bluetooth signal spectrum in 2.4 GHz ISM band

challenging spectrum sensing scenario with low SNR (negative in dBs), the spectral dynamic range is very small, and there are hardly any differences between FFT and AFB. Therefore, we show the results only for FFT. We assume 83.5 MHz sampling rate and 1024 subbands covering the ISM band. For 1 MHz sensing bandwidth of the Bluetooth signal, the sensing window consists of 11 subbands. In this example, we target at −5 dB SNR in spectrum sensing, false alarm probability of 0.1 and detection probability of 90 %.

With a simplified, non-frequency-selective model, the required sample complexity for the target SNR, P_D and P_{FA} is obtained from Eq. (3.7) as $\hat{N} = 89$. If the BT power was equally distributed within the 11 subbands, 8 samples from each of these subbands would almost reach the target. When using just the center subband for sensing, 89 is a lower bound for the time record length. Table 3.2 shows the number of required subband samples needed with the BT signal for different sensing bandwidths and with optimized and constant weights. For instance, when the optimum weight values are used for all 11 subbands of the Bluetooth signal, the required number of subband samples is calculated as 45, which corresponds to a lower time record length than the hoping interval (50 samples). Since most of the BT energy is concentrated on the center subbands, we may consider using only 3 or 5 subbands in sensing (0.25 or 0.42 MHz bandwidth, respectively), and the corresponding results are also seen in Table 3.2, both with optimum weighting and constant weights.

We can see that almost the same time record length can be used when sensing a single subband at the BT center frequency as when sensing the whole 1 MHz BT band with constant weights. With constant weights, 3 subbands is the optimum choice for

Table 3.2 Required time record length in subband samples for different weighting schemes in bluetooth sensing

Bluetooth SNR (dB)	Weight factors	11 subbands	5 subbands	3 subbands	1 subband
0	Constant	12	8	8	15
	Optimum	7	7	8	
−3	Constant	39	25	23	41
	Optimum	21	21	23	
−4	Constant	60	37	34	60
	Optimum	31	31	33	
−5	Constant	92	56	51	89
	Optimum	45	46	49	
−6	Constant	143	84	77	132
	Optimum	68	69	73	
−7	Constant	223	131	117	200
	Optimum	104	105	111	

all the considered SNR values. Using optimum weights, the sensing time can be reduced by about 10 %, and most of this benefit is gained by using only 5 subbands.

3.3.3.3 Simulation Results

Figure 3.11 shows the Receiver Operating Characteristic (ROC) curves in Bluetooth sensing for analytic and FFT based simulation models with constant weight values and optimum weight values. The time record length is chosen as 50 samples, due to the hopping limit corresponding to approximately 625 μs. The simulated performance is slightly worse than the analytical model. This is probably due to the poor frequency localization of FFT based spectrum analysis. To compensate the difference, the time record length should be increased by about 10 %. But we are still able to reach false alarm and detection probabilities of 0.1 and 0.9 at −5 dB SNR with 50 samples.

Figure 3.12 shows the detection probability with constant and optimum weight values as a function of the active Bluetooth SNR for AWGN channel. The time record length is again 50 samples and the desired false alarm and detection probabilities are chosen as 0.1 and 0.9, respectively. The constant weight case uses 3 subbands, which provides the best detection probability performance. With optimum weight values using 11 subbands, the highest detection probability performance is achieved again, but the benefit over the constant weight case is marginal. In this low dynamic range case, there is no big difference in detection probability between FFT and AFB.

3.3.3.4 Moving Average Processing for Bluetooth

In the previous discussion we have ignored the frequency hopping characteristic of the BT system. One critical issue is that if the time record for calculating the decision

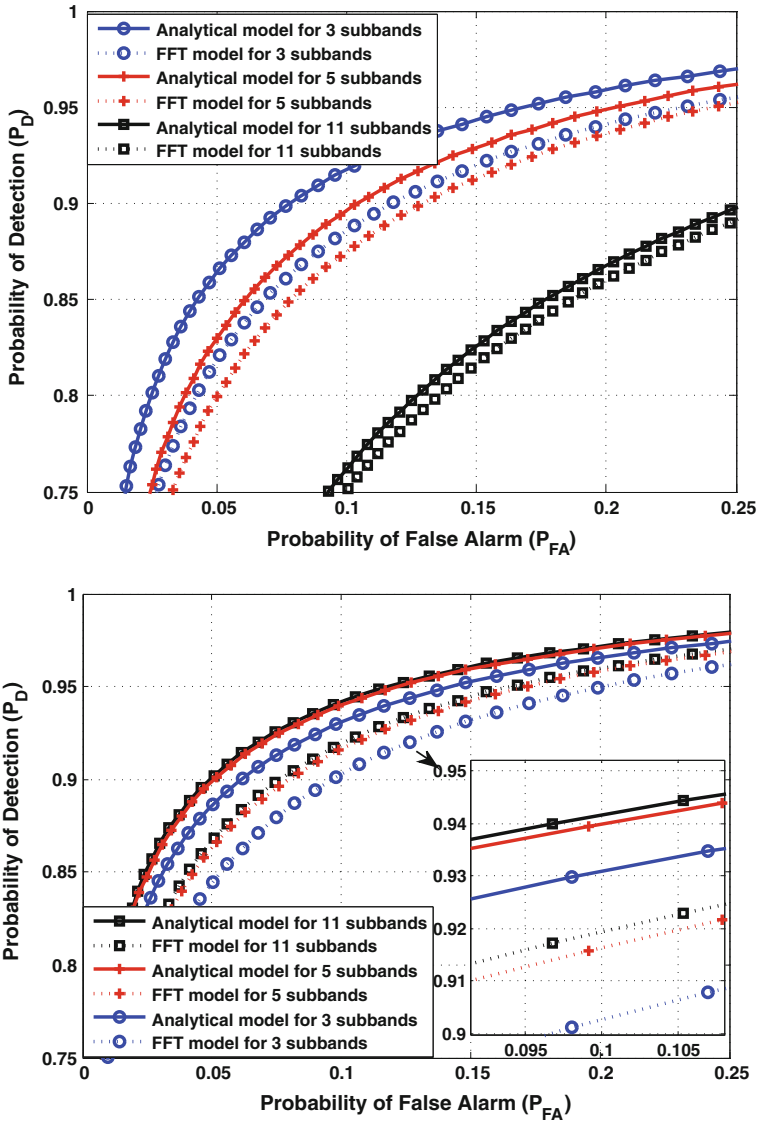

Fig. 3.11 ROC curves in Bluetooth sensing with time record length of 50 samples for constant weights (*upper*) and optimum weights (*lower*)

statistic is not aligned with the received BT burst, the detection performance will be greatly degraded, as discussed in Sect. 3.3.2.2. To avoid this problem, we can use the moving average (recursive running sum) approach in calculating the decision statistic.

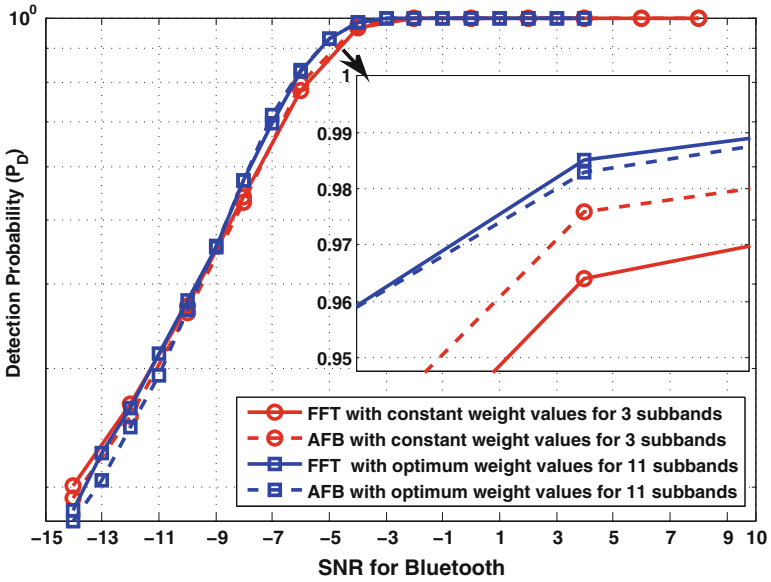

Fig. 3.12 Detection probability of Bluetooth signal with $P_{FA} = 0.1$ using time record length of 50 for constant and optimum weight values in AWGN including FFT and AFB based sensing

3.3.4 Effects of Fading Frequency Selective Channel

Most of the spectrum sensing publications, and also this report so far, have focused on the AWGN channel case. However, in most cognitive radio scenarios, the channel is frequency selective and fading. Next we discuss briefly about these issues.

The previous models based on Gaussian approximation of the PDF's can be extended to this situation as follows:

$$T(y)|_{\mathscr{H}_0} \sim \mathscr{N}\left(\sigma_n^2, \frac{1}{N_t N_f}\sigma_n^4\right) \tag{3.25a}$$

$$T(y)|_{\mathscr{H}_1} \sim \mathscr{N}\left(\sigma_{PU}^2 + \sigma_n^2, \frac{\beta_{ch}}{N_t N_f}(\sigma_{PU}^2 + \sigma_n^2)^2\right). \tag{3.25b}$$

where

$$\begin{aligned}\beta_{ch} &= \frac{1}{N_f}\sum_{k=1}^{N_f}\frac{(\sigma_k^2 + \sigma_n^2)^2}{(\sigma_{PU}^2 + \sigma_n^2)^2} = \frac{1}{N_f}\sum_{k=1}^{N_f}\frac{(1 + SNR_k)^2}{(1 + SNR)^2} \\ &= \frac{1}{N_f}\sum_{k=1}^{N_f}\frac{(1 + F_k^2 \cdot SNR)^2}{(1 + SNR)^2}.\end{aligned} \tag{3.26}$$

Here $SNR = \sigma_{PU}^2/\sigma_n^2$ is the overall PU SNR and $SNR_k = \sigma_k^2/\sigma_n^2$ are the subband-wise SNR's. F_k, $k = 1, \ldots, N_f$ are the channel amplitude responses of the subbands satisfying

$$\sum_{l=1}^{N_f} F_k^2 = N_f. \tag{3.27}$$

This condition basically means that the received PU signal power is assumed to be constant. In general, since spectrum sensing is commonly assumed to operate with low PU SNRs, and since the channel frequency response affects only the variance of the decision statistic under $\mathscr{H}_1$, the effect of the frequency selective channel on the energy detection performance is minor. We can extend the numerical results of Sect. 3.3.1 by calculating the β_{ch}-values for the ITU-R Vehicular A channel. Recalling that the theoretical β-values are in the range [1.13, 2.47], the corresponding range of β_{ch}-values is [1.03, 1.37] for 0 dB SNR and [1.01, 1.06] for −6 dB SNR.

In conclusion, we can say that the effect of channel frequency selectivity on spectrum sensing with clearly negative SNR's is in practice very small, and the sensing performance is determined by the temporal variations of the received PU signal power, i.e., the channel fading characteristics [3].

3.4 Concluding Remarks

We have explored the effects of different forms of frequency selectivity on energy detection performance. Using a generic Gaussian approximation based model, we first developed a simple metric for evaluating the effect of the sensing receiver frequency response on the required time record length. This metric can be used for optimizing the sensing filter, but the development of effective procedures for this purpose are left as topics for future studies. Next, we considered the effect of misaligned time-sensing window on the decision statistics and proposed moving average processing as a way to search for the best match in cases where the frequency range or burst timing of the PU signal is not known by the sensing receiver. Then we addressed the problem of known, non-flat PU spectrum, considering Bluetooth as an example case. Optimum weights were derived for FFT or filter bank based sensing where the PU signal band is divided into approximatively flat subbands. However, it was concluded that the use of constant weights is practically as good, if the range of subbands is chosen properly. Finally, the effect of stationary frequency selective channel was analyzed. It was found out that with typical PU SNRs, clearly negative in dB scale, the frequency selectivity has a minor effect on the sensing performance, while the energy detection performance is primarily determined by the temporal variations of the total received signal power.

References

1. Poor, H.: An Introduction to Signal Detection and Estimation, 2nd edn. Springer, Berlin (1994)
2. Urkowitz, H.: Energy detection of unknown deterministic signals. Proc. IEEE **55**(4), 523–531 (1967)
3. Digham, F., Alouini, M.S., Simon, M.: On energy detection of unknown deterministic signals over fading channels. IEEE Trans. Commun. **55**(1), 21–24 (2007)
4. Tandra, R., Sahai, A.: SNR walls for signal detection. IEEE J. Sel. Top. Sign. Proces. **2**(1), 4–17 (2008)
5. Cevher, V., Chellappa, R., McClellan, J.: Gaussian approximations for energy-based detection and localization in sensor networks. In: Proceedings of IEEE/SP 14th Workshop on Statistical Signal Processing (SSP '07), pp. 655–659. Madison, WI, USA (2007)
6. Taricco, G.: On the accuracy of the Gaussian approximation with linear cooperative spectrum sensing over Rician fading channels. IEEE Sign. Process. Lett. **17**(7), 651–654 (2010)
7. Taricco, G.: Optimization of linear cooperative spectrum sensing for cognitive radio networks. IEEE J. Sel. Top. Sign. Proces. **5**(1), 77–86 (2011)
8. Cabric, D., Tkachenko, A., Brodersen, R.: Spectrum sensing measurements of pilot, energy, and collaborative detection. In: Proceedings of the IEEE Military Communication Conference (2006)
9. Mariani, A., Giorgetti, A., Chiani, M.: Effects of noise power estimation on energy detection for cognitive radio applications. IEEE Trans. Commun. **59**(12), 3410–3420 (2011)
10. Bellanger, M., Ihalainen, T., Renfors, M.: Filter bank based cognitive radio physical layer. In: Proceedings of ICT Mobile Summit, Santander, Spain (June 2009)
11. Farhang-Boroujeny, B., Kempter, R.: Multicarrier communication techniques for spectrum sensing and communication in cognitive radios. IEEE Commun. Mag. **46**(4), 80–85 (2008)
12. Dikmese, S., Renfors, M.: Optimized FFT and filter bank based spectrum sensing for Bluetooth signal. In: Proceedings of Wireless Communications and Networking Conference (WCNC), pp. 792–797. Paris, France (2012).
13. INFSO-ICT-211887 Project PHYDYAS: Deliverable 5.1 Prototype filter and structure optimization (2009) http://www.ict-phydyas.org/userfiles/file/PHYDYASD5-1.pdf
14. Siohan, P., Siclet, C., Lacaille, N.: Analysis and design of OFDM-OQAM systems based on filterbank theory. IEEE Trans. Sign. Proces. **50**(5), 1170–1183 (2002)
15. Ihalainen, T., Viholainen, A., Hidalgo Stitz, T., Renfors, M.: Reappearing primary user detection in FBMC/OQAM cognitive radios. In: Proceedings of the CROWNCOM 2010, pp. 411–416. Cannes, France (June 2010)
16. Viholainen, A., Ihalainen, T., Hidalgo Stitz, T., Renfors, M., Bellanger, M.: Prototype filter design for filter bank based multicarrier transmission. In: Proceedings of the European Signal Processing Conference, Glasgow, Scotland (2009)
17. Ihalainen, T., Viholainen, A., Hidalgo Stitz, T., Renfors, M.: Spectrum monitoring scheme for filter bank based cognitive radios. In: Proceedings of Future Network and Mobile Summit, Florence, Italy (2010)
18. Penna, F., Pastrone, C., Spirito, M., Garello, R.: Energy detection spectrum sensing with discontinuous primary user signal. In: Proceedings of the IEEE International Conference Communications, Australia (2009)
19. Barry, J., Lee, E., Messerschmitt, D.: Digital Communication, 3rd edn. Kluwer, Dordrecht (2004)
20. Held, I., Chen, A.: Channel estimation and equalization algorithms for long range Bluetooth signal reception. In: Proceedings of IEEE Vehicular Technology Conference (VTC). Taipei, Taiwan (May 2010)

Chapter 4
Spectrum Sensing Algorithms for Cognitive TV White-Spaces Systems

Daniel Riviello, Sergio Benco, Floriana Loredana Crespi, Andrea Ghittino, Roberto Garello and Alberto Perotti

Abstract This chapter describes a set of spectrum sensing algorithms to be employed for the detection of Ortogonal Frequency Division Multiplexing transmissions in the TV bands (470–790 MHz), like DVB-T signals. Spectrum sensing techniques take a crucial role to support geo-referenced TV White-Spaces (TVWS) databases and to maintain them up-to-date over time. When considering a single-antenna spectrum sensing unit, very effective methods for detecting OFDM signals are based on DVB-T cyclic prefix and pilot pattern feature detection. Starting from these, further improvements can be obtained using multi-antenna techniques. This chapter shows performance analysis of feature-based single-antenna and multi-antenna techniques in order to derive trade-offs and conclusions.

D. Riviello (✉) · R. Garello
Politecnico di Torino, Turin, Italy
e-mail: daniel.riviello@polito.it

S. Benco · F. L. Crespi · A. Ghittino · A. Perotti
CSP-ICT innovation, Turin, Italy
e-mail: sergio.benco@csp.it

F. L. Crespi
e-mail: floriana.crespi@csp.it

A. Ghittino
e-mail: aghittino@gmail.com

R. Garello
e-mail: garello@polito.it

A. Perotti
e-mail: alberto.perotti@ieee.org

M.-G. Di Benedetto and F. Bader (eds.), *Cognitive Communication and Cooperative HetNet Coexistence*, Signals and Communication Technology,
DOI: 10.1007/978-3-319-01402-9_4,

Table 4.1 Main parameters of DVB-T

	2k mode	8k mode
Symbol duration (T_U)	224 μs	896 μs
Guard interval duration (Δ)	$7 - 56$ μs	$28 - 224$ μs
Number of active subcarriers	1705	6817
Subcarrier spacing (approx.)	4464 Hz	1116 Hz
CP duration ratio (Δ / T_U)	1/4, 1/8, 1/16, 1/32	
Constellations	QPSK, 16-QAM, 64-QAM	
Code rate	1/2, 2/3, 3/4, 5/6, 7/8	

4.1 Introduction

The huge interest in the TV White-Spaces (TVWS) has recently determined the development of novel Machine-to-Machine (M2M) standards such as the Weightless standard, secondary user rural broadband internet access (i.e. the IEEE 802.22 standard), and other proprietary protocols. In UK and USA the regulatory process is almost terminated, so that a huge number of large companies have demonstrated interest in the management of large TVWS databases. These databases are required to inform secondary users about primary transmitters presence and thus must be up-to-date over time and assessed by field test trials. In 2013 large-scale TVWS networks have been deployed for end-user testing purposes in the USA and South Africa. The crucial importance of precise and up-to-date information provided by such TVWS databases represents one of the weak points of the future TVWS broadband networks. In order to guarantee the highest primary detection sensitivity, future TVWS secondary networks must rely on master units able to perform either TVWS database access and advanced spectrum sensing. Currently, in the TVWS domain, the most relevant primary signal, is represented by ETSI Digital Video Broadcasting - Terrestrial (DVB-T) broadcast transmissions [1]. The transmitted DVB-T signal consists of a sequence of fixed-duration OFDM symbols. From a spectrum sensing point of view, important DVB-T parameters (see also Table 4.1) are represented by: channel bandwidth (that ranges from 5 to 8 MHz), the OFDM Cyclic Prefix (CP) length (that ranges from 1/32 to 1/4 of the OFDM symbol length), and the presence of OFDM pilots (continual and scattered ones). The presence of pre-determined patterns (and their periodic repetition) in the transmitted DVB-T signal determines the cyclo-stationary property shown by the OFDM signal.

In order to accurately assess the performance of the studied sensing algorithms in a realistic scenario consisting of real DVB-T primary signals, a SDR transmitter has been implemented using GNU Radio [2]. The GNU Radio framework can ease the development of DVB-T compliant DSP blocks that can be used to generate a DVB-T signal [1]. Two custom GNU Radio blocks have been developed: a DVB-T encoder and a DVB-T modulator. These blocks comply with a subset of the DVB-T physical layer parameters. The OFDM mode with 8k subcarriers and cyclic prefix 1/4 has

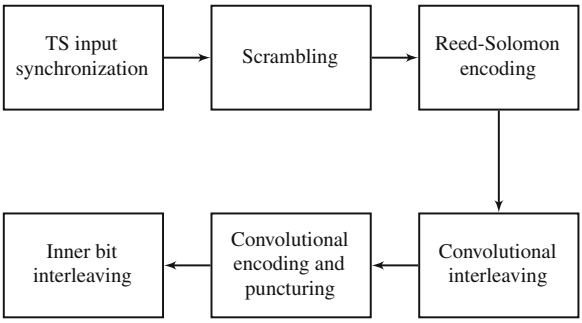

Fig. 4.1 Encoder block

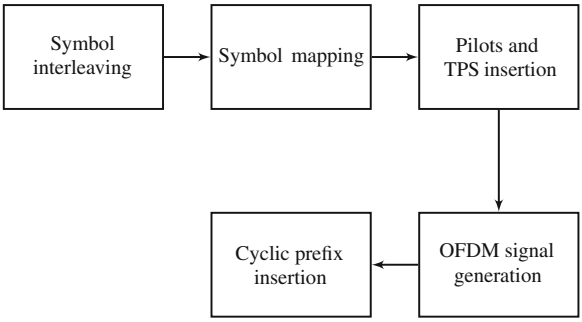

Fig. 4.2 Modulator block

been used in the spectrum sensing tests. The DVB-T encoder block receives the input Transport Stream (TS) containing a set of multiplexed TV and radio channels and, after encoding, sends the stream to the modulator. The output of the OFDM modulator is the sampled complex envelope (IQ stream) ready to be up-converted by a radio front-end device (in our case, a Universal Software Radio Peripheral, USRP). The SDR DVB-T transmitter has been successfully tested against commercial receivers to verify its correct behavior. The processing performed by the encoder and modulator blocks are summarized in Figs. 4.1 and 4.2. The whole project has been described in [3].

4.2 Single Antenna Spectrum Sensing Algorithms

An effective approach to performing spectrum sensing in the TVWSs consists in the exploitation of the intrinsic structure of OFDM signals. Spectrum sensing techniques able to detect a number of signal "features" in the received signal can be devised. Such features must uniquely characterize the DVB-T transmission in order to allow efficient signal detection.

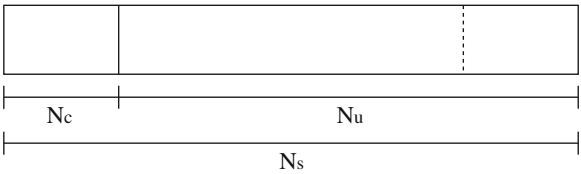

Fig. 4.3 OFDM symbol structure with cyclic prefix

Single-antenna feature-based detectors exhibit much better performance in case of unknown noise variance (unlike energy detection-based techniques), but require much longer sensing time for synchronization in order to achieve that performance. In the following subsection, these single-antenna CP-based feature detectors will be analyzed and their performance assessed.

4.2.1 Cyclic Prefix Based Detector

The cyclostationary properties of OFDM signals in DVB-T transmissions are due to the presence of a cyclic repetition of signal segments in the time domain.

We briefly recall the OFDM symbol structure in Fig. 4.3, where N_s is total number of samples per symbol, N_u is the number of useful samples and N_c the number of cyclic prefix samples.

The number of samples N_s, N_u and N_c can be obtained through the following relationship:

$$\begin{aligned} N_s &= N_{\text{mode}} \cdot (1 + \text{CP}) \cdot \frac{7}{8B} \cdot f_s \\ N_c &= N_{\text{mode}} \cdot \text{CP} \cdot \frac{7}{8B} \cdot f_s \\ N_u &= N_s - N_c \end{aligned} \tag{4.1}$$

where $N_{\text{mode}} = 8192$ for 8k mode or $N_{\text{mode}} = 2048$ for 2k mode. Moreover, CP is the ratio of cyclic prefix duration and the OFDM symbol duration, B is the signal bandwidth in MHz, and f_c is the sampling frequency, in Msamples/s.

As already mentioned, the detector operates asynchronously with the primary signal. The first detection technique we present is based on the computation of the autocorrelation (AC) of the received signal in order to detect its cyclostationary features. The autocorrelation function must be computed in the range $[0, N_s]$ in order to be sure to detect at least one cyclic repetition. The expression for the AC function used for detection is the following:

$$R_{xx}^{\mathrm{CP}}[n] = \frac{1}{N_c} \left| \sum_{k=0}^{N_c-1} x^*[n-k]x[n-k-N_u] \right|. \tag{4.2}$$

where $x[n]$ is the received primary signal, modeled as

$$x[n] = p[n] + w[n]. \tag{4.3}$$

Here, $p[n]$ is the transmitted primary signal and $w[n]$ is additive white Gaussian noise and the Signal-to-Noise Ratio (SNR) is equal to $E[|x[n]|^2]/E[|w[n]|^2]$. If detection is performed in very low SNR conditions, it may be necessary to improve the sensitivity by extending the observation window to K consecutive OFDM symbols (KN_s samples), thus the new AC function is

$$\widetilde{R}_{xx}^{\mathrm{CP}}[n] = \frac{1}{KN_c} \left| \sum_{j=0}^{K-1} \sum_{k=0}^{N_c-1} x^*[n-k-jN_s]x[n-k-jN_s-N_u] \right|. \tag{4.4}$$

which reduces to (4.2) for $K = 1$. The AC functions (4.2) and (4.4) exhibit a peak shape with the maximum value achieved synchronously with the end of the OFDM symbol. Based on this observation, the proposed test statistic for the CP-based detector based on (4.2) and (4.4) is:

$$T_{CP} = \frac{\max_i\{\widetilde{R}_{xx}^{\mathrm{CP}}[i]\}}{\mathbb{E}_j\{\widetilde{R}_{xx}^{\mathrm{CP}}[j]\}} \gtrless \theta_{CP}. \tag{4.5}$$

Here, the numerator denotes the peak of the whole autocorrelation function while the denominator denotes the temporal mean of the AC computed only at samples j, with $j \in J$, while θ_{CP} denotes the threshold. If we denote with $\varphi = \mathrm{argmax}_n(\widetilde{R}_{xx}^{\mathrm{CP}}[n])$ the peak of the AC function, the set J is defined as

$$\begin{aligned} &j \in J = N \backslash Q \\ &N = \{n \in \mathbb{N} \,:\, 0 \le n \le KN_s\} \\ &Q = \{q \in \mathbb{N} \,:\, \varphi - N_c \le q \le \varphi + N_c\}. \end{aligned} \tag{4.6}$$

Basically the mean of the autocorrelation is computed for all those samples whose "distance" from the instant with the peak (maximum of the AC) is larger than N_c samples. This way, neither part of the cyclic prefix is comprised in the summation of (4.4) and hence the estimation of the correlation noise is improved. The aforementioned mean is used to perform the calibration of the threshold θ_{CP} [4].

Figure 4.4 shows the amplitude of the CP-based autocorrelation function with an observation window of respectively $K = 1$ and $K = 10$ symbols with an infinite SNR. Results with realistic SNR values will be given later in the following sections. We observe that, for $K = 10$, the ratio of the peak value over the maximum value observed outside the cyclic prefix window is improved with respect to $K = 1$.

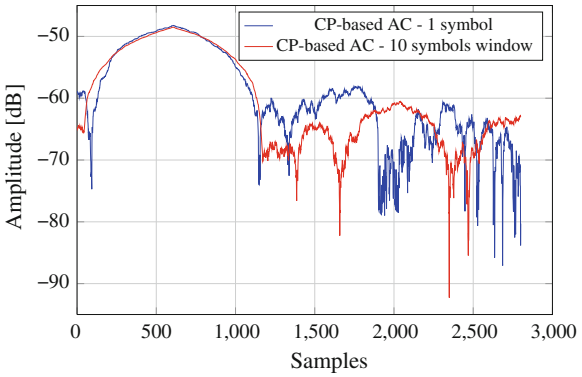

Fig. 4.4 Amplitude of the CP-based auto-correlation—1 symbol versus 10 symbols observation

4.2.2 Pilot-Based Detector

The DVB-T signal contains pilot subcarriers that can be used for channel estimation and signal synchronization at the receiver, therefore their values are known at the receiver. We can define an expression of the OFDM signal in terms of complex envelope where only the pilot carriers are active, while data and Transmission Parameter Signalling (TPS) carriers are set to zero:

$$s_p(t) = e^{-j2\pi f_c t} \sum_{p,l,k:(l,k)\in A} c_{p,l,k} \cdot \Psi_{p,l,k}(t) \tag{4.7}$$

where the set A defines the position of the continual and scattered pilot, while indices p, l and k correspond, respectively, to the DVB-T frame, to the OFDM symbol and to the subcarrier. $c_{p,l,k}$ is the pilot symbol and $\Psi_{p,l,k}(t)$ is the corresponding IFFT basis function. If such deterministic signal is sampled at the rate $1/T_s$, we obtain the sequence:

$$s_p[n] = s_p(nT_s), \qquad n = 0, 1, \ldots \tag{4.8}$$

We know that continual pilots have the same carrier positions for all symbols, while scattered pilots take the same position every four symbols. Thus, since every symbol contains $N_u + N_c$ samples, the sequence $s_p[n]$ is cyclic with period $4(N_u + N_c)$, which means that it is completely determined by the vector: $(s_p[0], s_p[1], s_p[2], \ldots, s_p[4(N_u + N_c) - 1])$. We can therefore apply a *matched filtering* approach to detecting the DVB-T incumbent by defining the following test statistic:

$$T_P = \max_{\tau\in\{0,1,\ldots,4(N_u+N_c)-1\}} \left| \sum_{k=0}^{M-\tau-1} s_p[k] x^*[k+\tau] \right| \tag{4.9}$$

where M is the total number of samples observed by the detector. As in the previous case, the determination of the threshold value may be performed empirically. In order to perform such computation, the detector is usually fed with pure white noise whose energy is computed and then the test statistic is derived. Once the desired false alarm probability is chosen, the threshold is set accordingly [5, 6].

4.3 Multi-Antenna Spectrum Sensing Algorithms

In this section we focus on a special class of detectors based on a multi-sensor/multi-antenna approach which will try to overcome the limitations of single sensor detectors. It is known that the performance of an energy detector drops if the actual value of the noise variance is different from the nominal value, the larger the uncertainty of the noise variance, the worse the performance. As far as the feature detectors are concerned, such approaches are quite sensitive to synchronization errors. In case of very low SNR the synchronization loops might not be able to provide the required accuracy for the carrier frequency and clock rate estimates. All these reasons motivate the search for asynchronous multi-sensor detectors robust to noise uncertainty.

Multiple-input multiple-output (MIMO) technology has reached considerable maturity, since it is already part of many wireless standards (LTE, IEEE 802.11, IEEE 802.16, etc.) it is very likely for future CR terminals to incorporate it. The basic idea is to exploit the fact that, if the channel is being used by the PU, then some spatial correlation should be present in signals at different antennas. On the other hand, when the signal is absent, all contributions will correspond to thermal noise, thus spatial correlation should be absent [7].

All the presented detectors are non parametric, i.e. they don't assume any prior knowledge of the signal and all their test statistics will be expressed as a function of the eigenvalues of the *sample covariance matrix* of the received signals, hence a new matricial system model is provided. All the eingenvalue-based algorithms are based on results from Random Matrix Theory (RMT) [8, 9]. Among them, we will present the energy detector for the multi-antenna eigenvalue-based case and methods based on Likelihood Ratio Tests (LRT) or generalized LRT.

The algorithms will be divided in two classes:

1. those that assume a known noise level;
2. those that estimate it from data.

4.3.1 System Model

We assume that the detector computes its test statistic from K sensors or antennas and N time samples. Let's denote with $\boldsymbol{y}[n] = \{y_1[n], \ldots, y_K[n]\}^T$ the $K \times 1$ received vector at time n, where the element $y_k[n]$ is the discrete baseband complex sample at receiver k.

Under hypothesis $\mathscr{H}_0$, the received vector consists of K complex Gaussian noise samples with zero mean and variance σ_v^2

$$\boldsymbol{y}[n]\,|_{\mathscr{H}_0} = \boldsymbol{v}[n] \tag{4.10}$$

where $\boldsymbol{v}[n] \sim \mathscr{N}_{\mathbb{C}}(\boldsymbol{0}_{K\times 1}, \sigma_v^2 \boldsymbol{I}_{K\times K})$. On the contrary, under hypothesis $\mathscr{H}_1$ the received vector contains signal plus noise

$$\boldsymbol{y}[n]\,|_{\mathscr{H}_1} = \boldsymbol{h}s[n] + \boldsymbol{v}[n] \tag{4.11}$$

where $s[n]$ is the transmitted signal sample, generally modeled as a Gaussian random variable with zero mean and variance σ_s^2, while for our simulation it will be drawn from a 8k DVB-T signal; $\boldsymbol{h}$ is a $K \times 1$ unknown complex channel vector. In this model a flat Rayleigh fading channel is considered, so basically every received signal is multiplied by a complex constant. For our simulation we considered also another channel model (the 6-path Typical Urban [10]), which will be described in the next section.

If we are under $\mathscr{H}_1$, the SNR is defined as follows:

$$\rho \triangleq \frac{\mathbb{E}\|x[n]\|^2}{\mathbb{E}\|w[n]\|^2} = \frac{\sigma_s^2 \|\boldsymbol{h}\|^2}{K\sigma_v^2} \tag{4.12}$$

The received samples are stored by the detector in the $K \times N$ matrix

$$\boldsymbol{Y} \triangleq \{\boldsymbol{y}[1], \ldots, \boldsymbol{y}[N]\} = \boldsymbol{h}\boldsymbol{s} + \boldsymbol{V} \tag{4.13}$$

where the $1 \times N$ signal vector $\boldsymbol{s}$ is defined as

$$\boldsymbol{s} \triangleq \{s[1], \ldots, s[N]\} \tag{4.14}$$

and the $K \times N$ noise matrix $\boldsymbol{V}$ as

$$\boldsymbol{V} \triangleq \{\boldsymbol{v}[1], \ldots, \boldsymbol{v}[N]\}. \tag{4.15}$$

At this point we can define the sample covariance matrix $\boldsymbol{R}$ as follows:

$$\boldsymbol{R} \triangleq \frac{1}{N}\boldsymbol{Y}\boldsymbol{Y}^H. \tag{4.16}$$

Let $\lambda_1 \geq \cdots \geq \lambda_K$ be the eigenvalues of $\boldsymbol{R}$ sorted in decreasing order.

The detector computes the test statistic T and compares it against a pre-defined threshold θ, if $T > \theta$ it decides for $\mathscr{H}_1$, otherwise $\mathscr{H}_0$. Usually, the decision threshold θ is determined as a function of the target false alarm probability. False alarm and detection probability are defined as follows:

$$P_d = \mathscr{P}(T > \theta \mid \mathscr{H}_1)$$
$$P_{fa} = \mathscr{P}(T > \theta \mid \mathscr{H}_0). \tag{4.17}$$

All the described test statistics are non-parametric, i.e. they do not assume any prior knowledge about the signal to be detected. In general all test have only two parameters: the number of samples N and the number of sensors K.

The methods are divided into two groups: methods for known or for unknown noise level. In the first group, the noise variance σ_v^2 is assumed to be known and appears explicitly in the test statistic. Methods belonging to the second group, on the contrary, do not require such information, i.e. the noise level is estimated in the test statistic.

4.3.2 Algorithms with known noise variance

1. **Energy Detection (ED)**: the test statistic is the average energy of the received samples, normalized by the noise variance [11, 12]:

$$T_{ED} = \frac{1}{KN\sigma_v^2} \sum_{k=1}^{K} \sum_{n=1}^{N} |y_k(n)|^2 = \frac{\|\boldsymbol{Y}\|_F^2}{KN\sigma_v^2} \tag{4.18}$$

   where $\|\cdot\|_F$ denotes the Frobenius norm. Note that it is possible to express T_{ED} in terms of the eigenvalues λ_i by exploiting the equivalence $\|\boldsymbol{Y}\|_F^2 = \mathrm{tr}(\boldsymbol{Y}\boldsymbol{Y}^H)$, thus obtaining

$$T_{ED} = \frac{1}{K\sigma_v^2}\mathrm{tr}(\boldsymbol{R}) = \frac{1}{KN\sigma_v^2} \sum_{i=1}^{K} \lambda_i. \tag{4.19}$$

2. **Roy's Largest Root Test (RLRT)**: this method tests the largest eigenvalue of the sample covariance matrix against the noise variance. The test statistic is

$$T_{RLRT} = \frac{\lambda_1}{\sigma_v^2}. \tag{4.20}$$

   The RLRT was originally developed in [13]. Performance analysis can be found, for example, in [14]. For Gaussian signals and not too low signal-to-noise ratio, the RLRT is the best test statistics in this class.
3. **Likelihood Ratio Tests (LRT)**: different LRT-based detectors were given in [15]. The complete, noise-dependent, log-likelihood ratio test statistic is given by

$$T_{LRT} = 2(N-1)\left[\log\left(\frac{\sigma_v^{2K}}{\det \boldsymbol{R}}\right) + \left(\frac{\mathrm{tr}\,\boldsymbol{R}}{\sigma_v^2} - K\right)\right]. \tag{4.21}$$

For this statistic, expressions of the false-alarm probability have been derived by means of numerical integration techniques. Performance analysis for this test can be found, for example, in [15].

4.3.3 Algorithms with unknown noise variance

1. **Eigenvalue Ratio Detector (ERD)**: the test statistic (also called maximum-minimum eigenvalue, or condition number test) is the ratio between the largest and the smallest eigenvalue of $\boldsymbol{R}$

$$T_{ERD} = \frac{\lambda_1}{\lambda_K}. \tag{4.22}$$

   A complete performance analysis can be found in [16, 17].
2. **Noise-independent LRT (LRT-)**: an alternative log-likelihood ratio was derived in [15], under the assumption of unknown noise variance:

$$T_{LRT-} = 2(N-1)\left[\frac{\frac{1}{K}\sum_{i=1}^{K}\lambda_i}{\left(\prod_{i=1}^{K}\lambda_i\right)^{1/K}}\right]^K. \tag{4.23}$$

   In statistics, this method has been known for many years as the *sphericity test* [18]. Performance analysis for cognitive radio applications plus an expression for the false alarm probability of this detector can be found in [15].
3. **Generalized Likelihood Ratio Test (GLRT)**: this method uses as test statistic the ratio

$$T_{GLRT} = \frac{\lambda_1}{\frac{1}{K}\mathrm{tr}(\boldsymbol{R})}. \tag{4.24}$$

   Performance analysis can be found for example in [19].
   It is interesting to note that the GLRT is equivalent (up to a nonlinear monotonic transformation) to [20]:

$$T_{GLRT'} = \frac{\lambda_1}{\frac{1}{K-1}\sum_{i=2}^{K}\lambda_i}. \tag{4.25}$$

   The denominator of $T_{GLRT'}$ is the maximum-likelihood (ML) estimate of the noise variance assuming the presence of a signal, hence the GLRT can be interpreted as a largest root test with an estimated $\hat{\sigma}_v^2$ instead of the true σ_v^2.

A comparative analysis of these algorithms based on numerical simulation for both known and unknown noise variance will be performed in the next section.

4.4 Channel Models

We consider three different channel models. For each of them we describe how the $K \times N$ matrix $\boldsymbol{Y}$ of received samples has been calculated.

4.4.1 Additive White Gaussian Noise Channel Model

In this case we generated a matrix $\boldsymbol{S}$ of size $K \times N$ where each row of $\boldsymbol{S}$ is equal to the $1 \times N$ signal vector $\boldsymbol{s}$. Hence:

$$\boldsymbol{Y} = \boldsymbol{S} + \boldsymbol{V}. \tag{4.26}$$

This channel model will be used to evaluate the performance of the CP-based detector and to compare it against the eigenvalue-based detectors with unknown noise variance.

4.4.2 Flat Rayleigh Fading Channel Model

With this model we assume that the coherence bandwidth, defined as the inverse of the delay spread, is much larger than the observed bandwidth. Under this assumption, our $1 \times N$ signal vector $\boldsymbol{s}$ is simply multiplied by a complex constant modeled as a Rayleigh random variable. We will have K random variables, one for each sensor, represented by the $K \times 1$ channel vector. Hence,

$$\boldsymbol{Y} = \boldsymbol{h}\boldsymbol{s} + \boldsymbol{V}. \tag{4.27}$$

In addition, the following normalization has been performed:

$$\sum_{i=1}^{K} h_i h_i^* = K \tag{4.28}$$

hence the energy of the channel vector is normalized to the number K of antennas.

4.4.3 Typical Urban 6-Path Channel Model

This channel models the terrestrial propagation in an urban area. It has been defined by COST 207 as a typical urban (TU6) profile and consists of 6 paths having wide dispersion in delay and relatively strong power [10].

Table 4.2 Typical urban profile (TU6)

Tap number	Delay τ_i (μs)	Average gain γ_i (dB)	Doppler spectrum
1	0.0	−3	Classical
2	0.2	0	Classical
3	0.5	−2	Classical
4	1.6	−6	Classical
5	2.3	−8	Classical
6	5.0	−10	Classical

This profile is a frequency- and time-selective Rayleigh fading channel model. Given $x(t)$ and $y(t)$ the input and output signal respectively, it can be expressed as follows:

$$y(t) = \sum_{i=1}^{M} \gamma_i e^{-j\theta_i} x(t - \tau_i) \tag{4.29}$$

where:

- M is the number of paths equal to 6;
- γ_i is the average path gain of the ith path (listed in Table 4.2);
- θ_i is the phase shift from scattering of the i'th path, modeled as a uniformly distributed random variable in $[-\pi, \pi]$;
- τ_i is the relative delay of the ith path (listed in Table 4.2);

where the classical doppler spectrum is defined as:

$$G(f; f_D) = \frac{1}{\sqrt{1 - (f/f_D)^2}}. \tag{4.30}$$

In our simulation the Doppler spread f_D has been set to 10 Hz, corresponding to a pedestrian mobile profile.

This channel has been implemented in our simulation setup as a *Finite Impulse Response* (FIR) filter. We generated K realizations of this channel and performed the convolution (filtering) with the input signal for each realization. We generated a $K \times N$ matrix $\boldsymbol{X}$ where each row corresponds to N samples of our K filtered signals. Hence the model yields:

$$\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{V}. \tag{4.31}$$

By storing all K channel realization in a $K \times M$ matrix $\boldsymbol{H}$, the following normalization has been performed:

$$\|\boldsymbol{H}\|_F^2 = \sum_{i=1}^{K} \sum_{j=1}^{6} h_{ij} h_{ij}^* = K \tag{4.32}$$

i.e. we performed the same normalization as the flat fading case.

4.5 Performance Assessment and Trade-Offs

The performance of the different sensing methods on the various channels have been evaluated by simulation (performed within Matlab). For each simulation we computed:

- the Receiver Operating Characteristic (ROC) curve obtained by plotting the detection probability versus the false alarm one;
- the detection probability as a function of the signal-to-noise ratio, by fixing the false alarm probability $P_{fa} = 10^{-2}$.

For our simulation we used the Monte Carlo method. In order to estimate the values of P_{fa} and P_d we performed for each SNR value $N_T = 10000$ trials.

For each trial we generated two instances of the matrix $\boldsymbol{Y}$: the first one has been computed as in (4.26), (4.27) or (4.31) (signal plus noise case); the second one instead as $\boldsymbol{V} = \boldsymbol{N}$ (only noise case). In this way, we computed two instances of the covariance matrix $\boldsymbol{R}$ and two test statistics for each algorithm: T_1 and T_0 respectively. Once all the trials have been performed, we generated a vector of threshold values from the smallest T_0 to the largest T_1 statistic. At this point we simply computed each ith element of the P_{fa} vector by counting how many T_0 values are greater than the ith threshold value. Similarly, each ith element of the P_d vector has been computed by counting how many T_1 values are greater than the ith threshold value. Each value of both P_{fa} and P_d vectors are finally divided by the number of trials.

A short description of the simulation algorithm in pseudocode follows:

```
1: N_B = 1000                                   ▷ N_B = number of threshold values
2: for all SNR values do
3:     for i = 1 → N_T do
4:         compute σ_s^2
5:         compute σ_v^2
6:         generate K × N random Gaussian noise matrix V as a function of SNR
7:         if chan = AWGN then
8:             Y_1 ← S + V                      ▷ Y_1 = signal+noise Y
9:         else if chan = flat-fading then
10:            Y_1 ← hs + V
11:        else                                 ▷ chan = TU6
12:            Y_1 ← X + V
13:        end if
14:        Y_0 ← V                              ▷ Y_0 = only-noise Y
```

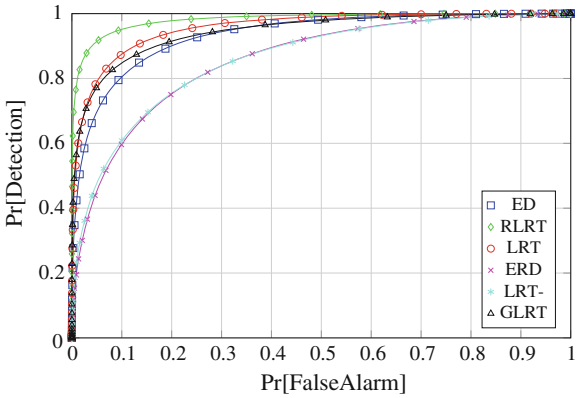

Fig. 4.5 Eigenvalue-based detectors, DVB-T 8k PU signal, flat fading channel, $N = 50$, $K = 10$, ROC curves (SNR $= -10$ dB)

compute $\boldsymbol{R}_1 \leftarrow (1/K)\boldsymbol{Y}_1\boldsymbol{Y}_1^H$ ▷ signal+noise
compute $\boldsymbol{R}_0 \leftarrow (1/K)\boldsymbol{Y}_0\boldsymbol{Y}_0^H$ ▷ noise-only
for all detectors **do**
compute signal+noise test statistic $T_1(i)$
compute only-noise test statistic $T_0(i)$
end for
end for
for all detectors **do**
create vector of thresholds $\boldsymbol{\theta}$ of N_B equally spaced values from $\min(\boldsymbol{T}_0)$ to $\max(\boldsymbol{T}_1)$
create vector of N_B elements $\boldsymbol{P_{fa}}$
create vector of N_B elements $\boldsymbol{P_d}$
for $i = 1 \rightarrow N_B$ **do**
for $j = 1 \rightarrow N_T$ **do**
if $T_0(j) > \theta(i)$ **then**
$P_{fa}(i) \leftarrow P_{fa}(i) + 1$
end if
if $T_1(j) > \theta(i)$ **then**
$P_d(i) \leftarrow P_d(i) + 1$
end if
end for
$P_{fa}(i) \leftarrow P_{fa}(i)/N_T$
$P_d(i) \leftarrow P_d(i)/N_T$
end for
end for
end for

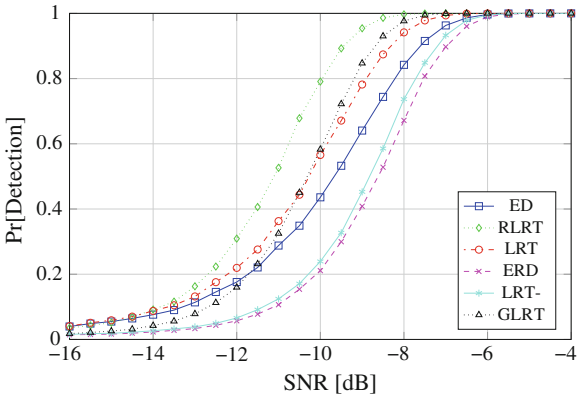

Fig. 4.6 Eigenvalue-based detectors, DVB-T 8k PU signal, flat fading channel, $N = 50$, $K = 10$, P_d versus SNR ($P_{fa} = 0.01$)

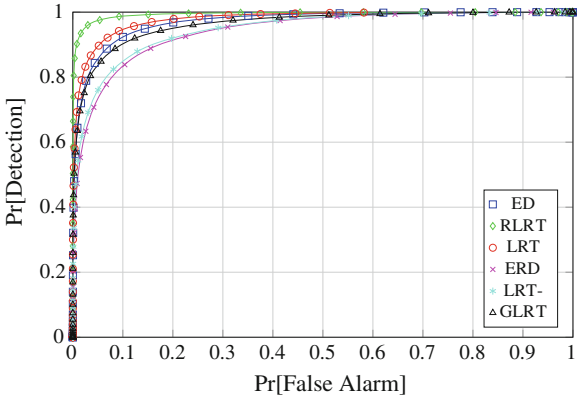

Fig. 4.7 Eigenvalue-based detectors, DVB-T 8k PU signal, flat fading channel, $N = 200$, $K = 4$, ROC curves (SNR $= -10$ dB)

4.5.1 Results

All the ROC performance curves have been evaluated at SNR $= -10$ dB. Such a challenging scenario corresponds to the so-called "hidden node problem" in the Wireless Regional Access Network (WRAN) cognitive radio scenario and has been chosen to emphasize the differences among the methods.

First of all, performance of eigenvalue-based algorithms has been evaluated and compared with different sets of parameters: by default we assumed $N = 50$ acquired samples for each antenna and $K = 10$ sensors.

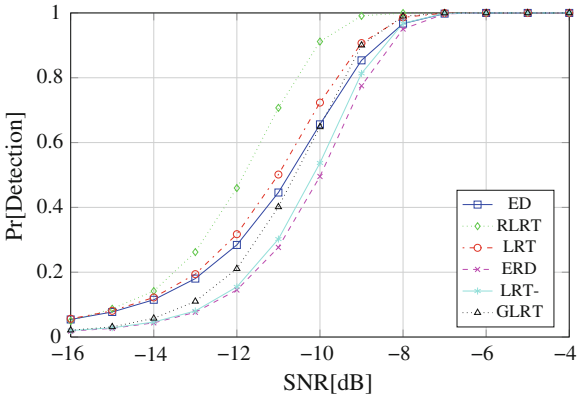

Fig. 4.8 Eigenvalue-based detectors, DVB-T 8k PU signal, flat fading channel, $N = 200$, $K = 4$, P_d versus SNR ($P_{fa} = 0.01$)

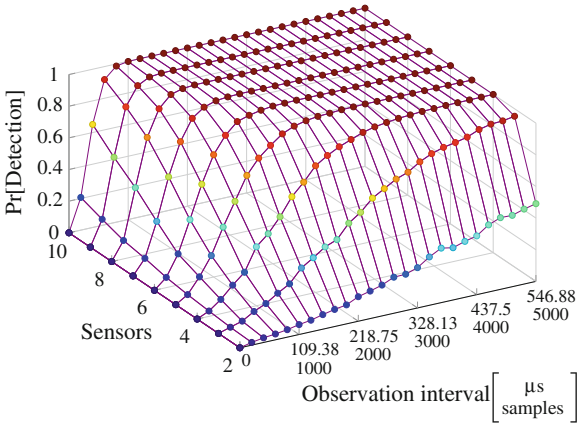

Fig. 4.9 GLRT detection probability as a function of time (samples) and sensors through flat-fading channel, SNR $= -15$ dB, $P_{fa} = 0.01$

4.5.1.1 Eigenvalue-Based Algorithms

Figures 4.5 and 4.6 refer to a DVB-T 8k signal with $N = 50$ samples and $K = 10$ sensors. In both examples we can observe that the best algorithm for known noise variance is the RLRT, while GLRT is the best under unknown variance. These results are in agreement with the results provided in the literature for Gaussian signals.

Secondly, in Figs. 4.7 and 4.8 the observation interval has been increased to $N = 200$, whereas the number of sensors reduced to $K = 4$.

Finally, in Fig. 4.9 we plot the detection probability of GLRT as a function of the observation interval (expressed both in time units and number of received samples

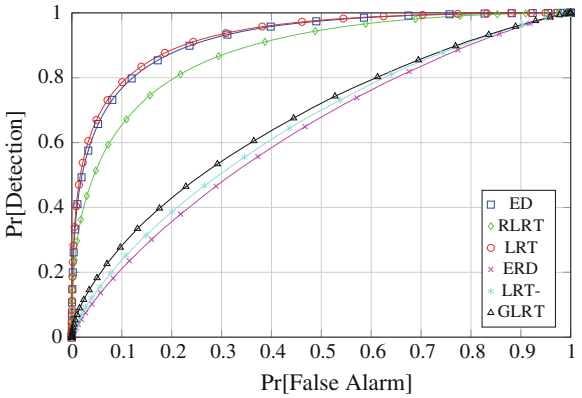

Fig. 4.10 Eigenvalue-based detectors, DVB-T 8k PU signal, TU6 channel model, $N = 50$, $K = 10$, ROC curves (SNR $= -10$dB)

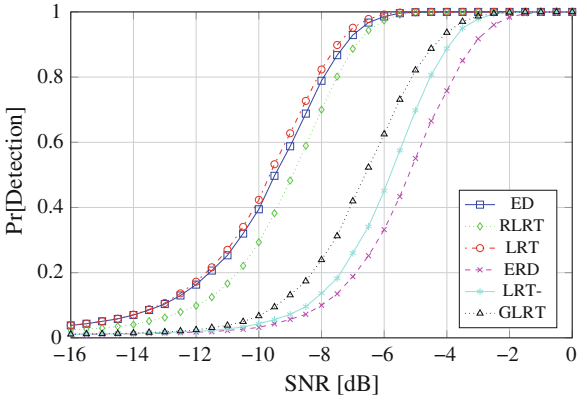

Fig. 4.11 Eigenvalue-based detectors, DVB-T 8k PU signal, TU6 channel model, $N = 50$, $K = 10$, P_d versus SNR, P_d versus SNR ($P_{fa} = 0.01$)

per sensor) and the number of sensors for a specific SNR value of -15 dB, while the false alarm probability remains fixed to 10^{-2}. The channel is Rayleigh flat-fading.

Under a more realistic model, the TU6 channel, the performance of the algorithms are different, as it can be observed in Figs. 4.10 and 4.11. We can see how both GLRT and RLRT lose their predominant position when the received model is different from the linear mixture one: simple energy detection becomes highly competitive in this case. The difference between algorithms with known and unknown noise variance is larger, too.

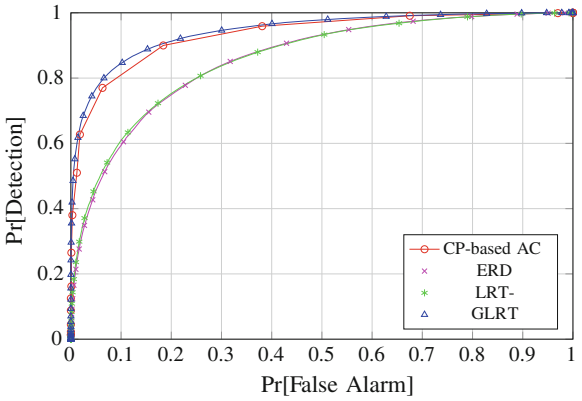

Fig. 4.12 CP-based versus eigenvalue-based (unknown σ_v^2) detectors, DVB-T 8k PU signal, flat fading channel, $N = 50$, $K = 10$, ROC curves (SNR $= -10$dB)

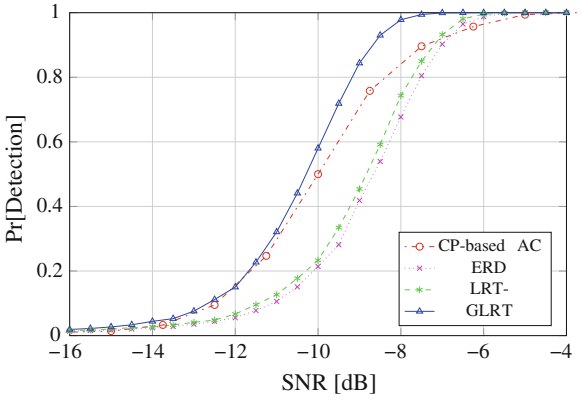

Fig. 4.13 CP-based versus eigenvalue-based (unknown σ_v^2) detectors, DVB-T 8k PU signal, flat fading channel,$N = 200$, $K = 4$, P_d versus SNR ($P_{fa} = 0.01$)

4.5.1.2 CP-Based Versus Eigenvalue-Based (unknown σ_v^2): AWGN Channel

In this last section we compare the eigenvalue-based algorithms for unknown noise variance against the technique exploiting the cyclic prefix autocorrelation of the received signal. Here, the AWGN channel model is adopted. In Figs. 4.12 and 4.13 we can observe that the performance of this algorithm is similar to that of the GLRT. This single-antenna algorithm does not require the computation of the sample covariance matrix eigenvalues, but resorts to a precise knowledge of the signal characteristics.

4.6 Conclusions

In this chapter, we analyzed and compared different sensing methods for DVB-T signals, a key element of cognitive systems operating in the TV white-spaces. First we considered feature-based techniques, which exploit OFDM peculiar characteristics. Then, we described multi-sensor algorithms that exploit the sample covariance matrix. The performance of these algorithms with real DVB-T signals have been assessed on different channel models. We believe that the presented results will be useful for researchers and designers who want to compare trade-off in terms of performance and efficiency of the different techniques in realistic conditions.

References

1. European Telecommunications Standards Institute, ETSI EN 300 744, Digital Video Broadcasting (DVB); framing structure, channel coding and modulation for digital terrestrial television, January (2009)
2. GNU Radio. http://gnuradio.org
3. Crespi, F.L., Maglioli, M., Benco, S., Perotti, A.: A real-time video broadcasting system based on the GNU Radio-USRP2 platform. Karlsruhe workshop on software radios, Karlsruhe (2012)
4. Benco, S., Crespi, F., Ghittino, A., Perotti, A.: Software-defined white space cognitive systems: implementation of the spectrum sensing unit. In: Proceedings of the 2nd Workshop of COST Action IC0902, Spain, Oct (2001)
5. Danev, D., Axell, E., Larsson, E.G.: Spectrum sensing methods for detection of DVB-T signals in AWGN and fading channels. In: Proceedings of IEEE 21st International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC), 2010
6. Cabric, D., Tkachenko, A., Brodersen, R.W.: Spectrum sensing measurements of pilot, energy and collaborative detection. In: Proceedings of IEEE Milcom 2006, 23–25 Oct 2006
7. Lopez-Valcarce, Vazquez-Villar, Sala: Multiantenna spectrum sensing for cognitive radio: overcoming noise uncertainty Cognitive Information Processing (CIP), In: 2nd International Workshop on 14–16 June 2010, pp. 310–315 2010
8. Tulino, A.M., Verdú, S.: Random matrix theory and wireless communications. Found. Trends Commun. Inf. Theory **1**(1), 1–182 (2004)
9. Anderson, G.W., Guionnet, A., Zeitouni, O.: An Introduction to Random Matrices. Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge (2009)
10. European Telecommunications Standards Institute: ETSI TR 102 377, Digital Video Broadcasting (DVB). DVB-H Implementation Guidelines, June (2009)
11. Urkowitz, H.: Energy detection of unknown deterministic signals. In: Proceedings of the IEEE, 1967
12. Tandra, R., Sahai, A.: Fundamental limits on detection in low SNR. In: Proceedings of WirelessComm05 Symposium on, Signal Processing, June (2005)
13. Roy, S.N.: On a heuristic method of test construction and its use in multivariate analysis. Ann. Math. Statist. **24**(20), 220–238 (1953)
14. Wei, L., Tirkkonen, O.: Cooperative spectrum sensing of OFDM signals using largest eigenvalue distributions. In: Proceedings of IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), 2009
15. Zhang, Q.: Advanced detection techniques for cognitive radio. In: International Conference on Communications (ICC), 2009

16. Zeng, Y., Liang, Y.-C.: Maximum-minimum eigenvalue detection for cognitive radio. In: The 18th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communication (PIMRC), 2007
17. Zeng, Y.H., Liang, Y.-C.: Eigenvalue based spectrum sensing algorithms for cognitive radio. IEEE Trans. Commun. **57**(6), 1784–1793 (2009)
18. Krzanowski, W.J.: Principles of Multivariate Analysis: A User's Perspective. Oxford University Press, Oxford (2000)
19. Bianchi, P., Najim, J., Alfano, G., Debbah, M.: Asymptotics of eigenbased collaborative sensing. In: Proceedings of IEEE Information Theory Workshop (ITW), 2009
20. Nadler, B., Penna, F., Garello, R.: Performance of eigenvalue-based signal detectors with known and unknown noise level. In: International Conference on Communications (ICC), 2011

Chapter 5
New Blind Free-Band Detectors Exploiting Cyclic Autocorrelation Function Sparsity

Ziad Khalaf and Jacques Palicot

Abstract In this chapter, we will firstly show that the Cyclic Autocorrelation Function (CAF) of a lineary modulated signal is a sparse function in the cyclic frequency domain. Then using this property we propose a new CAF estimator, using compressed sensing technique with the Orthogonal Matching Pursuit (OMP) algorithm. This new proposed estimator outperforms the classic estimator used in [1] under the same conditions, using the same number of samples. Furthermore, since our estimator does not need any information, we claim that it is a blind estimator whereas the estimator of [1] is clearly not blind because it needs the knowledge of the cyclic frequency. Many cases will be analysed: with and without the impact of a propagation channel at the reception. Using this new CAF estimator we propose two blind free bands detectors in the second part of this chapter. The first one is a soft version of the algorithm proposed in [2], that assumes that two estimated CAF of two successive packets of samples, should have close cyclic frequencies. The second one [3] uses Symmetry Property of the Second Order Cyclic Autocorrelation. Both methods outperform the cyclostationnarity detector of Dantawate-Giannakis of [1]. The second method outperforms the first one. Finally we study the complexity of the new proposed detectors and compare it to the complexity of [1].

5.1 Introduction

The wireless communication systems continue to grow [4] to become essential nowadays. This growth was accompanied by an increase in the demand of the spectrum resources needed by the wireless technology. This high demand made the spectrum

Z. Khalaf (✉) · J. Palicot
Supélec/IETR, avenue de la boulaie, Cesson Sevigné, France
e-mail: ziad.khalaf@supelec.fr

J. Palicot
e-mail: jacques.palicot@supelec.fr

M.-G. Di Benedetto and F. Bader (eds.), *Cognitive Communication and Cooperative HetNet Coexistence*, Signals and Communication Technology,
DOI: 10.1007/978-3-319-01402-9_5,

resources scarce. To solve this problem of spectrum scarcity, Joseph Mitola III [5] in 2000 introduced the idea of *Cognitive Radio* that relies on the idea of dynamic spectrum allocation. *Cognitive Radio* is widely expected to be the next *Big Bang* in the future wireless communication [6].

In 2002, the FCC (Federal Communications Commission), regulator and spectrum management in the United States, published a report [7] on the use of frequencies in which it is noted that, 70 % of cases, the spectrum is under-utilized according to time or space. The problem of spectrum scarcity is in fact an artificial problem mainly due to the current policy of static spectrum management is responsible. This policy of spectrum management is managed by the World Administrative Radio Conference (WARC), which updates the radio regulations that assign the use of radio spectrum in the world. To solve the current problem of spectrum management, new dynamic radio spectrum access approaches are developed, where opportunistic access is the most common because it tackles the cause of the shortage of frequencies. In fact, this approach proposes a new user called Secondary Users (SU) that can access the frequency resources allocated to Primary Users (PU) when they are not using them. Thus, the spectral efficiency is increased by allowing the transmission of SU over the free detected bands. These systems are called cognitive radio because, in addition to the autonomous detection of free bands, they must also be able to change their transmission parameters to meet the needs of the users and the constraints of availability of frequencies and the resources (band, signal to noise ratio (SNR)...).

Cognitive Radio (CR) and their resulting Cognitive Wireless Networks (CWN) have become one of the most studied in the wireless communications paradigms. The CR as initially defined in the work of Mitola [5] is a radio that can find, collect and learn from their environment and take action to make life easier for the user and to optimize the resources. An important application, often associated with cognitive radio is the Dynamic Spectrum Access (DSA). The dynamic spectrum access is the re-use of radio frequencies dismissed by the PU, by SU, i.e. when the PU does not use these frequencies at a given moment or in a given region of space [8], leaving part of the spectrum temporarily empty and available for secondary users. Work on dynamic spectrum access has become so important that the DSA is often regarded as the essential feature of cognitive radio.

As part of this work we are interested in the spectrum sensing problem which i.e. the detection of PU in the licensed spectrum, in the context of cognitive radio. We are not interested in a particular band (GSM or TV, for example), or to a particular system. The objective of this chapter is to propose effective methods of PU detection using the minimum *a priori* information about the signal to be detected and having short observation time.

In Sect. 5.2 we exploit the sparsity property of the Cyclic Autocorrelation Function to propose a new blind estimator based on the compressed sensing to estimate the Cyclic Autocorrelation Vector (CAV) which is a particular vector of the CAF for a fixed lag τ. Two metrics will be used to evaluate this estimate. The first, the mean squared error (MSE) compares the estimated vector with the theoretical reference obtained using the CAF. The second metric, denoted MSE_{α_f}, compares the estimated value of the cyclic frequency with its theoretical value. We show by simulation that

the new estimator gives better results than those obtained with the conventional (unbiased) estimator used in [1], which is not blind, in the same conditions and using the same number of samples. In Sect. 5.3 the proposed estimator in Sect. 5.2 is used to propose two blind detectors using fewer samples than the second order temporal detector of [1] requires, which is based on the classical estimator of the CAF. The first algorithm uses only the sparsity property of the CAV while the second detector uses in addition to the sparsity, the symmetry property of the CAV, allowing to get better detection performance. Both detectors, besides being blind, are more efficient than the non-blind detector of [1] in the case of a small number of samples. The performance of these detectors in several simulation environments are evaluated by comparing them to the cyclostationary detector. Finally, the complexity of these new detectors are studied and compared to that of non-blind detector of [1]. Finally Sect. 5.4 concludes this work and propose future perspectives.

5.2 A New Cyclic Autocorrelation Estimator Using Compressed Sensing

As written in [8], *in the most realistic case, when the transmitted sequences by the primary network are not a priori known by the secondary network, it is always possible to isolate some intrinsic properties of the transmitted signals in order to identify them. In particular, any redundancy character in an informative signal allows to differentiate it from a purely non colored Gaussian signal. This is always true, in principle, to a telecommunications signal* [9]. *We note in particular, the case of the OFDM modulation that conventionally uses a large cyclic prefixes, generating a temporal redundancy of the transmited signal*.

In this chapter, we will exploit the sparsity property of the CAF to propose a new estimator based on compressed sensing, to estimate the Cyclic Autocorrelation Vector (CAV), which is a particular vector of the CAF for a for a fixed delay τ. As mentioned in the introduction two metrics will be used to evaluate this estimation. The first, the Mean Square Error (MSE) compares the estimated CAV with the theoretical reference obtained using the CAF. The second metric, denoted MSE_{α_f}, compares the estimated value of the theoretical cyclic frequency. The simulation results of this new estimator show lower values of *MSE* and the MSE_{α_f} than those obtained with the conventional estimator (5.7) used in [1] under the same conditions and using the same number of samples. Several cases will be analyzed: at the reception using a filter at the transmission side with and without the addition of a propagation channel.

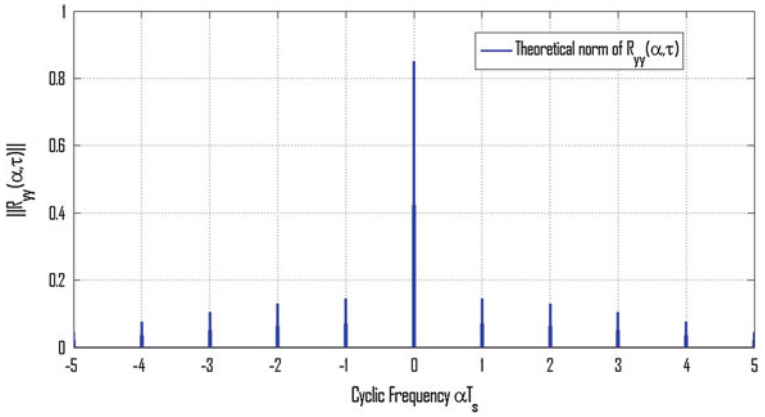

Fig. 5.1 The theoretical norm of the cyclic autocorrelation function of a BPSK

5.2.1 CAF and Sparsity

As presented in [10], the theoretical expression of the CAF of a linearly modulated signal, with the symbol period, T_s and a given lag τ, is given by Eq. (5.1).

$$R_{xx}(\alpha,\tau) = \begin{cases} \frac{\sigma_d^2}{T_s} e^{-j2\pi\alpha\varepsilon} \int_{-\infty}^{\infty} g(t-\frac{\tau}{2}) g^*(t+\frac{\tau}{2}) e^{-j2\pi\alpha t} dt \ \alpha = \frac{k}{T_s},\ k \in Z \\ 0 \qquad \text{elsewhere} \end{cases} \tag{5.1}$$

With ε an unknown delay, $g(t)$ the temporal impulse response of the transmission filter, $g^*(t)$ denotes the complex conjugate of $g(t)$ and σ_d^2 the power of a symbol. From Eq. (5.2) it is clear that the CAF is a sparse function in the cyclic frequencies domain α because it takes non-zero values for the values of α which are multiple integer of the fundamental cyclic frequency.

To simplify, we take σ_d^2 equal to 1, thus the expression of the CAF can be given by:

$$R_{xx}(\alpha,\tau) = \begin{cases} \frac{1}{T_s} e^{-j2\pi\alpha\varepsilon} \int_{-\infty}^{\infty} g(t-\frac{\tau}{2}) g^*(t+\frac{\tau}{2}) e^{-j2\pi\alpha t} dt \ \alpha = \frac{k}{T_s},\ k \in Z \\ 0 \qquad \text{elsewhere} \end{cases} \tag{5.2}$$

We can also check the sparsity property of the CAF by looking at Fig. 5.1 which shows the norm of the cyclic autocorrelation function of a BPSK signal for a fixed lag τ.

We define the Cyclic Autocorrelation Vector (CAV) as a particular vector of the CAF for a fixed lag $\tau = \tau_0$ and over a cyclic frequency domain $[\alpha_{min}, \alpha_{max}]$. The CAV is given by:

$$\mathbf{r}_{yy}^{(\tau_0)} = [R_{yy}(\alpha_{min},\tau_0), R_{yy}(\alpha_{min}+\delta_\alpha,\tau_0), \ldots, R_{yy}(\alpha_{max},\tau_0)]^T \tag{5.3}$$

with δ_α represents the resolution step.

From (5.2) it is easy to derive the theoretical expression for the norm of the CAF, using a rectangular temporal window $g(t)$ defined as follows:

$$g(t) = \begin{cases} 1 \ |t| \leq \frac{T_s}{2} \\ 0 \text{ elsewhere} \end{cases} \tag{5.4}$$

thus, the obtained result is as follows:

$$||R_{yy}(\alpha, \tau)||_2 = \begin{cases} |\frac{(T_s-\tau)}{T_s} sinc(\alpha(T_s - \tau))| \\ 0 \text{ for } \alpha \neq \frac{k}{T_s}, \ k \in Z \end{cases} \tag{5.5}$$

Equation (5.5) will be used as our theoretical reference in the calculation of the MSE of the vector obtained by estimating the CAV with several techniques in the rest of the work; we note then:

$$R_{yy}^{||\mathbf{ref}||}(\alpha, \tau) = ||R_{yy}^{\mathbf{ref}}(\alpha, \tau)||_2 = ||R_{yy}(\alpha, \tau)||_2 \tag{5.6}$$

5.2.2 Classic Estimation of the Cyclic Autocorrelation Vector

In order to estimate the CAV as defined in (5.3), we must first estimate the CAF (see Sect. 5.2.1) on the various sets (α_{min}, τ_0), $(\alpha_{min} + \delta_\alpha, \tau_0)$, ..., (α_{max}, τ_0).

A classical estimation of the CAF of a process $y(t)$ on a set (α, τ) can be made using the unbiased estimator used in [1]:

$$\hat{R}_{yy}^{\mathbf{cl}}(\alpha, \tau) \cong \frac{1}{N} \sum_{k=0}^{N-1} y(kT_e) y(kT_e + \tau) e^{-j2\pi\alpha k T_e} \tag{5.7}$$

with T_e is the sampling period. Therefore the vector $\hat{\mathbf{r}}_{yy}^{(\tau_0)}(\alpha)$ of $\widetilde{N}$ elements represents the conventional estimator of the CAV noted $\hat{CAV}_{classic}$ is given by Eq. (5.8):

$$\hat{VAC}_{classic} = \hat{\mathbf{r}}_{yy}^{(\tau_0)}(\alpha) = [\hat{R}_{yy}^{\mathbf{cl}}(-\alpha_{max}, \tau_0), \hat{R}_{yy}^{\mathbf{cl}}(-\alpha_{max} + \delta_\alpha, \tau_0), \ldots, \hat{R}_{yy}^{\mathbf{cl}}(\alpha_{max}, \tau_0)]^T \tag{5.8}$$

with $\delta_\alpha = \dfrac{2 \cdot \alpha_{max}}{\widetilde{N}} = \dfrac{1}{\widetilde{N} \cdot T_e}$ represents the resolution step. Note that $\mathbf{r}_{yy}^{(\tau_0)}(\alpha)$ can also be estimated using the FFT operator applied to the product $y(kT_e) \cdot y(kT_e + \tau)$ [11]. We define:

$$f_\tau(kT_e) = y(kT_e) y(kT_e + \tau) \tag{5.9}$$

By replacing in (5.7), we can write:

$$\hat{R}_{yy}^{\mathbf{cl}}(\alpha,\tau) \cong \frac{1}{N}\sum_{k=0}^{N-1} f_\tau(kT_e)e^{-j2\pi\alpha kT_e} \tag{5.10}$$

We define the vector $\mathbf{f}_\tau$ as follows:

$$\mathbf{f}_\tau = [f_\tau(0), f_\tau(1\cdot T_e), \ldots, f_\tau((\widetilde{N}-1)\cdot T_e)]^T \tag{5.11}$$

By observing Eq. (5.10), we can see that the vector $\hat{\mathbf{r}}_{yy}^{(\tau_0)}(\alpha)$ which is the estimated CAV of the signal $y(t)$ in the domain $[-\alpha_{max}, \alpha_{max}]$ is the Discrete Fourier Transform (DFT) of the vector $\mathbf{f}_{\tau_0}$ multiplied by $\frac{1}{\widetilde{N}}$:

$$\hat{\mathbf{r}}_{yy}^{(\tau_0)}(\alpha) = \frac{1}{\widetilde{N}} DFT(\mathbf{f}_{\tau_0}) \tag{5.12}$$

So $\mathbf{r}_{yy}^{(\tau_0)}(\alpha)$ can also be estimated using the FFT operator according to (5.12) we can write:

$$\hat{\mathbf{r}}^{FFT} = \frac{1}{\widetilde{N}} FFT(\mathbf{f}_{\tau_0}) \tag{5.13}$$

This estimator (5.13) will not be used in our simulations, because by using it, the number of samples N used to build the vector $\mathbf{f}_{\tau_0}$ is almost equal to the number of elements $\widetilde{N}$ of the estimated vector $\hat{\mathbf{r}}^{FFT}$, $(N = \widetilde{N} + \lceil \tau_0/T_e \rceil \cong \widetilde{N})$. But in this chapter it is sometimes interesting to vary the number of samples N to estimate a vector of size $\widetilde{N}$ fixed, making the estimator $\hat{CAV}_{classic}$ best suited.

5.2.3 *Blind Estimation of the Cyclic Autocorrelation Vector Using Compressed Sensing*

In this section we introduce a new method based on compressed sensing to blindly estimate the CAV i.e. without the need to know *a priori* value of the fundamental cyclic frequency in opposition to the classical method. In addition, using the same number of samples, with the compressed sensing we show that a better estimation is obtained.

Equation (5.12) will be used to justify the choice of the dictionary $\mathbf{A}$ in the sparse representation of the CAV as will be explained below.

It is proposed in this work to exploit the sparsity property of the CAV $\mathbf{r}_{yy}^{(\tau_0)}(\alpha)$ as it does contain non-zero elements only for few specific values of α. We can therefore reconstruct the vector $\hat{\mathbf{r}}_{yy}^{(\tau_0)}(\alpha)$ on the discrete interval of size $\widetilde{N}$ having δ_α as a resolution step which is equal to $\frac{2\cdot\alpha_{max}}{\widetilde{N}}$, using only n samples (or observations) instead of $\widetilde{N}$ $(n < \widetilde{N})$ with the classical method given by (5.12).

We therefore propose to apply a sparse representation, which is based on the representation of the first n elements of $\mathbf{f}_{\tau_0}$ over a complex value redundant base formed of $\widetilde{N}$ equidistant cisodes equispaced in frequency by a step δ_α, or equivalently over the columns (atoms) of the matrix $\mathbf{A}$ called the dictionary, which is a sub-matrix formed by the n first lines of $\mathbf{F}^*$, the complex conjugate of the Fourier square matrix $\mathbf{F}$ of dimension $\widetilde{N}$ (as $\hat{\mathbf{r}}_{yy}^{(\tau_0)}(\alpha)$ is simply a scaled version of the DFT of $\mathbf{f}_{\tau_0}$ from (5.12), $\mathbf{A}$ is then constructed using the lines of $\mathbf{F}^*$). We also note that all atoms should be normalized to 1, so they all get the same chance of being selected by applying any algorithm of inverse reconstruction.

We recall that the element (p, q) of the Fourier matrix $\mathbf{F}$ of order $\widetilde{N}$ is equal to $e^{-2i\pi(p-1)(q-1)/\widetilde{N}}$. The matrix $\mathbf{F}$ is a complex symmetric matrix ($\mathbf{F} = \mathbf{F}^T$). We also note that $\mathbf{F}\mathbf{F}^* = \widetilde{N}\mathbf{I}_{\widetilde{N}}$ with $\mathbf{I}_{\widetilde{N}}$ the identity matrix. So the DFT of a signal $\mathbf{y}$ of size $\widetilde{N}$ is given by $\bar{\mathbf{y}} = \mathbf{F}\mathbf{y}$ and the Inverse Discrete Fourier Transform (IDFT) of $\bar{\mathbf{y}}$ is given by $\mathbf{y} = (1/\widetilde{N})\mathbf{F}^*\bar{\mathbf{y}}$.

We define $\mathbf{b}^{(\tau_0)}$ the vector consisting of the first n elements of $\mathbf{f}_{\tau_0}$, so the problem is to solve the following inverse problem:

$$\mathbf{A}\mathbf{r}^{(\tau_0)} = \mathbf{b}^{(\tau_0)} \tag{5.14}$$

The solution $\hat{\mathbf{r}}^{(\tau_0)}$ of (5.14), which is a vector of $\widetilde{N}$ elements, represents the estimation of the CAV over the interval $[-\alpha_{max}, +\alpha_{max}]$. For complexity reasons we use the Orthogonal Matching Pursuit (OMP) algorithm [12] in order to solve the problem (5.14). We note the estimated CAV using the compressed sensing $\hat{CAV}_{CS}$, which is equal to $\hat{\mathbf{v}}_l$ the solution of the OMP after his last iteration l and is given by the following equation:

$$\hat{CAV}_{CS} = \hat{\mathbf{v}}_l \tag{5.15}$$

It is therefore noted that the estimator (5.15) is a blind estimator and therefore the atoms in the dictionary are not necessarily selected to be an integer divisor of the fundamental cyclic frequency. A sparse approximation is then made; the more δ_α is small the more the obtained approximation is better, unlike the classical estimator $\hat{CAV}_{classic}$ given by (5.8), that requires *a priori* knowledge of the fundamental cyclic frequency in order to estimate the CAV on the exact values of cyclic frequencies otherwise a vector composed of estimation noise is obtained and the spikes representing the cyclic frequencies are not observed.

5.2.3.1 Model Checking

It is found that the $\hat{CAV}_{CS}$ of $\widetilde{N}$ elements (obtained with the estimator (5.15)) can be reconstructed from only $n < \widetilde{N}$ samples for a given lag τ_0 by applying the technique of compressed sensing, in fact we plot on Fig. 5.2 the CAV of a BPSK modulation without filtering at the transmission side by using:

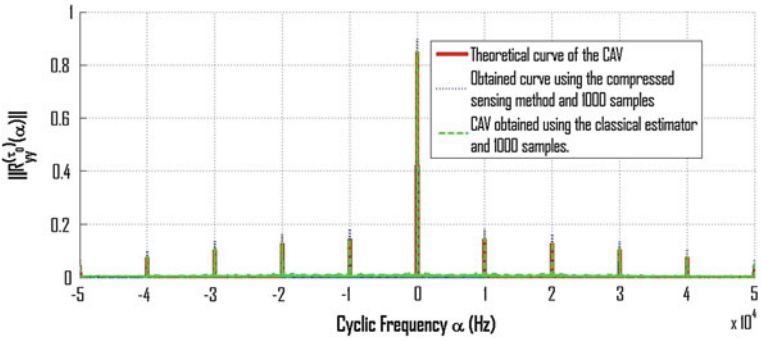

Fig. 5.2 The norm of the CAV of a modulated signal using a BPSK, with a fundamental cyclic frequency $\alpha_f = \frac{1}{T_s} = 10^4$ Hz and $\tau_0 = 3 \cdot T_e$, obtained by using (5.8), also compared to the theoretical curve (5.5), and to the norm of the CAV obtained after using the estimator (5.15). For both estimators, (5.8) and (5.15), 1000 samples have been used

- the sparse representation using the estimator (5.15) with a fixed number of iterations, using a dictionary $\mathbf{A}$ of size $(n, \widetilde{N})$ with $n = 1000$ and $\widetilde{N} = 4000$,
- the classical estimator (5.8) using 1000 samples,
- the theoretical curve representing the norm of the CAV given by (5.5).

In our example we have chosen arbitrarily and without loss of generality, $\tau_0 = 3 \cdot T_e$, with T_e the sampling period. The fundamental cyclic frequency of the BPSK used in the simulation of the transmitted signal is equal to $\alpha_f = \frac{1}{T_s} = 10^4$ Hz and $T_s = 20T_e$ represents the symbol period of the BPSK modulation. It can be seen in Fig. 5.2 that when we use the same number of samples (1000) to estimate the CAV using the OMP, we obtain better estimates than the curve obtained using (5.8) compared to the theoretical curve (5.5). One can check on Fig. 5.3 (a scaled version of Fig. 5.2), that the estimation noise is obtained using (5.8), by cons it is not observed any estimation noise on the curve obtained with the compressed sensing method. In addition, a very good estimate of the position of the spikes that are located on integer multiples of the fundamental cyclic frequency are observed using the compressed sensing. This result validates the sparse model of the CAV for telecommunications signals.

5.2.4 Performance Analyses of the New Blind Estimator

In this section we define two metrics, the *MSE* and the MSE_{α_f} in order to compare the quality of our proposed estimator $\hat{CAV}_{CS}$ and the classical estimator $\hat{CAV}_{classic}$. The first metric (the *MSE*) compares the estimated *CAV* with the theoretical reference given by Eq. (5.5). The second metric (MSE_{α_f}) compares the estimated value of the cyclic frequency with the theoretical value of the cyclic frequency.

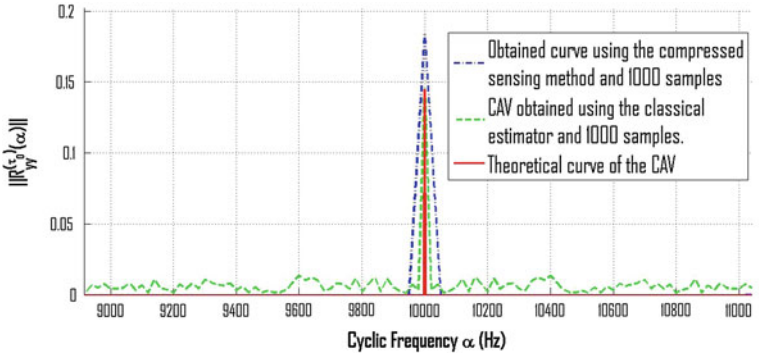

Fig. 5.3 A scaled version of the Fig. 5.2, in order to show the estimation noise obtained when estimator (5.8) is used

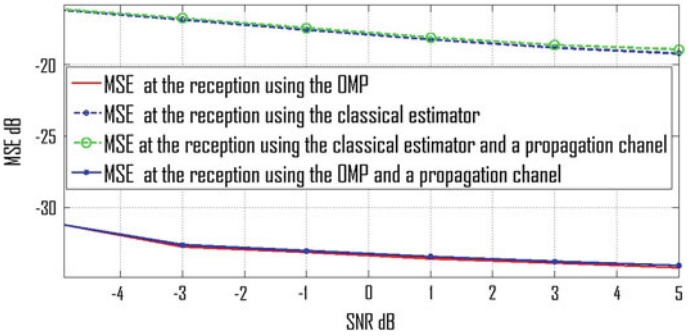

Fig. 5.4 The MSE at the reception as function of the *SNR* for both estimation methods using 300 samples with and without a propagation channel

5.2.4.1 Mean Square Error (First Criterion)

To evaluate the error of the estimation of the CAV which is represented by the vector $\hat{\mathbf{r}}_{yy}^{(\tau_0)}(\alpha)$ of $\widetilde{N}$ elements; a criterion is to calculate the MSE of $||\hat{\mathbf{r}}_{yy}^{(\tau_0)}(\alpha)||_2$ compared with the $\widetilde{N}$ elements of the reference $R_{yy}^{||\mathbf{ref}||}(\alpha, \tau_0)$ defined in (5.6), the $\widetilde{N}$ elements correspond to the theoretical values obtained using (5.5). We can then write:

$$MSE = \frac{1}{\widetilde{N}} \sum_{\alpha_i \in \mathscr{B}_\alpha} (||\hat{\mathbf{r}}_{yy}^{(\tau_0)}(\alpha_i)||_2 - R_{yy}^{||\mathbf{ref}||}(\alpha_i, \tau_0))^2 \tag{5.16}$$

with $\mathscr{B}_\alpha = \{\alpha | \alpha = k \cdot \delta_\alpha; \alpha \in [\alpha_{min}, \alpha_{max}]\}$ which means that the MSE is calculated on a $\mathbf{w} = [\alpha_{min}, \alpha_{max}]$. We note that in our study tude $\alpha_{min} = -\alpha_{max}$. We also note that the number of atoms $\widetilde{N}$ of the dictionary $\mathbf{A}$ is equal to 4000 in all the simulations of this chapter unless otherwise is stated.

5.2.4.2 MSE at the Reception

To observe the evolution of the MSE at the reception (with filtering at the transmission) as function of the SNR we plot in Fig. 5.4 the curve representing the MSE as a function of SNR for a fixed number of samples equal to 300 by averaging the value of the MSE over 1000 independent tests and this for both estimation methods. As expected the higher the SNR is, the smaller the MSE becomes regardless of the used estimation method. We introduce ΔMSE defined as the difference between the MSE of the classical method and the MSE of the OMP method. It is noted that for all values of SNR, ΔMSE is always greater than 10 dB highlighting the advantage of using the sparse representation method to estimate the CAV in the presence of noise and a filter at the transmission.

5.2.4.3 The MSE at the Reception with a Propagation Channel

In this part the filtered signal at the transmission passes through a propagation channel then marred by an additive noise. The pattern of transmission can be expressed by the input output relationship of the channel which is written, by noting $C(.)$ the corresponding function to the channel and $\mathbf{b}$ a white Gaussian noise:

$$\mathbf{y} = C(\mathbf{x}) + \mathbf{b} \tag{5.17}$$

The transformation C in our study will be taken as a linear filtering with time invariant and finite impulse response. The channel used in the simulations of this chapter is described in detail in the Appendix.

A. Simulation results of the MSE at the reception using a propagation channel
In this section we calculate the MSE at the reception for 300 samples using the channel defined in the Appendix, in the simulations. A Nyquist filter is used at the transmitter with a roll-off factor $\beta = 0.5$. The result is shown in Fig. 5.4 for both estimators (5.8) and (5.15). A small increase in the MSE is observed compared to the case that there is no propagation channel, (cf. Fig. 5.4). In addition we note that our new proposed estimator based on compressed sensing is always better than the classical estimator even in the most realistic case.

5.2.5 *Analyzing the Error in the Estimation of the Position of the Cyclic Frequency: The MSE_{α_f}*

The MSE defined in the previous section can be misleading in the evaluation of the estimated CAV, in fact after getting the CAV using the OMP, the vector obtained after performing a limited number of iterations contains by defaults too many zero elements, so it is highly correlated with the theoretical equation of CAV which

contains many zero elements as well. This large number of zero elements induces a lower MSE and this regardless of the position of the cyclic frequencies because their number is relatively small. To solve this problem and to better compare the different estimators of the CAV in a more precise manner, the estimation error on the position of the cyclic frequency (or estimated) should also be evaluated. We consider two cases with filtering at the reception with and without the addition of a propagation channel.

5.2.5.1 The Evaluation Parameter

To evaluate the estimation of the value of the cyclic frequency we evaluated the average of the Mean Square Error noted MSE_{α_f} between the estimated value and the theoretical value of the cyclic frequency. The normalized MSE_{α_f} is defined as follows:

$$MSE_{\alpha_f} = \sum_{i=1}^{N_{trial}} \frac{(\hat{\alpha}_i - \alpha_f)^2}{T_s^2 \cdot N_{trial}} \tag{5.18}$$

We recall that $\alpha_f = \frac{1}{T_s}$ is the fundamental cyclic frequency of the BPSK used in the simulation of the transmitted signal. A perfect estimator has a MSE_{α_f} perfectly equal to zero. It should be noted that the bias of the estimator (5.15) depends directly on the resolution step δ_α, in fact, in the most favorable case (sufficient number of samples and without noise at the reception) the bias is always less than δ_α as the atoms takes discrete values and therefore the estimate is discrete. Therefore, the more δ_α is small the smaller is the bias. In the particular case where δ_α is equal to an integer divisor of the fundamental cyclic frequency a perfectly zero bias can be achieved (under certain conditions on the number of samples and the SNR). In our simulation examples and to make a fair comparison between the two methods (5.8) and (5.15), the step δ_α in equal to an integer divisor of the cyclic frequency since the new estimator is compared to the non-blind estimator (5.8) and therefore the choice of the atoms must exactly match the values of the discrete interval $I'_{ALPHA} = [\frac{\alpha_f}{2}, \frac{3.\alpha_f}{2}]$ which is exactly centered on the value of the cyclic frequency. That is why under certain conditions, ideal estimates with a bias perfectly zero can be achieved.

5.2.5.2 Comparison of the Different Estimators Using the MSE_{α_f} Parameter

To compare the performance of two estimators (5.8) and (5.15) using the MSE_{α_f} criterion, several comparisons are made:

- calculation of the MSE_{α_f} at the reception, using (5.8) and (5.15), with filtering at the transmission.
- calculation of the MSE_{α_f} at the reception, using (5.8) and (5.15), with filtering at the transmission and a propagation channel (most realistic case).

Before making these comparisons we will explain in detail in 5.2.5.2. A and B how the estimate of the cyclic frequency $\hat{\alpha}$ is made with each of the two estimation methods (5.15) and (5.8) respectively.

A. The estimation of the cyclic frequency using compressed sensing
In this sub section is described how to estimate the cyclic frequency using the technique of compressed sensing (5.15). The cyclic frequency chosen for this comparison is the fundamental cyclic frequency $\alpha_f = \frac{1}{T_s} = 10$ kHz, the step δ_α of the dictionary A is equal to 50 Hz, the number of atoms in the dictionary is equal to 4000 representing the interval I_{ALPHA} from $\alpha_{min} = -100$ kHz till $\alpha_{max} = 100$ kHz. After solving the inverse problem using a fixed number of iterations of the OMP, it is considered that the value $\hat{\alpha}$, which represents the estimate of α_f, is the value corresponding to the abscissa of the atom of the $\hat{CAV}_{CS}$ having the largest norm and located within the range of $I'_{ALPHA} = [\frac{\alpha_f}{2}, \frac{3.\alpha_f}{2}] = [5000, 15000]$, which is centered on α_f and of width α_f. We note that the bounds of the interval I'_{ALPHA} are $\frac{\alpha_f}{2}$ and $\frac{3.\alpha_f}{2}$ since all values outside this range will be closer to a cyclic frequency other than α_f such as $\alpha = 0, \alpha = 2\alpha_f, \alpha = 3\alpha_f$... So it is not appropriate to consider that the external value of the interval I'_{ALPHA} is an estimate of α_f.

B. The estimation using the classical estimator
To compare the classical estimation method (5.8) with the one that uses the compressed sensing we apply the estimator (5.8) on the same discrete interval I'_{ALPHA} used with (5.15) and this is to give both estimation methods the same resolution and thus to make a fair comparison, then the abscissa of the norm of the $\hat{CAV}_{classic}$ with a maximum value is considered as the estimated value of the cyclic frequency. In fact, the theoretical norm of the CAV is maximal at the position of the cyclic frequency (and zero elsewhere). These estimates are repeated on a large number of tests in order to calculate the MSE_{α_f}, using (5.18).

5.2.5.3 MSE_{α_f} at the Reception

To simulate the results at the reception we fixed a total number of samples equal to 1000 for each estimator, then we add an additive Gaussian noise. 1000 independent tests where made in the case of each estimator for each value of the signal to noise ratio SNR. We obtain the curve representing the MSE_{α_f} as a function of the SNR in Fig. 5.5. We therefore conclude that with additive Gaussian noise and filtering issue the MSE_{α_f} obtained with the method using compressed sensing is always lower than that the one obtained with the classical estimator (5.8).

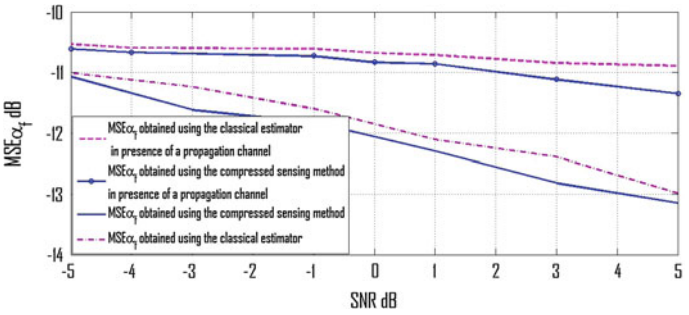

Fig. 5.5 The curves of the MSE_{α_f} obtained using the estimator (5.8) and the compressed sensing method for 1000 samples as function of the SNR with and without a propagation channel

5.2.5.4 MSE_{α_f} at the Reception with the Addition of a Propagation Channel

In this subsection the filtered signal at the transmission is convoluted with a propagation channel then an additive noise is added. The pattern of the transmission can be expressed by the input output channel relationship which is written by noting $C(.)$ the corresponding function to the channel and $\mathbf{b}$ a white Gaussian noise:

$$\mathbf{y} = C(\mathbf{x}) + \mathbf{b} \tag{5.19}$$

The transformation C in our study will be taken as a linear filtering with finite impulse response and time-invariant described in the Appendix. We calculate the MSE_{α_f} using a Nyquist filter at the transmission with $\beta = 0.5$ convolved with the channel defined in the Appendix, for many SNR by setting the number of samples to 1000, and this for both estimation methods. In Fig. 5.5 we observe the simulation results. An increase of the MSE_{α_f} is observed after the introduction of the propagation channel compared to the results without the propagation channel. In addition we note that the method using the compressed sensing is better than the conventional estimator which allows us to conclude that in the more realistic case the new proposed method is always more advantageous than the classical method without omitting that the classical method must be repeated over several "candidates" of α_f against the new totally blind method, which is performed only once.

5.3 Reduced Complexity Blind Detection Using Sparsity

In the previous section we introduced a new estimator of the cyclic autocorrelation for a given lag which is based on the compressed sensing. It has been verified by simulation that the two criteria already defined (the MSE and the MSE_{α_f}) for the evaluation of the quality of the estimate of the CAV are always better for our new estimator compared to the classical estimator used in [1], even in the most realistic

case when using a filter at the transmission and a propagation channel. In this section we will use the estimator that we have proposed previously. We propose two blind detectors using fewer samples than the second order temporal detector of [1] requires which is based on the classical estimator of the CAF. Both detectors are blind they are also more efficient than the non-blind detector of [1].

Mainly two detection methods that are based on the new estimator will be proposed, the first is based on the comparison of the cyclic frequencies obtained in two consecutive slots, and therefore in this case it is assumed that two consecutive slots both belong either to H_0 or to H_1 with H_1 and H_0 denote the binary hypotheses that a primary user is present and absent, respectively. This last assumption is realistic because the size of the used slots is very small. The second detection method, in addition to the sparsity property of the CAF will use its symmetry property. This symmetry, which is exploited in addition to the sparsity property of the CAF, will improve the detection performance compared to the first method that only exploits the sparsity property.

5.3.1 Slot Comparison Methods (Hard and Soft)

5.3.1.1 Slot Comparison Methods: The Hard Version

To choose between H_0 and H_1, the Slot Comparison Method (SCM) takes two consecutive slots of size n_s for each slot. Note that this method can be applied using several slots, but for complexity reasons of we will use only 2 slots. Both slots are denoted $\mathbf{s}_1 = [y(0), y(1), \ldots, y(n_s - 1)]^T = [y_1(0), y_1(1), \ldots, y_1(n_s - 1)]^T$ and $\mathbf{s}_2 = [y(n_s), y(n_s + 1), \ldots, y(2n_s - 1)]^T = [y_2(0), y_2(1), \ldots, y_2(n_s - 1)]^T$. It is assumed that two consecutive slots belong to the same event (both are either in H_0 or H_1). The idea of this test is to estimate the CAVs $\hat{\mathbf{r}}^{(\tau_0)}_{y_1y_1}(\alpha)$ and $\hat{\mathbf{r}}^{(\tau_0)}_{y_2y_2}(\alpha)$ of size $(\widetilde{N}, 1)$ by exploiting their sparsity property (using only n_s samples for each estimate) then by comparing the obtained cyclic frequencies for each slot (or equivalently by comparing the indexes of the elements with a maximal norm in the CAV. We note by index the position of the non zero element in the CAV which is an integer between 1 and the maximal size of the estimated CAV with the new estimator). Note that the index representing the cyclic frequency $\alpha = 0$ is not taken into account because under H_0 or under H_1, a non-zero value will be obtained for $\alpha = 0$ because it represents the value of the classical autocorrelation of the received signal for $\tau = \tau_0$. Then if the obtained indexes of the cyclic frequencies $\hat{\mathbf{r}}^{(\tau_0)}_{y_1y_1}(\alpha)$ and $\hat{\mathbf{r}}^{(\tau_0)}_{y_2y_2}(\alpha)$ are identical or close, H_1 is retained otherwise H_0 is chosen as the noise have no cyclic frequency.

To estimate $\hat{\mathbf{r}}^{(\tau_0)}_{y_1y_1}(\alpha)$, (resp. $\hat{\mathbf{r}}^{(\tau_0)}_{y_2y_2}(\alpha)$) one must solve the system $\mathbf{A}\mathbf{r}^{(\tau_0)}_1 = \mathbf{b}^{(\tau_0)}_1$ (resp. $\mathbf{A}\mathbf{r}^{(\tau_0)}_2 = \mathbf{b}^{\tau_0}_2$) given by (5.14) using the OMP. We recall that $\mathbf{b}^{(\tau_0)}_1$ (resp. $\mathbf{b}^{(\tau_0)}_2$) is constructed using the n_s first elements of the vector $\mathbf{f}^{(1)}_{\tau_0}$, (resp. $\mathbf{f}^{(2)}_{\tau_0}$) as defined in (5.11) with the samples of the slot $\mathbf{s}_1$ (resp. $\mathbf{s}_2$). Then after obtaining $\mathbf{b}^{(\tau_0)}_1$ and

$\mathbf{b}_2^{(\tau_0)}$ using the $2n_s$ samples of the slots $\mathbf{s}_1$ and $\mathbf{s}_2$, we solve both inverse problems $\mathbf{A}\mathbf{r}_1^{(\tau_0)} = \mathbf{b}_1^{(\tau_0)}$ and $\mathbf{A}\mathbf{r}_2^{(\tau_0)} = \mathbf{b}_2^{(\tau_0)}$ using the OMP.

The obtained vectors $\hat{\mathbf{r}}_1^{(\tau_0)}$ and $\hat{\mathbf{r}}_2^{(\tau_0)}$ that respectively represents $\hat{\mathbf{r}}_{y_1y_1}^{(\tau_0)}(\alpha)$ and $\hat{\mathbf{r}}_{y_2y_2}^{(\tau_0)}(\alpha)$ are compared together. More precisely, what should be compared are the positions (or indexes) of the elements of the vectors $\hat{\mathbf{r}}_1^{(\tau_0)}$ and $\hat{\mathbf{r}}_2^{(\tau_0)}$ that correspond to the estimated cyclic frequency (the index of the estimated cyclic frequency corresponds to the index of the element of the estimated vector $\hat{\mathbf{r}}^{(\tau)}$ having the highest norm. We note $index_1$ and $index_2$ the indexes corresponding to the cyclic frequencies of $\hat{\mathbf{r}}_1^{(\tau_0)}$ and $\hat{\mathbf{r}}_2^{(\tau_0)}$ respectively). We recall that we do not take into account the indexes that are corresponding to the cyclic frequency $\alpha = 0$, since under both hypotheses (H_0 or H_1) the cyclic frequency zero, represents the classique autocorrelation of the received signal and takes a non zero value, making it impossible to perform the test at the cyclic frequency zero. Then if $|index_1 - index_2| < k$ we consider that $\mathbf{s}_1$ and $\mathbf{s}_2$ have the same cyclic frequency and the boolean variable Δ is then fixed to 1. Otherwise if $|index_1 - index_2| \geq k$, Δ is then set to zero. We note that by choosing the variable k, which can take any integer value between 1 and $\widetilde{N}$ (the size of the vector $\hat{\mathbf{r}}^{(\tau_0)}$), we can fix the performances of the detector as function of the false alarm rate. We note that $\hat{\mathbf{r}}_{y_1y_1}^{(\tau_0)}(\alpha)$ and $\hat{\mathbf{r}}_{y_2y_2}^{(\tau_0)}(\alpha)$ have to be estimated several time over many values of τ_i, $i \in \{1, 2, \ldots, M\}$ to increase the chances of detection, since it is not necessary that $R(\alpha, \tau)$ takes a non zero value for any value of τ, even if α is a cyclic frequency of the received signal. In fact if we take a look at the expression of the norm of the CAF given by (5.5), we can conclude that it is equal to zero, whatever the value of α is (α non zero), for the following values of τ:

$$\tau = T_s + \frac{k\pi}{\alpha},\ k \in Z \tag{5.20}$$

By repeating the same test over many τ_i (by comparing all the $\hat{\mathbf{r}}_1^{(\tau_i)}$ and $\hat{\mathbf{r}}_2^{(\tau_i)}$ together one by one) we obtain M boolean values (Δ_1,…,Δ_M), and the final decision will be taken using the rule of decision OR:

$$\Gamma = \sum_{i=1}^{M} \Delta_i \tag{5.21}$$

We also note that when a boolean variable takes the value 1 the algorithm stops and takes the decision, minimizing the complexity of calculation. Here are the algorithm lines:

$\mathbf{s}_1 \leftarrow [y_1(0), \ldots, y_1(n_s - 1)]^T$
$\mathbf{s}_2 \leftarrow [y_2(0), \ldots, y_2(n_s - 1)]^T$
for $i = 1$ M **do**
  $\mathbf{b}_1^{(\tau_i)} \leftarrow$ calculation of the n_s elements of $\mathbf{f}_{\tau_i}^{(1)}$
  $\mathbf{b}_2^{(\tau_i)} \leftarrow$ calculation of the n_s elements of $\mathbf{f}_{\tau_i}^{(2)}$
  $\hat{\mathbf{r}}_1^{(\tau_i)} \leftarrow \text{OMP}(A, \mathbf{b}_1^{(\tau_i)})$
  $\hat{\mathbf{r}}_2^{(\tau_i)} \leftarrow \text{OMP}(A, \mathbf{b}_2^{(\tau_i)})$
  $index_1 \leftarrow index(max(|\hat{\mathbf{r}}_1^{(\tau_i)}|))$
  $index_2 \leftarrow index(max(|\hat{\mathbf{r}}_2^{(\tau_i)}|))$
  note: $index_1$ and $index_2$ are chosen without taking into account the indexes of the zero cyclic frequency
  if $|index_1 - index_2| < k$ **then**
    $\Delta_i = 1$
  else
    $\Delta_i = 0$
  end if
  $\Gamma \leftarrow \sum_{l=1}^{i} \Delta_l$
  if $\Gamma \geq 1$ **then**
    H_1 is chosen
  end if
end for
H_0 is chosen

5.3.1.2 Slot Comparison Methods: The Soft Version

After introducing the SCM, we propose in this section an enhanced method of this last one that we will call SCMS (Slot Comparison Method Soft version). There is no added complexity on the SCMS method compared to the SCMS. The only added modification is going to be only on the decision making rule. In fact rather than using the boolean variables Δ_i and to make the choice H_1 if one of these variables is "true", the idea of the SCMS is to save the differences between the indexes δ_i of the cyclic frequencies of $\mathbf{s}_1$ and $\mathbf{s}_2$. This difference δ_i will be saved for each lag τ_i, with i starting from 1 to M. Finally we calculate the mean value of the variables δ_i using:

$$\bar{\delta} = \frac{1}{M} \sum_{i=1}^{M} \delta_i \tag{5.22}$$

And finally the final decision H_1 will be taken if $\bar{\delta}$ is lower than k, and k starting from 1 to $\widetilde{N}$. It is then evident that the smaller k is the smaller p_{fa}. These are the algorithm lines of the SCMS:

$\mathbf{s}_1 \leftarrow [y_1(0), \ldots, y_1(n_s - 1)]^T$
$\mathbf{s}_2 \leftarrow [y_2(0), \ldots, y_2(n_s - 1)]^T$
for $i = 1$ to M **do**
  $\mathbf{b}_1^{(\tau_i)} \leftarrow$calculation of the n_s first elements of $\mathbf{f}_{\tau_i}^{(1)}$
  $\mathbf{b}_2^{(\tau_i)} \leftarrow$calculation of the n_s first elements of $\mathbf{f}_{\tau_i}^{(2)}$
  $\hat{\mathbf{r}}_1^{(\tau_i)} \leftarrow \text{OMP}(A, \mathbf{b}_1^{(\tau_i)})$
  $\hat{\mathbf{r}}_2^{(\tau_i)} \leftarrow \text{OMP}(A, \mathbf{b}_2^{(\tau_i)})$
  $index_1 \leftarrow index(max(|\hat{\mathbf{r}}_1^{(\tau_i)}|))$
  $index_2 \leftarrow index(max(|\hat{\mathbf{r}}_2^{(\tau_i)}|))$
  note: $index_1$ and $index_2$ are chosen whiteout taking into account the indexes of the zero cyclic frequency
  $\delta_i \leftarrow |index_1 - index_2|$
end for
$\bar{\delta} \leftarrow \frac{1}{M} \sum_{i=1}^{M} \delta_i$
if $\bar{\delta} < k$ **then**
  H_1 is chosen
else
  H_0 is chosen
end if

5.3.1.3 Performance Detection of the SCM, SCMS and the Cyclostationary Detector

In order to observe the performance gain of the SCMS over the SCM and the cyclostationary detector, we plot the ROC curves on Fig. 5.6 of these different detectors using the same signal as in Sect. 5.2.1 with $M = 5$, $n_s = 200$ (i.e. $n_{tot} = 2n_s = 400$ is the total number of samples used globally for each detector), and the $SNR = 0$ dB. The result is shown on Fig. 5.6. It clearly shows the improved detection performance using the soft method against the hard method (SCM) while keeping the same complexity with the same blind property. For these reasons, in the remainder of this chapter, the SCMS is used for the remaining analyzes and comparisons.

Now we compare our proposed test (the SCMS) to the second order detector of cyclostationarity in the time domain published in [1] requiring the *a priori* knowledge of the cyclic frequency of the transmitted signal (in this case $\alpha_f = \frac{1}{T_s}$). To make a fair comparison we used 400 samples for cyclostationary detector, and using the same set of delays τ_i i.e. ($M = 5$). Then we plot on the same Fig. 5.6, the results obtained with the cyclostationary detector for the same SNR. The conclusion is that under the same conditions, except that the cyclostationary detector has additional information that is the *a priori* knowledge of the cyclic frequency α_f, our new detector (SCMS) outperforms the cyclostationary detector. Note that for a robust detection, cyclostationary detector requires a large number of samples [13], which explains the poor performance of cyclostationary detector obtained with a small number of samples.

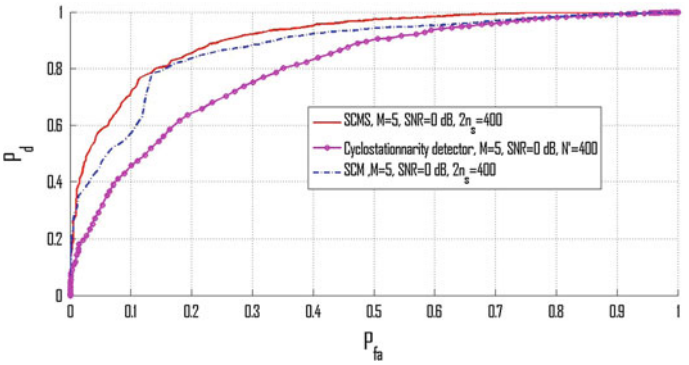

Fig. 5.6 The ROC curves of the SCM, SCMS and the cyclostationary detector for a total number of samples equal to 400 and an $SNR = 0$ dB

5.3.1.4 The Effect of the Propagation Channel

To study the impact of the propagation channel on the performance of the SCMS and the cyclostationary detector, we evaluate the performances of these two detectors based on the SNR with and without a propagation channel. Note that in both cases a filter at the transmission is used to simulate a realistic situation.

We choose to see the impact of the Rayleigh channel of unit variance (each sample is multiplied by a coefficient which follows a Rayleigh distribution with unit variance and these coefficients are i.i.d., this model of Rayleigh channel is used throughout this chapter). The Rayleigh channel is selected because it is a very severe channel that is normally used to represent in digital communications, situations where the transmitter is not in direct vision with the receiver.

In Fig. 5.7 we plot the probability of correct detection for a false alarm set to 10 % as function of the SNR for the SCMS and the cyclostationarity detector using a total of 400 samples and with $M = 5$ for both methods. We observe a degradation of the two detection methods after the introduction of the Rayleigh channel inducing a loss of 4 dB in term of SNR for the SCMS. In fact, before the introduction of channel the 100 % detection probability is achieved for $SNR = 4$ dB, and after its introduction this probability is reached for $SNR = 8$ dB. Regarding the cyclostationarity method a slight degradation is also observed, also the cyclostationary detector is already degraded due to the low number of used samples.

5.3.2 The Symmetry Method

In this section we propose another blind detection method based not only on the sparsity property of the CAF but also on its symmetry properties with respect to the axis $\alpha = 0$ for a given lag τ.

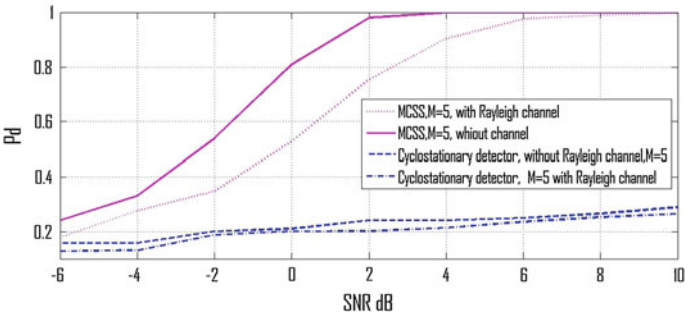

Fig. 5.7 The probability of detection as a function of the SNR for a given false alarm rate set to 10 % for the SCMS and the cyclostationary detector. Two cases are considered: with and without a propagation channel for both detection methods

5.3.2.1 Symmetry Property

We recall from [24] that for real values of $y(t)$, $R_{yy}(\alpha, \tau)$ represents the following symmetry properties:

$$R_{yy}(\alpha, -\tau) = R_{yy}(\alpha, \tau) \tag{5.23}$$

$$R_{yy}(-\alpha, \tau) = R^*_{yy}(\alpha, \tau) \tag{5.24}$$

with $R^*_{yy}(\alpha, \tau)$ represents the complex conjugate of $R_{yy}(\alpha, \tau)$. By taking the two members of the Eq. (5.24) we obtain:

$$||R_{yy}(-\alpha, \tau)||_2 = ||R_{yy}(\alpha, \tau)||_2 \tag{5.25}$$

We can also verify the property of the Eq. (5.25) by looking on Fig. 5.1. We can observe on Fig. 5.1 the symmetry of the CAF around $\alpha = 0$.

The main idea of this detector is to partially estimate the CAV using few samples with the OMP. After the partial reconstruction of the CAV, the property of symmetry around $\alpha = 0$ is tested. If the obtained CAV checks approximately (5.25) then H_1 is selected, otherwise it is H_0. It is important to note that under H_0 the CAV is theoretically symmetrical and therefore satisfies the property (5.25) for real values of $y(t)$ under H_1 and H_0, but under H_0, when few iterations are used for the reconstruction of the CAV with the OMP, the probability of obtaining a symmetric CAV is very low. This last point will be explained in details later.

5.3.2.2 The Symmetry Method SM

Let $\mathbf{b}^{(\tau_0)}$ the vector defined previously (cf. Sect. 5.3.1), constructed using the first n elements received from $y(t)$. We fix l (odd) the number of iteration of the OMP in order to solve $\mathbf{b}^{(\tau_0)} = \mathbf{A}\mathbf{r}^{(\tau_0)}$. Then after obtaining the solution vector $\mathbf{r}_l^{(\tau_0)}$, it will

be composed of null elements except l non zero elements (equal to the number of iterations of the OMP). We define by $IND_{sym}^{(\tau_0)}$ the index of symmetry of the vector $\mathbf{r}_l^{(\tau_0)}$. To calculate the index $IND_{sym}^{(\tau_0)}$ we ignore the element of $\mathbf{r}_l^{(\tau_0)}$ corresponding to $\alpha = 0$ (having the largest amplitude), then $IND_{sym}^{(\tau_0)}$ is obtained by calculating the mean value of the absissa of the $l-1$ remaining elements. The more $IND_{sym}^{(\tau_0)}$ is closer to sero the more the symmetry is considered ideal. An ideal symmetry is obtained when the estimation is perfect i.e. for $IND_{sym}^{(\tau_0)} = 0$. $IND_{sym}^{(\tau_0)}$ is obtained using the following equation:

$$IND_{sym}^{(\tau_0)} = \frac{1}{l-1} \sum_{j=2}^{l} \alpha_j \tag{5.26}$$

Before taking the final decision we note that $\mathbf{r}_l^{\tau_i}$ should be estimated over different values τ_i, $i \in \{1, 2, \ldots M\}$, and this to increase the probability of detection since it is not necessary that $R(\alpha, \tau)$ takes zero value for any value of τ even if α is a cyclic frequency of the received signal (cf. Eq. 5.20). The algorithm should then calculates M different values of $IND_{sym}^{(\tau_i)}$ (using M times the OMP algorithm) and the final decision will be made using the equivalent index obtained by combining all the indexes making then a soft decision:

$$IND_{sym}^{(equ)} = \frac{1}{M} \sum_{i=1}^{M} |IND_{sym}^{(\tau_i)}| \tag{5.27}$$

Finally $IND_{sym}^{(equ)}$ will be compared to a positive threshold ξ to decide between H_0 or H_1. We note that the more ξ is large the more the probabilities of detection and false alarm increases and vice versa. These are the algorithm lines of the SM:

$\mathbf{s} \leftarrow [y_1(0), \ldots, y_1(n_{tot} - 1)]^T$
for $i = 1$ to M **do**
  $\mathbf{b}^{(\tau_i)} \leftarrow$calculation of the n_{tot} first elements of $\mathbf{f}_{\tau_i}$
  $\hat{\mathbf{r}}_l^{(\tau_i)} \leftarrow \text{OMP}(A, \mathbf{b}^{(\tau_i)})$
  $\alpha_1 \leftarrow$ the number 1 selected atom of $\hat{\mathbf{r}}_l^{(\tau_i)}$
  .
  .
  $\alpha_l \leftarrow$ the number l selected atom of $\hat{\mathbf{r}}_l^{(\tau_i)}$
  $IND_{sym}^{(\tau_i)} = \frac{1}{l-1} \sum_{j=2}^{l} \alpha_j$
end for
$IND_{sym}^{(equ)} = \frac{1}{M} \sum_{i=1}^{M} |IND_{sym}^{(\tau_i)}|$
if $IND_{sym}^{(equ)} < \xi$ **then**
  H_1 is chosen
else
  H_0 is chosen
end if

5.3.2.3 The Choice of the Number of Iterations

The objective of the detection algorithm is not to make a complete reconstruction of the CAV. The goal is only to detect the existence of a signal in the band. For this reason it is wise to choose l in order to maximize the detection performance and minimize the complexity of the algorithm at the same time. In the following subsections we will explain in detail why it is wise to choose an odd number of iterations l on the one hand and a minimum of iterations on the other hand.

A. An odd number of iterations

The number of iterations l must be an odd number in order to evaluate correctly the $|IND_{sym}^{\tau_0}|$, since as it was mentioned the algorithm needs to verify the symmetry around $\alpha = 0$. The element in $\mathbf{r}_l^{\tau}$ having the highest amplitude corresponds to $\alpha = 0$, and is obtained with high probability after the first iteration of the OMP (since for $\alpha = 0$, the CAV have the highest amplitude cf. (5.5)). Then the $l-1$ (even) non zero elements remaining represents the cyclic frequencies which theatrically are pairwise symmetric with respect to $\alpha = 0$. In fact, for each cyclic frequency α must be another cyclic frequency on $-\alpha$ and having the same norm.

B. A minimal number of iterations

Two main reasons make sense to choose a minimum number of iterations ($l = 3$). The first reason is related to the complexity of the OMP, which directly depends on the number of iterations l, and therefore choosing a minimum value of l is equivalent to minimizing the complexity of the detection algorithm. Note that the complexity of this algorithm is the same as that of the OMP that is given by [14] multiplied by M, and is given by $O(M \cdot n_{tot} \cdot l \cdot \widetilde{N})$, with n_{tot} is the total number of samples used to detect, and $\widetilde{N}$ is the total number of atoms used to define the dictionary. The second major reason to choose a minimal l is to minimize the false alarms. Although Eq. (5.25) holds under H_0 and H_1 symmetry method can always be used to distinguish between H_0 and H_1, in fact under H_1 when the OMP is applied, it is likely that the symmetric atoms around $\alpha = 0$ and close to the value of the fundamental cyclic frequency will be selected consecutively after consecutive iterations as they have the same norm (which is different from zero because of the symmetry in fact by definition the OMP selects the most correlated atoms with the residual vector after each iteration). In contrast, under H_0 the CAV is zero, because the noise does not have cyclic frequencies except for $\alpha = 0$ (takes the value of the classical autocorrelation). For this reason under H_0 atoms are not selected with a specific order and minimizing the probability of having a symmetric reconstructed CAV $\mathbf{r}_l^{\tau}$ with a low value of l. We note that under H_0, the number of iterations l increases, a more complete reconstruction of $\mathbf{r}_l^{\tau}$ will be made and the probability of generating a symmetric vector increases inducing more false alarms. This last result was shown by simulation in [15]. For these reasons it is desirable to work with a minimum number of iterations l to minimize the complexity and maximize the performance of detection for a given false alarm rate.

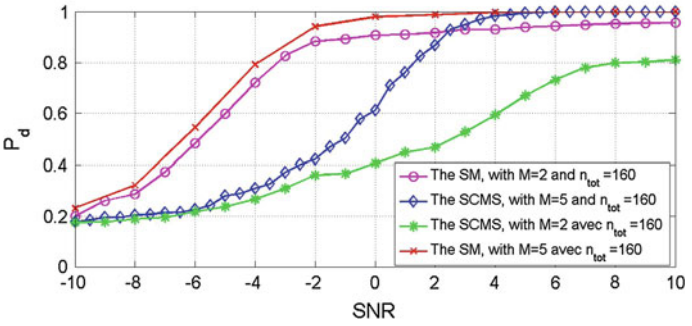

Fig. 5.8 The probability of detection as a function of the SNR for a given false alarm rate equal to 15 % with a total number of samples equal to 160 for the SCMS and the SM

5.3.2.4 Simulation Results and Performance Comparison of the SM

In this section we use the same BPSK modulation used previously. A small dictionary **A** of size (160, 512) is used (i.e. $n_{tot} = 160$). The number of iterations l is fixed to 3 for the previously mentioned reasons. We compare the performances of the SM detector with the SCMS that was previously proposed. Figure 5.8 show the probability of detection as function of the SNR for a fixed false alarm fixed to 15 % by using a total number of samples equal to 160 for both methods in order to make a fair comparison. Two different group of lag τ_i are used in the simulations and this for both detection methods. The first group is formed by only two values of τ ($M = 2$), and the other is constituted of five different values ($M = 5$). We can conclude firstly that whatever the group of the used lags ($M = 2$ or $M = 5$) the SM outperforms in performance the SCMS. We can also verified that when the number of lags M increase that the detection performance increases for both detection methods (as expected). Finally we can observe that for $M = 5$, the SM achieve 90 % of detection versus only 42 % for the SCMS, and this is in the same simulations conditions and for the same complexity.

5.3.2.5 The Influence of the Transmission Channel

To study the impact of the propagation channel on the performance of the SM we evaluate the performance of this method as function of the SNR, with and without a propagation channel. Note that in both cases a transmission filter is used in order to be in a realistic situation. We selected the same channel used in Sect. 5.3.1.4, which is a Rayleigh channel with unit variance. On Fig. 5.9 we plot the probability of correct detection for a fixed false alarm set to 10 % as a function of the SNR for the SM with and without propagation channel, using a total of 160 samples where $M = 5$ for both cases. We observe the degradation of the SM after the introduction of the Rayleigh channel that induced a loss of 2 dB in SNR for the SM. In fact, before the

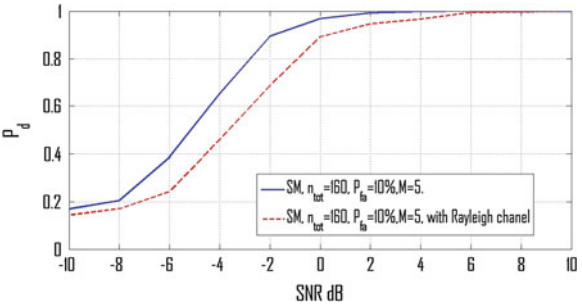

Fig. 5.9 The probability of detection of the SM as function of the *SNR* for a false alarm rate equal to 10 % and a sample number equal to 160. We distinguish two case: with and without a propagation channel

introduction of the channel, the detection probability 100 % is achieved for an SNR of 4 dB, and after its introduction this probability is reached for an SNR of 6 dB.

5.3.2.6 Complexity Analysis and Observation Time for the SCMS, the SM and the Cyclostationnary Detector

The complexity of the new SCMS detector is the same as that of the OMP which is given in [14], multiplied by $2 \cdot M$, knowing the OMP is used $2 \cdot M$ times. We note that the complexity of the OMP is equal to $O(l_1 \cdot l_2 \cdot l_3)$, where l_1, l_2, and l_3 are the number of lines of the dictionary, the number of atoms, and the number of iterations respectively. We found then that the complexity of the SCMC is then equal to $O(2 \cdot n_s \cdot M \cdot S \cdot \widetilde{N})$, with S the number of iterations used by the OMP. In practice S is equal to 3, because after filtering only the zero cyclic frequency and the two fundamental cyclic frequency at $\pm\alpha_f$ appears in the CAF for a given delay τ.

We calculated the complexity of the second order time domain cyclostationarity detector [1], we find that the result is equal to $O(M \cdot N' \cdot (L+1) + 4 \cdot M \cdot L^2 + 8 \cdot M^3 + 6 \cdot M^2 + 2 \cdot M) \cong O(M \cdot N' \cdot (L+1) + 4 \cdot M \cdot L^2)$, where L (odd) is the size of the spectral window used in the cyclic test. We note that in all the simulations of this chapter L is equal to 41) and N' represents the total number of samples used by the cyclostationarity detector.

Finally the complexity of the SM, is none other than the complexity of the OMP, multiplied by M, because the OMP is called M times before choosing between H_0 or H_1. So the complexity of the SM is given by $O(M \cdot n_{tot} \cdot \widetilde{N} \cdot l)$, with $\widetilde{N}$ the total number of atoms used in the dictionary to solve the inverse problem and l is the number of iterations of the OMP. So the expressions of the complexities of SM and SCMS are the same.

Figure 5.10 shows the probability of detection for a P_{fa} equal to 10 % for an SNR of 0 dB as function of the total number of samples for the three detectors (SM, SCMS and the cyclostationnarity detector) under the same conditions and using a filter at

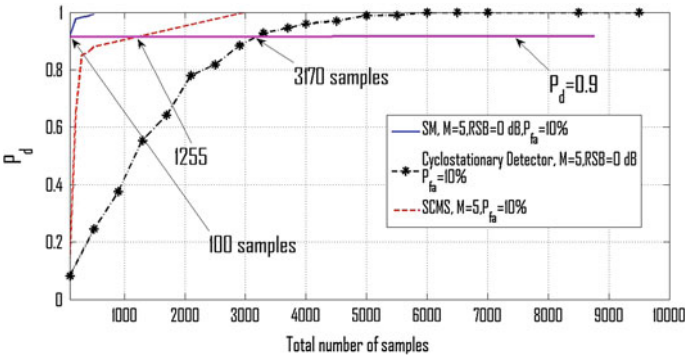

Fig. 5.10 The probability of detection P_d for a given false alarm set to 10 % for an $SNR = 0$ dB, as function of the number of samples (observation time) for the SM, SCMS and the cyclostationarity detector

Table 5.1 Comparison of the number of samples and complexities between the SM, the SCMS and the cyclostationary detector for the same (P_d, P_{fa}) performance

Detection method	SM	SCMS	Cyclostationary Detector
Complexity	$O(M \cdot n_{tot} \cdot l \cdot \widetilde{N})$	$O(2 \cdot M \cdot n_s \cdot S \cdot \widetilde{N})$	$O(M \cdot N' \cdot (L+1) + 4 \cdot M \cdot L^2)$
Needed samples to achieve: $(P_d, P_{fa}) = (0.9, 0.1)$	100	1255	3170
log (Number of operations)	13.55	16.08	13.45
a priori information	Blind method	Blind method	$\alpha_f = \frac{1}{T_s}$

the transmission. It is clear that the SM outperforms the two other methods and the SCMS outperforms the cyclostationarity detector. For example, for a probability of detection equal to 90 % the SM reaches a gain in the observation time equal to 12 compared to the SCMS, and a gain of 31 compared to the cyclostationnarity detector.

Table 5.1 shows a comparison between the complexity (number of operations) and the number of samples (observation time) corresponding to the SCMS the SM and the cyclostationary methods. For a probability of detection equal to 90 % the SM reaches a gain in the observation time equal to 12 compared to the SCMS, and a gain of 31 compared to the cyclostationnarity detector.

5.4 Conclusion

In this work, we proposed a new estimator which estimates the CAV of the received signal. This new estimator uses the compressed sensing to make the estimation. In fact it uses sparsity property of the cyclic autocorrelation vector in the cyclic frequencies domain. We defined two metrics to assess the quality of estimation of this estimator. The first metric calculates the mean square error between the estimated

cyclic autocorrelation vector and its corresponding theoretical value. The second metric calculates the mean square error of the estimated position of the cyclic frequency in the cyclic autocorrelation vector. These metrics are used in order to make a comparison between the unbiased estimator of the cyclic autocorrelation vector and our new estimator. In the different simulation cases, our new estimator showed better performance than the classical unbiased estimator. We note in addition that the proposed new estimator based on the compressed sensing, is a blind estimator and does not require any knowledge on the value of the cyclic frequency of the received signal in opposition to the classical unbiased estimator. In the future it would be interesting to calculate theoretical values of these new metrics for different estimators depending on the number of samples and the SNR. Also it would be interesting to calculate the analytical expression of the Cramer-Rao bound of our new estimator.

In the second part of this work we proposed two types of blind detectors that are based on our new estimator of the cyclic autocorrelation vector. The first detector, the SCMS takes two consecutive slots of the same size, assuming that these two slots belong to the same hypothesis. Then from these two slots, the new blind estimator estimates the two corresponding CAV. Then by comparing these two vectors, if they contain close cyclic frequencies it is stated that the band is busy, otherwise it is declared free because the noise does not contain cyclic frequencies. The second detector, the SM, in addition to the sparsity property, uses the property of symmetry of the cyclic autocorrelation vector. If the estimated vector exhibits symmetry around the axis $\alpha = 0$, the band is declared occupied. For a small number of samples, the performance of the SCMS, have exceeded those of the time second order cyclostationnary detector (which is not blind) proposed in [1] and in all the simulations conditions. Also for the same detection performance the SCMS needs 2.5 less times smaller observation time than that [1] with a slightly greater complexity. As for the SM, for a small number of samples, it was shown that it is not more complex than the SCMS. In addition, the MS outperforms the SCMS (thus also exceeds the detector [1]) and this in all the simulation conditions. The better performance of the MS compared to SCMS are due to the use of the symmetry property that is used as additional *a priori* information. We have seen that the MS can achieve a gain in the observation time equal to 30 compared to the cyclostationary detector of [1].

In this chapter, we focused on the evaluation of the observation time, simulations and analysis of the complexity of the proposed new blind methods. In future work, it would be useful to find explicit expressions for the detection thresholds depending on the desired probability of false alarm. In addition it is very important to implement these architectures and to test them on real signals to evaluate the use of these detectors and compare their total execution time (which is equal to the computation time added to the observation time),with the total execution time of the cyclostationary detector. Also it would be interesting to study these new detectors in a MIMO context using multiple antennas. The collaborative aspect using these new detectors should also be studied. Finally it would also be useful in the future to offer new detection algorithms based on the new estimator of this chapter.

Table 5.2 Attenuation and delay of the channel response h

Trajet	1	2	3	4	5	6
Attenuation in dB	0	−1	−9	−10	−15	−20
Delay in T_e	0	4	6	8	14	20

Appendix

In this Appendix we will describe in a little more detail by giving the coefficients of the channel used in the simulations of Sect. 2.

One way to represent the impulse response of a multipath channel, is by a discrete number of pulses as follows:

$$c(t, \tau) = \sum_{i=1}^{l} a_i(t)\delta(\tau - \tau_i) \tag{5.28}$$

where $a_i(t)$ is the attenuation at the time t of the traject i, and τ_i represents the delay of the traject i. For a time-invariant channel the response is:

$$c(\tau) = \sum_{i=1}^{l} a_i\delta(\tau - \tau_i) \tag{5.29}$$

with coefficients a_i invariant in time. In the simulations the used coefficients a_i are given in the Table 5.2.

From Table 5.2 we can conclude the impulse response of the channel using the z transform:

$$C(z) = 1 + 0.79 \cdot z^{-4} + 0.12 \cdot z^{-6} + 0.1 \cdot z^{-8} + 0.03 \cdot z^{-14} + 0.01 \cdot z^{-20} \tag{5.30}$$

References

1. Dandawate, A.V., Giannakis, G.B.: Statistical tests for presence of cyslostationarity. IEEE Trans. Inf. Theory **42**, 2355–2369 (1994)
2. Khalaf, Z., Nafkha, A., Palicot, J.: Blind Spectrum Detector for Cognitive Radio Using Compressed Sensing. IEEE GLOBECOM, Houston (2011)
3. Khalaf, Z., Nafkha, A., and Palicot, J.: Blind spectrum detector for cognitive radio using compressed sensing and symmetry property of the second order cyclic autocorrelation. In: 7th International Conference on Cognitive Radio Oriented Wireless Networks (CROWNCOM), Stockholm, Sweden, 18–20 June 2012
4. Axell, E., Leus, G., Larsson, E.G., Poor, H.V.: Spectrum sensing for cognitive radio: State-of-the-art and recent advances. IEEE Sig. Process. Mag. **29**, 101–116 (2012)

5. Mitola, J.: Cognitive radio: an integrated agent architecture for software defined radio. Ph.D. thesis, Royal Institute of Technology (2000)
6. Zeng, Y., et al.: A review on spectrum sensing for cognitive radio: challenges and solutions. EURASIP J. Adv. Sig. Process. **2010**, 1–15. Hindawi Publishing Corp., New York (2010)
7. Federal Communications Commission (FCC): Spectrum policy task force, ET Docket No. 02–135, Nov 2002
8. Palicot, J. (Surpvised by): De la radio logicielle à la radio intelligente, Collection Télécom, Lavoisier Librairie (2010)
9. Bic, J.C., Duponteil, D., Imbeaux, J.C.: Element de communications numériques. Dunod, Paris (1986)
10. Gardner, W.A.: Statistical Spectral Analysis: A Nonprobabilistic Theory. Prentice-Hall, Englewood Cliffs (1987)
11. Giannakis, G.B.: Cyclostationary signal analysis. In: Madisetti, V.K., Williams D. (eds.) Statistical Signal Processing Section of Digital Signal Processing Handbook. CRC Press, Boca Raton(1999) (Chap. 17)
12. Mallat, S., Zhang, Z.: Matching pursuit in a time-frequency dictionary. Trans. Sig. Process. **41**, 3397–3415 (1993)
13. Lu, L., Wu, H.A.: Novel robust detection algorithm for spectrum sensing. IEEE J. Sel. Areas Commun. **29**, 305–315 (2011)
14. Doostan, A., Owhadi, H.: A non-adapted sparse approximation of PDEs with stochastic inputs. J. Comput. Phys. **230**(8), 3015–3034 (2011)
15. Khalaf, Z.: Contributions à l'étude de détection des bandes libres dans le contexte de la radio intelligente. Ph.D. thesis, Supélec, Feb 2013
16. Khalaf, Z., Palicot, J.: Exploiting Sparse Property of the Cyclic Autocorrelation Function for Cyclostationary Process Compressed Sensing. WSR, Karlsruhe (2012)
17. Donoho, D.L., Huo, X.: Uncertainty principles and ideal atomic decomposition. IEEE Trans. Inf. Theory **47**(11), 2845–2862
18. Gribonval, R., Nielsen, M.: Sparse representations in unions of bases. IEEE Trans. I.T. **49**(12), 3320–3325 (2003)
19. Fuchs, J.J.: More on sparse representations in arbitrary bases. IEEE Trans. I.T. **50**(6), 1341–1344 (2004). In: 13th IFAC SYSID, 1357–1362, Rotterdam, 2003
20. Fuchs, J.J.: Sparse representations and realization theory. Proceedings of the 17th International Symposium on Mathematical Theory of Networks and Systems, Kyoto, Japan, July 2006
21. Fuchs, J.J.: Identification of real sinusoids in noise, the Global Matched Filter approach. In: 14th Ifac-Ifors Symposium on Identification and System Parameter Estimation, pp. 1127–1132, Saint-Malo, France, July 2009
22. Davis, G., Mallat, S., Avellaneda, M.: Greedy adaptive approximation. J. Constr. Approx. **13**, 57–98 (1997)
23. Mallat, S.: A Wavelet Tour of Signal Processing: The Sparse Way. Academic Press, San Diego (2009)
24. William, A.: Gardner, Statistical Spectral Analysis: A Non-Probabilistic Theory. Prentice Hall, Englewood Cliffs (1988)
25. Donoho, D.L.: For most large underdetermined systems of linear equations the minimal l1-norm solution is also the sparsest solution. Commun. Pure Appl. Math. (Wiley Online Library) **59**(7), 907–934 (July 2006)
26. J-L Lacoume, Amblard, P.-O., Comon, P.: Statistiques d'ordres supérieurs pour le Traitement du Signal. MASSON, Paris (1997)
27. Fuchs, J.J.: Sparse representations and realization theory. In: Proceedings of the 17th International Symposium on Mathematical Theory of Networks and Systems, Kyoto, Japan, July 2006

Chapter 6
Intersystem Coexistence and Cooperation Through Control Channels

Andreas Georgakopoulos, Dimitrios Karvounas, Vera Stavroulaki, Kostas Tsagkaris and Panagiotis Demestichas

Abstract The emerging wireless world is expected to be characterized from the demand for new, diversified applications/services, the expanded use of wireless and the need for increased efficiency in resource provisioning and utilization. Therefore, the resource usage in an opportunistic manner would provide a solution towards this direction. Opportunistic networks (ONs) and Cognitive Management Systems (CMSs) for cellular extensions are one of the emerging communication paradigms in wireless mobile communications. For the cooperation of CMSs, specific mechanisms need to be defined in order to increase the accuracy of obtained knowledge on the context of the operational environment. Also, a cooperation mechanism is required for efficient coordination between the infrastructure and the devices in the scope of an ON. Therefore, Control Channels (CCs) are required for the exchange of information and the coordination between CMSs. Consequently, this work focuses on the definition of the needed interfaces that are introduced in order to enable communication between the cognitive management systems and also to the related groups of information that is needed to be conveyed between these systems in order to ensure proper interaction. Finally, an indication of the amount of information conveyed through CCs is provided in order to be able to assess the impact to the network of control-related transmitted information.

A. Georgakopoulos (✉) · D. Karvounas · V. Stavroulaki · K. Tsagkaris · P. Demestichas
University of Piraeus, Piraeus, Greece
e-mail: andgeorg@unipi.gr

D. Karvounas
e-mail: dkarvoyn@unipi.gr

V. Stavroulaki
e-mail: veras@unipi.gr

K. Tsagkaris
e-mail: ktsagk@unipi.gr

P. Demestichas
e-mail: pdemest@unipi.gr

M.-G. Di Benedetto and F. Bader (eds.), *Cognitive Communication and Cooperative HetNet Coexistence*, Signals and Communication Technology,
DOI: 10.1007/978-3-319-01402-9_6,

6.1 Introduction

The emerging wireless world is expected to be characterized from the demand for new, diversified applications/services, the expanded use of wireless and the need for increased efficiency in resource provisioning and utilization. Therefore, the resource usage in an opportunistic manner would provide a solution towards this direction. Opportunistic networks (ONs) and cognitive management systems (CMSs) for cellular extensions are one of the emerging communication paradigms in wireless mobile communications. Operator-governed ONs use the basic concepts of opportunistic networking and extend them by proposing coordination mechanisms which cooperate with the infrastructure. ONs enable pervasive communications in an environment where disconnection and reconnection may frequently occur and link performance and availability are highly dynamic and they make use of available network resources such as spectrum, nearby nodes etc. in an opportunistic manner. Operator-governance in such cases is realized through the use of CMSs which acquire operators policies and proceed with the establishment of an operator-governed ON.

Control Channels (CCs) have been identified as a key feature required for supporting CMSs in their operation, through the provision of the information and knowledge [1–4]. It should be noted that, by processing the information acquired through CCs, there can be more elaborate knowledge that can be generated through machine learning, and, therefore, there can be efficient intersystem coexistence and cooperation and a reactive or proactive response to situations. Moreover, it can be seen that the CC concept comprises information on all the layers of the protocol stack, e.g., ranging from spectrum sensing and spectrum awareness to various application and user requirements.

Investigation of ONs in literature is not a new topic. Specifically authors in [5] provide a conceptual point of view of the opportunistic networking paradigm by focusing on Delay-Tolerant Networks (DTNs). Also, the idea of an expanding network by dynamically adding new nodes has been examined by authors in [6]. The coexistence of ONs with network infrastructure is empirically analyzed in [12], where it is concluded that after a certain point, the benefits of additional infrastructure deployments are minor and the utility of the system remains stable. On the other hand, the OneFIT project [14, 15] has proceeded to the definition of the operator-governed ONs by extensively studying and evaluating various facets of ONs under specific scenarios and operating conditions. Finally, the possibility of extending ad-hoc networks with the support of infrastructure, (i.e., hybrid networks according to [13]) has been investigated as a way of improving the connectivity in large scale ad-hoc networks, while authors in [7] provide solutions with respect to data storage, carrying and forwarding in an ON.

In addition, authors in [8] presented a technique based on a fittingness factor for the selection of the spectrum to be assigned to a set of radio links that belong to an ON, among pairs of terminals and infrastructure nodes is considered. In [9], a modular decision flow for the selection of Radio Access Technology (RAT), frequency, and bandwidth for operator-governed ONs, was presented, in order to ensure fair

operation for the whole ON and adequate quality levels for each user. Moreover, authors in [10] studied the probability of finding a supporting device that can relay traffic to the infrastructure through ON with respect to the range of the air interface and to the density of the supporting devices. Furthermore, authors in [11] presented a decision making system that can select the most suitable techniques for obtaining spectrum availability information in ONs from cognitive control channels, databases, and spectrum sensing techniques.

To this extent, this work supports the idea that CMSs are a vital component of operator-governed ONs and evaluates the impact to the network of the conveyed information through control channels which are used for the cooperation of the CMSs.

The rest of the chapter is structured as follows. Section 6.2 deals with basic scenarios while Sect. 6.3 analyzes the main functional blocks, including an overview of CMSs. Section 6.4 provides essential information related to the cooperation between CMSs through the CCs, while Sect. 6.5 includes an indicative evaluation of the conveyed information through CCs. Finally, the chapter concludes at Sect. 6.6.

6.2 Scenarios

In order to be able to show the benefits of intersystem coexistence and cooperation through CCs, we have to define a few scenarios that make use of the CCs as enablers for conveying information so as to proceed to ON establishment [17]. One of the scenarios deals with the opportunistic coverage extension which offers extension of the infrastructure coverage of a macro Base Station (BS). In this scenario, it shall be assumed that an area is not well-covered, so an operator-governed ON is needed to be established upon operator's request in order to provide connectivity to non-covered places. For the realization of the scenario, a device which is out of the infrastructure's coverage will search for neighboring, ON-enabled devices which have direct access to a BS and are willing to help, in order to connect to it. This device will act as an intermediate node between the out-of-infrastructure coverage device and the infrastructure (BS). Eventually, the benefits from such an approach would affect various stakeholders. For example, users who cannot access directly the infrastructure, can now be served via intermediate users under operator governance of the created ON. On the other hand, operators shall experience increased revenue from the fact that extra users can now be served through their infrastructure (without having to establish permanent, new infrastructure elements in order to cover small portions of non-covered areas).

Another situation would involve a device which experiences low quality of communication (e.g., decreased bitrates, increased communication delay) due to the congestion of infrastructure elements (BSs). Therefore, the redirection of traffic through an operator-governed ON that avoids the congested element would be a feasible solution. This is the opportunistic capacity extension scenario which enables devices to maintain the required level of communication by redirecting ON-enabled

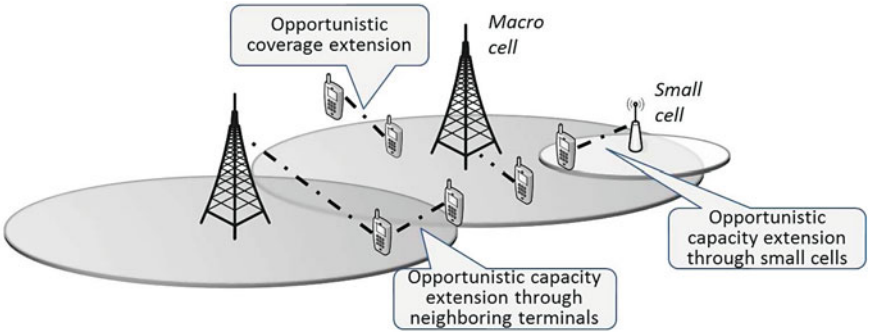

Fig. 6.1 Envisaged scenarios

devices to neighboring, non-congested BSs. Alternatively, users can be assigned to available small cells in the problematic area.

From the opportunistic capacity extension, operators and end-users would actually benefit from the fact that users can now be served with better quality of communication, since excessive amount of users in a BS is redirected to neighboring non-congested BSs. Therefore, previously congested BS is now relieved and even the users who remain attached to this BS are benefited from the congestion resolution. Additionally, reduction of load and energy consumption will be observed to the infrastructure elements.

Figure 6.1 illustrates situations which may occur in a cellular network, without the use of operator-governed ONs. Consequently, there are some terminals with ON capabilities which are out of infrastructure coverage and other terminals which are in coverage and are willing to help. All these challenges are tackled with the formation of operator-governed ONs. To that respect, terminals with ON capabilities have the ability to connect to each other and form an ON in order to gain access to the infrastructure. Also, users located to the congested BS can exploit the opportunity of nearby terminals or small cells in order to redirect traffic to alternate BSs with better quality of communication (e.g., higher bitrate etc.). In order to make all these solutions possible, novel functional entities are proposed in the following sections which enable the creation, monitoring, management and termination of operator-governed ONs.

6.3 Functional Architecture Overview and Cognitive Management Systems

In order to achieve operator-governance of ONs, decisions and policies from the operator are utilized. Specific management entities need to be introduced in order to be able to obtain this kind of input and proceed to decision making and enforcement.

To that respect, the management and control functionalities for ONs shall be an addition to existing functionalities in operators' networks.

Therefore, the proposed Functional Architecture (FA) is an extension of an existing architecture as defined in ETSI [1, 19] and is about "Functional Architecture for the Management and Control of Reconfigurable Radio Systems". New building blocks (which are also defined as CMSs) need to be added to the previously defined FA so as to facilitate the cognitive management and control of operator-governed ONs as proposed in [16, 18]. These are:

- The Cognitive Management System for the Coordination of the Infrastructure (CSCI) which is responsible for the detection of situations where an ON would be useful (prior to the formation of the ON);
- The Cognitive System for the Management of the Opportunistic Network (CMON) which is responsible for the creation, maintenance and termination of a given ON based on the context and policy information provided by the CSCI.

CMSs such as CSCI and CMON, capitalize on management and knowledge mechanisms as envisaged in cognitive radio principles [20]. To this extent, as already defined in [16, 18] the CSCI is responsible for the ON suitability determination phase, which is executed before the actual ON creation and deals with issues like "is it possible for an ON to be established right now, under the current operating conditions, current neighboring nodes etc.?" If the answer to this is positive, then the decision is forwarded to the CMON which will deal with the creation, maintenance and termination of the ON.

Apart from the CSCI and CMON functional blocks, further, legacy functionalities include the following [1, 19]:

- Dynamic Spectrum Management (DSM);
- Dynamic, Self-Organizing Network Planning and Management (DSONPM);
- Joint Radio Resources Management (JRRM);
- Configuration Control Module (CCM).

Figure 6.2 provides an overview of the aforementioned main functional blocks. CMSs are realized through the introduction of CSCI and CMON to the functional architecture. For the communication between the functional entities specific interfaces have been proposed [16, 18]. These interfaces include the CI interface which stands for the Coordination with the Infrastructure and connects different CSCI instances; the OM interface which stands for the Opportunistic Management and connects different CMON instances; the CS interface which enables the connection of the CSCI/CMON with the DSM; the CD interface which deals with the communication among the DSONPM and CSCI/CMON and can be used by the CSCI/CMON to retrieve information on the configuration of the operators network; the OJ interface which is located among the JRRM and the CSCI/CMON; and the OC interface which is located among the CCM and the CSCI/CMON.

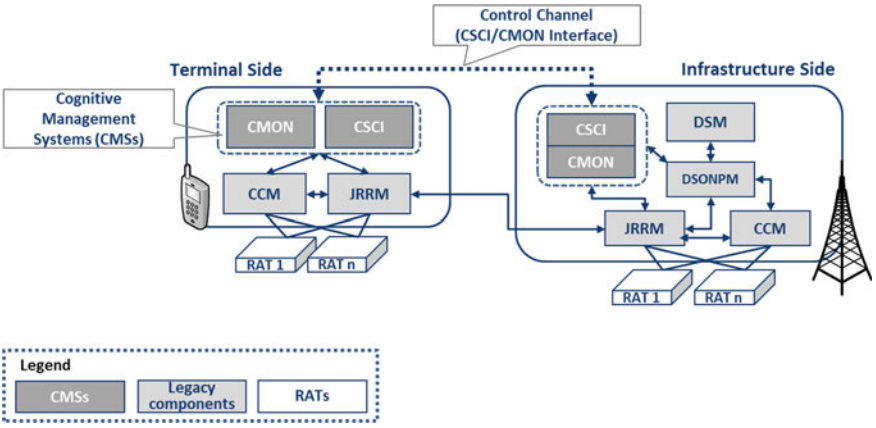

Fig. 6.2 Main functional blocks and interfaces

6.4 Cooperation Between Cognitive Management Systems Through Control Channels

For the cooperation of CSCIs and CMONs, specific mechanisms need to be defined in order to increase the accuracy of obtained knowledge on the context of the operational environment. Also, a cooperation mechanism is required for efficient coordination between the infrastructure and the devices in the scope of an ON. Therefore, CCs are required for the exchange of information and the coordination between CMSs.

Such control channels can be based on the exploitation and evolution of two concepts: the cognitive pilot channel (CPC) and the cognitive control radio (CCR) (as defined by ETSI). The CPC can be seen as an enabler for providing information from the network to the terminals and vice-versa (e.g., available RATs, spectrum, bands etc.) [14].

Specific data structures have been developed so as to formulate the structure of the information that needs to be exchanged through CCs for supporting the various scenarios described before by mainly containing information on the context of operation, on the profiles of the involved nodes, on the policies to be obeyed, on the decisions made, as well as on knowledge derived for all the above [21].

Profiles: Profiles are divided into terminal, base station and user profiles. Terminal and BS profiles include "General capabilities" like ID of the node, type of the node, IP address of the node etc.; "Communication capabilities" like communication interface capabilities, available/supported RATs, available/supported spectrum bands, sensing techniques etc.; "Computing capabilities" like CPU, memory data etc.; "Storage capabilities" like available size of caching/storage etc.; "Energy capabilities" like total capacity and characteristics of available batteries etc. and finally "ON capabilities", e.g., maximum number of allowed/supported nodes to be used in an ON etc. In addition, "User Profiles" include information related to behavior aspects of users

(e.g., how many times do they use an application etc.), the subscribed applications and the associated user class.

Context: Accordingly context information is divided into terminal and base station context. Terminal and BS context include "General status" such as ON node's current location, timestamp of collected context, mobility characteristics (applicable only to moving terminals), etc.; "Communication status" such as current RAT utilized for the communication, demand and quality of communication offered per application (i.e., requested and offered bitrate, delay etc.), user class etc.; "Computing status" such as the current utilization of the CPU/ memory etc.; "Storage status" such as current utilization of storage/caching space; "Energy status" such as the current level of the battery and finally "ON status" such as lifetime of the current ON, amount of data transmitted through the current node which participates in an ON etc.

Information on Decisions: Information on decisions is divided into ON decisions, infrastructure decisions and terminal decisions. Specifically, ON decisions may include decisions related to path selection (ON nodes and links utilized in an ON) as well as spectrum selection (central frequency, bandwidth, sensing technique utilized, transmission constraints etc.). Additionally, infrastructure and terminal decisions cover aspects on communication (e.g., RAT to be operated etc.), storage (e.g., amount of storage space to be utilized etc.) and computing (e.g., percentage of CPU or memory to be used etc.).

Knowledge: Knowledge is related to acquired context and decisions made. To that respect, through registered values obtained during the time of operation, knowledge mechanisms can support decisions by assessing context, previous decision made and the performance achieved through the decision. In the case that the performance was satisfactory and similar context is observed then the same decision can be enforced. Apparently, this would lead to faster decision making for similar contexts of operation.

Policies: Policies represent rules of the network operator that can be imposed for certain reasons. To that respect, network operator policies shall include communication-related policies (e.g., what are the allowed interfaces etc.); computing-related policies (e.g., what is the allowed CPU/memory usage etc.); storage-related policies (e.g., what is the allowed storage size) and energy-related policies (e.g., what is the allowed energy consumption).

Figure 6.3 that follows provides an overall representation of the information conveyed.

6.5 Evaluation of Conveyed Information Through Control Channels

The following section provides evaluation of the information conveyed through CCs. The analysis is conducted for the CSCI/CMON interfaces between different nodes. Specific test cases are considered for the evaluation as provided in Table 6.1. Testcases take into account the following attributes:

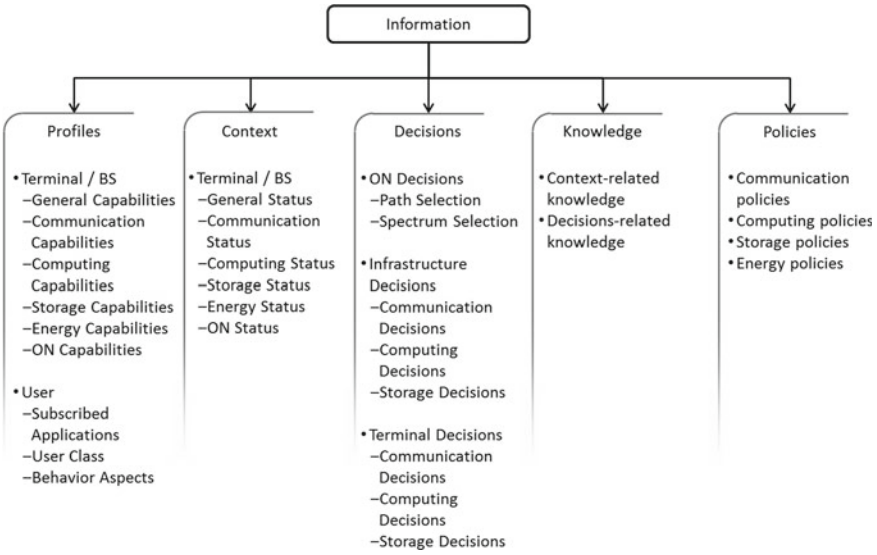

Fig. 6.3 Information conveyed through Control Channels

Table 6.1 Considered test cases

Attribute	Value for case 1	Value for case 2	Value for case 3
Total BSs	7	7	7
Non-congested BSs	6	6	6
Congested BSs	1	1	1
Terminals in non-congested BSs	15	20	25
Terminals in congested BS	40	80	160
Terminals switching to ONs	12	24	48
No. of created ONs	12	24	48
No. of links per ON	2	2	2
No. of interfaces in BSs	1	1	1
No. of interfaces in terminals	2	2	2
No. of RATs (per interface)	1	1	1

- Number of BSs;
- Non-congested BSs;
- Congested BSs;
- Terminals in non-congested BSs;
- Terminals in congested BS;
- Terminals switching to ONs;
- Number of created ONs;
- Number of links per ON;
- Number of interfaces in BSs;
- Number of interfaces in terminals;

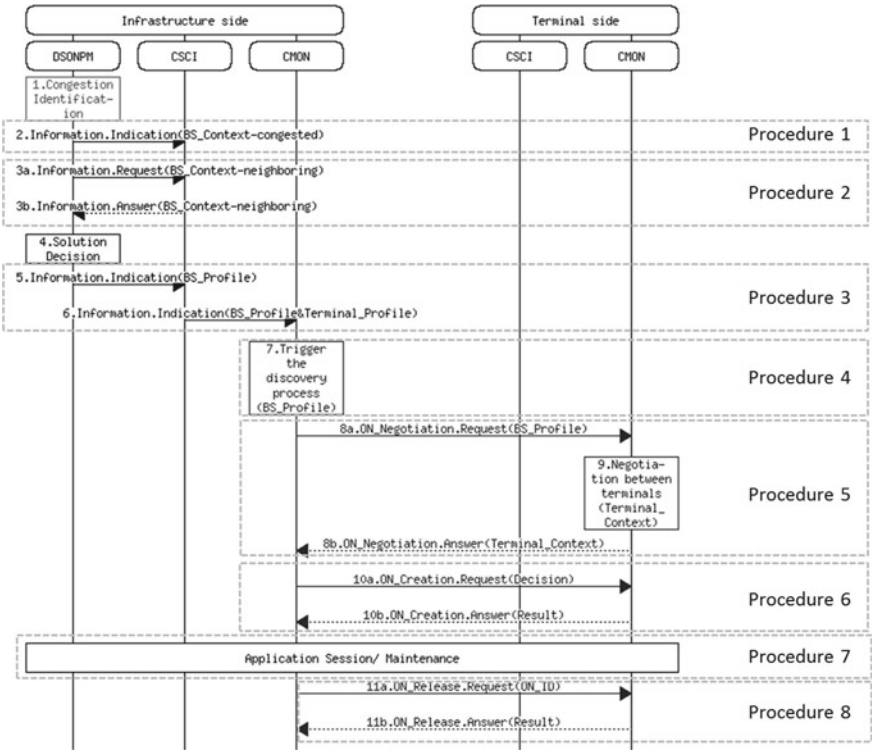

Fig. 6.4 Mapping of messages and related information to procedures

- Number of RATs (per interface).

Evaluation is taking place by considering a scenario related to capacity extension. Also, for the exchange of messages an agent-based, Java prototype has been developed by using the Java Agent Development Platform (JADE) [22]. JADE acts as a middleware on top of a customized version of the Opportunistic Network Environment (ONE) [23] simulator, so as to realize transmission of messages among functional entities. Customization of the ONE has provided the flexibility to include also communication with infrastructure (simulated base stations). Therefore, a specific message sequence chart (MSC) is provided in order to show which kind of messages are being considered. Also messages are grouped onto specific procedures. For the procedures 1–6 and 8, triggered-based events are considered (i.e., messages are being exchanged only upon request/trigger and not periodically). On the contrary, for procedure 7 a periodic exchange of messages is also considered (i.e., messages can be exchanged on periodic basis e.g., every 1 s etc.). The considered procedures are illustrated in Fig. 6.4.

The chart in Fig. 6.5 shows the impact of each procedure 1–6 and 8 which involve trigger-based messages. Periodic messages are not exchanged during these procedures, because the procedures are executed only when they are instructed from the

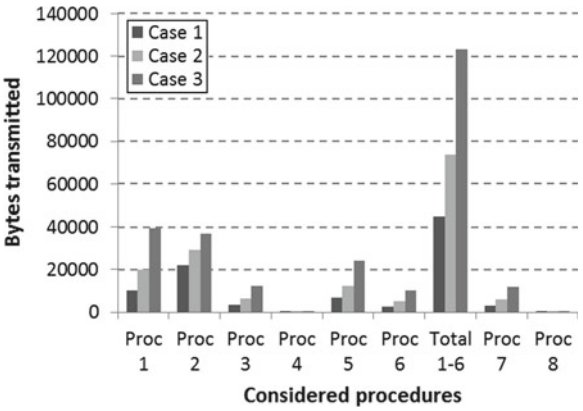

Fig. 6.5 Load associated with specific procedures

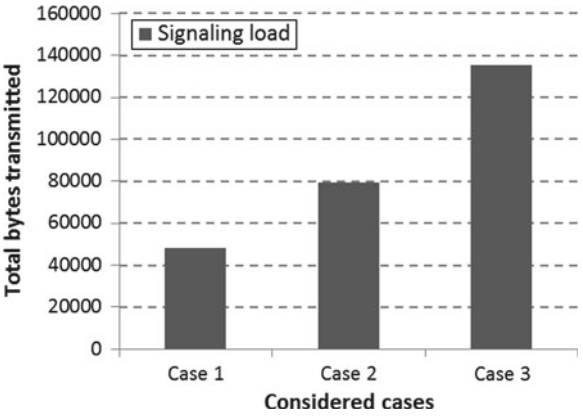

Fig. 6.6 Total signaling load for each case

CMSs. Each procedure is evaluated separately for each one of the testcases. Also, the chart in Fig. 6.6 illustrates the total signaling load for each testcase.

The chart in Fig. 6.7 provides an estimation of the periodic load which is considered to take place during procedure 7. Specifically, during this procedure, terminal context is considered to be sent periodically (every 1, 5 or 30 s) in order to know the status of the nodes involved in the ON (e.g., their current location, current energy level, current links etc.). It is observed that as long as the intervals of transmission are more closely defined, the load per second (bytes/s) is higher (for each case considered).

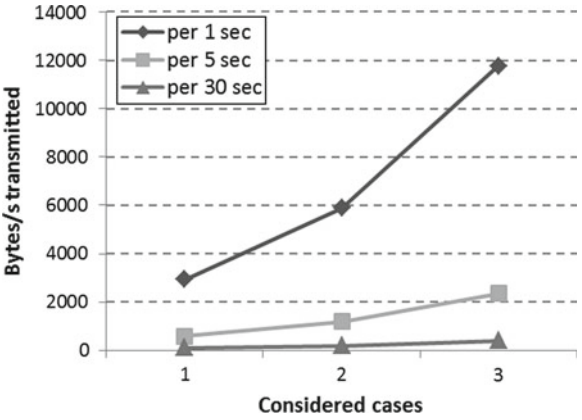

Fig. 6.7 Calculation of periodically exchanged load associated with procedure 7

6.6 Conclusions

The emerging wireless world introduces new challenges and requirements that need to be tackled in order to allow key players to be innovative and stay competitive in the wireless telecommunications market. As a result, solutions comprising opportunistic usage of available resources for addressing the new challenges are proposed in this work. The design, development and validation of global and cost-efficient solutions to address the above appear to be of major importance. Novel functional architectures of communication systems become necessary. The definition of new functional blocks and interfaces is presented in this work. New functional blocks comprise CMSs which are responsible for the creation, maintenance and termination of ONs. The cooperation of CMSs through CCs provides solutions to problems such as coverage and capacity extension through the creation of ONs. Finally, the impact to the network of the conveyed information through control channels is evaluated in order to be able to use it as an indication. It is concluded that control messages related to the operation of ONs are rather small in terms of size.

Acknowledgments This work was performed in the framework of the European-Union funded project OneFIT (www.ict-onefit.eu). The project was supported by the European Community's Seventh Framework Program (FP7). The views expressed in this document do not necessarily represent the views of the complete consortium. The Community is not liable for any use that may be made of the information contained herein. Also, this work is benefited from discussions in COST Action IC0902 "Cognitive Radio and Networking for Cooperative Coexistence of Heterogeneous Wireless Networks".

References

1. ETSI TR 102 682 V1.1.1: Reconfigurable Radio Systems (RRS); Functional Architecture (FA) for the Management and Control of Reconfigurable Radio Systems (2009)
2. ETSI TR 102.683 V1.1.1: Reconfigurable Radio Systems (RRS); Cognitive Pilot Channel (CPC) (2009)
3. ETSI TR 102 684 V1.1.1: Reconfigurable Radio Systems (RRS); Feasibility Study on Control Channels for Cognitive Radio Systems (2012)
4. Stavroulaki, V., Tsagkaris, K., Demestichas, P., Gebert, J., Mueck, M., Schmidt, A., Ferrus, R., Sallent, O., Filo, M., Mouton, C., Rakotoharison, L.: Cognitive control channels: from concept to identification of implementation options. IEEE Commun. Mag. **50**(7), 96–108 (2012)
5. Pelusi, L., Passarella, A., Conti, M.: Opportunistic networking: data forwarding in disconnected mobile ad-hoc networks. IEEE Commun. Mag. **44**(11), 134–141 (2006)
6. Lilien, L., Gupta, A., Yang, Z.: Opportunistic networks for emergency applications and their standard implementation framework. In: Proceedings of IEEE International Performance, Computing and Communications Conference (IPCCC 2007), Louisiana (2007)
7. Huang, C.-M., Lan, K.-C., Tsai, C.-Z.: A survey of opportunistic networks. In: Proceedings of International Conference on Advanced Information Networking and Applications-Workshops (AINAW 2008), Okinawa (2008)
8. Bouali, F., Sallent, O., Prez-Romero, J., Agusti, R.: Exploiting knowledge management for supporting spectrum selection in cognitive radio networks. In: Proceedings of 7th International Conference on Cognitive Radio Oriented Wireless Networks (CrownCom) 2012, Stockholm (2012)
9. Sarvanko, H., Mustonen, M., Matinmikko, M., Hoyhtya, M., Del Ser, J.: Spectrum band and RAT selection for infrastructure governed opportunistic networks. In: Proceedings of 17th International Workshop on Computer-Aided Modeling Analysis and Design of Communication Links and Networks (CAMAD) 2012, Barcelona (2012)
10. Gebert, J., Fuchs, R.: Probabilities for opportunistic networking in different scenarios. In: Proceedings of 21st Future Network and Mobile Summit 2012, Berlin (2012)
11. Matinmikko, M., Mustonen, M., Rauma, T., Del Ser, J.: Decision-making system for obtaining spectrum availability information in opportunistic networks. In: Proceedings of 4th International Conference on Cognitive Radio and Advanced Spectrum Management (CogArt) 2011. Barcelona (2011)
12. Hui, P., Lindgren, A., Crowcroft, J.: Empirical evaluation of hybrid opportunistic networks. In: Proceedings of Communication Systems and Networks and Workshops, (COMSNETS 2009), Bangalore (2009)
13. Dousse, O., Thiran, P., Hasler, M.: Connectivity in ad-hoc and hybrid networks. In: Proceedings of IEEE International Conference on Computer Communications, (INFOCOM 2002), New York (2002)
14. Stavroulaki, V., Tsagkaris, K., Logothetis, M., Georgakopoulos, A., Demestichas, P., Gebert, J., Filo, M.: Opportunistic networks: an approach for exploiting cognitive radio networking technologies in the future internet. IEEE Veh. Technol. Mag. **7**(2), 52–59 (2011)
15. FP7/ICT project OneFIT (Opportunistic networks and Cognitive Management Systems for Efficient Application Provision in the Future InterneT, ICT-2009-257385), Jul 2010–Dec 2012. http://www.ict-onefit.eu. Accessed April 2013
16. Gebert, J., Georgakopoulos, A., Karvounas, D., Stavroulaki, V., Demestichas P.: Management of opportunistic networks through cognitive functionalities. In: Proceedings of 9th International Conference on Wireless On-Demand Network Systems and Services (WONS), Courmayeur, pp. 113–118 (2012)
17. OneFIT project Deliverable 2.1: Scenarios, technical challenges and system requirements. www.ict-onefit.eu. Accessed Oct 2010
18. OneFIT project Deliverable 2.2.2/6.4: Functional and system architecture-version 2.0. www.ict-onefit.eu. Accessed Dec 2012

19. Mueck, M., Piipponen, A., Kalliojarvi, K., Dimitrakopoulos, G., Tsagkaris, K., Demestichas, P., Casadevall, F., Perez-Romero, J., Sallent, O., Baldini, G., Filin, S., Harada, H., Debbah, M., Haustein, T., Gebert, J., Deschamps, B., Bender, P., Street, M., Kandeepan, S., Lota, J.: Hayar. A: ETSI reconfigurable radio systems: status and future directions on software defined radio and cognitive radio standards. IEEE Commun. Mag. **48**(9), 78–86 (2010)
20. Mitola III, J., Maguire Jr. G.-Q.: Cognitive radio: making software radios more personal. IEEE Pers. Commun. **6**(4), 13–18 (1999)
21. Georgakopoulos, A., Demestichas, P., Stavroulaki, V., Tsagkaris, K., Bantouna, A.: Mechanisms for information and knowledge sharing in wireless communication systems. In: Proceedings of International Symposium on Wireless Communication Systems (ISWCS), Paris, pp. 411–415 (2012)
22. Java Agent DEvelopment Platform (JADE). http://jade.tilab.com
23. Keranen, A., Ott, J., Teemu, K.: The ONE simulator for DTN protocol evaluation. In: Proceedings of 2nd International Conference on Simulation Tools and Techniques (SIMUTools 2009), Rome (2009)

Chapter 7
Cooperative Spectrum Sensing with Censoring of Cognitive Radios in Fading Channel Under Majority Logic Fusion

Srinivas Nallagonda, Sanjay Dhar Roy, Sumit Kundu, Gianluigi Ferrari and Riccardo Raheli

Abstract In a cooperative spectrum sensing (CSS) scheme, the detection of the presence of activity of a primary user (PU) is improved by the fact that several cognitive radio (CR) users send, through reporting channels (R-channels), their sensed information on the activity of this PU to a common base station (BS). The benefits are particularly relevant in scenarios where the sensing channels (S-channels) towards the PU of interest of CR users are affected by severe fading or shadowing. However, in a CSS scheme with R channels affected by fading or shadowing as well, there may be erroneous reception, at the BS, of decisions from CR users: this can be counter-acted by using censoring of CR users. In this chapter, we discuss the performance of CSS with censoring of CR users based on their R-channels' statuses. Two schemes of censoring are considered: (i) rank-based censoring, where a pre-defined number of CR users, associated with the best R-channels, are selected; and (ii) threshold-based censoring, where CR users, whose R-channel fading coefficients exceed a pre-determined threshold, are selected. The performance of both censoring schemes is evaluated considering two different R-channel fading conditions: (i) Rayleigh fading and (ii) Nakagami-*m* fading. In both cases, majority logic fusion is considered at the BS (also denoted re-interpreted as fusion center, FC). The impact of various network parameters—such as censoring threshold, number of CR

S. Nallagonda (✉) · S. D. Roy · S. Kundu
ECE Department, NIT, Durgapur, West Bengal, India
e-mail: srinivas.nallagonda@gmail.com

S. D. Roy
e-mail: s_dharroy@yahoo.com

S. Kundu
e-mail: sumit.kundu@ece.nitdgp.ac.in

G. Ferrari · R. Raheli
Deparment of Information Engineering, University of Parma, Parma, Italy
e-mail: gianluigi.ferrari@unipr.it

R. Raheli
e-mail: raheli@unipr.it

M.-G. Di Benedetto and F. Bader (eds.), *Cognitive Communication and Cooperative HetNet Coexistence*, Signals and Communication Technology,
DOI: 10.1007/978-3-319-01402-9_7,

users, average S-and R-channels' SNRs, channel estimation (CE) quality, and fading severity—on the performance of the considered CSS schemes will be evaluated in terms of missed detection and total error probabilities.

7.1 Introduction

Cognitive radio (CR[1]) has been proposed [1] as a promising technique to solve the conflicts between spectrum scarcity and spectrum under utilization. CR systems allow CR users to share the spectrum with primary users (PUs) either opportunistically or without creating any intolerable interference to PU. Spectrum sensing is an important feature of CR technology since it is necessary to detect the presence of PUs accurately and quickly in order to find availability of unused spectrum, i.e., the spectrum holes. Accurate sensing of spectrum holes is a hard task because of the time-varying nature of wireless channels [2], including fading and shadowing. Due to severe multipath fading in sensing channel (S-channel) between a PU and a CR user, the CR user may fail to detect the presence of the PU. The detection/sensing performance can be improved, by limiting the negative impact of fading, if different CR users are allowed to cooperate by sharing their detected information on the activity status of PUs: this is the essence of cooperative spectrum sensing (CSS). Therefore, CSS improves the detection performance when all CR users sense the PU individually and send their sensing information in the form of 1-bit binary decisions (1 or 0) via ideal (noiseless) reporting channels (R-channels) to a fusion centre (FC)—the FC corresponds to the base station (BS). In CSS schemes, the local decisions on PUs's activity status sent by several CR users are combined at FC to obtain a global decision. In general, the sensing information reported to the FC by several CR users can be combined in two different ways: through (i) soft or (ii) hard combining. According to a soft combining approach, CR users transmit the entire local sensing samples or the complete local test statistics which are combined using any one of possible diversity combining technique such as likelihood ratio test (LRT), maximal ratio combining (MRC), and equal gain combining (EGC) [3–5]. In [3] the authors consider soft information combining of the signals received via multiple antennas of a single CR. In [4], the LRT fusion is discussed in case of wireless sensor networks. In [6], an optimal soft combination scheme based on neyman-pearson (NP) criterion is proposed to combine the weighted local observations. The proposed scheme reduces to EGC at high SNR and reduces to MRC at low SNR. In the presence of hard combining, CR users make a local decision (hard decision on the PU activity status) and transmit the one bit decision for hard combining. A hard decision combining fusion rule—such as OR-logic, AND-logic, and majority-logic—is implemented at FC to make the final decision on the presence or absence of a PU [7–9].

[1] Note that with the generic term CR we also refer to a secondary (cognitive) user (SU). The context eliminates any ambiguity.

In many wireless applications, it is of great interest to check the presence and availability of an active communication link when the signal is unknown. In such scenarios, one appropriate choice consists in using an energy detector (ED) which measures the energy in the received waveform over an observation time window [10, 11]. The existing literature energy detector-based on single CR user [11, 12] and cooperative CR users [13–16] spectrum sensing, typically assumes popular fading models such as Rayleigh and Nakagami-*m* (*m* being the fading severity parameter). In these cases, R-channels are assumed to be ideal and S-channels are considered as Rayleigh and Nakagami-*m* fading channels. However, in many practical situations R-channels may not be noiseless (ideal) channels. Though most works on spectrum sensing assume noiseless R-channels [7–9, 12–16], the presence of fading in R-channels is likely to affect the decisions sent by CR users where the FC is far from CR users. If the R-channel connecting a CR user to the FC is heavily faded, the decision received at the FC is likely to be erroneous with respect to that transmitted by the CR user. If this is the case, it is better to stop transmitting decisions from such CR user and, thus, the use of censoring is expedient. The CR users whose R-channels are estimated as reliable by the FC are censored, i.e., they are allowed to transmit. The CR users which are not participating in improving the detection performance may be stopped, so that the system complexity can be reduced and the detection performance can be improved. This will further reduce the energy consumption for an energy-constrained network. Therefore, censoring of CR users is necessary to improve the performance of CSS. The R-channels are considered as noisy and Rayleigh faded in [17, 18], in the context of a sensor network where sensors report their decisions to a FC. Censoring of sensors, as proposed in [19, 20], and channel-aware censoring of sensors, as discussed in [21], can be well applied in the context of energy detection based CSS.

In our present discussion, we consider both R-channel and S-channel to be (i) Rayleigh faded and (ii) Nakagami-*m* faded. Similar fading scenario is considered in S-channel and R-channel i.e., both S-channel and R-channel as Rayleigh faded or Nakagami-*m* faded. Though all the CR users detect PUs using energy detectors, only those CR users censored based on quality of R-channels are allowed to transmit. The censoring decision is taken by FC based on estimation of R-channel. In [23], the performance of CSS systems with censoring of CR users under both majority-logic fusion and maximal ratio combining (MRC) fusion has been evaluated only in Rayleigh faded environments, considering CR users' censoring on the basis of the qualities of their R-channels. Using minimum mean square estimation (MMSE)-based estimation of the R-channels, the FC selects the subset of CR users among all the available ones (say K out of N) which have the highest channel coefficients, i.e., the CR users associated with best estimated channel coefficients are selected. However, in an alternative censoring scheme, based on channel thresholding, is considered and analyzed in [24] in the context of distributed detection in a (non-cognitive) sensor network where a number of sensors observe a common binary phenomenon. In [25], the performance of CSS schemes with channel thresholding-based censoring of CR users with Rayleigh fading and majority-logic fusion at the FC is evaluated. The investigation of majority-logic fusion schemes where both

S- and R-channels are Nakagami-m faded is an interesting research extension. The Nakagami-m distribution provides flexibility in describing the fading severity of the channel and encompasses special cases such as Rayleigh fading (for $m = 1$) [5].

In the current chapter, we consider the same system model of [23, 25] and evaluate the performance of CSS with censoring of CR users based on quality of R-channels. More precisely, we analyze the performance of CSS schemes with censored CR users in Nakagami-m faded environments (with special case given by Rayleigh faded environment), considering a network of N CR users. Each CR user makes local observation on the activity of the PU using energy detectors. We consider two schemes on channel quality-based censoring. The first scheme consists of rank-based censoring: using MMSE-based estimation of the R-channels, the FC selects the subset of CR users among all the available ones (say K out of N) which have the highest channel coefficients, i.e., the CR users associated with best estimated channel coefficients are selected. The second censoring scheme is threshold-based: a CR user is selected to transmit its decision if the estimated R-channel fading coefficient exceeds a given threshold (denoted as censoring threshold and indicated as C_{th}). The channel estimation is either perfect (no estimation error) or imperfect (with an estimation error). Accordingly, for each censoring strategy, there are two possibilities, namely perfect or imperfect channel estimation. The FC employs coherent reception to fuse the binary local decisions received from the censored CR users, in order to obtain a final decision regarding the presence or absence of PUs. Low complexity majority-logic fusion of the decisions received from the selected CR users is considered in present case. The overall probability of missed detection is selected as the key performance metric and is evaluated, through simulations, under several channel and network conditions.

The main contributions of this chapter can be summarized as follows.

- Closed-form expressions of the estimation error variances for Rayleigh and Nakagami-m fading channels are presented. These expressions are expedient to evaluate the performance of CSS with censoring based on imperfect channel estimation.
- The performance, in terms of missed detection and total error probabilities under both perfect and imperfect channel estimation strategies, is investigated. The effects of Nakagami-m fading, S- and R-channel SNRs on the performance of the considered CSS schemes are investigated.
- The impact of the R-channel estimation error on the detection performance in the considered fading scenarios is evaluted.
- Direct performance comparisons between perfect and imperfect channel estimation schemes, for various values of the main channel and network parameters, are carried out.
- In threshold-based censoring scenarios, novel analytical expressions, as functions of the censoring threshold C_{th}, for the selection of CR users are derived in Rayleigh and Nakagami-m fading channels. In particular, the probability mass functions (PMF) of the number of censored CR users is analyzed.

- The impact of the number of available CR users and of the average R-channel SNRs on the average missed detection and average total error probabilities of CSS schemes is investigated.
- In threshold-based censoring schemes, the impact of the censoring threshold on the average missed detection and average total error probability, with the derivation of an optimized censoring threshold.
- The performances of several hard-decision fusion strategies are also evaluated and compared with each other under various fading channels.

The rest of the chapter is organized as follows. In Sect. 7.2, the basics of CSS are introduced. In Sect. 7.3, the performance of CSS in faded environments (Rayleigh and Nakagami-*m*) under several hard decision fusion rules is studied. In Sect. 7.4, two different censoring methods such as Rank-based and threshold-based censoring have been analyzed under both perfect and imperfect channel estimation schemes. Finally, conclusions are drawn in Sect. 7.5.

7.2 Cooperative Spectrum Sensing

Detection of PU by a single CR user may not be accurate due to impairment in S-channel or hidden node problem which necessitates the use of cooperation among many CR users. In such cases, as anticipated in Sect. 7.1, detection/sensing performance can be improved, by alleviating the effects of fading, if different CR users are allowed to cooperate by sharing their detection information, i.e., considering CSS. Therefore, CSS improves the detection performance where all CR users employ identical EDs and sense the PU individually and send their sensing information in the form of 1-bit binary decisions (1 or 0) via R-channels to FC. The hard decision combining fusion rule (OR, AND, and majority-logic fusion rules) is performed at FC using a counting rule to make the final decision regarding the presence or absence of a PU [7–9, 12, 15]. In case of soft decision combining, the CR users can transmit the entire local sensing samples or the complete local test statistics to FC. Existing receiver diversity techniques [3–5], such as LRT, EGC, and MRC, can be utilized at the FC for soft combining of local observations or test statistics. The performance of CSS with hard decision fusion in faded environments is investigated in the next section.

7.3 Impact of Fading on Cooperative Spectrum Sensing

The energy detection method is the common method for detection of unknown signals in noise [10, 11]. The block diagram of an energy detector is shown in [10, 11] which consist of one band pass filter (BPF), one signal squarer, one integrator and one decision device. The input BPF selects the center frequency and the corresponding

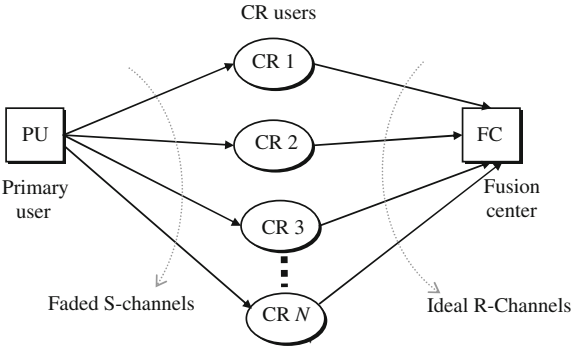

Fig. 7.1 Cooperative spectrum sensing network: illustrative scenario

bandwidth of interest (with width W). The output of the BPF filter is passed to a squaring device to measure the received energy. Then an integrator is placed to determine the observation interval, T. Finally, the output of the integrator, denoted as Y, is compared with a detection threshold to decide on the presence/absence of a PU signal. We assume that all CR users use the same energy detector and the identical threshold (denoted as λ). The received signal $x_i(t)$ at the input of the i-th CR user can be expressed as

$$x_i(t) = \begin{cases} n(t) & : H_0 \\ h_i(t)s(t) + n(t) & : H_1 \end{cases} \tag{7.1}$$

where $s(t)$ is the PU signal with energy E_s and $n(t)$ is the noise waveform. The noise $n(t)$ is modeled as a zero-mean white Gaussian random process. The Rayleigh faded S-channel coefficient for the i-th CR user is denoted as $h_i(t)$. H_1 and H_0 are the two hypotheses associated with presence and absence of a PU respectively. Each CR user has an energy detector to detect on the presence or absence of a useful signal.

We consider a network of N CR users sensing the spectrum of a PU, as shown in Fig. 7.1. Each CR user makes its own decision regarding the presence of the PU, and forwards the binary decision (1 or 0) to FC for data fusion. We consider only one FC and all CR users are equipped with single antenna. The PU is located far away from all CR users. All CR users are assumed to be relatively close to each other. The distance between any two CR users is shorter than the distance between a PU and a CR user and the distance between a CR user and the FC. For simplicity, we assume that the average SNR in the S-channel is the same for each CR user. We consider that the S-channels are faded, while the R-channels are ideal channels (noiseless). According to the sampling theorem, the noise process can be expressed as follows [26]:

$$n(t) = \sum_{j=-\infty}^{\infty} n_j \, \mathrm{sinc}(2Wt - j) \tag{7.2}$$

where $\text{sinc}(x) = \sin(\pi x)/(\pi x)$ and $n_j = n(j/(2W))$. One can easily show that

$$n_j \sim \mathcal{N}(0, N_{01}W); \forall j \tag{7.3}$$

where N_{01} is the one-sided noise power spectral density, W is the one-sided bandwidth and $\mathcal{N}(\mu, \sigma^2)$ is a Gaussian distribution with mean μ and variance σ^2.

When the PU is absent (i.e., H_0 is true), each CR user receives only the noise signal at the input of the ED and the noise energy can be approximated, over the time interval $(0, T)$, as follows [10, 11]:

$$\int_0^T n^2(t)\,\mathrm{d}t = \frac{1}{2W}\sum_{j=1}^{2u} n_j^2 \tag{7.4}$$

where u is the time-bandwidth product. If we define $n_j' = n_j/\sqrt{N_{01}W}$, the decision statistic at i-th CR user, denoted as Y_i in case of H_0, can be written as [10, 11]:

$$Y_i = \sum_{j=1}^{2u} n_j'^2. \tag{7.5}$$

In particular, Y_i is the sum of the squares of $2u$ standard Gaussian variates with zero mean and unit variance. Therefore, Y_i has a central χ^2 distribution with $2u$ degrees of freedom.

The same approach can be applied in the presence of the signal $s(t)$ of a PU, by replacing $\{n_j\}$ in (7.4) with $n_j + s_j$, where $s_j = s(j/(2W))$. In this case, the decision statistic Y_i has a non-central χ^2 distribution with $2u$ degrees of freedom and non-centrality parameter $2\gamma_{s,i}$ [10, 11]. More precisely:

$$Y_i \sim \begin{cases} \chi^2_{2u} & : H_0 \\ \chi^2_{2u}(2\gamma_{s,i}) & : H_1 \end{cases} \tag{7.6}$$

In a non-faded environment, the detection and false alarm probabilities for the i-th CR user can be expressed as follows [7, 16]:

$$P_{d,i} = \Pr[Y_i > \lambda | H_1] = Q_u\left(\sqrt{2\gamma_{s,i}}, \sqrt{\lambda}\right) \tag{7.7}$$

$$P_{f,i} = \Pr[Y_i > \lambda | H_0] = \Gamma(u, \lambda/2)/\Gamma(u) \tag{7.8}$$

where $\gamma_{s,i}$ is the instantaneous S-channel SNR, $\Gamma(\cdot)$ is the incomplete gamma function [27], and $Q_u(\cdot, \cdot)$ is the generalized Marcum Q-function of order u [28]. The expression for the probability of false alarm ($P_{f,i}$) for the i-th CR user, as given in Eq. (7.8), remains the same when fading is considered in the S-channel, owing to the independence of $P_{f,i}$ from the SNR $\gamma_{s,i}$. For a chosen value of $P_{f,i}$, the corresponding detection threshold λ can be set following Eq. (7.8). The ED thus compares Y_i with

its preset detection threshold λ and takes a hard binary decision about the presence of a PU.

When h_i is time-varying, because of fading, Eq. (7.7) returns the probability of detection as a function of the instantaneous SNR $\gamma_{s,i}$. In this case, the average probability of detection at the i-th CR user can be derived by averaging (7.7) over fading statistics [7–9] and can be given the following expression:

$$\bar{P}_{d,i} = \int_0^\infty Q_u\left(\sqrt{2x}, \sqrt{\lambda}\right) f_\gamma(x)\, \mathrm{d}x \tag{7.9}$$

where $f_\gamma(x)$ is the probability density function (pdf) of γ under fading.

7.3.1 Rayleigh Fading Channel

If the received signal amplitude at the i-th CR user has a Rayleigh distribution, then the SNR ($\gamma_{s,i}$) has the following exponential pdf [5, 11]:

$$f_\gamma(\gamma_{s,i}) = \frac{1}{\bar{\gamma}_s} \exp\left(-\frac{\gamma_{s,i}}{\bar{\gamma}_s}\right);\ \gamma_{s,i} \geq 0 \tag{7.10}$$

where $\bar{\gamma}_s$ is the average SNR of the S-channel. The average P_d at the i-th CR user in this case, $\bar{P}_{d,i,Ray}$ can now be evaluated by substituting (7.10) in (7.9), thus obtaning:

$$\bar{P}_{d,i,Rayl} = \exp\left(-\frac{\lambda}{2}\right) \sum_{k=0}^{u-2} \frac{1}{k!}\left(\frac{\lambda}{2}\right)^k + \left(\frac{1+\bar{\gamma}_s}{\bar{\gamma}_s}\right)^{u-1} \times \left(\exp\left(-\frac{\lambda}{2(1+\bar{\gamma}_s)}\right) - \exp\left(-\frac{\lambda}{2}\right) \sum_{k=0}^{u-2} \frac{1}{k!}\left(\frac{\lambda\bar{\gamma}_s}{2(1+\bar{\gamma}_s)}\right)^k\right). \tag{7.11}$$

7.3.2 Nakagami-m Fading Channel

If the received signal amplitude at the i-th CR user follows a Nakagami-m distribution, then $\gamma_{s,i}$ has the following gamma pdf [5, 11]:

$$f_\gamma(\gamma_{s,i}) = \left(\frac{m}{\bar{\gamma}_s}\right)^m \frac{\gamma_{s,i}^{m-1}}{\Gamma(m)} \exp\left(-\frac{m\gamma_{s,i}}{\bar{\gamma}_s}\right);\ \gamma_{s,i} \geq 0 \tag{7.12}$$

where m is the Nakagami fading parameter. The average probability of detection at the i-th CR user in the case of Nakagami-m channel $\bar{P}_{d,i,Nak}$ can be evaluated by substituting (7.12) in (7.9), obtaining:

$$\bar{P}_{d,i,Naka} = \alpha \left[G_1 + \beta \sum_{n=1}^{u-1} \frac{(\lambda/2)^n}{2n!} {}_1F_1\left(m; n+1; \frac{\lambda \bar{\gamma}_s}{2(m+\bar{\gamma}_s)}\right)\right] \tag{7.13}$$

where ${}_1F_1(\cdot;\cdot;\cdot)$ is the confluent hypergeometric function [27, Sect. 9.2]

$$\alpha = \frac{1}{\Gamma(m) 2^{m-1}} \left(\frac{m}{\bar{\gamma}_s}\right)^m \tag{7.14}$$

$$\beta = \Gamma(m) \left(\frac{2\bar{\gamma}_s}{m+\bar{\gamma}_s}\right)^m \exp\left(-\frac{\lambda}{2}\right) \tag{7.15}$$

and

$$G_1 = \frac{2^{m-1}(m-1)!}{\left(\frac{m}{\bar{\gamma}_s}\right)^m} \frac{\bar{\gamma}_s}{m+\bar{\gamma}_s} \exp\left(-\frac{m\lambda}{2(m+\bar{\gamma}_s)}\right) \left[\left(1+\frac{m}{\bar{\gamma}_s}\right) \left(\frac{m}{m+\bar{\gamma}_s}\right)^{m-1} \times L_{m-1}\left(-\frac{\lambda\bar{\gamma}_s}{2(m+\bar{\gamma}_s)}\right) + \sum_{n=0}^{m-2} \left(\frac{m}{m+\bar{\gamma}_s}\right)^n L_n\left(-\frac{\lambda\bar{\gamma}_s}{2(m+\bar{\gamma}_s)}\right)\right] \tag{7.16}$$

where $L_n(\cdot)$ is the Laguerre polynomial of degree n [27, Sect. 8.970]. We can also obtain an alternative expression for $\bar{P}_{d,i,Ray}$ by setting $m = 1$ in (7.13)—this expression is numerically equivalent to the one in (7.11). As already discussed in Sect. 7.3, all CR users in the network use identical EDs (with the same threshold λ) which make hard binary decisions and transmit them to the FC via noiseless R-channels.

Assuming independent decisions, the fusion rule according to which k-out of-N CR users are needed to make a final decision on the presence/absence of a PU can be characterized by a binomial distribution based on Bernoulli trials, where each trial represents the decision process of each CR user. The generalized formula for the overall probability of detection, according to a generic k-out of-N rule, is given by [8, 29]:

$$Q_d = \sum_{l=k}^{N} \binom{N}{l} \bar{P}_d^l (1-\bar{P}_d)^{N-l} \tag{7.17}$$

where $\bar{P}_d$ is the average probability of detection for each individual CR user as defined by generalized Eq. (7.9). The overall probability of detection under OR-fusion rule (i.e., 1 out of N rule) can be evaluated by setting $k = 1$ in Eq. (7.17):

$$Q_{d,OR} = \sum_{l=1}^{N} \binom{N}{l} \bar{P}_d^l (1-\bar{P}_d)^{N-l} = 1 - (1-\bar{P}_d)^N. \tag{7.18}$$

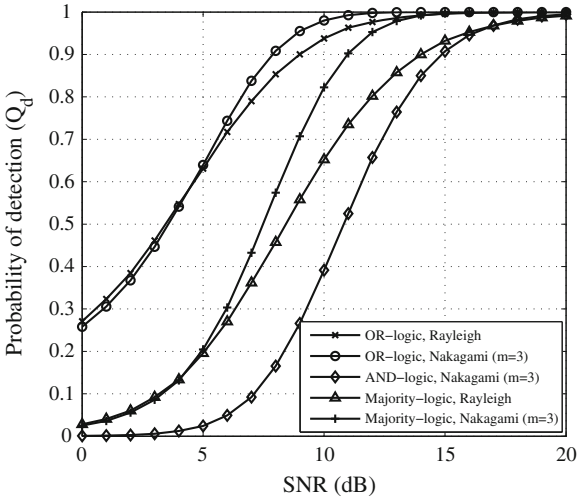

Fig. 7.2 Probability of detection as a function of the average S-channel SNR ($\bar{\gamma}_s$), considering various fusion rules (OR, AND, majority logic). Both Rayleigh and Nakagami-*m* fading scenarios are considered. In all cases, $N = 3$ CR users, $Q_f = 0.1$, and $u = 5$

The performance with AND-fusion rule (i.e., N out of N rule) can be evaluated by setting $k = N$ in Eq. (7.17):

$$Q_{d,\,AND} = \sum_{l=N}^{N} \binom{N}{l} \bar{P}_d^l (1 - \bar{P}_d)^{N-l} = \bar{P}_d^N. \tag{7.19}$$

Finally, for the case of majority-fusion rule, or simply for $(N/2+1)$ out of N rule, the probability of detection, denoted as $Q_{d,Maj}$, can be evaluated by setting $k = \lfloor N/2 \rfloor$ in Eq. (7.17).

The overall probability of false alarm (Q_f) for the considered fusion rules (OR, AND, and Majority fusion rules) can be evaluated by replacing $\bar{P}_d$ with P_f in Eqs. (7.17), (7.18) and (7.19), respectively. It is of interest to observe that the probability of false alarm (P_f) is independent of the SNR (γ_s), so that it remains same for all fading channels. It may also be observed that in order to evaluate Q_d for a specific fading channel, we need to consider the appropriate expression for $\bar{P}_d$ (namely, $\bar{P}_{d,i,Ray}$ or $\bar{P}_{d,i,Naka}$ in Eqs. (7.17) to (7.19)) to obtain the performance in Rayleigh or Nakagami-*m* channels, respectively.

In Fig. 7.2, the probability of detection Q_d is shown, as a function of the S-channel SNR, considering AND, OR, and majority logic hard decision fusion rules. For each fusion rule, Nakagami-*m* fading channel is considered. The OR and majority fusion rules for Rayleigh fading channel are also shown for comparison purposes. In all cases, there are $N = 3$ cooperating CR users, $Q_f = 0.1$, and $u = 5$. In the case of CSS in a Nakagami-*m* fading channel, for a particular value of the average SNR

(namely, 6 dB), the probability of detection is above 0.82, 0.36 and 0.01 for the OR, majority logic, and AND fusion rules, respectively. We can say that OR fusion rule performs better than the AND and the majority logic fusion rules. In the presence of Rayleigh fading, the CSS with OR fusion rule outperforms the schemes with the other fusion rules. Furthermore, in all cases of logic fusions we observe that the performance of CSS in Nakagami-*m* fading channel is better than the performance in Rayleigh fading channel—this is expected, as the Nakagami-*m* (with $m = 3$) fading is less severe than Rayleigh fading. Therefore, in the presence of such a Nakagami-*m* fading, the number of reliable S-channels is higher than the number in the Rayleigh fading case.

7.4 Censoring of CR User

As already discussed in Sect. 7.3, the performance of CSS, considering S-channels as noisy-faded and R-channels as ideal, has been well studied. However, in many practical situations R-channels may not be noiseless (ideal) channels. Though most works on spectrum sensing assume noiseless R-channels [7–9, 12–16, 29], the presence of fading in R-channels is likely to affect the decisions sent by CR users where the FC is far from CR users. If the R-channel connecting a CR user to the FC is heavily faded, the decision received at the FC is likely to be an erroneous version of that transmitted by the CR user. In such cases, it is better to stop transmitting decisions from this CR user and, thus, censoring is expedient in these scenarios. The CR users whose R-channels are estimated as reliable by the FC are censored, i.e., they are allowed to transmit. The CR users which are not participating in improving the detection performance may be stopped, so that system complexity can be reduced and the detection performance can be improved. Therefore, censoring of CR users is necessary to improve the performance of CSS. The cooperative spectrum sensing network with censoring of CR users is shown in Fig. 7.3. We assume that both S-and R-channels are modeled as noisy and faded.

As anticipated in Sect. 7.1, in this section we study the performance of two censoring schemes, namely: (i) rank-based censoring and (ii) threshold- based censoring. The FC employs coherent reception to fuse binary local decisions received from censored CR users to obtain a final decision regarding the presence or absence of PUs. The overall probabilities of missed detection and total error are selected as the key performance metrics and are evaluated, through simulations, under several channel and network conditions. In [29], it is shown that the total error probability (given as the sum of the probabilities of missed detection and false alarm) is a decreasing function of number of available CR users in the network when majority-logic fusion is performed at the FC. One can easily expect that as the number of available CR users in the network increases, the performance of majority-logic fusion, in terms of total error probability, is better than that with AND-logic (where the probability of missed detection is a decreasing function of the probability of false alarm) and OR-logic (where the missed probability of detection is a decreasing function of the

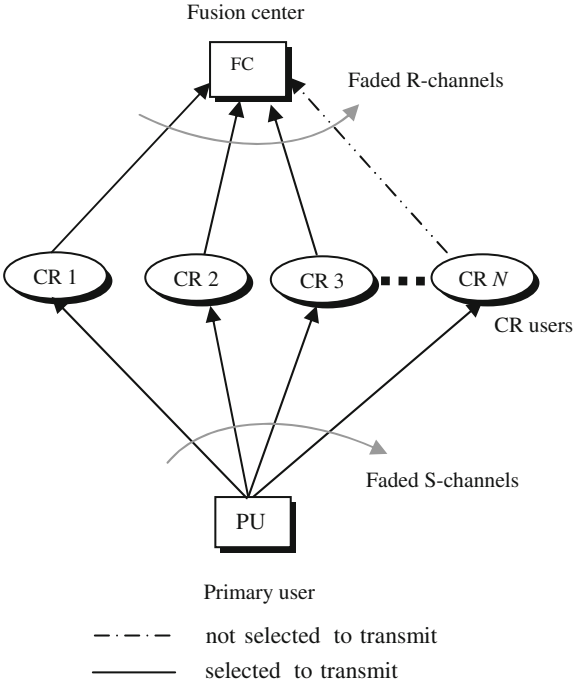

Fig. 7.3 Cooperative spectrum sensing network with censoring: illustrative scenario

probability of false alarm) fusions. This is why, in the current section, low-complexity majority-logic fusion of the decisions received from the selected CR users is considered. A CR user takes an individual hard binary decision and, if censored, transmits its decision, using binary phase shift keying (BPSK) as modulation format, to the FC over the corresponding faded R-channel.

Transmissions between the CR users and the FC are carried out in two phases. In the first transmission phase, each CR user sends one training symbol to enable the FC to estimate all fading channel coefficients between FC and N number of CR users corresponding to N participating CR users. Minimum mean square estimation (MMSE) of the R-channel coefficients is obtained at the FC using training symbols sent by the CR users to the FC. The signal from the k-th CR user received at the FC is:

$$y_k = s_k h_k + n_k; \quad k \in \{1, 2, \ldots, N\} \tag{7.20}$$

where S_k is BPSK symbol $(\sqrt{E_b}, -\sqrt{E_b})$ indicating H_1 and H_0, respectively. The R-channel fading coefficient is denoted as h_k and $n_i \sim CN(0, \sigma_n^2)$ is the sample of AWGN. The complex Gaussian channel noise samples $\{n_k\}$ and faded R-channel coefficients $\{h_k\}$ are mutually independent. We assume that the FC estimates the k-th CR user's fading coefficient h_k according to an MMSE estimation strategy on the basis of the observable y_k as follows [22–24]:

$$\hat{h}_k = E\left[h_k | y_k\right] = \frac{\sqrt{E_b}}{E_b + \sigma_n^2} y_k. \tag{7.21}$$

We model the k-th R-channel estimation error ($\tilde{h}_k$) as the difference between the actual and the estimated k-th R-channel coefficients, i.e., $\tilde{h}_k = h_k - \hat{h}_k$, where h_k is the actual k-th R-channel coefficient while $\hat{h}_k$ is its estimate. The channel estimation is either perfect (with no estimation error) or imperfect (with estimation error). Accordingly, two censoring schemes are considered: one is based on perfect channel estimation ($\hat{h}_k = h_k$) while the other is based on imperfect channel estimation ($\hat{h}_k = h_k - \tilde{h}_k$). After the first phase, K (out of N) CR users, selected on the basis of rank-based censoring (the selected CR users are associated with the best K estimated channel coefficients) and threshold-based censoring (the selected K CR users have estimated channel coefficients exceeding the predefined threshold C_{th}). The FC informs the selected CR users via one-bit feedback (we assume that feedback channels are error-free). In the second transmission phase, the K selected CR users send their local binary BPSK modulated decisions to the FC over the corresponding R-channels. The fading coefficients of R-channels are assumed to be fixed over a symbol transmission time, as the channel is assumed to be slowly faded.

The signal, received from the k-th selected CR user, at the FC is [22–24]:

$$y_{k,d} = m_k h_k + n_{k,d}; \qquad k \in \{1, 2, \ldots, K\} \tag{7.22}$$

where the channel noise $n_{k,d} \sim CN(0, \sigma_n^2)$ and $m_k \in \{\sqrt{E_b}, -\sqrt{E_b}\}$ is the BPSK modulated binary decisions.

Since the communication channel is noisy and affected by fading, a decision received by the FC might differ from the one sent by the corresponding CR user. The decision received from the k-the selected CR user is

$$u_k = \begin{cases} 1 & \text{if the received decision in favor of } H_1 \\ 0 & \text{if the received decision in favor of } H_0 \end{cases} \tag{7.23}$$

where $k \in \{1, 2, \ldots, K\}$. The FC finally makes a global decision according to the following general majority logic-like fusion rule $u_0 = \Gamma(u_1, \ldots, u_K)$ [17]:

$$u_0 = \Gamma\left(u_1, \ldots, u_k\right) = \begin{cases} H_1 & \text{if } \sum_{k=1}^{K} u_k > \frac{K}{2} \\ H_0 & \text{if } \sum_{k=1}^{K} u_k < \frac{K}{2} \\ H_0 \text{ or } H_1 & \text{if } \sum_{k=1}^{K} u_k = \frac{K}{2}. \end{cases} \tag{7.24}$$

In other words, if the number of decisions in favor of H_1 is larger than the number of decisions in favor of H_0, the FC takes a global decision in favor of H_1 and vice versa. Sometimes, if the number of decisions in favor of H_1 is equal to the number

of decisions in favor of H_0, then the FC flips a coin and takes a decision in favor of either H_0 or H_0.

7.4.1 Rank-Based Censoring

According to this censoring scheme, K (out of N) CR users—those with the best estimated channel coefficients (i.e., the highest ones)—are selected, as already discussed above section.

7.4.1.1 Rank-Based Censoring in Rayleigh Faded Channel

The R-channel coefficient h_k (for the k-th selected CR user) is modeled as a zero-mean complex Gaussian random variable with variance $\sigma^2 = 1 (h_k \sim CN(0, \sigma^2))$, as in [22], and $n_k \sim CN(0, \sigma_n^2)$. The complex Gaussian channel noise samples $\{n_k\}$ and Rayleigh faded R-channel coefficients $\{h_k\}$ are mutually independent. For the k-th Rayleigh faded R-channel, the fading coefficient ($h_k = \alpha_k exp(j\theta_k)$, where $\theta_k \sim U(-\pi, \pi)$ can be expressed, in terms of h_{kI} and h_{kQ}, as

$$h_k = h_{kI} + jh_{kQ} \tag{7.25}$$

where $h_{kI} = \alpha_k \cos\theta_k$ and $h_{kQ} = \alpha_k \cos\theta_k$.

The amplitude $|h_k|$ is Rayleigh distributed only when $h_{kI}, h_{kQ} \sim CN(0, \sigma_n^2/2)$ [5]. The estimated k-th R-channel coefficient can be obtained by substituting (7.20) in (7.21), obtaining:

$$\begin{aligned} \hat{h}_k &= \frac{\sqrt{E_b}}{E_b + \sigma_n^2} \left(\sqrt{E_b} h_k + n_k \right) \\ &= \frac{E_b}{E_b + \sigma_n^2} h_k + \frac{\sqrt{E_b}}{E_b + \sigma_n^2} n_k \end{aligned} \tag{7.26}$$

where $h_k = h_{kI} + h_{kQ}$, with $h_{kI}, h_{kQ} \sim \mathcal{N}(0, 1/2)$ given by (7.26) (assuming normalized fading power $\mathbb{E}(\alpha^2) = 1$), and n_k is also complex Gaussian, i.e., $n_k = n_{kI} + n_{kQ}$ where $n_{kI}, n_{kQ} \sim \mathcal{N}(0, \sigma_n^2/2)$. From (7.26), the estimation error coefficient for the k-th R-channel $\tilde{h}_k = h_k - \hat{h}_k$ can be expressed as

$$\begin{aligned} \tilde{h}_k &= h_k \left(1 - \frac{E_b}{E_b + \sigma_n^2} \right) - \frac{\sqrt{E_b}}{E_b + \sigma_n^2} n_k \\ &= h_k \frac{\sigma_n^2}{E_b + \sigma_n^2} - \frac{\sqrt{E_b}}{E_b + \sigma_n^2} n_k. \end{aligned} \tag{7.27}$$

As seen from (7.27), the term $\tilde{h}_k$ is a complex quantity and can also be written in terms of real and imaginary parts, i.e.,

$$\tilde{h}_k = \tilde{h}_{kI} + j\tilde{h}_{kQ} \tag{7.28}$$

where

$$\tilde{h}_{kI} = \frac{\sigma_n^2}{E_b + \sigma_n^2} h_{kI} - \frac{\sqrt{E_b}}{E_b + \sigma_n^2} n_{k,I}$$
$$\tilde{h}_{kQ} = \frac{\sigma_n^2}{E_b + \sigma_n^2} h_{kQ} - \frac{\sqrt{E_b}}{E_b + \sigma_n^2} n_{k,Q}.$$

From the theory of Gaussian random variables, it is well known that if $Z = aX + bY$ where $X \sim \mathcal{N}(m_X, \sigma_X^2)$ and $Y \sim \mathcal{N}(m_Y, \sigma_Y^2)$ then

$$Z \sim \mathcal{N}(m_Z, \sigma_Z^2); \quad m_Z = am_X + bm_Y, \sigma_Z^2 = a^2\sigma_X^2 + b^2\sigma_Y^2. \tag{7.29}$$

This implies that both $\tilde{h}_{kI}$ and $\tilde{h}_{kQ}$ that appear in (7.28) can be written in terms of their means and variances as follows:

$$\tilde{h}_{kI} \sim \mathcal{N}\left(0, \left(\frac{\sigma_n^2}{E_b + \sigma_n^2}\right)^2 \frac{1}{2} + \frac{E_b}{(E_b + \sigma_n^2)^2} \frac{\sigma_n^2}{2}\right) \tag{7.30}$$

$$\tilde{h}_{kQ} \sim \mathcal{N}\left(0, \left(\frac{\sigma_n^2}{E_b + \sigma_n^2}\right)^2 \frac{1}{2} + \frac{E_b}{(E_b + \sigma_n^2)^2} \frac{\sigma_n^2}{2}\right). \tag{7.31}$$

The mean and variance of $\tilde{h}_k$ are 0 and $\sigma^2_{\tilde{h},Rayl}$, respectively, i.e., $\tilde{h}_k \sim \mathcal{N}(0, \sigma^2_{\tilde{h},Rayl})$ [21, 22], where

$$\begin{aligned}\sigma^2_{\tilde{h},Rayl} &= 2\left[\left(\frac{\sigma_n^2}{E_b + \sigma_n^2}\right)^2 \frac{1}{2} + \frac{E_b}{(E_b + \sigma_n^2)^2} \frac{\sigma_n^2}{2}\right] \\ &= \frac{\sigma_n^4 + E_b\sigma_n^2}{(E_b + \sigma_n^2)^2} = \frac{\sigma_n^2}{E_b + \sigma_n^2} \\ &= \left(1 + \frac{E_b}{\sigma_n^2}\right)^{-1} = \frac{1}{1 + \bar{\gamma}_R}.\end{aligned} \tag{7.32}$$

The k-th Rayleigh faded R-channel estimation error coefficient can be generated using the following distribution:

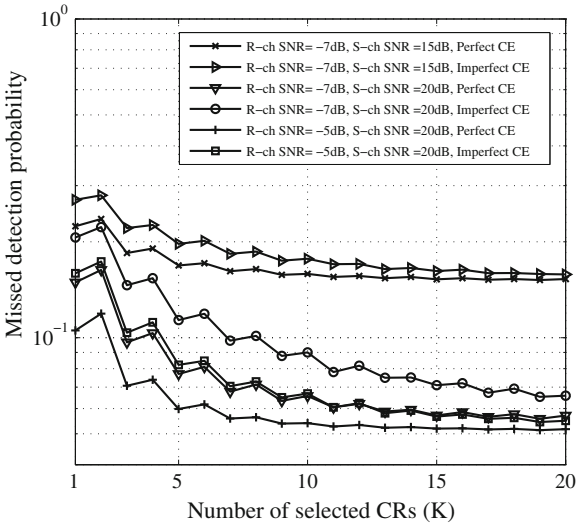

Fig. 7.4 Performance of CSS network with censoring of CR users under perfect and imperfect channel estimation for various values of average R-channel SNR ($\bar{\gamma}_R$) and average S-channel SNR ($\bar{\gamma}_s$) in Rayleigh fading (CE stands for channel estimation, $N = 20$, $P_f = 0.05$ and $u = 5$)

$$|\tilde{h}_k| = \sqrt{\tilde{h}_{kI}^2 + \tilde{h}_{kQ}^2}; \qquad \tilde{h}_{kI} \sim \mathcal{N}\left(0, \frac{\sigma_{\tilde{h},Rayl}^2}{2}\right), \quad \tilde{h}_{kQ} \sim \mathcal{N}\left(0, \frac{\sigma_{\tilde{h},Rayl}^2}{2}\right). \tag{7.33}$$

The following results are obtained using MATLAB-based simulations for both perfect and imperfect channel estimation schemes. S-channels and R-channels are both considered to be Rayleigh faded. The missed detection (Q_m) and the total error ($Q_m + Q_f$) probabilities are evaluated considering the impact of several network parameters, such as the probability of false alarm (P_f) in each CR user, the average R-channel SNR ($\bar{\gamma}_R$), and the average S-channel SNR ($\bar{\gamma}_s$).

In Fig. 7.4, the probability of missed detection is shown as a function of K. The performance of CSS with censoring under both perfect and imperfect channel estimation schemes is evaluated. Two values of S-channel average SNR (15 dB, 20 dB) and two values of R-channel average SNR (−5 dB, −7 dB) are considered. With both perfect and imperfect channel estimation, the probability of missed detection reduces for increasing values of the number of selected CR users, as well as of the S- and R-channel SNRs. The probability of incorrect reception from CR users at the FC reduces with higher R-channel SNR. As expected, for a given value of the R-channel SNR, Q_m is higher with imperfect channel estimation, as channel-based censoring leads to the selection of a group of CR users which may not be the best ones due to error in channel estimation. Furthermore, according to (7.32), an increase in the R-channel SNR leads to a decrease in estimation error variance and this, in

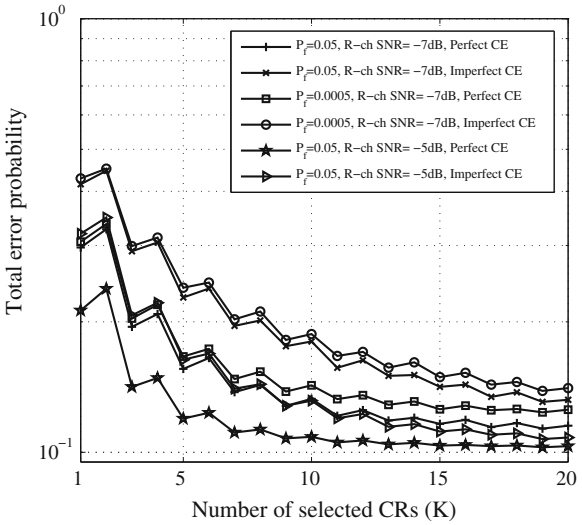

Fig. 7.5 Total error probability versus the number of selected CR users (K) under perfect and imperfect channel estimation for various values of ($\bar{\gamma}_R$) and P_f in Rayleigh fading ($\bar{\gamma}_s = 20\text{dB}$, $N = 20$, and $u = 5$)

turn, reduces the average estimation error. A reduced estimation error leads to a further reduction of Q_m. In particular, in the case of imperfect channel estimation with $K = 10$, Q_m decreases by 25.77 % when the R-channel SNR increases from -7 to -5 dB. Similarly, in the case of perfect channel estimation, Q_m decreases by 17.80 % for the same values of K and R-channel SNR. Higher values of the S-channel SNRs improves the detection of the PU at the CR user. For example, as the S-channel SNR increases from 15 to 20 dB, and $K = 10$, Q_m decreases by 58.57 and 49.15 % in case of perfect and imperfect channel estimation, respectively. Under perfect channel estimation, by censoring, the FC selects CR users with best R-channel coefficients which means that decisions sent by selected CR users to the FC have low probability of getting flipped. As the FC uses a majority-logic fusion, it achieves a floor in the missed detection performance at a certain number of CR users, i.e., no further improvement in detection performance is obtained by increasing the number of CR users beyond this.

In Fig. 7.5, the total error probability is shown as a function of the number of selected CR users for various values of the probability of false alarm (P_f), and R-channel SNR. The number of available CR users is 20 and S-channel SNR is fixed at 20 dB. As the R-channel SNR increases from -7 to -5 dB, the total error probability reduces for both the cases of perfect channel and imperfect channel estimation. Higher R-channel SNR reduces probabilities of incorrect reception from CR users at the FC. As expected, for a given value of the R-channel SNR, the total error probability is higher with imperfect channel estimation, as channel-based censoring leads to the selection of a group of CR users which may not be the best

ones due to error in channel estimaton. Furthermore, according to (7.32), an increase in the R-channel SNR leads to a decrease in the estimation error variance $\sigma^2_{\tilde{h},Rayl}$ and this, in turn, reduces the average estimation error. A reduced estimation error leads to a further reduction of the total error probability as CR users with higher reliability in R-channels are selected. In particular, in case of perfect channel estimation, when the selected number of CR users is 10 and the R-channel SNR increases from -7 to -5 dB, the total error probability reduces by 18.28 %. However, in case of imperfect channel estimation, the total error probability decreases by 26.97 % for the same values of network parameters. Further the impact of P_f is also depicted in Fig. 7.5. Different P_f corresponds to setting of different threshold levels (λ) at an energy detector. In all cases, two values of P_f, namely 0.05 and 0.0005, are considered. As P_f increases from 0.0005 to 0.05, the total error probability decreases for both perfect and imperfect channel estimation. This is because as P_f increases from 0.0005 to 0.05 the value of detector threshold is lowered from 31 to 18. Thus number of decisions of CR users in favor of hypothesis H1 increases. So the total error probability decreases. For example, in case of imperfect channel estimation, as P_f increases from 0.0005 to 0.05 with the number of selected CR users is 10, the total error probability decreases by 3.84 %, while in case of perfect channel estimation, the total error probability decreases by 6.86 %.

7.4.1.2 Rank-Based Censoring in Nakagami-*m* Faded Channel

The Nakagami-*m* distribution can be obtained from the Gamma distribution [3]. More precisely, if $X \sim gamma(r, s)$ then the k-th CR users' Nakagami-*m* fading channel coefficient (h_k) is obtained by setting $r = m, s = \Omega/m$(Ω is the Nakagami-*m* fading power which is normalized to unity, i.e., $\Omega = 1$) and considering $h_k = \sqrt{X}$. The estimated k-th Nakagami-*m* faded R-channel coefficient can be expressed as $\hat{h}_k = h_k - \tilde{h}_k$ and, taking into account the independence between h_k and $\tilde{h}_k$, it follows that $\sigma^2_{\hat{h},Naka} = \sigma^2_{h,Naka} + \sigma^2_{\tilde{h},Naka}$, where $\sigma^2_{\hat{h},Naka}$ is the estimated variance of Nakagami-*m* fading coefficient ($\hat{h}_k$), $\sigma^2_{h,Naka}$ is the actual variance of h_k and $\sigma^2_{\tilde{h},Naka}$ is the error variance of $\tilde{h}_k$. The analytical expressions for actual, estimated and error variance of Nakagami-*m* distribution can be derived using $\sigma^2_{h,Naka}$ as

$$\sigma^2_{h,Naka} = \Omega\left[1 - \frac{1}{m}\left(\frac{\Gamma(m+\frac{1}{2})}{\Gamma(m)}\right)^2\right] \tag{7.34}$$

$$\sigma^2_{\hat{h},Naka} = \hat{\Omega}\left[1 - \frac{1}{m}\left(\frac{\Gamma(m+\frac{1}{2})}{\Gamma(m)}\right)^2\right] \tag{7.35}$$

where $\hat{\Omega}$ is the estimated Nakagami-m fading power (which is not equal to 1). It can be shown that the error variance for Nakagam-m fading channel ($\sigma^2_{\tilde{h},Naka}$) is given as:

$$\sigma^2_{\tilde{h},Naka} = \frac{1}{(1+\bar{\gamma}_R)^2}\left[\Omega + \bar{\gamma}_R - \frac{\Omega}{m}\left(\frac{\Gamma(m+\frac{1}{2})}{\Gamma(m)}\right)^2\right]. \tag{7.36}$$

Setting $m = 1$ and assuming $\Gamma(m+1/2) \cong 1$ in (7.36), one derives the expression for estimated error variance in Rayleigh channel ($\sigma^2_{\tilde{h},Rayl}$), which matches with Eq. (7.32).

From (7.34), (7.35) and (7.36), it follows that

$$\hat{\Omega} = \frac{1 + \left(\frac{1}{1+\bar{\gamma}_R}\right) - \frac{1}{m}\left(1 + \left(\frac{1}{1+\bar{\gamma}_R}\right)^2\right)\left(\frac{\Gamma(m+\frac{1}{2})}{\Gamma(m)}\right)^2}{1 - \frac{1}{m}\left(\frac{\Gamma(m+\frac{1}{2})}{\Gamma(m)}\right)^2}. \tag{7.37}$$

The estimated Nakagami-m faded coefficient for the k-th CR user, in the case of imperfect channel estimation, can be generated using $\hat{h}_k = \sqrt{\left(gamma(m, \hat{\Omega}/m)\right)}$.

The following results are obtained using MATLAB-based simulations for both perfect and imperfect channel estimation schemes. The performance of CSS has been evaluated in Nakagami-m faded environment. S-channel and R-channel fading are considered to be same, i.e., Nakagami-m fading in S-channel and Nakagami-m fading in R-channel with same Nakagami parameter. The missed detection probability (Q_m) is evaluated by varying the Nakagami fading parameter m and the average R-channel SNR $\bar{\gamma}_R$.

In Fig. 7.6, the missed detection probability is analyzed as a function of the number of selected sensors. The impact of the Nakagami fading parameter (m) and of the average R-channel SNR, with both perfect and imperfect channel estimation, is analyzed. Two values of R-channel SNR (-5 and -7 dB) and three different values of m (1, 2 and 3) are considered for this figure. The performance with $m = 1$ corresponds to that of Rayleigh fading as in Fig. 7.4. For increasing values of K, of the R-channel SNR, and of the parameter m, the missed detection probability (Q_m) decreases at the FC significantly, for both perfect and imperfect channel estimation. When the R-channel SNRs increase, the noise effect reduces in the R-channel so that the FC receives a larger number of correct decisions and this leads to a reduction in the missed detection probability. Higher values of the R-channel SNR reduces the probability of incorrect reception from CR users at the FC. As seen earlier in Fig. 7.4, for a given value of the R-channel SNR, the missed detection probability is higher with imperfect channel estimation than with perfect channel estimation. Furthermore, according to (7.36), an increase in the R-channel SNR leads to a decrease in the estimation error variance $\sigma^2_{\tilde{h},Naka}$ and this, in turn, reduces the average estimation

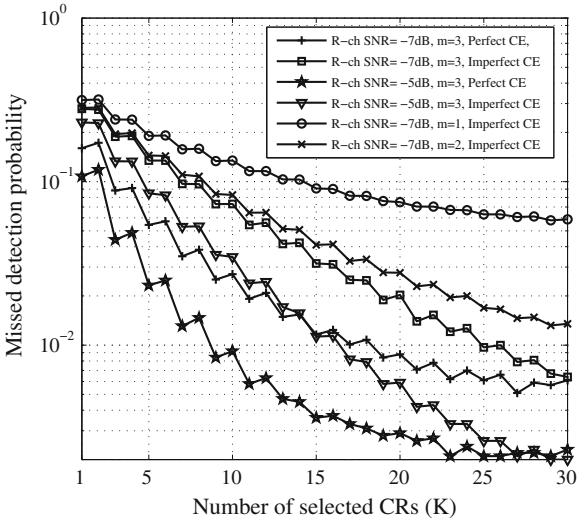

Fig. 7.6 Performance of CSS with censoring of CR users under perfect and imperfect channel estimation for various values of average R-channel SNRs ($\bar{\gamma}_R$) in Nakagami-m faded environment, impact of fading parameter (m) on imperfect channel estimation is also depicted ($\bar{\gamma}_s = 20\,\text{dB}$, N = 30, $P_f = 0.05$ and $u = 5$)

error. A reduced estimation error leads to a further reduction of the missed detection probability. When fading parameter increases from 1 to 3, the fading severity in the channel decreases so that the FC receives more correct decisions which lead to further reduction in missed detection probability.

7.4.2 Threshold-Based Censoring

In this censoring scheme, a CR user (say the k-th) is selected for transmission if the amplitude of the corresponding estimated R-channel fading coefficient $\hat{h}_k$ is above C_{th}. This approach involves two transmission phases: in the first phase, the FC estimates the R-channel corresponding to each CR user; in the second phase, the FC censors a CR user if the corresponding estimated channel coefficient exceeds a chosen threshold.

7.4.2.1 Threshold Based Censoring in Rayleigh Faded Channel

If the amplitude of the estimated R-channel fading coefficient is a Rayleigh distributed random variable with parameter σ. The probability of selecting a CR user is [24, 25]:

$$p = \Pr\left(|\hat{h}_k| > C_{th}\right) = \exp\left(-\frac{C_{th}^2}{2\sigma^2}\right). \tag{7.38}$$

The probability of selecting K CR users from N available CR users can then be expressed as follows [24, 25]:

$$P(K) = \binom{N}{K} p^k (1-p)^{N-K} \tag{7.39}$$

where p is the probability of selecting a CR user which is obtained from Eq. (7.38).

Let $P_m(\text{error}|K)$ indicate the conditional missed detection probability when decisions from K CR users are fused. Given $P(K)$, the probability of selecting K CR users in (7.39), the average probabilities of missed and false detection can be expressed as follows [24, 25]:

$$\bar{Q}_m = P\,(\text{missed detection}) = \sum_{K=0}^{N} P_m(\text{error}|K)P(K) \tag{7.40}$$

$$\bar{Q}_f = P\,(\text{false detection}) = \sum_{K=0}^{N} P_f(\text{error}|K)P(K). \tag{7.41}$$

Therefore, the average total error probability (an error occurs either with a missed detection or a false detection) can be expressed as follows:

$$\bar{Q} = \bar{Q}_m + \bar{Q}_f. \tag{7.42}$$

The average missed detection probability ($\bar{Q}_m$) and the average false alarm probability ($\bar{Q}_f$) are functions of the chosen censoring threshold C_{th}, as the PMF $\{P(K)\}$ of the number of censored CR users depends on C_{th}.

The following results are obtained, as in the previous sections, using MATLAB-based simulations. The performance of CSS for both perfect and imperfect channel estimation cases been evaluated in Rayleigh faded environments considering the impact of various network parameters, such as the censoring threshold (C_{th}), the number of available CR users (N), and the average R-channel SNRs ($\bar{\gamma}_R$).

In Fig. 7.7, the binomially distributed PMF of the number of selected CR users is shown, for various values of the censoring threshold C_{th}, under both cases of perfect and imperfect channel estimation schemes in Rayleigh faded channel. It can be observed that for small values of the censoring threshold a larger number of CR users are likely to be selected, while the PMF tends to concentrate on small values for higher values of the censoring threshold for both the channel estimation (CE) cases. For example, for a censoring threshold of 0.8, it is seen that $K = 16$ CR users have highest probability (0.13) of being selected under perfect channel estimation scheme. Similarly, in case of imperfect channel estimation scheme it is found that $K = 21$ CR users have highest probability (0.16) of being selected for the same

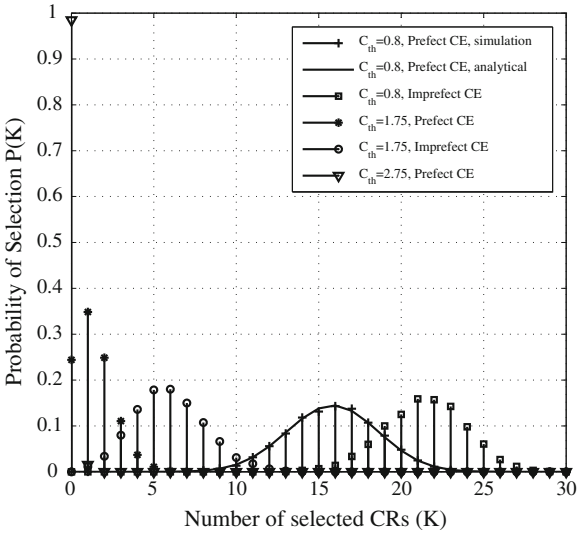

Fig. 7.7 PMF of the number of censored CR users for different censoring thresholds (C_{th}) under both perfect and imperfect channel estimation in Rayleigh faded channel

value of C_{th}. It can also be observed that as the censoring threshold increases, the PMF moves towards the origin for both channel estimation schemes. This is due to the fact that increasing the censoring threshold decreases the number of selected CR users. The obtained results show that the PMF of the number of selected CR users under imperfect channel estimation shifts to the right side of the PMF of the number of selected CR users under perfect channel estimation for a particular value of censoring threshold. According to Eq. (7.32), in the case of imperfect channel estimation, depending on the estimation error a larger number of CR users can be selected, for a fixed value of R-channel SNR, with respect to the case with perfect channel estimation. The binomially distributed PMF of the number of selected CR users, as obtained through simulations, matches exactly with result obtained based on the analytical expression given in Eqs. (7.38) and (7.39).

Figure 7.8 shows the impact of censoring threshold on the average missed detection probability, under perfect and imperfect channel estimation. Two different values of the average R-channel SNR (-5 and -7 dB) and two values of the available number of CR users (i.e. $N = 10$ and 30) are considered. It can be seen from the figure that as the censoring threshold increases, the average missed detection probability attains a minimum value in corespondence to an "optimal" censoring threshold, beyond which it increases and finally saturates to 0.5. The optimum censoring threshold is found to be different for the cases with perfect and imperfect channel estimation strategies and it depends on the number N of CR users and on the average R-channel SNR. For example, in the case of perfect channel estimation, as seen from the figure that an optimum censoring threshold is found to exist near 0.5 for $N = 10$ and average

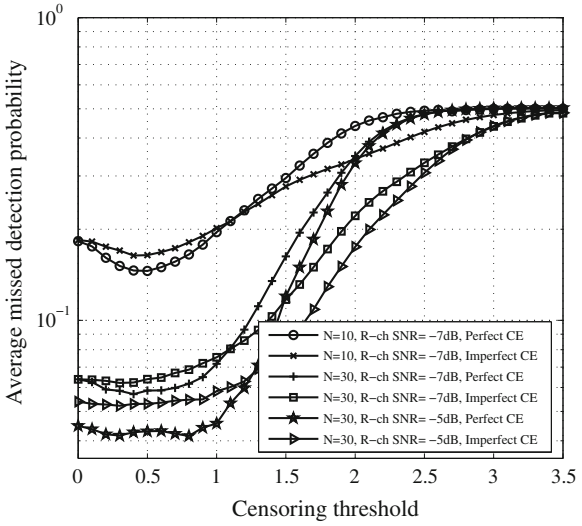

Fig. 7.8 Average missed detection probability as a function of C_{th} for various values of N and $\bar{\gamma}_R$ under perfect and imperfect channel estimation in Rayleigh fading channel ($\bar{\gamma}_s = 20$ dB, $P_f = 0.05$, and $u = 5$)

R-channel SNR $= -7$ dB. Similarly, in case of imperfect channel estimation as seen from the figure an optimum censoring threshold is found to exist near 0.4 for $N = 10$ and average R-channel SNR $= -7$ dB. This behavior of the average missed detection probability is due to the changing PMF of the number of censored CR users for various values of the censoring threshold. For very small values of the threshold, even unreliable links tend to be selected, and the average probability of missed detection is rather high. On the other hand, as the censoring threshold is increased to a very high level, no CR user is selected to transmit, i.e. $P(0) = 1$, and the FC takes a decision by flipping a fair coin resulting in an average missed detection probability of 0.5. Therefore, there exists an optimal value of the censoring threshold, in correspondence to which the average probability of missed detection is minimized. Further, as expected, it can be seen that a larger number of CR users, as well as a higher average R-channel SNR, leads to a reduced average missed detection probability in correspondence to the optimized censoring threshold.

Figure 7.9 shows the impact of censoring threshold on the average total error probability (sum of average missed detection and average false alarm probabilities) under perfect and imperfect channel estimation. As censoring threshold increases, the average total error probability attains a minimum value at an 'optimal' censoring threshold level and thereafter increases with further increase in censoring threshold to finally attain a value of 1.0 (average missed detection probability reaches a value of 0.5 and average false alarm probability reaches a value of 0.5). There exists an optimal value of the censoring threshold, in correspondence to which the average total error probability is minimized. It can be seen that a high value of R-channel

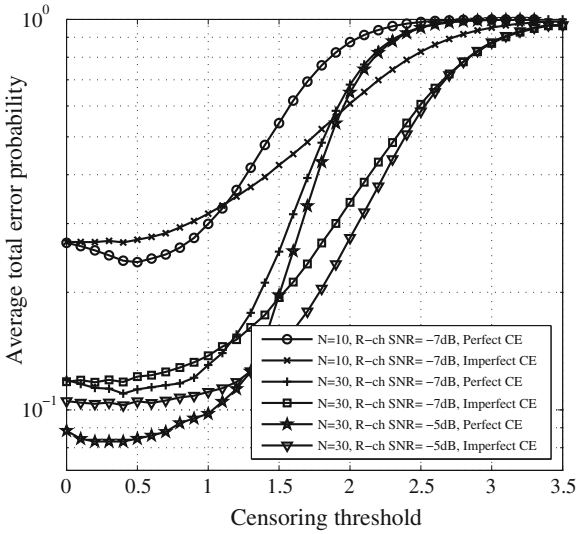

Fig. 7.9 Average total error probability as a function of C_{th} for various values of N and $\bar{\gamma}_R$ under perfect and imperfect channel estimation in Rayleigh fading channel ($\bar{\gamma}_s = 20$ dB, $P_f = 0.05$, and $u = 5$)

SNR as well as higher number of CR users leads to a reduced average total error probability in correspondence to the optimized C_{th} for both perfect and imperfect channel estimation cases. The optimum censoring threshold is found to be different for perfect and imperfect cases. For example, in perfect channel estimation case, the optimum censoring threshold is found to exist near 0.3 for $N = 30$ and average R-channel SNR of 5 dB. Similarly, in imperfect channel estimation case, the optimum censoring threshold is found to exist near 0.4 for $N = 30$ for the same value of R-channel SNR.

7.4.2.2 Threshold-Based Censoring in Nakagami-*m* Faded Channel

If the amplitude of estimated R-channel fading coefficient is a Nakagami-*m*-distributed random variable, the probability of selecting a CR user can be expressed as follows:

$$p = \Pr\left(|\hat{h}_k| > C_{th}\right) = 1 - \frac{\gamma\left(m, \frac{m}{\Omega} C^2{}_{th}\right)}{\Gamma(m)} \tag{7.43}$$

where $\gamma(s, x) = \int_0^x t^{s-1} e^{-t}\, \mathrm{d}t$ is the lower incomplete gamma function. The performance in Nakagami-*m* faded R-channels can be evaluated by substituting the expression of p given by (7.43) into (7.39), (7.40), (7.41) and (7.42). More details are presented in the following.

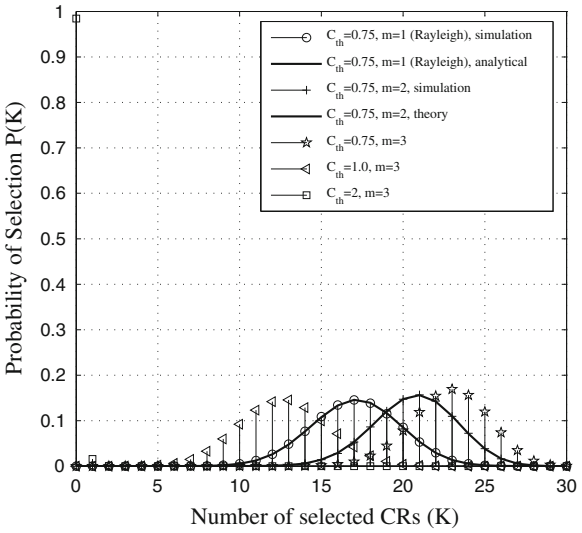

Fig. 7.10 PMF of the number of censored CR users for different censoring thresholds (C_{th}) under perfect channel estimation in Nakagami-m faded channel

As before, the following results are obtained using MATLAB-based simulations. The performance of CSS has been evaluated for both perfect and imperfect channel estimation schemes in Nakagami-m faded environments for various network parameters, such as the Nakagami fading parameter, the censoring threshold (C_{th}), the number of available CR users (N), and the average R-channel SNRs ($\bar{\gamma}_R$).

In Fig. 7.10, the binomially-distributed PMF of the number of selected CR users is shown for various values of the censoring threshold C_{th}. The impact of the Nakagami fading parameter m on the PMF is investigated. It can be observed that for small values of the censoring threshold, larger numbers of CR users are likely to be censored, while the PMF tends to concentrate on small values for higher values of the censoring threshold as observed in case of Rayleigh fading case in Fig. 7.7. It is also observed that when m increases, larger numbers of CR users are likely to be censored. The binomially distributed PMF of the number of selected CR users as obtained based on our simulation testbed matches exactly with result obtained based on the analytical expression given in Eqs. (7.39) and (7.43), which validates our simulation testbed. The binomially distributed PMF of the number of selected CR users as obtained for $m = 1$ matches exactly with result obtained for Rayleigh (Fig. 7.7) under perfect channel estimation.

Figure 7.11 shows the effects of Nakagami fading parameter, number of available CR users in the network, and R-channel SNR on the average missed detection probability under both perfect and imperfect channel estimations. We observe that for a fixed value of C_{th} when fading parameter as well as R-channel SNR increase, the average probability of missed detection decreases for both perfect channel estimation and imperfect channel estimation. When R-channel SNR increases, the effect of noise reduces in the channel so that the FC receives more correct decisions which leads to reduction in average missed detection probability. As expected, for a given value of

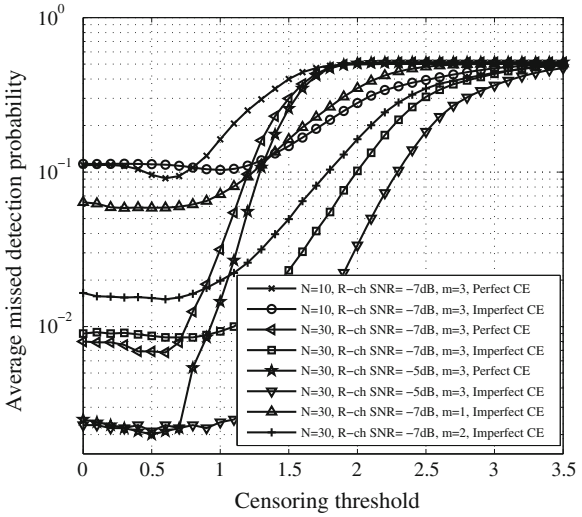

Fig. 7.11 Average missed detection probability as a function of C_{th} for various values of N, m and $\bar{\gamma}_R$ under perfect and imperfect channel estimation in case of Nakagami-m fading ($\bar{\gamma}_s = 20$ dB, $P_f = 0.05$, and $u = 5$)

the R-channel SNR, the missed detection probability is higher with imperfect channel estimation. Furthermore, according to (7.36), an increase in the R-channel SNR leads to a decrease in the estimation error variance $\sigma^2_{\tilde{h},Naka}$ and this, in turn, reduces the average estimation error. A reduced estimation error leads to a further reduction of the average missed detection probability. When fading parameter increases from 1 to 3, the fading severity in the R-channel as well as in S-channel decreases so that the FC receives more correct decisions which leads to reduction in average missed detection probability. We observe that the results obtained for fading parameter $m = 1$ match exactly with the results obtained for Rayleigh fading as shown in Fig. 7.8. As in the case of Rayleigh faded channel, an optimal censoring threshold exists in present Nakagami-m fading case also, which minimizes the average probability of missed detection. Further this optimum threshold also depends on the number of CR users, fading parameter (m), average R-channel SNR, and channel estimation schemes i.e., perfect and imperfect estimation.

In Fig. 7.12, the impact of censoring threshold, number of available CR users and R-channel SNR on the average total error probability (sum of average missed detection and average false alarm probabilities) is shown for Nakagami-m fading. The performance comparison between perfect and imperfect channel estimation is evaluated. It is seen from the figure that as C_{th} increases, the average total error probability attains a minimum value at an optimal C_{th} level and thereafter increases with further increase in C_{th} to finally attain a value of 1.0 (both average missed detection probability and average false alarm probability reach a value of each 0.5). The optimal value of C_{th}, in correspondence to minimum average total error probability is found to depend on channel and network parameters.

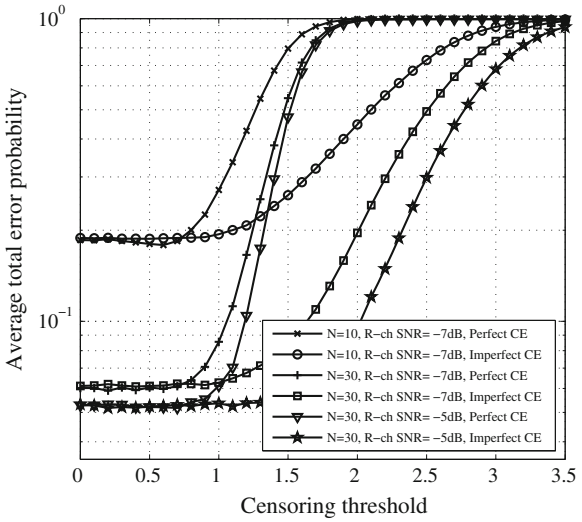

Fig. 7.12 Average total error probability as a function of C_{th} for various values of N, and $\bar{\gamma}_R$ under perfect and imperfect channel estimation in case of Nakagami-m fading ($\bar{\gamma}_s = 20$ dB, $m = 3$, $P_f = 0.05$, and $u = 5$)

7.5 Conclusion

In this chapter, the performance of cooperative spectrum sensing (CSS) using energy detection with and without censoring in Rayleigh and Nakagami-m faded channels has been investigated. The performance of a few hard decision fusion rules (OR-logic, AND-logic, and majority-logic) has been analyzed in a comparative way, considering meaningful performance metrics and evaluating the impact of several system parameters. Our results show that the CSS using energy detection and no censoring achieves highest probability of detection with OR-logic fusion, with respect to majority-logic and AND-logic fusions, under the same average SNR conditions in both Rayleigh and Nakagami-m fading channels. We have also investigated the performance of CSS with CR users censored on the basis of the quality of the R-channels, considering both Rayleigh and Nakagami-m faded channels. The performance with perfect and imperfect channel estimation has been analyzed, in a comparative way, under majority-logic fusion. Our results show that missed detection and total error probabilities reduce for increasing values of the number of selected CR users, regarless of the channel estimation quality (perfect or imperfect). However, in the presence of perfect channel estimation no further improvement, in terms of missed detection and total error probabilities, is obtained by increasing the number of CR users beyond a given limit. The Nakagami-m fading parameter and the R-/S-channel SNRs have a significant impact on the missed detection probability. With Rayleigh fading and majority-logic fusion, as the false alarm probability at each CR user increases, the total error decreases for both perfect and imperfect channel estimation. The censoring threshold for the selection of CR users has a significant impact on the average

missed detection probability. Depending on the configuration of relevant network parameters, such as the available number of CR users and the average R-channel SNRs, there exists an optimal censoring threshold, which corresponds to the minimum average missed detection and total error probabilities, for both perfect and imperfect channel estimation. The framework presented in this paper is useful in designing a cooperative spectrum sensing scheme able to prolong, by minimizing the number of "useless" transmission acts, the lifetime of an energy-constrained cognitive radio network.

References

1. Haykin, S.: Cognitive radio: brain-empowered wireless communications. IEEE J. Sel. Areas Commun. **23**(2), 201–220 (2005)
2. Cabric, S. D., Mishra, S. M., Brodersen, R. W.: Implementation issues in spectrum sensing for cognitive radios. In: Proceedings of the 38th Asilomar Conference on Signals, Systems, and Computers (ACSSC), vol. 1, pp. 772–776. Pacific Grove (2004)
3. Pandharipande. A., Linnartz. J. P. M. G.: Performance analysis of primary user detection in a multiple antenna cognitive radio. In: Proceedings of the IEEE International Conference on Communications (ICC), pp. 6482–6486. Glasgow (2007)
4. Niu, R., Chen, B., Varshney, P.K.: Decision fusion rules in wireless sensor networks using fading statistics. In: Proceedings of the 37th Annual Confrence on Information Sciences and Systems (CISS), Johns Hopkins University, Baltimore (2003)
5. Simon, M.K., Alouini, M.-S.: Digital Communication Over Fading Channels, 2nd edn. John Wiley and Sons, New Jersey (2004)
6. Ma, J., Zhao, G., Li, Y.: Soft combination and detection for cooperative spectrum sensing in cognitive radio networks. IEEE Trans. Wireless Commun. **7**(11), 4502–4507 (2008)
7. Ghasemi, A., Sousa, E. S.: Collaborative spectrum sensing for opportunistic access in fading environments. In: Proceedings of the IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN), pp. 131–136. Baltimore (2005)
8. Duan, J., Li, Y.: Performance analysis of cooperative spectrum sensing in different fading channels. In: Proceedings of the IEEE International Conference on Computer Engineering and Technology (ICCET), vol. 3, pp. 64–68. Chengdu (2010)
9. Nallagonda, S., Roy, S. D., Kundu, S.: Performance of cooperative spectrum sensing in fading channels. In: Proceedings of the IEEE International Conference on Recent Advances in Information Technology (RAIT), pp. 202–207. ISM, Dhanbad (2012)
10. Urkowitz, H.: Energy detection of unknown deterministic signals. Proc. of IEEE **55**(4), 523–531 (1967)
11. Digham, F. F., Alouini, M. -S., Simon, M. K.: On the energy detection of unknown signals over fading channels. In: Proceedings of the IEEE International Conference on Communications (ICC), pp. 3575–3579. Anchorage (2003)
12. Nallagonda, S., Suraparaju, S., Roy, S. D., Kundu, S.: Performance of energy detection based spectrum sensing in fading channels. In: Proceedings of the IEEE International Conference on Computer and Communication Technology (ICCCT), pp. 575–580. MNIT, Allahabad (2011)
13. Ghasemi, A., Sousa, E.S.: Opportunistic spectrum access in fading channels through collaborative sensing. IEEE J. sel. Areas Commun **2**(2), 71–82 (2007)
14. Ghasemi, A., Sousa, E. S.: Impact of user collaboration on the performance of opportunistic spectrum access. In: Proceedings of the IEEE Vehicular Technology Conference (VTC), Montreal (2006)
15. Zhang, W., Mallik, R., Letaief, K. B.: Cooperative Spectrum Sensing Optimization in Cognitive Radio networks. In: Proceedings of the IEEE International Conference on Communications (ICC), pp. 3411–3415. Beijing (2008)

16. Nallagonda, S., Roy, S.D., Kundu, S.: Performance of cooperative spectrum sensing in Log-normal Shadowing and fading under fusion rules. International Journal of Energy, Information and Communications, Science and Engineering Research Support Center (SERSC). Korea **3**(3), 15–28 (2012)
17. Ferrari, G., Pagliari, R.: Decentralized binary detection with noisy communication links. IEEE Trans. Aerosp. Electron. Syst. **42**(4), 1554–1563 (2006)
18. Chen, B., Jiang, R., Kasetkasem, T., Varshney, P.: Channel aware decision fusion in wireless sensor networks. IEEE Trans. Signal Process **52**(12), 3454–3458 (2004)
19. Rago, C., Willett, P., Bar-Shalom, Y.: Censoring sensors: a low-communication-rate scheme for distributed detection. IEEE Trans. Aerosp. Electron. Syst. **32**(2), 554–568 (1996)
20. Appadwedula, S., Veeravalli, V.V., Jones, D.L.: Energy-efficient detection in sensor networks. IEEE J. Sel. Areas Commun. **23**(4), 693–702 (2005)
21. Ahmadi, H. R., Vosoughi, A.: Channel aware sensor selection in distributed detection systems. In: Proceedings of the IEEE International Workshop On Signal Processing Advances in Wireless Communications (SPAWC), pp. 71–75. Perugia (2009)
22. Ahmadi, H. R., Vosoughi, A.: Impact of channel estimation error on decentralized detection in bandwidth constrained wireless sensor networks. In: Proceedings of the IEEE Conference on Military Communications (MILCOM), pp. 1–7. San Diego (2008)
23. Nallagonda, S., Roy, S.D., Kundu, S.: Performance evaluation of cooperative spectrum sensing with censoring of cognitive radios in Rayleigh fading channel. Wireless Pers. Commun. **70**(4), 1409–1424 (2013)
24. Kundu, C., Kundu, S., Ferrari, G., Raheli, R.: Majority logic fusion of censored decisions in wireless sensor networks with Rayleigh fading. In Proceedings of the IEEE Natitional conference on Comuunications (NCC), pp. 1–5. IIT, Kharagpur (2012)
25. Nallagonda, S., Roy, S. D., Kundu, S., Ferrari, G., Raheli, R.: Cooperative spectrum sensing with censoring of cognitive radios in Rayleigh fading under Majority Logic Fusion. In: Proceedings of the IEEE Natitional conference on Comuunications (NCC), pp. 1–5. IIT, Delhi (2013)
26. Shannon, C.E.: Communication in the presence of noise. Proc. IRE **37**(1), 10–21 (1949)
27. Gradshteyn, I.S., Ryzhik, I.M.: Table of Integrals, Series and Products, 7th edn. Academic Press/ Elsevier, San Diego (2007)
28. Nuttall, A.H.: Some integrals involving the Q_M function. IEEE Trans. Inf. Theory **21**(1), 95–96 (1975)
29. Zhang, W., Mallik, R., Letaief, K.B.: Optimization of cooperative Spectrum Sensing with energy detection in Cognitive Radio networks. IEEE Trans. wireless Commun. **8**(12), 5761–5766 (2009)

Chapter 8
Medium Access Control in Cognitive Impulse Radio UWB Networks

Luca De Nardis, Guido Carlo Ferrante and Maria-Gabriella Di Benedetto

Abstract Impulse Radio Ultra Wide Band (IR-UWB) is a candidate technology in the deployment of cognitive underlay networks. Medium Access Control protocols for IR-UWB networks were however conceived in the past moving from considerations related to performance of the UWB networks, rather than from the need to coexist with other wireless systems. This chapter analyses existing MAC protocols for low rate IR-UWB networks, and focuses on two specific protocols: the $(UWB)^2$ MAC, and the MAC of the IEEE 802.15.4a standard that leveraged the access strategy proposed by $(UWB)^2$. Characteristics of the two MAC protocols are reviewed, and the performance of the $(UWB)^2$ MAC is analysed by means of computer simulations, adopting an accurate model for Multiple User Interference. Results confirm the suitability of the $(UWB)^2$ protocol for low rate IR-UWB networks. Finally, the chapter discusses potential improvements and adaptations to be introduced for $(UWB)^2$ to meet the coexistence requirements imposed by operation of the UWB network in a cognitive fashion.

8.1 Introduction

The increasing interest towards the creation of smart environments, where people and objects interact thanks to seamless wireless connectivity, is leading to a renewed attention towards low data rate and low cost networks for mixed indoor/outdoor communications, capable of supporting sensor and ad-hoc networking. The interest

L. De Nardis (✉) · G. C. Ferrante · M.-G. Di Benedetto
DIET Department, Sapienza University of Rome,Via Eudossiana 18, Rome, Italy
e-mail: lucadn@newyork.ing.uniroma1.it

G. C. Ferrante
e-mail: ferrante@newyork.ing.uniroma1.it

M.-G. Di Benedetto
e-mail: dibenedetto@newyork.ing.uniroma1.it

M.-G. Di Benedetto and F. Bader (eds.), *Cognitive Communication and Cooperative HetNet Coexistence*, Signals and Communication Technology,
DOI: 10.1007/978-3-319-01402-9_8,

towards low data rate networks led, back in 2003, to the definition of the IEEE 802.15.4 standard for low rate, low complexity, and low power wireless networks [1]. The 802.15.4 standard also formed the basis for the definition of the ZigBee industrial standard, providing a comprehensive solution for low data rate networking spanning from physical layer to application [2].

IEEE 802.15.4/ZigBee lacked, however, a feature that was deemed important in a broad range of applications for low rate networks, namely, the capability of accurately determine the mutual distance, and in turn the position, of network terminals and of objects and individuals associated to them. This issue was addressed by the IEEE 802.15.4a Task Group [3], which defined an amendment to the original standard [4], adopting the Impulse Radio Ultra Wide Band (IR-UWB) technology as an alternative solution for the physical layer [5, 6]. IR-UWB was selected due to its interesting properties, in particular:

- an inherently high temporal resolution that provides good robustness in the presence of multipath, thereby allowing communication despite obstacles and Non-Line-Of-Sight (NLOS) propagation conditions;
- the capability of providing accurate ranging, thanks to its high temporal resolution, and thus to support the operation of positioning algorithms.

IR-UWB is also characterized by the capability of operating with extremely low power spectral density values, thanks to its ultra wide frequency band. This feature resonates with a recent trend in wireless communications, that called for more flexibility and efficiency in the use of the radio spectrum by means of reuse and coexistence between multiple radio systems and eventually led to the concept of cognitive radio [7]. UWB is a perfect candidate for the deployment of so called underlay cognitive radio networks, capable of operating simultaneously with other systems, thanks to its high coexistence capabilities. Higher layer design should however take into account the specific features of UWB in order to take full advantage of the possibilities offered by this technology. This is true in particular for the Medium Access Control sub-layer. Since the early stage of research on UWB networks, two different approaches were followed in MAC design: a conservative approach, that aimed to adopt existing solutions originally proposed for narrowband and wideband/CDMA networks, and a more innovative one that proposed solutions specifically designed to take advantage of the features of the IR-UWB signal. Algorithms following the former avenue of research included solutions derived from DS-CDMA networks, such as in [8, 9], or TDMA-based schemes, such as in the protocol proposed for the planned sister standard to 802.15.4a, IEEE 802.15.3a, focusing on high speed, short range UWB systems, to be derived from the 802.15.3 PAN standard [10], but never released and eventually withdrawn. More innovative proposals considered the impulsiveness of the IR-UWB signal and its impact of the performance of multiple access, such as the protocol proposed in [11], aiming at the minimization of transmitted power and maximization of transmission covertness for military applications, and the Uncoordinated Baseborn Wireless medium access control for UWB networks ($((UWB)^2$) protocol, originally proposed in [12], considering both communications and ranging/positioning aspects, and proposing an uncoordinated approach inspired

to the Aloha protocol. The $(UWB)^2$ idea was proposed jointly with a Multi User Interference (MUI) model specific to IR-UWB systems, named Pulse Collision model. The $(UWB)^2$ approach for propagation over AWGN channels was validated in [13]. In [14] the $(UWB)^2$ was revisited by redefining the structure of both control and DATA MAC Protocol Data Units (MACPDUs) based on the PDU structure of the original IEEE 802.15.4 MAC standard in order to guarantee compatibility of the new MAC protocol with both distributed and centralized network topologies defined in the 802.15.4 standard, and the performance analysis of the $(UWB)^2$ protocol originally carried out in [12] was extended to the case of multipath-affected channels, for both indoor and outdoor channel scenarios in Line Of Sight (LOS) and Non-Line Of Sight (NLOS) conditions. Channel parameters were obtained from the channel model proposed within the 802.15.4a Task Group, and a set of channel realizations were considered for each selected scenario. Finally, [14] included MUI in the performance analysis based on an enhanced version of the Pulse Collision model specific for IR-UWB [13], which took into account multipath. This MUI model was used to analyze performance of $(UWB)^2$ by simulation, as a function of channel, network size, and user bit rates.

The MAC defined in the IEEE 802.15.4a standard released in 2007 took an intermediate approach: it leveraged some of the innovative ideas proposed in the above protocols, such as the support for ranging and adoption of an Aloha-like medium access approach, as proposed in $(UWB)^2$, but framed such ideas in a typical Personal Area Network MAC protocol, not dissimilar from the pre-existing 802.15.3/802.15.4 MAC protocols.

In the above framework, this chapter focuses on the analysis of the $(UWB)^2$ MAC, and discusses how this protocol might be evolved to take into account the specific requirements of underlay UWB networks, and recent advances in UWB technology, such as the introduction of Time Reversal [15, 16] for increased spatial and temporal focus of UWB emissions.

The chapter is organized as follows. Section 8.2 summarizes the $(UWB)^2$ MAC protocol and the IEEE 802.15.4a MAC, as the only standard currently defined for low rate UWB communications. Section 8.3 reviews the Pulse Collision MUI model, used then in Sect. 8.4 to evaluate the performance of $(UWB)^2$ in presence of multipath and variable traffic conditions. Section 8.5 discusses potential extensions and evolution of the $(UWB)^2$ MAC so to support cognitive UWB networks by taking into account coexistence issues and novel UWB features such as Time Reversal algorithms. Section 8.6 draws conclusions.

8.2 $(UWB)^2$ and IEEE 802.15.4a MAC Protocols

8.2.1 The $(UWB)^2$ MAC

The high temporal resolution of IR-UWB signals has the beneficial side effect of reinforcing robustness to MUI, in particular for low data rate applications [5]. As a

consequence, access to the medium in low data rate UWB networks can be based on a most straightforward solution, Aloha [12, 17], by which devices transmit in an uncoordinated fashion. Thanks to the resilience to MUI offered by impulse radio, correct reception for multiple simultaneous links can be obtained. An Aloha-like approach may also favor lowering costs, since it does not rely on specific physical layer (PHY) functions, such as carrier sensing, and may thus be adapted with little effort to different PHYs.

As for the duty cycle of emitted signals, low data rate scenarios usually lead to an average Pulse Repetition Period (PRP), the average time between two consecutive pulses emitted by a device, on the order of $10^{-4} - 10^{-5}$ s, with an average duration of emitted pulses typically on the order of 10^{-10} s. Theoretically, the duty cycle can thus be as low as 10^{-6}. However, a detailed analysis of this issue requires introducing the channel model in order to take into account propagation effects on pulse duration.

When Time Hopping (TH) is the selected coding technique, TH Code Division Multiple Access (TH-CDMA) is a natural choice for multiple access. The adoption of TH-CDMA can introduce an additional degree of freedom, since the effect of pulse collisions is further reduced by the adoption of different codes on different links. Two factors cooperate in determining the robustness to MUI: low duty cycle of emitted signals, and association of different TH-Codes with different links.

$(UWB)^2$ is a multi-channel MAC protocol that is based on the combination of Aloha with TH-CDMA [12]. $(UWB)^2$ adopts the combination of a common code for signaling, where terminals share the same code, and code collisions are avoided thanks to phase shifts between different links, and Transmitter codes for data transfers, where each terminal has a unique code for transmitting, and the receiver switches to the code of the transmitter for receiving a packet.

The $(UWB)^2$ protocol applies the multi-code concept to the specific case of a TH-IR UWB system. $(UWB)^2$ adopts a Hybrid scheme based on the combination of a common control channel, provided by a Common TH code, with dedicated data channels associated to Transmitter TH codes. The adoption of a Hybrid scheme can be motivated as follows:

1. It simplifies the structure and the operation of the receiver, since data packet transmissions (and corresponding TH codes) are first communicated on the control channel.
2. It provides a common channel for broadcasting. This is a fundamental property for supporting the operation of higher layers protocols. Broadcast messages are for example required for routing and distributed positioning protocols.

As regards code assignment, several solutions have been proposed in the literature for distributed code assignment [18, 19]; as an alternative, a function creating a unique association between MAC ID and Transmitter Code can be adopted, similarly to the approach adopted in Bluetooth.

$(UWB)^2$ does not assume that synchronization between transmitter and receiver is available at the beginning of packet transmission, due to clock drifts in each terminal that, in case of low data rate networks, may lead to significant differences between times measured at each terminal. As a consequence, a synchronization trailer long

enough to guarantee a minimum required synchronization probability is included into the packet structure. The length of the trailer depends on current network conditions, and is provided to the MAC by the synchronization logic.

$(UWB)^2$ also supports the ranging feature provided by UWB. Distance information between transmitter and receiver can be in fact collected during control packets exchanges. Such information can enable optimizations of several MAC features, and allow the introduction of new functions at other layers, such as distributed positioning. Procedures adopted in $(UWB)^2$ for transmitting and receiving packets were defined in [12] and are described later in the Chapter. The procedures have two main objectives:

1. To exchange information required to allow transmitter and receiver to move to a dedicated code for subsequent data transmissions, such as the adopted synchronization trailer, hopping sequence and length.
2. To perform ranging. Since no common time reference is available, a three-way handshake is required to collect distance information at both ends of the link by estimating the round-trip-time of signals in the air.

It is assumed that, at each terminal T, MAC Protocol Data Units (MACPDUs) resulting from the segmentation/concatenation of MAC Service Data Units (MACSDUs) are stored in a transmit queue. It is also assumed that T is able to determine how many MACPDUs in the queue are directed to a given receiver R.

8.2.1.1 Transmission Procedure

The transmission procedure is triggered by the arrival of one or more MACPDUs in the transmit queue. The procedure comprises the following steps:

1. The ID of the intended receiver R is extracted from the first PDU in the queue.
2. T determines the number $N_{PACKETS}$ of MACPDUs in the queue that have R as their destination.
3. T checks if other MACPDUs were sent to R in the last T_{ACTIVE} seconds. If this is the case, T considers R as an *Active receiver*, and moves to step 5 below.
4. If R is not an Active receiver, T generates a Link Establish (LE) packet composed of the following fields:

   - *SyncTrailer* Used for synchronization;
   - *TxNodeID* The MAC ID of transmitter T;
   - *RxNodeID* The MAC ID of receiver R;
   - TH_{Flag} This flag is set to true if the standard TH code associated to *TxNodeID* is going to be used for the transmission of data PDUs. Conversely, if the transmitter is proposing to use a different TH code the flag is set to false.
   - *THCode* (optional)—This field is used when the TH_{Flag} is set to false to provide the information on the TH-code to be adopted for DATA pdus. In the original definition of $(UWB)^2$ no guidelines were provided on how to select the TH code;

as discussed in Sect. 8.5, TH code selection might take into account coexistence issues.

5. Terminal T sends the LE packet and waits for a Link Confirm (LC) response packet from R.
6. If the LC packet is not received within a time T_{LC}, the LE packet is re-transmitted for a maximum of N_{LC} times, before the transmission of the MACPDU is considered as failed.
7. Following the reception of a LC packet, T switches to the TH code declared in the LE packet and transmits the data PDU structured as detailed in Sect. 8.2.1.3.
8. Once the transmission is completed, T checks again the status of the data queue, and repeats the procedure until all MACPDUs in the transmit queue are served.

8.2.1.2 Reception Procedure

A terminal *R* in Idle state listens to the Common TH code. When a SyncTrailer is detected, *R* executes the following procedure:

1. *R* checks the *RxNodeID* field. If the value in the field is neither the MAC ID of *R* nor the broadcast ID, the reception is aborted and the reception procedure ends.
2. if the *RxNodeID* field contains the MAC ID of *R*, since *R* is assumed in *Idle* state, MACPDUs directed to this terminal will necessarily be LE packets; otherwise, if the $RxNodeID$ field contains the broadcast ID, *R* processes the payload and the resulting MACSDU is passed to higher layers.
3. Following the reception of the LE packet, *R* creates a LC packet, composed of:
   - *SyncTrailer* Used for synchronization:
   - *TxNodeID* The MAC ID of of transmitter *T*;
   - *RxNodeID* The MAC ID of receiver *R*;
4. *R* sends the LC packet and moves in the *Active* state, listening on the TH code indicated in the LE packet. If no data packet is received within a time T_{DATA} the receiver falls back to *Idle* state and the procedure ends.
5. When a data packet is received, *R* processes the payload, and extracts $N_{PACKETS}$ from the header. If $N_{PACKETS} > 0$, *R* remains in *Active* state, since at least $N_{PACKETS}$ more data packets are expected to be received from T. If $N_{PACKETS} = 0$, *R* goes back to the *Idle* state.

Note that multiple reception procedures could be run at the same time, assuming that the receiver is capable of listening to multiple TH-codes at the same time. Finally, the exchange of LE/LC packets can also be triggered on a periodic basis for the purpose of updating distance information, based on a three-way exchange involving LE, LC and DATA PDUs.

8.2.1.3 The $(UWB)^2$ MACPDU Format

The format of the MACPDU originally proposed in [12] was revisited and modified in order to take into account the characteristics of the future IEEE 802.15.4a PHY in [14]. The revised MACPDU is composed of a header, a payload, and a trailer. The standard header, shared by all PDUs and up to 23 bytes long, is derived from the 802.15.4 header and is organized as follows:

- frame control (2 bytes)
- sequence number (1 byte)
- destination PAN identifier (2 bytes)
- destination address (2/8 bytes)
- source PAN identifier (2 bytes)
- source address (2/8 bytes)

In the case of LE control packets (link set-up phase of $(UWB)^2$), the header includes the following additional fields:

- Time Hopping flag (1 bit), used to inform destination whether the standard Time Hopping code or a different one is going to be adopted in the DATA transmission;
- Time Hopping code (0/2 bytes), used for communicating the TH code to the destination (e.g., by including the code identifier, assuming that all nodes share a common codebook).

In the case of DATA PDU, the header contains the 1 byte additional field NPDU that indicates to the destination the number of additional DATA PDUs that will be sent from the source. If NPDU is different from 0, the destination will keep on listening on the DATA TH code and wait for additional DATA PDUs. The length of the payload is set to 0 for LE and LC PDUs, while the ACK PDU has a 2 byte payload containing the status of the corresponding DATA PDU. Finally, DATA PDUs have a payload length of up to 103 bytes. All PDUs include a 2 byte trailer consisting of a CRC code evaluated on the entire PDU. The above PDU structure leads to a maximum PDU length of 129 bytes, corresponding to the case of a DATA PDU with full header (24 bytes), full payload (103 bytes), and the 2 byte trailer.

8.2.2 The IEEE 802.15.4a MAC

The MAC of the 802.15.4a standard shares most of its features with the original 802.15.4 one. Two classes of devices are defined: Full-Function Devices (FFD), that implement all network functionalities, and Reduced-Function Devices (RFD), that only support a reduced set of functionalities. RFD devices can be dedicated to simple tasks, such as measuring and reporting the variations of a physical parameter (e.g. temperature, wind or humidity).

RFD and FFD devices organize in Personal Area Networks (PANs). A PAN is controlled by a coordinator, in charge of setting up and maintaining the PAN. Only

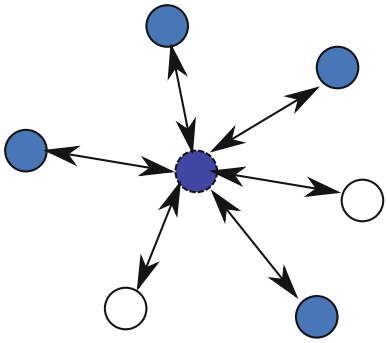

Fig. 8.1 Example of star topology (*Dark blue circle* PAN coordinator; *light blue circles* FFD devices; *white circles* RFD devices)

FFD devices can become PAN coordinators, while RFD devices can associate to an existing PAN by communicating with the PAN coordinator.

A PAN can be organized according to either of the two following network topologies:

- *star topology*—In this topology, devices only establish links with the coordinator, which is thus involved in all communications. The star topology is thus better suited for network architectures where a device is connected to a power outlet, and can thus act as the coordinator for a long time without exhausting its battery. Star topology is presented in Fig. 8.1.
- *peer-to-peer topology*—In this topology, FFD devices can communicate directly as long as they are within physical reach, while RFD devices, due to their limitations, are still limited to communications with the coordinator. Peer-to-peer topology is presented in Fig. 8.2.

The peer-to-peer topology is best suited for more complex topologies, thanks to its higher flexibility. The standard does not however include the definition of algorithms to create and manage large network topologies e.g. involving multiple hops.

8.2.2.1 Access Strategies

Medium access is one of the aspects where the 802.15.4a standard introduced significant variations with respect to 802.15.4. In the following the original 802.15.4 MAC is first described, and, next, the modifications introduced in 802.15.4a are discussed.

Access to the medium within a PAN is based on a combination of random access and scheduled access. Medium access within a PAN is controlled by the coordinator, that selects one of two alternative operation modes:

- *beacon-enabled*
- *nonbeacon-enabled*

In the *beacon-enabled* mode, the PAN coordinator broadcasts a periodic beacon containing information about the PAN. The period between two consecutive beacons

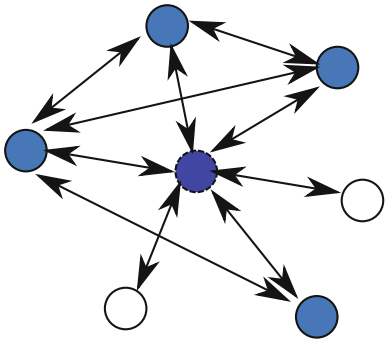

Fig. 8.2 Example of peer-to-peer topology (*Dark blue circle* PAN coordinator; *light blue circles* FFD devices; *white circles* RFD devices)

defines a superframe, characterized by a structure divided in 16 slots. The first slot is always occupied by the beacon, while the other slots are used for data communication by means of either random or scheduled access. Random access slots form the so-called Contention Access Period (CAP), while scheduled access slots, referred to in the standard as Guaranteed Time Slots (GTS), form the Contention Free Period (CFP) of the superframe. GTS slots can be assigned by the PAN coordinator to devices running applications with low latency requirements. The beacon contains information related to PAN identification, synchronization, and superframe structure.

The beacon-enabled modality is only adopted when the PAN has a star topology, and foresees two data transfer modes:

1. *Transfer from a device to the coordinator*—a device willing to transfer data to the coordinator uses a slotted Carrier Sensing Multiple Access with Collision Avoidance (CSMA-CA). A description of the CSMA-CA protocol adopted in the standard is provided later in this section for both slotted and unslotted versions. The coordinator may confirm the successful data reception with an optional acknowledgement packet within the same slot.
2. *Transfer from the coordinator to a device*—when the coordinator has data pending for a device, it announces so in the beacon. The interested device selects a free slot and sends a data request to the coordinator, indicating that it is ready to receive the data. Slotted CSMA-CA is adopted to send the request. When the coordinator receives the data request message, it selects a free slot and sends data using slotted CSMA-CA as well.

An example of superframe with both CAP and CFP is shown in Fig. 8.3. In the *nonbeacon-enabled* mode there is no explicit synchronization provided by the PAN coordinator. This mode is particularly suited for PANs adopting the peer-to-peer topology, but can be adopted in PANs adopting a star topology as well.

Peer-to-peer topology allows for a third transfer mode in addition to the two defined above: the *peer-to-peer data transfer*, in which devices exchange data without involving the PAN coordinator, thus allowing more complex topologies and larger networks. On the other hand, due to the lack of a superframe, the nonbeacon-enabled

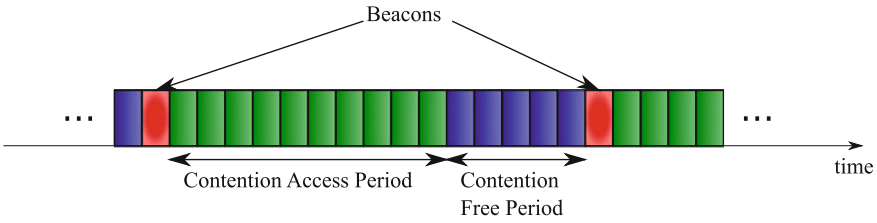

Fig. 8.3 Example of superframe in beacon-enabled modality

modality does not allow the use of GTS slots, and random access is adopted for medium sharing in all transfer modes.

As described above, in the random access phase devices adopt a protocol to access the medium, either slotted on unslotted depending on the selected PAN operation mode. It should be noted, however, that although the access protocol is referred as CSMA-CA, the CSMA-CA implementation in 802.15.4 significantly differs from the one in IEEE WLAN standards, such as the 802.11 family, where the access protocol foresees the use of Request To Send (RTS) and Clear To Send (CTS) packets [20]; in the CSMA-CA protocol in 802.15.4 no Collision Avoidance packets are used, and the implementation is actually closer to a CSMA scheme. The protocol can be summarized as follows.

In a nonbeacon-enabled network, when a device needs to send data it picks a random backoff delay, defined as a multiple of a backoff time unit. When the backoff delay expires, the device performs a Clear Channel Assessment (CCA) operation, listening to the channel in order to determine if it is idle. If this is the case the device immediately transmits the data packet; otherwise the device repeats the procedure with a new, larger backoff delay.

In a beacon enabled network the devices use a slotted version of the previous protocol to access the medium in the CAP portion of the superframe. The main differences compared to the unslotted version are the following:

1. the backoff delay unit is set to the duration of a slot in the superframe, and the beginning of the random delay interval always coincides with the beginning of a slot.
2. at the end of the random backoff delay the device performs a CCA operation at the beginning of the next slot; if the channel is idle, however, the device does not transmit the data packet immediately, but repeats the CCA for a number of slots defined by the value of a parameter called Contention Window. If the channel is idle for all the slots within the Contention Window the device transmits the data. If during one of the slots in the Contention Window the channel is detected to be busy, the device repeats the procedure by picking a new, larger backoff delay.

The most significant difference in the 802.15.4a MAC is the introduction of Aloha as an alternative channel access strategy, based on research results such those obtained for the $(UWB)^2$ showing that, thanks to the MUI robustness guaranteed by the IR-UWB signal structure, the Aloha approach provides satisfactory throughput in UWB

networks [12] for light and medium traffic loads, thus allowing to eliminate the collision avoidance and the corresponding additional access delay. The robustness of the Aloha approach adopted in the 802.15.4a standard is furthermore increased by the possibility of adopting a Time Hopping scheme, introducing a different delay on each burst in a packet and thus further reducing the probability of packet loss due to collisions. The CSMA-CA access was however kept as an option, in order to address high-density and high-traffic scenarios where the MUI robustness of the IR-UWB may prove not sufficient to provide good performance, and to support alternative physical layers.

8.3 BER Evaluation Under the Pulse Collision Model

8.3.1 System Model

We assume IR-UWB transmissions with Pulse Position Modulation (PPM) and TH coding. Signals generated at TX are described as follows:

$$s_{TX}(t) = \sqrt{E_{TX}} \sum\nolimits_j p_0\left(t - jT_S - \theta_j - \varepsilon b_{\lfloor j/N_s \rfloor}\right), \tag{8.1}$$

where $p_0(t)$ is the energy normalized waveform of the transmitted pulses, E_{TX} is the transmitted energy per pulse, T_S is the average pulse repetition period, $0 \leq \theta_j < T_S$ is the TH time shift of the j-th pulse, ε is the PPM shift, b_x is the x-th bit of a binary source sequence $\mathbf{b}$, N_S is the number of pulses transmitted for each bit, and $\lfloor x \rfloor$ is the inferior integer part of x. Propagation for link m occurs over a multipath-affected channel with impulse response given by:

$$h^{(m)}(t) = X^{(m)} \sum_{l=0}^{L^{(m)}} \sum_{k=0}^{K} \alpha_{k,l}^{(m)} \delta\left(t - \Delta t^{(m)} - T_l^{(m)} - \tau_{k,l}^{(m)}\right), \tag{8.2}$$

where $X^{(m)}$ is the amplitude gain, $L^{(m)}$ is the number of clusters, K is the number of paths that are considered within each cluster, $\delta(t)$ is the Dirac function, $\Delta t^{(m)}$ is the propagation delay, $T_l^{(m)}$ is the delay of the l-th cluster with respect to $\Delta t^{(m)}$, $\tau_{k,l}^{(m)}$ is the delay of the k-th path relative to the l-th cluster arrival time, and $\alpha_{k,l}^{(m)}$ is the real-valued tap weight of the k-th path within the l-th cluster. Tap weights are energy-normalized and thus verify:

$$\sum_{l=0}^{L^{(m)}} \sum_{k=0}^{K} \left(\alpha_{k,l}^{(m)}\right)^2 = 1. \tag{8.3}$$

For all channel parameters in Eq. 8.2, statistical characterization is suggested in [21] for 9 different propagation environments, i.e., (i) residential LOS, (ii) residential NLOS, (iii) office LOS, (iv) office NLOS, (v) outdoor LOS, (vi) outdoor NLOS, (vii) industrial LOS, (viii) industrial NLOS, and (ix) open outdoor environment NLOS (farm, snow-covered open area). For link m, both channel gain $X^{(m)}$ and propagation delay $\Delta t^{(m)}$ depend on distance of propagation $D^{(m)}$ between TX and RX. For $X^{(m)}$, in particular, one has:

$$X^{(\mathrm{m})} = 1 \Big/ \sqrt{10^{(PL^{(\mathrm{m})}/10)}}, \tag{8.4}$$

where $PL^{(m)}$ is the path loss in dB, which can be modelled as indicated in [21]. Reference TX and RX are assumed to be perfectly synchronized. The channel output is corrupted by thermal noise and MUI generated by N_i interfering and asynchronous IR-UWB devices. The received signal at RX input writes:

$$s_{RX}(t) = r_u(t) + r_{mui}(t) + n(t), \tag{8.5}$$

where $r_u(t)$, $r_{mui}(t)$, and $n(t)$ are the useful signal, MUI, and thermal Gaussian noise with double-sided power spectral density $N_0/2$, respectively. By denoting as 0 the reference link between TX and RX, the useful signal $r_u(t)$ writes as follows:

$$\begin{aligned} r_u(t) = \sqrt{E_0} \sum\nolimits_j \sum_{l=0}^{L^{(0)}} \sum_{k=0}^{K} \alpha_{k,l}^{(0)} \cdot \\ \cdot p_0\left(t - jT_S - \theta_j^{(0)} - \varepsilon b_{\lfloor j/N_S \rfloor} - \Delta t^{(0)} - T_l^{(0)} - \tau_{k,l}^{(0)}\right), \end{aligned} \tag{8.6}$$

where $E_0 = \left(X^{(0)}\right)^2 E_{TX}$ is the total received energy per pulse.
As for $r_{mui}(t)$, we assume all interfering signals to be characterized by the same T_S; thus:

$$\begin{aligned} r_{mui}(t) = \sum_{n=1}^{N_i} \sqrt{E_n} \sum\nolimits_j \sum_{l=0}^{L^{(n)}} \sum_{k=0}^{K} \alpha_{k,l}^{(n)} \cdot \\ \cdot p_0\left(t - jT_S - \theta_j^{(n)} - \varepsilon b^{(n)}_{\left\lfloor j/N_S^{(n)} \right\rfloor} - \Delta t^{(n)} - T_l^{(n)} - \tau_{k,l}^{(n)}\right), \end{aligned} \tag{8.7}$$

where index n represents the wireless link between the n-th interfering device and RX. In Eq. 8.7, $E_n = \left(X^{(n)}\right)^2 E_{TX}$, and $\Delta t^{(n)}$ are the received energy per pulse and the delay for link n. The terms $\theta_j^{(n)}$, $b_x^{(n)}$ and $N_S^{(n)}$ in Eq. 8.7 are the time shift of the j-th pulse for user n, the x-th bit, and the number of pulses per bit, respectively for user n. Both TH codes and data bit sequences are assumed to be randomly generated and correspond to pseudonoise sequences, that is, $\theta_j^{(n)}$ terms are assumed to be independent random variables uniformly distributed in the range $[0, T_S)$, and $b_x^{(n)}$ values are assumed to be independent random variables with equal probability to be 0 or 1. Based on the above assumptions, the N_i relative delays $\Delta t^{(0)} - \Delta t^{(n)}$,

with $n = 1, \ldots, N_i$ may be reasonably modelled as independent random variables uniformly distributed between 0 and T_S.

As well known, the optimum receiver structure for Eq. 8.6 consists of a RAKE receiver composed of a parallel bank of correlators, followed by a combiner that determines the variable to be used for the decision on the transmitted symbol. Each correlator of the RAKE is locked on one of the different replicas of the transmitted waveform $p_0(t)$. The complexity of such a receiver increases with the number of multipath components that are analyzed and combined before decision, and can be reduced by processing only a sub-set of the components that are available at the receiver input [22]. Such a reduction, however, entails a decrease in the available useful energy in the decision process, together with a consequent decrease in receiver performance. As a result, system designers have the possibility to trade the cost of the devices with the performance of the physical layer. For some application scenarios, for example, it might be better to have very cheap devices with modest performance with respect to high-priced terminals with better performance. In the examined scenario, we adopt a basic IR receiver that analyzes a single component of the received signal. This basic receiver is composed by a coherent correlator followed by an ML detector [22]. In every bit period, the correlator converts the received signal in Eq. 8.5 into a decision variable Z that forms the input of the detector. Soft decision detection is performed. For each pulse, we assume that the correlator locks onto the multipath component with maximum energy. By indicating with l_M and k_M the cluster and the path of the maximum energy multipath component for the reference user, the input of the detector Z for a generic bit b_x is as follows:

$$Z = \int_{xN_ST_S+\Delta T^{(0)}}^{(x+1)N_ST_S+\Delta T^{(0)}} s_{RX}(t)\, m_x\left(t - \Delta T^{(0)}\right) dt \tag{8.8}$$

where

$$\Delta T^{(0)} = \Delta t^{(0)} + T_{l_M}^{(0)} + \tau_{k_M,l_M}^{(0)} \tag{8.9}$$

and where

$$m_x(t) = \sum_{j=xN_S}^{(x+1)N_S} \left(p_0\left(t - jT_S - \theta_j\right) - p_0\left(t - jT_S - \theta_j - \varepsilon\right)\right). \tag{8.10}$$

By introducing Eq. 8.5 into Eq. 8.8, we obtain: $Z = Z_u + Z_{mui} + Z_n$, where Z_u is the signal term, Z_{mui} is the MUI contribution, and Z_n is the noise contribution, which is Gaussian with zero mean and variance $\sigma_n^2 = N_S N_0 \xi(\varepsilon)$, where $\xi(\varepsilon) = 1 - R_0(\varepsilon)$, and where $R_0(\varepsilon)$ is the autocorrelation function of the pulse waveform $p_0(t)$ [22]. Bit b_x is estimated by comparing the Z term in Eq. 8.8 with a zero-valued threshold according to the following rule: when Z is positive decision is 0, when Z is negative decision is 1.

8.3.2 BER Estimation Under the Pulse Collision Approach

According to Sect. 8.3.1, the average probability of error on the bit at the output of the detector for independent and equiprobable transmitted bits is: $BER = Prob\{Z < 0|b_x = 0\} = Prob\{Z_{mui} < y\}$, where $y = Z_u + Z_n$ is a Gaussian random variable with mean:

$$\mu_y = N_S \xi(\varepsilon) \sqrt{\left(\alpha^{(0)}_{l_M,k_M}\right)^2 E_0} = N_S \xi(\varepsilon) \sqrt{E_u} \tag{8.11}$$

and variance $\sigma_y^2 = N_S N_0 \xi(\varepsilon)$. The quantity E_u in Eq. 8.11 indicates the amount of useful energy conveyed by the maximum multipath contribution. The average BER at the receiver output can be evaluated by applying the Pulse Collision (PC) approach [23]. First, we compute the conditional BER for a generic y value, i.e., $Prob\{Z_{mui} < y|y\}$, and we then average over all possible y values, that is:

$$BER = \int_{-\infty}^{+\infty} Prob\{Z_{mui} < -y|y\}\, p_Y(y)\, dy. \tag{8.12}$$

Next, we expand the conditional BER in order to take into account collisions between pulses of different transmissions. In a bit period, the number of possible collisions at the input of the reference receiver, denoted with c, is confined between 0 and $N_S N_i$, with N_S pulses per bit and N_i interfering users. Thus:

$$BER = \sum_{c=0}^{N_S N_i} P_C(c) \int_{-\infty}^{+\infty} Prob\{Z_{mui} < -y|y, c\}\, p_Y(y)\, dy, \tag{8.13}$$

where $P_C(c)$ is the probability of having c collisions at the receiver input. For independent interferers, $P_C(c)$ can be expressed through the binomial distribution:

$$P_C(c) = \binom{N_S N_i}{c} (P_0)^c (1 - P_0)^{N_S N_i - c}, \tag{8.14}$$

where P_0 is the basic collision probability, which is defined as the probability that an interfering device produces a non-zero contribution within a single T_S. Given the receiver structure in Eq. 8.8, we approximate P_0 as follows:

$$P_0 = \frac{T_m + \varepsilon + \tau_{MAX}}{T_S}. \tag{8.15}$$

where T_m is the time duration of the pulse waveform $p_0(t)$, and τ_{MAX} is the maximum among the values of the root mean square delay spread for the Ni channels between the interfering devices and RX. Note that Eq. 8.15 provides acceptable P_0 values if $T_S > T_m + \varepsilon + \tau_{MAX}$, which is reasonable for LDR systems with long pulse repetition

periods. This condition guarantees that no Inter Frame Interference (ISI) is present at the receiver, even in the presence of multipath propagation.

As regards $Prob(Z_{mui} < y|y, c)$, we adopt the linear model introduced in [23], that is:

$$Prob\ (Z_{mui} < -y \mid y, c) = \begin{cases} 1 & for \quad y \leq -\zeta\ (n) \\ 1 - \frac{P_C(c)}{2}\left(1 + \frac{y}{\zeta(c)}\right) & for \quad \zeta\ (n) < y \leq 0 \\ \frac{P_C(c)}{2}\left(1 - \frac{y}{\zeta(c)}\right) & for \quad 0 < y \leq \zeta\ (n) \\ 0 & for \quad y > \zeta\ (n)\,, \end{cases} \tag{8.16}$$

where $\zeta(c)$ indicates the maximum interference contribution that can be measured at the output of the correlator. Based on [23], we propose here the following approximation for $\zeta(c)$:

$$\zeta(c) = \sum_{j=1}^{N_i}\left(\left\lceil \frac{c-j+1}{N_i} \right\rceil \sqrt{E_{\text{int}}^{(j)} \frac{T_m + \varepsilon}{\tau_{rms}^{(j)}}}\right), \tag{8.17}$$

where $\{E_{int}(1), E_{int}(2), \ldots, E_{int}(N_i)\}$ are the interfering energies $\{E_1, E_2, ..., E_{N_i}\}$ of Eq. 8.7, sorted in descending order so that $E_{int}(j) \geq E_{int}(j+1)$ for $j = 1, \ldots, N_i - 1$. The expression in Eq. 8.17 indicates that the value of the maximum interference contribution at the receiver output is computed privileging dominating interferers, that is, those users with the highest interfering energies. Note that in Eq. 8.17 we multiply the value of j-th interfering energy $E_{int}(j)$ by $(T_m + \varepsilon)/\tau_{rms}(j)$. This operation indicates that only part of the energy associated with a colliding pulse contributes to Z in Eq. 8.8, corresponding to the ratio between the correlator window $(T_m + \varepsilon)$ and the length of the pulse at the receiver, approximated by $\tau_{rms}(j)$. By introducing Eq. 8.16 into Eq. 8.13, one has:

$$\begin{aligned} BER \approx\ & \tfrac{1}{2} erfc\left(\sqrt{\tfrac{1}{2}\tfrac{N_S E_u}{N_0}\xi\ (\varepsilon)}\right) \\ & + \sum_{c=0}^{N_i N_S} \frac{P_C(c)^2}{2}\, \Omega\left(\frac{N_S E_u}{N_0}\xi\ (\varepsilon)\,,\ \frac{\varsigma(c)^2}{N_S N_0 \xi\ (\varepsilon)}\right), \end{aligned} \tag{8.18}$$

where

$$\begin{aligned} \Omega\ (A, B) =\ & \tfrac{1}{2} erfc\left(\sqrt{\tfrac{A}{2}} - \sqrt{\tfrac{B}{2}}\right) \\ & + \tfrac{1}{2} erfc\left(\sqrt{\tfrac{A}{2}} + \sqrt{\tfrac{B}{2}}\right) - erfc\left(\sqrt{\tfrac{A}{2}}\right). \end{aligned} \tag{8.19}$$

The first term in Eq. 8.18 only depends on the signal to thermal noise ratio at the receiver input, while the second one accounts for MUI. The proposed approach was demonstrated to guarantee high accuracy in estimating receiver performance for impulse-based transmissions, even in the presence of scarcely populated systems, systems with dominating interferers, or low-rate systems [23].

Table 8.1 Channel scenario characteristics

Path loss at d = 1 m	Path loss exponent
43.9 dB	1.79

Table 8.2 Simulation settings

Parameter	Setting
L	3
Number of nodes	20
Area	30 m × 30 m
Network topology	Random node positions
Channel model	See Eq. 8.2 and [21]
Application-layer user bit rate R	40 kb/s
Transmission rate	966 kb/s
Power	36.5 μW (FCC limit for Bandwidth = 0.5 GHz)
Packet traffic model	Poisson generation for data connection, Constant Bit Rate flow during connection, uniform distribution for destination node
Average connection interarrival time	From 100 to 6.25 s
DATA MACPDU length	824 bits (+208 bits for header and trailer)
Interference model	Pulse collision (see Sect. 8.3)
Physical layer settings	$N_s = 4$, $T_s = 258.8$ ns $T_m = 2$ ns, Reed Solomon (43, 51) FEC

8.4 Performance Analysis

The $(UWB)^2$ protocol described in Sect. 8.2.1 was tested by simulation. Simulation results were averaged over L different simulation runs. In each simulation run, N nodes were randomly located inside a square region with area A. Next, a realization of the channel impulse response, path loss, and delay spread was generated for each pair of nodes, with characteristics depending on the considered propagation scenario. These quantities were used by the interference module for introducing errors on the received packets, according to the MUI model described in Sect. 8.3.2. The CM1 scenario defined within IEEE 802.15.4a was considered during simulations, corresponding to indoor propagation in residential environments in Line-Of-Sight (LOS) conditions [21]. The corresponding settings for the path loss at a reference distance and path loss exponentare presented in Table 8.1.

During all simulations, the maximum size of 1288 bits was adopted for the PHYPDU. This value was obtained by considering as PHY payload a full size MACPDU of 129 bytes coded with a Reed Solomon (43, 51) Forward Error Correction code, in compliance with the specifications for the UWB PHY of the IEEE 802.15.4a standard [24]. The 1224 coded bits were then combined with a PHY synchronization trailer of length 64 bits, leading to a size of 1288 bits for each PHYPDU. Table 8.2 presents the main simulation settings.

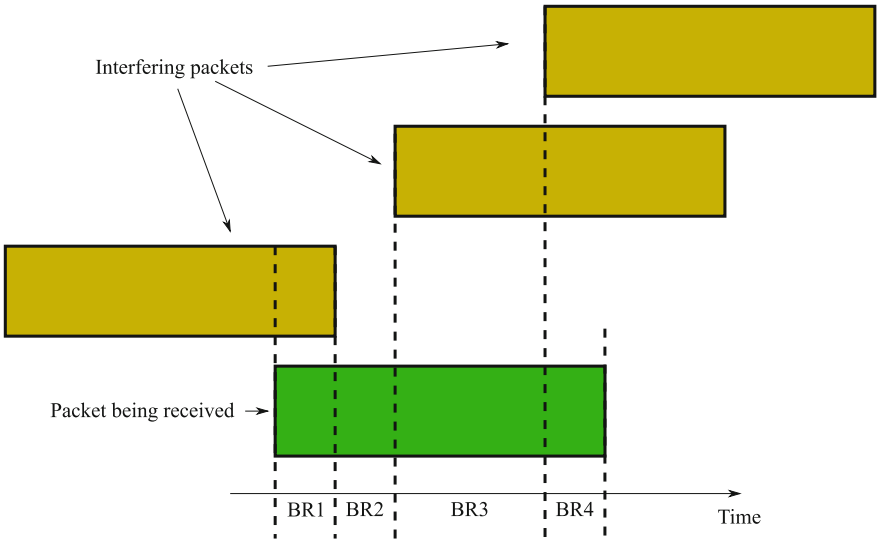

Fig. 8.4 Example of bit region identification during a packet reception in OMNeT++; 4 Bit Regions (BR1–BR4) are identified based on the variations in the set of interfering packets

8.4.1 Interference Modeling

The impact of interference was modeled building on an approach recently proposed in [25], that guarantees an accuracy significantly higher than previous solutions available in OMNeT++ frameworks, such as INET [26] and MixiM [27], in order to ensure a correct analysis of the impact of the proposed optimization on network performance. The approach in [25] keeps track of all transmitted packets and, for each packet reception, determines the interference level in every symbol period. Consecutive bits subject to the same interference are grouped into so called bit regions: Figure 8.4 shows an example of packet reception where four different regions are identified due to varying interference conditions. In [25] average Bit Error Probability (BEP) was then evaluated for each bit region according to the Standard Gaussian Approximation. In this work, oppositely, for each bit region the BEP was evaluated by adopting the Pulse Collision model, and the number of bit errors per region was randomly determined according to the BEP. Again according to [25] the total number of bit errors generated in the packet was compared with the maximum number of errors admitted for the packet as determined by the adoption of a Reed-Solomon code with a coding rate $RS_{rate} = 0.843$ (corresponding to the RS(51, 43) code proposed in the IEEE 802.15.4a standard) in order to decide if the packet is correctly received or discarded.

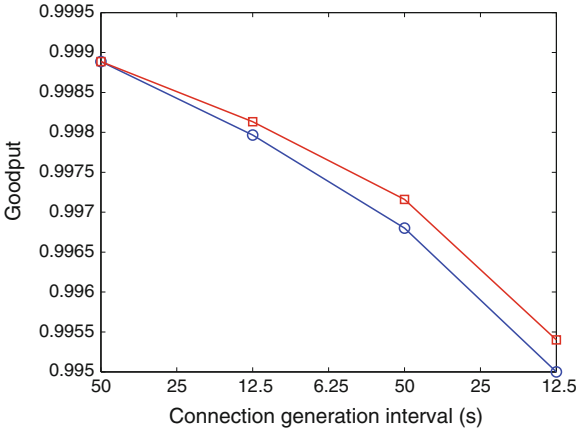

Fig. 8.5 Goodput as a function of the connection generation interval for Pure Aloha and Slotted Aloha access strategies (*Circles* Pure Aloha; *Squares* Slotted Aloha)

8.4.2 Simulation Results

Performance of $(UWB)^2$ was analyzed as a function of:

- application connection generation interval
- access strategy (pure vs. slotted).

As originally observed in [14], the comparison between pure and slotted Aloha was motivated by the fact that, as well known, in narrowband networks, slotted Aloha guarantees a higher (up to two times) throughput with respect to pure Aloha, thanks to a reduced probability of packet collision. The goal of this comparison was to verify whether this performance gap can be also observed in low bit rate UWB networks, where the negative impact of packet collisions is mitigated by the high processing gain.

Figure 8.5 presents the goodput, defined as the ratio between delivered packets and transmitted packets, for both Pure Aloha and Slotted Aloha access strategies as a function of the connection generation interval.

Figure 8.5 shows that in all cases goodput is greater than around 99.5 % for both Pure and Slotted Aloha. Figure 8.6 presents the delay measured in the same simulations. Figure 8.6 shows that the delay experienced by DATA PDUs for Pure Aloha is close to its minimum possible value, given by the DATA PDU transmission time. On the other hand, the additional delay in PDU transmission introduced by the slotted time axis leads to higher delays as shown in the same Fig. 8.6. These results are justified by the high robustness of IR-UWB to MUI, as shown by the comparison of Figs. 8.7 and 8.8, showing the percentage of packets subject to MUI and the Packet Error Rate (PER), respectively, for both Pure Aloha and Slotted Aloha. Results show that although the interference due to collisions increases with the offered traffic, until

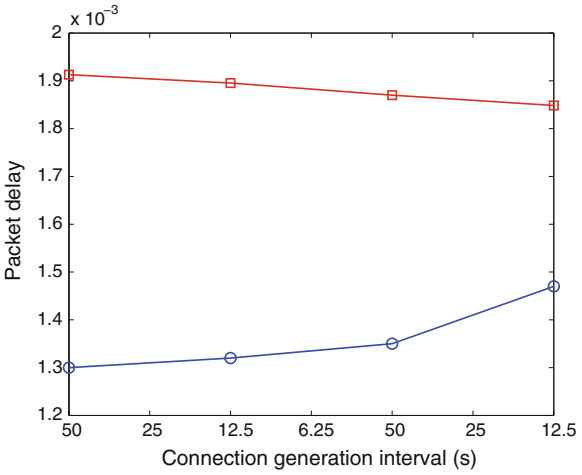

Fig. 8.6 Delay as a function of the connection generation interval for Pure Aloha and Slotted Aloha access strategies (*Circles* Pure Aloha; *Squares* Slotted Aloha)

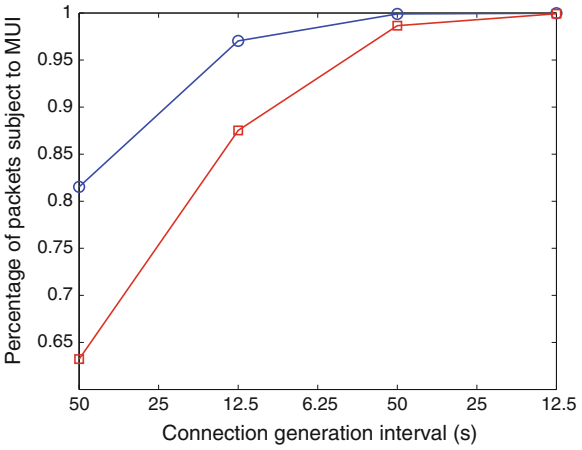

Fig. 8.7 Percentage of packets subject to MUI as a function of the connection generation interval for Pure Aloha and Slotted Aloha access strategies (*Circles* Pure Aloha; *Squares* Slotted Aloha)

almost 100 % of them are subject to MUI, the PER increases only slightly. In addition results confirms, as expected, that Slotted Aloha reduces the number of collisions.

It can be expected that, as the number of users and/or the offered traffic increases, the packet delay will increase faster in the case of the Pure Aloha mode, due to the higher maximum goodput guaranteed by the Slotted Aloha mode. However, in the conditions analysed in this work, that correspond to a significant traffic load for a low data rate network, the Pure Aloha mode is still the best solution, as it provides high

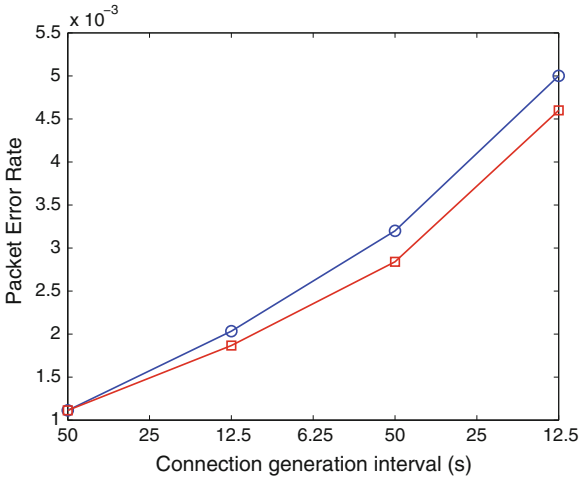

Fig. 8.8 Packet Error Rate as a function of the connection generation interval for Pure Aloha and Slotted Aloha access strategies (*Circles* Pure Aloha; *Squares* Slotted Aloha)

throughput while leading to lower delays. This confirms and reinforces the original findings presented in [12, 14].

8.5 MAC for Cognitive UWB Underlay Networks

The $(UWB)^2$ MAC was originally conceived taking into account the characteristics of TH-IR-UWB signals, with the goals of maximizing the performance of low rate UWB networks and providing support for the ranging feature. The evolution towards cognitive UWB networks to be deployed according the underlay approach calls however for a revision of this MAC protocol for IR-UWB networks, so to address the coexistence issues posed by the presence of other systems co-located with the UWB network. Two main research areas related to the further development of $(UWB)^2$ were identified, and will be discussed in the following subsections:

- Introduction of specific coexistence-related information in control MAC packets, to be piggybacked during LE/LC exchanges or distributed by means of dedicated broadcast packets.
- Support for novel features developed for IR-UWB technology, in particular with regards to the Time Reversal technique.

8.5.1 Information Exchanges Related to Coexistence

The LE/LC exchange part of the $(UWB)^2$ may be adapted so to support the exchange of coexistence-related information. Information elements that might be exchanged between transmitter and receiver include as an example sensing data and decisions related to the presence of devices belonging to other networks. Although the LE/LC exchange is a one-to-one exchange between transmitter and receiver, it is worth noting that it takes place on the common code that all devices can access at any time. A device may thus acquire information regarding all of its neighbours by extracting it from LE/LC packets exchanged by them, even if not directly involved in the exchange. In addition, in very low traffic scenarios, where LE/LC exchanges may not be frequent enough, broadcast packets may be added to the protocol definition.

A second aspect of $(UWB)^2$ that may be adapted to improve coexistence of low rate cognitive UWB networks with narrowband and wideband systems is the hybrid multi-code approach adopted in the MAC. It is well known that codes may be used to shape the spectrum of a UWB signal, allowing, to some extent, to introduce spectrum holes in specific frequency bands, guaranteeing improved coexistence with systems operating at those frequencies [28, 29]. This solution might be adopted in $(UWB)^2$, taking into account measurements taken by both transmitter and receiver in the selection at the transmitter side of the TH code that minimizes the interference to devices operating in the neighbourhood of the transmitter-receiver pair.

8.5.2 Support for IR-UWB Novel Features

Time Reversal (TR) is a coding technique that consists in pre-filter the transmitted signal with an impulse response consisting in a temporally inverted copy of the channel impulse response between transmitter and receiver [15]; the precoding filter compensates for temporal and spatial dispersion introduced by the channel, leading to temporal and spatial focusing of the signal. The TR technique has been adopted in the acoustics field in the last twenty years both for communications [15, 30] and positioning, in combination with Direction of Arrival estimation algorithms based on subspace decomposition [31, 32].

More recently, TR was proposed as a solution for UWB communication systems given the strong similarities between IR-UWB and acoustic signals [11, 16]. The adoption of TR may significantly increase the coexistence capabilities of UWB systems: on one hand, temporal correlation leads in fact to higher efficiency in the use of available power, guaranteeing the same received power at the intended receiver using a lower transmitted power; on the other, spatial focusing leads to higher energy in the point in space occupied by the intended receiver, thus further reducing the power in other locations. TR may thus help in meeting coexistence requirements in the deployment of cognitive UWB networks. In further analogy with acoustical signals, TR can also improve DOA estimation accuracy in UWB systems, as discussed in [33].

The effective introduction of TR in UWB networks depends on support by the MAC layer, that should provide the information required to implement it. An aspect in particular that requires the support of the MAC is related to the spatial focusing property previously described, and the relation between the topology of primary and secondary network. As mentioned before, spatial focusing leads to higher received power at the position occupied by the receiver; more in general, this will correspond to a higher power in a small region centered around the position of the intended receiver; a primary receiver in close proximity to a UWB reciever would thus potentially be subject to higher interference. In order to address this issue, an exclusion zone should be defined around the position of the intended receiver, and TR should only be adopted if no victim receivers lie within such exclusion zone. The MAC entity running at the receiver would thus be in charge of determining whether a potential victim receiver lies within the exclusion zone, and to communicate this information to the transmitter.

8.6 Conclusions

This book chapter focused on the analysis of MAC issues in cognitive underlay IR-UWB networks. Existing MAC protocols were reviewed, and the performance of the $(UWB)^2$ MAC, originally proposed in [12], was evaluated by means of computer simulations in the OMNeT++ simulation environment, with accurate real-time evaluation of Multi User Interference measured according to the Pulse Collision model described in [14]. Performance evaluation confirmed the suitability of the multi-channel Aloha approach proposed in [12] for low data rate UWB networks, thanks to the inherent robustness to MUI guaranteed by the impulsive nature of the IR-UWB signal.

The chapter focused next on discussing the adaptation of the $(UWB)^2$ MAC for operation of cognitive UWB underlay networks in coexistence-aware application scenarios. Both support for the exchange of data and coexistence-aware selection of MAC parameters and introduction of novel features recently proposed for the IR-UWB technology, in particular Time Reversal, were addressed. Research aspects identified as a result of this discussion will be the subject of future work.

Acknowledgments Part of this work was supported by COST Action IC0902 Cognitive Radio and Networking for Cooperative Coexistence of Heterogeneous Wireless Networks and by the ICT ACROPOLIS Network of Excellence, FP7 project n. 257626.

References

1. IEEE: IEEE 802.15.4 MAC standard. http://www.ieee.org/ (2006)
2. Kinney, P.: ZigBee technology: wireless control that simply works. http://www.zigbee.org/imwp/idms/popups/pop_download.asp?contentID=812 (2003)

3. IEEE: IEEE 802.15.TG4a official web page. http://www.ieee802.org/15/pub/TG4a.html (2007)
4. IEEE 802 part 15.4: Wireless medium access control (mac) and physical layer (phy) specifications for low-rate wireless personal area networks (wpans)—amendment 1: Add alternate phys. http://www.ieee.org/ (2007)
5. Di Benedetto, M.G., Giancola, G.: Understanding Ultra Wide Band Radio Fundamentals. Prentice Hall, NJ (2004)
6. De Nardis, L., Maggio, G.M.: Low Data Rate UWB Networks, pp. 315–339. Wiley, New York (2006)
7. Mitola, J., Maguire, G.Q.: Cognitive radio: making software radios more personal. IEEE Pers. Commun. **6**(4), 13–18 (1999)
8. Cuomo, F., Martello, C., Baiocchi, A., Capriotti, F.: Radio resource sharing for Ad Hoc networking with UWB. IEEE J. Sel. Areas Commun. **20**(9), 1722–1732 (2002)
9. Cuomo, F., Martello, C.: Improving wireless access control schemes via adaptive power regulation. 8th International Conference on Personal Wireless Communications, pp. 114–127. Venice, Italy (2003)
10. IEEE 802.15.3 MAC standard. http://www.ieee.org/
11. Abiodun, E.A., Qiu, R.C., N., G.: Demonstrating time reversal in ultra-wideband communications using time domain measurements. 51st International Instrumentation Symposium (2005)
12. Di Benedetto, M.G., De Nardis, L., Junk, M., Giancola, G.: $(UWB)^2$: Uncoordinated, Wireless, Baseborn medium access control for UWB communication networks. J. Mob. Netw. Appl. **10**(5), 663–674 (2005)
13. De Nardis, L., Giancola, G., Di Benedetto, M.G.: Performance analysis of uncoordinated medium access control in low data rate UWB networks. In: 1st IEEE/CreateNet International Workshop on "Ultrawideband Wireless Networking", within the 2nd International Conference on Broadband Networks (2005)
14. Di Benedetto, M.G., De Nardis, L., Giancola, G., Domenicali, D.: The Aloha access $(UWB)^2$ protocol revisited for IEEE 802.15.4a. ST J. Res. **4**(1), 131–142 (2007)
15. Fink, M.: Time-reversal waves and super resolution. In: Journal of Physics: Conference Series 124, 4th AIP International Conference and the 1st Congress of the IPIA (2008)
16. Xiao, S., Chen, J., Liu, X., Wang, B.Z.: Spatial focusing characteristics of time reversal uwb pulse transmission with different antenna arrays. Prog. Electromagn. Res. B **2**, 223–232 (2008)
17. De Nardis, L., Di Benedetto, M.G.: Medium access control design for UWB communication systems: review and trends. J. Commun. Netw. **5**(4), 386–393 (2003)
18. Sousa, E.S., Silvester, J.A.: Spreading code protocols for distributed spread-spectrum packet radio networks. IEEE Trans. Commun. **COM-36**(3), 272–281 (1988)
19. Garcia-Luna-Aceves, J.J., Raju, J.: Distributed assignment of codes for multihop packet-radio networks. In: IEEE Military Communications Conference, vol. 1, pp. 450–454 (1997)
20. IEEE 802.11 standard. http://www.ieee.org/
21. IEEE 802.15.4a channel model final report, rev.1 (november 2004).ftp://ieee: wireless@ftp.802wirelessworld.com/15/04/15-04-0662-00-004a-channel-model-final-report-r1.pdf (2004)
22. Di Benedetto, M.G., Giancola, G., Di Benedetto, M.D.: Introducing consciousness in UWB networks by hybrid modelling of admission control. Mob. Netw. Appl. **11**(4), 521–553 (2006)
23. Giancola, G., Di Benedetto, M.G.: A novel approach for estimating multi user interference in impulse radio UWB networks: the pulse collision model. Sig. Process. Spec. Issue Sig. Process. UWB Commun. **86**(9), 2185–2197 (2006)
24. Maggio, G.M.: 802.15.4a uwb-phy. Technical Report. IEEE 15–05-0707-01-004a, IEEE 802.15.4a (2005)
25. Akhtar, A.M., De Nardis, L., Nakhai, M.R., Holland, O., Di Benedetto, M.G., Aghvami, A.H.: Multi-hop cognitive radio networking through beamformed underlay secondary access. In: IEEE International Conference on Communications. Budapest, Hungary (2013)
26. The INET framework. http://inet.omnetpp.org/index.php?n=Main.HomePage
27. The MIXIM OMNeT++ modeling framework. http://mixim.sourceforge.net/

28. Sablatash, M.: Mitigation of interference by ultra wide band radio into other communication services: evolution to cognitive ultra wide band radio. In: Canadian Conference on Electrical and Computer Engineering, pp. 1345–1348 (2007)
29. Ohkuni, K., Hayasi, M., Kohno, R.: A study on interference mitigation method with spectrum shaping code in ds-uwb radar. In: 9th International Conference on Intelligent Transport Systems Telecommunications (ITST), pp. 239–242 (2009)
30. Derode, A., Roux, P., Fink, M.: Robust acoustic time reversal with high-order multiple scattering. Phys. Rev. Lett. **75**, 4206–4209 (1995)
31. Prada, C., Manneville, S., Spoliansky, D., Fink, M.: Decomposition of the time reversal operator: detection and selective focusing on two scatterers. J. Acoustic. Soc. Amer. **99**(4), 2067–2076 (1996)
32. Prada, C., Thomas, J.L.: Experimental subwavelength localization of scatterers by decomposition of time reversal operator interpreted as covariance matrix. J. Acoustic. Soc. Amer. **114**(1), 235–243 (2003)
33. De Nardis, L., Fiorina, J., Panaitopol, D., Di Benedetto, M.G.: Combining uwb with time reversal for improved communication and positioning. Springer Telecommun. Syst. (2011)

Chapter 9
Integration of Heterogeneous Spectrum Sensing Devices Towards Accurate REM Construction

Liljana Gavrilovska, Vladimir Atanasovski, Valentin Rakovic and Daniel Denkovski

Abstract This chapter introduces a recently developed generic REM construction architecture capable of integrating heterogeneous spectrum sensing devices, combining the spectrum sensing and the database approach for accurate radio environmental mapping. It elaborates on the required interfaces and data structures, architectural components and toolboxes for spectrum data collection, storage, processing and usage. Different spectrum sensing devices possess diverse sensing, processing and hardware capabilities in terms of sensitivity, data resolution, gains, sweeping time, processing power, antennas etc. This yields various practical implementation challenges in order to facilitate their integration into a single REM construction platform. The practical challenges vary from code compatibility and processing limitations up to device calibration. This chapter places a particular focus on the device calibration procedure as a quintessential part of the integration process and discusses in details its theoretical and practical aspects. Furthermore, the book chapter elaborates on a prototype implementation based on the developed REM architecture and several types of spectrum sensing devices: USRP2, SunSPOTs and TI eZ430 RF2500. All heterogeneous devices are upgraded with custom developed software for interfacing to the REM prototype and providing versatile spectrum measurement capabilities based on different energy detection techniques. The performances of the developed prototype and the gains of using a larger scale heterogeneous measurement platform for REM constitution are validated in terms of Radio Interference Field (RIF) estimation

L. Gavrilovska (✉) · V. Atanasovski · V. Rakovic · D. Denkovski
Faculty of Electrical Engineering and Information Technologies,
Ss. Cyril and Methodius University, Skopje, Macedonia
e-mail: liljana@feit.ukim.edu.mk

V. Atanasovski
e-mail: vladimir@feit.ukim.edu.mk

V. Rakovic
e-mail: valentin@feit.ukim.edu.mk

D. Denkovski
e-mail: danield@feit.ukim.edu.mk

M.-G. Di Benedetto and F. Bader (eds.), *Cognitive Communication and Cooperative HetNet Coexistence*, Signals and Communication Technology,
DOI: 10.1007/978-3-319-01402-9_9,

via spatial interpolation, source localization, propagation model estimation and statistical analysis of spectrum occupancy. All presented evaluations, discussions and conclusions stem from the authors' own practical work in the field.

9.1 Introduction

Wireless devices, applications and services attract an increased interest due to their ubiquitous and pervasive vision for anytime and anywhere connectivity. This imposes serious challenges in front of the spectrum management aspect of wireless networks resulting in a necessity for more spectrum resources and a necessity for more optimized usage of current spectrum resources. Many empirical studies [1, 2] show that there exists a handful of available spectrum for secondary use making the Dynamic Spectrum Access (DSA) [3, 4] and Cognitive Radio (CR) [3–6] preferred approaches towards more optimal and more efficient spectrum utilization in future wireless networks. However, recent advances in the field also show that the *available spectrum for secondary use* is time, location and scenario dependent [7]. Therefore, the appropriate solutions for more efficient spectrum management must encompass appropriate business models and scenarios that first and foremost protect the primary users in the network and, at the same time, satisfy the requirements of the secondaries in the most appropriate manner.

The answer to the previously raised problems lies in environment-aware, self-configurable and learning-capable techniques for optimization of wireless systems. These techniques inevitably yield a mindset shift from *theoretically possible*, but not quantifiable and tangible solutions, to real-world experimentation and operation tools with distinct potentials and benefits in real-world scenarios. This is where the notion of *Radio Environmental Maps (REMs)* finds its place and serves its envisioned purpose.

REMs were originally envisioned as a two-dimensional representation of the radio field strength [8]. Today, they are foreseen as a *rich hierarchical database or knowledge base* that stores various kinds of *radio environmental information*, which can be subsequently used for a plethora of optimization procedures in different secondary spectrum access scenarios [9–15]. The REM stored information can be either directly measured (empirical) or indirectly derived (through modeling) and classified as:

- *Static information*, e.g. locations of transmitters and/or receivers, terrain model etc. and
- *Dynamic information*, e.g. propagation environment, up-to-date spectrum measurements, users activity patterns etc.

As a result, REMs represent a powerful enabler and/or facilitator for reliable DSA and, more generally, for improving the environmental awareness and spectral efficiency of wireless networks. The range of REM beneficiaries is manifold:

- Regulators and dedicated public bodies, which can use the REMs for large-scale estimation of spectrum usage in order to track compliance to regulations, estimate frequency planning effectiveness etc.;
- Cellular network operators, which can interpret measured results through drive tests or mobile subscribers in order to perform Minimization of Drive Tests (MDTs) [16], network planning and fault detection etc. and
- Consumers, which can perform self-optimization and/or learning of patterns and habits leading to higher QoS, lower prices etc.

The focal point of REMs is their *ability* to *measure*, *collect* and *efficiently represent* radio environmental information. This subsequently represents a tremendous aid in the optimization of wireless networks technology chain. The measurement capability within REM solutions requires the usage of various market available and custom built devices capable of performing various, often user-defined and user-specific, measurements. The variety of measurement devices imposes their inevitable *heterogeneity* as a potential problem regarding their mutual operation, reliability and accuracy. Therefore, the integration of various spectrum sensing devices in a single REM solution is intertwined with *specific implementation problems* such as calibration, operating ranges, accuracy etc.

This chapter discusses the most important aspects of the integration of heterogeneous spectrum sensing devices in a single REM construction architecture. The analysis tackles the actual REM construction approach along with the associated problems stemming from the heterogeneity of the spectrum sensing devices within the REM architecture. It pinpoints the most optimal way, regarding architecture complexity and performance reliability, to accurate REM construction and the need for intelligent management of available spectrum sensing resources leading to increased reliability and accuracy of the constructed REM even with the usage of low-end devices.

9.2 How to Construct Accurate REM?

REMs represent a *technology enabler* for DSA. They are capable of increasing the level of radio environmental awareness, thus facilitating various resource management procedures within wireless networks. Their strongest asset is the ability to provide different accurate and up-to-date information on the radio environment to all interested stakeholders. Therefore, the REM construction approach must encompass a strong and unambiguous capability for reliable radio field estimation.

REMs operation requires a clearly defined and distinct feasibility of a certain practical DSA solution. REMs are enablers rather than solutions in wireless networks and must be treated as such under all possible circumstances. The feasibility of a certain DSA solution can be seen from a [1]:

- Technical side (i.e. realistic assumptions of the behavior of all network entities such as primaries, secondaries, transmitters, receivers etc. and their mutual interactions such as interference, blocking etc.) and
- Commercial side (i.e. quantification of the available spectrum usage and development of appropriate business models and scenarios).

In any case, the feasibility of a DSA solution defines the *requirements* that must be met by the enabling REM. Both technical and commercial aspects of the feasibility directly influence the design and the applicability of the REM. Having these boundaries in mind, the crucial question now becomes the *design of an efficient, practically deployable and operational REM*, which must be capable of monitoring and interpreting the REM-related information in the most optimal way.

The following section discusses the possible approaches towards accurate REM construction. It pinpoints their advantages and disadvantages and elaborates on the most important REM architecture design concepts.

9.2.1 Possible Approaches

The focal aspect of every REM design is the capability to accurately assess the spectrum occupancy in order to accurately decide on possible spectrum opportunities. As a result, the spectrum opportunity detection techniques directly influence and dictate the possible REM construction approaches. Currently, there are two distinct spectrum opportunity detection techniques [7], i.e.:

- *Sensing-based* spectrum opportunity detection and
- *Database-based* spectrum opportunity detection.

The sensing-based technique requires that secondary devices employ a certain detection method in order to assess the monitored spectrum. It is a device-centric method not dependent on any network infrastructure and information exchanges with the network. The secondary devices independently or cooperatively decide upon spectrum opportunities using detection methods ranging from blind to signal specific [17]. However, most practical detection techniques usually rely on blind techniques, specifically energy detection and its derivatives (e.g. FFT (Fast Fourier Transform) Averaging Ratio—FAR [18], Higher-Order-Statistics—HOS [19] etc.) due to their simplicity for practical deployments and real-time operation.

The database-based technique requires that secondary wireless devices report their *location* to a *centralized database* and obtain *available spectrum information*. This is a network-centric method where all related calculations are performed within the centralized database framework residing in the network infrastructure. The secondary devices do not need a sensing capability in this case.

Possible REM construction approaches directly stem from the spectrum opportunity detection techniques. Therefore, it is necessary to carefully analyze and scrutinize their advantages and disadvantages in order to find the optimal way to construct a

Table 9.1 Comparison between sensing-based and database-based opportunity detection techniques

	Sensing-based	Database-based
Implementation complexity	Variable	High
Location of detection decision	Distributed	Centralized
Suitable for	Dynamic radio environments	Static radio environments
Advantages	Plenty of market available devices	Robustness
Disadvantages	Heterogeneity (requires calibration); Inability to control the aggregate interference	Obsolete data for dynamic radio environments
Possibility for practical deployments	Yes	Yes

practically feasible REM solution. For instance, the major drawbacks of the sensing-based techniques lie in the heterogeneity of the market available devices for spectrum sensing and the inability to control the aggregate interference. The heterogeneity is reflected in a variety of possible devices in different price and complexity ranges leading to necessary trade-offs with the corresponding solutions reliability and accuracy (i.e. the higher the price the more accurate and reliable the spectrum sensing solution will likely be). These issues can be partly overcome with accurate calibration of the various devices and even turn the drawback into a practical advantage for applications with limited resources on the sensing side [10]. Additionally, there are complex sensing methods found in the literature that significantly increase the precision and the reliability, but these methods often require offline calculations limiting their practical real-time operation in an integrated REM solution. However, the inability to control the aggregate interference using only spectrum sensing methods is the major limiting factor for practical deployments. Therefore, the regulators prefer the database-based technique [20], which is more robust and rugged completely stripping the spectrum sensing requirement. Table 9.1 compares the major characteristics of the sensing-based and the database-based spectrum opportunity detection techniques.

It is evident that both techniques (and REM construction approaches) exhibit advantages under certain circumstances. The sensing-based technique is more flexible and dynamic providing real-time radio environmental information even for low cost devices (with appropriate calibration). Additionally, most of the wireless devices on the market inherently possess the capability to perform spectrum sensing by measuring the RSS values. The database-based technique can provide accurate and reliable information for static radio environments (e.g. TVWS—TV white spaces) and provide extensive set of spectrum related information real-time (e.g. historical spectrum occupancy, locations of primary static transmitters and receivers etc.).

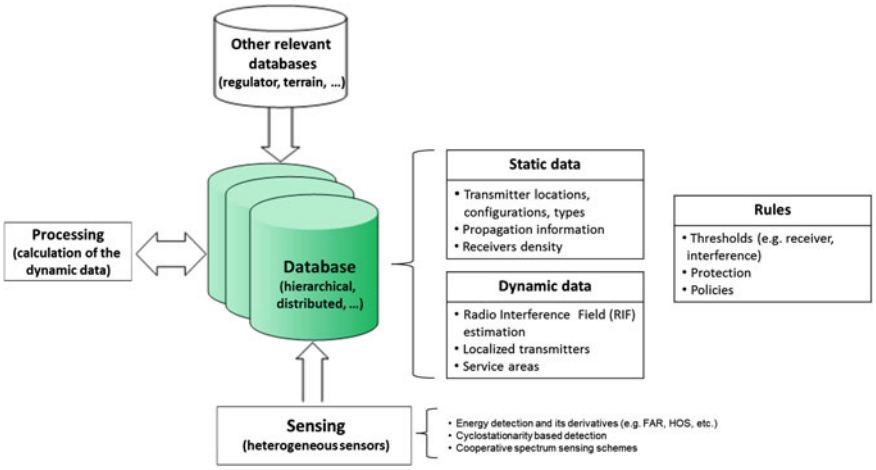

Fig. 9.1 Generic REM construction approach

It is clear that the *optimal way* towards accurate REM construction should embrace the *benefits of both approaches* [2]. The *synergy* between them allows using their *complementarity* in order to derive a practically deployable and accurate REM architecture for various applications [9, 10]. Figure 9.1 depicts a possible REM construction approach based on the joint sensing-based and database-based detections combined in a single architecture.

More details on the actual data model and the necessary architectural components descriptions are given in the following text.

9.2.2 Architectural Cornerstones

The REM architecture design, besides the functional architectural blocks, requires appropriate *data model* capable of capturing all possible parameters of interest in various scenarios of subsequent REM usage. The data model allows the REM to serve its purpose of technology enabler for DSA and must be open for future extensions. The minimal data model set for practical REM development comprises:

- Transmitter locations, configurations and types;
- Radio Interference Fields (RIFs);
- Service areas;
- Receiver information;
- Propagation information;
- Activity/usage information;
- Received signal strength distributions.

This information should be sufficient for real-time and offline calculations of the most important radio environment parameters for majority of dynamic and static DSA solutions. However, the REM architecture must be easily extensible in terms of new data of interest for a future particular scenario.

The architectural design essentially incorporates two major architectural components, i.e.:

- *REM backend* and
- *REM users*.

The REM backend is responsible for acquiring, storing and processing of radio environmental information important for subsequent usage by the REM users. It comprises:

- *Sensing Devices (SDs)* that perform spectrum measurements using various techniques (e.g. blind detection, signal specific detection, cooperative detection etc.);
- *Spectrum Data Server(s) (SDS)* that store the measured and the processed REM information and
- *REM Processing Center (RPC)* that calculates the REM and other various important radio environmental parameters. The most important features of the RPC are data fusion, spatial interpolation of sparse RSS measurements, statistical analysis of the radio environment including propagation estimation, localization of transmitters and basic RRM (Radio Resource Manager) functionalities (this list is not exhaustive and should be easily updated in future if needed).

The REM users represent various entities that benefit from the radio environmental information. These can be RRM blocks within a DSA solution, policy manager, GUIs for REM visualization and other entities using the radio environment information for some optimization procedure. The REM users are diverse and the REM architecture must be capable of closely following their requirements in various scenarios of interest.

There are also other important aspects towards design and development of REM construction architecture. All previously mentioned architectural blocks must communicate using *transparent interfaces* and *corresponding protocols*. Also, the RPC must be easily extendable in order to accommodate future possible REM needed calculations. These aspects ensure that the REM architecture is *scalable*, *modular* and *easily extensible* towards future envisioned usage.

9.2.3 Integration

Figure 9.2 depicts the integration of all previously discussed aspects of REM construction in a single and unified architecture [9, 10]. It clearly shows the synergy between the sensing-based approach (using heterogeneous SDs) and the database-based approach (using a centralized storage).

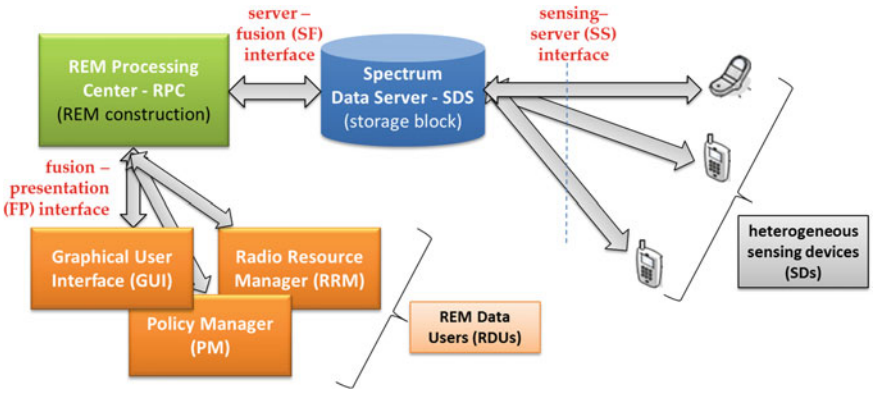

Fig. 9.2 General REM construction architecture

The most important features of a general REM construction architecture can be summarized as heterogeneity of sensing devices, modularity and scalability/extensibility. The following sections will discuss in more details these features and show how to integrate heterogeneous market available devices in the relevant REM sensing pool while providing reliable and accurate management of the REM information.

9.3 Heterogeneity in the Spectrum Sensing Pool

The SDs represent the ensemble of network entities that are responsible for and contribute to REMs with active spectrum monitoring functionalities. They are the cornerstone of the REM architecture, acquiring the measurements information from the surrounding environment and hence, facilitating the up-to-date tracking and surveillance of various radio environment events. Such events can refer to dynamic changes in the propagation phenomena, appearance of transmitters in the targeted area and tracking of their respective service areas, monitoring the changes of the signal strength of respective radio emissions, as well as the imposed (aggregate) interference in the measurement area.

With respect to the REM data model and architectural requirements the most important tasks of the SDs in the REM context are the signals detection, identification and classification. Tightly related to their hardware capabilities, the spectrum sensing devices can perform blind or feature based signal detection [17]. While the blind detection is preferable due to the low implementation complexity and feasibility to most of the radio hardware, the feature based detection implies the need of more complex radio hardware/software. Furthermore, the feature based detection requires the prior knowledge on the underlying signal models. This requirement cannot always be complied in practice, especially in dynamic and diverse (in terms of lower layer

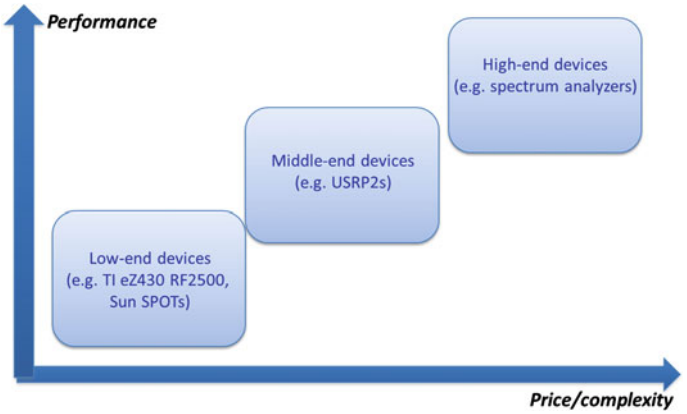

Fig. 9.3 Classification of sensing devices and their price/performance tradeoff

implementations) radio networks deployments, such as the cognitive radio networks. In such deployments, besides the signal detection capabilities, the sensing devices can/should provide additional signal identification and classification capabilities. The measured signals should be associated with the specific radio technology, activity times, as well as other signal specific metrics, such as RSS, Signal-to-Noise (and/or Interference) Ratios etc. All these sensing inputs can be of a great importance towards the accurate data processing and construction of different types of REMs.

9.3.1 Classification of Heterogeneous Sensing Devices

The spectrum sensing process can often combine multiple and heterogeneous types of sensing hardware involved in the REMs construction. The REM architecture envisions and allows for the usage of the measurement capabilities of existing infrastructure components such as base stations, terminals etc., or a dedicated sensor network. These different types of SDs should be seamlessly integrated into the REM architecture, i.e. the referred should support heterogeneous SD deployments. However, the different sensing devices posses diverse sensing, processing and hardware capabilities, in terms of *sensitivity, dynamic range, time/frequency versatility, data resolution, gains, sweeping time, processing power, antenna(s), measurement metrics* etc. Due to the diversity of the available hardware and its capabilities, the implementation of heterogeneous sensing devices into a single REM statistical-inference platform is far from a trivial process and implies many practical implementation challenges: device interfacing, time/frequency/power calibration, data formats unification etc.

With respect to the sensing/processing capabilities and the price, the market available measurement devices can be classified as high-, mid- and low-end devices [10] (Fig. 9.3, Table 9.2). The high-end devices, i.e. the signal/spectrum analyzers, offer

Table 9.2 Classification of sensing devices and their performances

	High-end	Mid-end	Low-end
Examples	Signal/spectrum analyzers, base stations etc.	Software Defined Radios, mobile terminals etc.	Low price, light weight wireless sensors etc.
Price	High (10 k euros and above)	Medium (1 k up to 10 k euros)	Low (up to 1 k euros)
Time/frequency/power resolution/versatility	High	Medium/high	Low
Sensing performances (accuracy, reliability, dynamic range, sensitivity)	High	Medium	Low
On-board processing capabilities and memory	High	Medium	Low
Sensing software	Available	Available/custom made	Custom made
Detection capabilities	IQ detection	IQ detection possible	Energy detection
Power consumption	High	Medium	Low
Mobility	Low	Medium	High
Large scale deployments possibility	Low	Medium	High

the highest performance and capabilities at the highest price. Besides the basic energy detection capabilities, they are also capable to perform IQ (In-phase Quadrature) detection of the emitted signals, and hence, feature detection, identification and classification techniques are also feasible using these radio devices. The mid-end sensing devices (Software Defined Radios, mobile terminals, etc.) are a trade-off between price and performance while offering a set of advanced measurement capabilities. Although they usually possess lower sensitivity and lower time/frequency/power resolutions compared to the high-end devices, they can also perform IQ detection, and therefore, can be viable solutions for the implementation of more advanced feature detection and identification techniques. The low-end devices, (e.g. low cost sensors) represent low-price spectrum sensing solutions. They possess limited sensing capabilities and often require custom development of sensing software to extract the Received Signal Strength (Indicator), RSS(I), as a common metric. Therefore, the only feasible REM techniques are the energy based ones. However, due to the low price of the devices, large-scale sensor deployments are possible allowing for experimentation in the area of cooperative energy detection, RSS-based single and multi-source localization etc.

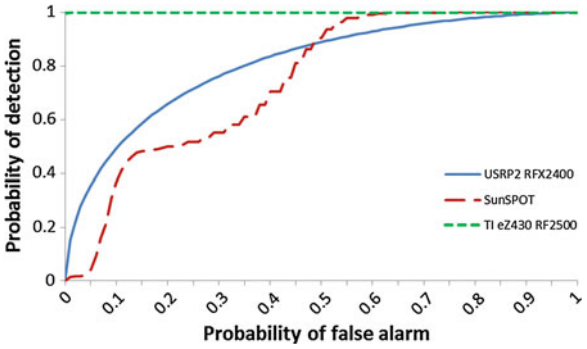

Fig. 9.4 ROC comparison for the different SDs and input power of −96 dBm. SunSPOT and TI eZ430 RF2500 perform 100 samples averaging in dBm domain, while USRP2 performs 100 samples averaging in mW domain

The heterogeneity of the market available devices represents both a challenge and an advantage in a variety of scenarios. The challenge lies in the necessity for calibration using a benchmark device, while the advantage lies in the flexibility for small-scale and large-scale practical deployments depending on the scenarios needs.

9.3.2 Implemented Heterogeneous Sensing Devices

The practical realization (prototype) [9, 10] of the generic REM architecture comprises several sensing device solutions, i.e. a high-end Anritsu MS2690A signal analyzer [21], mid-end USRP2 SDR devices [22] and two low-end sensing devices: Texas Instruments (TI) eZ430 RF2500 [23] and Sun SPOTs sensors [24]. In order to be adapted to the envisioned REM evaluation applications, the devices sensing performances need to be extensively assessed and an appropriate input/output power calibration is required [10].

The calibration process involves generation and reception of various signals with different power levels and known constant amplitudes by both a high-end spectrum analyzer and the measurement device. The Anritsu MS2690A signal analyzer can be used as a signal generator as well, emitting signals with different waveforms and power levels. In addition to the high sensing precision and reliability, this provides that the Anritsu MS2690A can be used as a reference for the calibration of the mid- and low-end spectrum sensing devices.

Figure 9.4 plots the ROC curves for all inspected mid- and low-end sensing devices for an input power of −96 dBm [10]. The low cost devices use their optimized sensing cases, i.e. the SunSPOT and TI eZ430 RF2500 perform 100 samples averaging in dBm domain. The best detection performances are offered by the TI eZ430 RF2500 spectrum sensor, once again proving that the referred has a high sensitivity when

used in proper manner. However, it should be noted that the USRP2 did not use the highest gain setting, while the other two devices were optimized for sensitivity. When the USRP2 uses the highest gain setting it would most probably provide the best detection performances [25].

Table 9.3 summarizes the hardware capabilities of the implemented low- and mid-end SDs in the REM prototype, compared to the reference Anritsu MS2690A, serving as a rationale for the obtained sensing performances.

The REM prototype, developed based on the generic REM architecture, can be further extended with additional sensing devices. Legacy radio hardware such as mobile terminals, laptop computers etc., can be also incorporated in the REM prototype providing large-scale deployment possibilities. However, the SDs sensing performances, accuracy and reliability need to be extensively assessed and an appropriate device calibration is required to obtain reliable and accurate REMs.

9.4 Management of REM Information

The REM backend framework is responsible for gathering, storing and processing of the REM information. The REM backend is also responsible for providing REM information to the REM users (Fig. 9.5). This allows various instantiations of the REM concept within different scenarios depending on the actual REM user profile. This section describes the implementation aspects of all backend components and the different interfaces, which enable the management of REM information and form the cornerstone of the REM concept.

As already introduced in Sect. 9.2, the REM backend framework comprises of three main components (Fig. 9.5), i.e. Spectrum Data Server(s) (SDS), Sensing Devices (SDs) and a REM Processing Center (RPC). This section provides information and details on the REM information management process regarding the three REM backend *components* and their interconnecting *interfaces*.

9.4.1 REM Backend Components and Interfaces

This section describes the REM information management based on the implementation of the SDs, the SDS and the RPC along with their associated functionalities and features.

9.4.2 Spectrum Data Server (SDS)

The Spectrum Data Server is the main component in charge of storing and managing the raw and the processed REM information as well as the active SD information. The storage database and its structure are depicted in Fig. 9.6, and consist of the following tables:

Table 9.3 Sensing device hardware capabilities

Performance metric	Anritsu MS2690A	USRP2	TI eZ430 RF2500	Sun SPOT
Frequency bands	50 Hz–6 GHz	Daughterboard dependent	2.4–2.485 GHz	2.253–2.74 GHz
Resolution bandwidths	30 Hz–20 MHz	195 kHz–25 MHz	58 kHz–812.5 KHz	2 MHz (IEEE 802.15.4 compliant)
Frequency steps	Various	Various	25 kHz–405 kHz	1 MHz
Frequency switching delays	Negligible	$<200\,\mu s$	$<809\,\mu s$	1 ms
Sampling type	IQ	IQ	RSSI	RSSI
Sampling period	>15 ns (bandwidth dependent)	>40 ns (decimation dependent)	<310 s (bandwidth dependent)	$128\,\mu s$
Sensitivity	−155 dBm/Hz	−164 dBm/Hz (for RFX2400)	−83 dBm (812 kHz); −104 dBm (203kHz)	−91 dBm (2 MHz)
Data resolution	32 bits (for each, I and Q)	14 bits (for each, I and Q)	8 bits	8 bits
On-board processing capabilities	1.5 GHz Intel Celeron M	50 MHz 32bit RISC CPU	16 MHz 16bit RISC CPU	180 MHz 32bit ARM920T
On-board memory	1 GB RAM memory	1 MB SRAM	1 K RAM/ 32 K ROM	512 K RAM/4 M Flash
Spurious Free Dynamic Range (SFDR)	88.5 dBc (at 10 kHz RBW)	88 dB (with manual gain control)	N/A	N/A
Power consumption	Not an issue	Not an issue	400 nA (Deep sleep)/ 13.3 mA (Active)	$33\,\mu A$ (Deep sleep)/ 104 mA (Active)

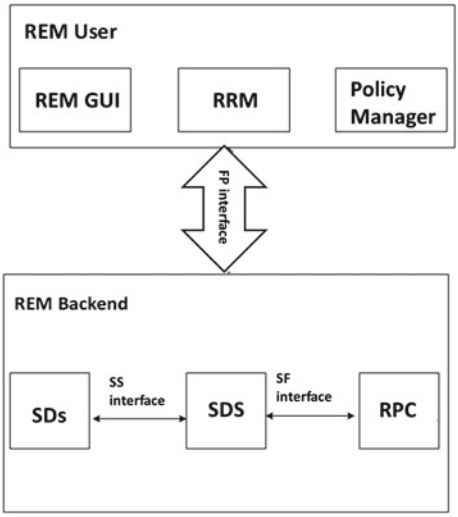

Fig. 9.5 REM backend architecture

- The *SDs* table (i.e. *Sniffers*) is used to keep sensor information such as SD addresses (IP and MAC), SD location (the current location of the SD expressed in Cartesian coordinate system), SD status (Active, Idle or Off) and SD capabilities (referring

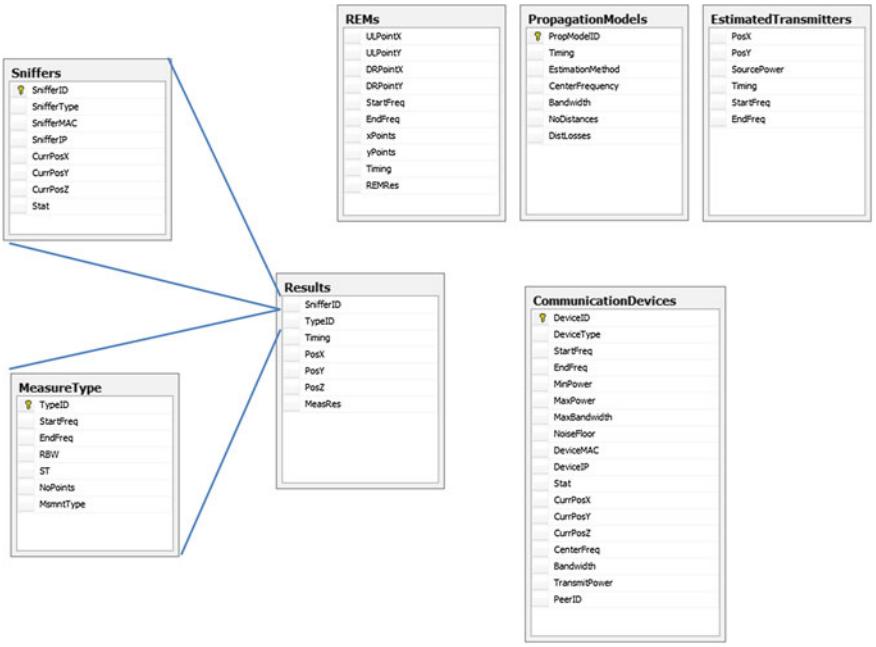

Fig. 9.6 SDS data storage structure

to the different types of SDs, i.e. USRP2, TI eZ430, SCALDIO). This table carries valuable information enabling the REM backend to be aware of and track the active SDs in the network.

- The *Measure type* table is used to store the measurement types that can be performed by the sensor devices specified by start and end frequency, resolution bandwidth, number of measurement points, sweep time and measurement type (referring to the min, mean and max hold detectors). Each SD is associated to an entry in this table corresponding to its current measurement configuration.
- The *Results* table stores the raw measurement results as Binary Large Object (BLOB) data from the active SDs in the network. The BLOB data approach enables the database to store large quantities of measurement information without affecting the overall performance of the SDS component. Besides the measurement data, this table also stores information about the location (useful for the case of Radio Interference Field construction) and time the data was collected, as well as the SD and measurement configuration that were used to perform the given measurement.
- The *REMs* table stores pixel images of the Radio Interference Field Estimation (RIFE) maps for a given frequency range and time. Moreover, the RIFE maps are also stored as BLOBs in the SDS in order to minimize the quantity of the stored data.
- The *Propagation Models* table is used to store propagation model information of the vicinity. The propagation model information comprises of information data such as center frequency (carrier frequency for the given propagation model), bandwidth, type of the model (static (i.e. predefined) or estimated (i.e. estimated by the SDs)), the number of distances i.e. points on which the propagation model is valid and the results (defined as the pathloss in dependence of the distance) stored as BLOB data.
- The *Estimated Transmitters* table stores information regarding the active transmitting devices in the area regarding their location (expressed in Cartesian coordinate system), operating frequency (i.e. frequency range), time (time at which they were detected as active) as well as the estimated transmit power. The detection method of the transmitters can be either a predefined administrator input (e.g. information of TV broadcast towers) or estimated input using the SDs to detect the active transmitter (e.g. a self-organizing networks use case, like the Femtocell scenario).
- The *Communication Devices* table stores information regarding the active network devices that utilize the stored REM information, i.e. the active REM users regarding their addresses (IP and MAC), location (the current location of the device expressed in Cartesian coordinate system), capabilities (frequency range, max bandwidth, noise floor, minimal and maximal transmit power), current configuration (current transmit power and bandwidth), status (idle, receiver, transmitter, transceiver), device type (LTE femtocell, LTE TWVS) etc. Similar to the SDs, this table facilitates the REM backend to be aware and serve the active REM users in the network.

The presented SDS structure only reflects the generic data storage format and architecture regarding the REM information management. The SDS is considered

to be modular and can be either upgraded with additional information fields and tables or demoted by excluding some of the generic fields and tables depending on the specific requirements of the underlying use-cases and scenarios. For example, in scenarios where the database storage has low memory requirements it is essential to construct the SDS by utilizing only a subset of all information fields in order not to overload the database storage capacity.

9.4.3 Sensing Devices (SDs)

The REM backend supports several different types of SDs (like, USRP2, TI eZ430, SCALDIO, etc.), which were presented and elaborated in more details in the previous section. Regarding the REM information management, the SDs are developed to support multiple functionalities like device *registration/deregistration* to the SDS, real-time periodic and triggered *spectrum measurements* as well as remote and on-the-fly *reconfiguration*.

All of the SDs, regardless of their type, exploit the same unified protocol and interface between the SD-SDS in order to exchange the required information with the SDS. The SD-SDS protocol, i.e. the SS interface, defines the communication over the referred interface including several protocol messages: *registration message, measurement report message, reconfiguration message, measurement request message, fusion report message and de-registration message*. Upon activation, the SDS address and TCP port are provided to the SD in order to create a TCP connection between them. Moreover, the initial configuration parameters of the SD are specified at the startup. After the initialization, the SD registers to the SDS by sending *registration message* containing information about the *SDs type* (e.g. SCALDIO or USRP2, mobile terminal etc.), *SDs address* (i.e. MAC and IP), *measurement parameters* (e.g. detection type, measurement mode, resolution bandwidth, start frequency, end frequency, sweep time, etc.) (Fig. 9.7). If the SD entry already exists in the SDS, then it is updated with the renewed information contained in the last registration message. If not, a new SD entry is created. While active, the SDs upload dynamic information in the SDS using the *measurement report message*, as described in Fig. 9.7. The SDS can send a *reconfiguration message* at any time during the operation mode of the SD (Fig. 9.7). The RPC can initiate sending of this message when the actual SDs configurations cannot support the required data processing tasks (e.g. the band of interest differs in the RPC from the one that the SDs are monitoring). In the same fashion the RPC can trigger a *measurement request message*, which initiates an asynchronous measurement process of the SDs. Every SD is obligated to de-register from the SDS before going offline using the *deregistration message*. This action is required so the SDS can be aware of the active SDs set in the network.

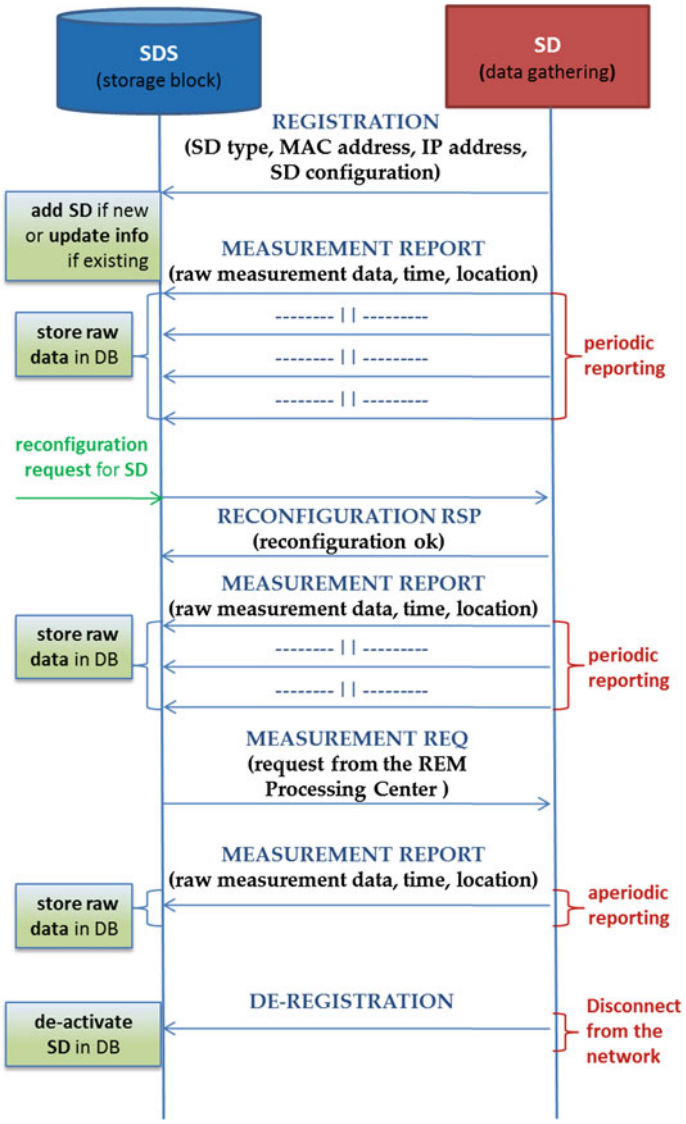

Fig. 9.7 Message sequence chart (MSC) of the SD-SDS protocol operation

9.4.4 REM Processing Center (RPC)

The main goal of the RPC is to perform the required processing tasks in the REM backend (e.g. interpolation and RIF Estimation, node localization, propagation model estimation, radio environmental statistics etc.). It is designed to be scalable and

extensible due to the need of different processing modules, which depend on the underlying scenario and use case. The prototype RPC implementation includes three processing modules (also denoted as toolboxes):

- *Spatial interpolation toolbox*. The spatial interpolation toolbox is in charge for the data fusion and construction of the RIFE maps. It utilizes the aspects of spatial interpolation, like pure IDW or IDW modified Shepards method techniques [26, 27] and other more complex techniques such as Kriging [28], radial basis function [29], polyharmonic spline, thin plate spline [30], triangulated irregular network, Gradient plus Inverse Distance Squared [31] etc. The toolbox is built on two generic processing functions:
  - Interpolate function—calculates the RIFE map and uses multiple input parameters in terms of the size of the interpolation area, number of interpolation points, fusion algorithm parameters (where the fusion algorithm parameters reflect the characteristics of the measured radio environment).
  - Interpolate point function—calculates the RIF output at a single interpolation point with known Cartesian coordinates. The Interpolate function utilizes the output parameters of this function in order to calculate the complete RIFE map.

  The spatial interpolation toolbox can select the interpolation technique of interest based on the requirements of the given scenario. For example, if the scenario requires highly precise RIFE maps, regardless of the real-time operation of the RPC and number of SDs, the toolbox can utilize the Kriging approach (e.g. a *Minimisation of Drive Tests* [16] scenario). In scenarios where the SDs are sparse and scattered and real-time operation of the RPC is of crucial importance, the best solution is to exploit the IDW approaches (e.g. a *Self Organizing Femtocell scenario*).
- *Statistical analyses toolbox*. This toolbox processes radio environmental statistical information such as the propagation model estimation, the empirical probability models (empirical probability density functions—epdf and cumulative density functions—ecdf), duty cycles (i.e. historical spectrum occupancy) etc. The toolbox utilizes multiple processing functions in order to process the required statistical information:
  - Propagation model function—estimates the path loss model in the area of interest. It returns the estimation of the pathloss as a function of the distance. The function utilizes the duration of the estimation period, center frequency and channel bandwidth as input parameters in the estimation process.
  - Empirical density function—calculates the empirical probability models i.e. the epdf or the ecdf function depending on the request.
  - Duty cycle function—calculates the duty cycle of the radio environment at a given time and location of interest.

  The statistical analyses toolbox can be easily upgraded with additional functionalities which can give better insight of activity/usage of the radio environment

Table 9.4 RPC interface functions

Name	Description
Get active SDs set	Returns the active SD set and their configuration and capabilities
Extract measurement data form single SD	Returns measurement data entry or entries from a single SD
Extract measurement data form multiple SD	Returns measurement data entry or entries from all active SD
Extract specific raw measurement data	Returns measurement data entry or entries from all active SD regarding a specific frequency range, location area and time period
Reconfigure SD	Reconfigures a given SD for a specified SD ID, with specified parameters for detection type, measurement mode, resolution bandwidth, start frequency and end frequency, sweep, time, etc.
Request measurements	Request measurements for a specified SD or group of SDs

in terms of its spatio-temporal and spectral statistics (e.g. Semi-Markov On-Off periods, activity models for certain bands of interest).

- *Transmitter localization toolbox*. The transmitter localization toolbox is responsible for performing an estimation of the transmitter location and estimation of its transmit power. The toolbox comprises of one generic processing function which returns information about the location of the transmitter (in Cartesian coordinate system) as well as its estimated transmit power based on variety of algorithms, which can perform single-source as well as multi-source localization [32, 33].

The RPC modules are connected to external interfaces, i.e. SDS-RPC interface, RPC-RDU interface, in order to communicate with the other architectural blocks of the REM. The communication between the SDS and the RPC is defined by the SDS-RPC protocol i.e. the SF interface. Since the referred entities are collocated together, their interactions are defined by several interface functions initiated by the RPC, Table 9.4.

The execution of these procedures can be triggered by the RDU, when a specific calculation or data processing is required. For example, if a REM calculation is required at a location area, in a specified time period and frequency band, the active SDs data is filtered within the specified boundaries using the *extract specific raw measurement data* function and is combined using some of the interpolation methods mentioned above. Whenever a required calculation is not supported because of insufficient amount of data a reconfiguration of the active SDs can be initiated by the RPC calling the *reconfigure SD* function. An example of a set of interactions considering the RPC is illustrated on the message sequence chart in Fig. 9.8, in the following section.

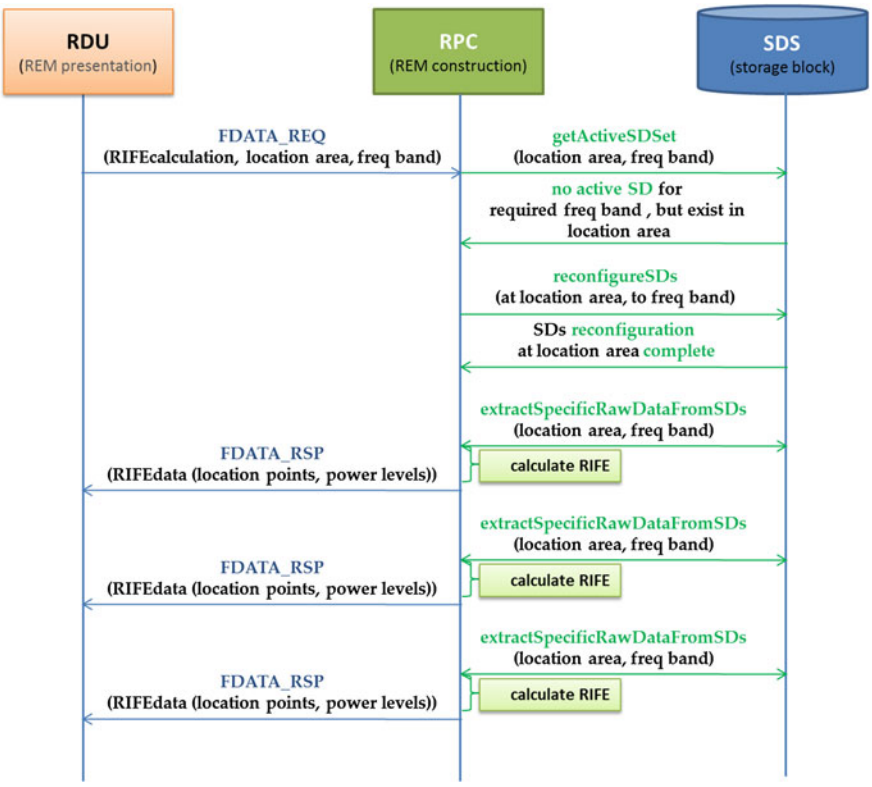

Fig. 9.8 Message sequence chart (MSC) illustrating the operation of SDS-RPC protocol as well as RPC-RDU protocol

9.4.5 REM Users and Interfaces

The REM users represent various cognitive and non-cognitive network entities capable of exploiting the REM information in order to increase the secondary system performance. These entities can vary from network management nodes such as RRM blocks within a DSA solution [9, 10] up to spectrum policy managers. The REM users can also be exemplified as visual representation tools of the REM information (i.e. GUIs for REM visualization). As depicted in Fig. 9.2, the RDU has a direct communication only with the RPC via the RPC-RDU protocol, i.e. the FP interface. For the most simplistic case, when the RDU is represented with a REM GUI, the RPC-RDU protocol is defined with two messages that are used for exchanging the information about the fused (processed) data sent over the FP interface, i.e. the *FData_req* and *FData_rsp* messages. As depicted in Fig. 9.8, the RDU requests the fused data, i.e. RIFE map from the RPC by sending the *FData_req* message. After receiving the *FData_req* message, the RPC starts to calculate the requested data

based on the parameters in the message. When the calculation is finished the RPC sends the processed data to the RDU in the *FData_rsp* message.

The *FData_req* message carries parameters that are needed by the RPC for the fusion process, such as start and end frequency (i.e. the frequency band), the location area for which the REM is being used. The last information carried by *FData_req* message is about the nature of the requested REM. For example, the give RDU can request a real time REM (in this case a real time RIFE map) or a long term RIFE that has been processed for a larger period of time (hours, days, etc.). The scenario of message exchange in order to provide the required RIFE map derivation is presented on the MSC in Fig. 9.8. The notion from the example can be extended to more generic cases when the requested data can be a structure of REM information (e.g. RIFE map + Transmitter localization + epdf, etc.).

This section provided a high level overview of the REM information management via the REM backend architecture. Moreover, the section pinpointed the implementation facets of the REM backend components and interfaces and discussed about its generic and flexible nature. This generic and flexible nature enables the REM backend (as well as the REM information management) to be further developed and enhanced regarding the specific considerations that arise from the underlying use-cases and scenarios.

9.5 Conclusions

Future wireless networks inevitably yield increased network performance and more optimized usage of available network resources due to the ever-increasing users needs. The problem of practical DSA and CR deployments is the limitation in terms of available spectrum dependency on time, location and scenario and lack of clearly defined commercial business model and scenario. The REM technology proves to be a viable solution towards embodying *self-x characteristics* in future wireless networks. The *symbiosis* of *spectrum sensing* and *database* represents an anchor for collection, storage and processing of valuable radio field information leading to potential benefits for a manifold of wireless stakeholders.

This chapter discussed the details of a developed and operable REM prototype that can be used in various scenarios. For instance, the presented REM backend can provide sufficient radio environment information in mobile networks in order to Minimize Drive Tests (MDTs) [16], which is currently being tackled by 3GPP and other standardization bodies. Also, the presented REM technology can be utilized to facilitate the process of LTE implementation in TVWS [13]. Finally, REMs can be used to perform dedicated spectrum monitoring on certain bands of interest and secure valuable information such as transmitters locations, spectrum utilization, potential spectrum holes etc. The unique feature of the presented REM approach here (i.e. the synergy between sensing-based and database-based spectrum opportunity detection) allows its usage both in static and dynamic scenarios and also allows for *extensibil-*

ity and *accommodation* towards future possible regulations (e.g. implementation of explicit regulatory rules in the RPC).

However, there are several aspects of the REM prototype and the REM approach in general that need to be settled prior to its more widespread practical implementation. The spectrum sensing portion of the architecture requires research and development of possible *self-* and *auto-calibration techniques* of the heterogeneous market-available devices. This can ensure transparent integration of future spectrum sensing solutions. Furthermore, it is expected that there will be low-power, low-cost, powerful and reliable spectrum sensing solutions to be deployed in handheld devices, which stresses the importance of the calibration further on. Finally, the practical implementation of spectrum sensing solution will require methods other than pure energy detection (e.g. cyclostationarity-based detection etc.). The database portion of the REM architecture can benefit from novel techniques and methods for optimal data organization and datamining. This can guarantee quick and on-time REM information fetch and applicability of the REM architecture in various scenarios. Finally, the overall REM technology should be upgraded with more reliable RIF synthesizing methods, algorithms to track moving and multiple transmitters etc.

The REM technology is a valuable aid in the process of optimization of future wireless networks. If designed appropriately, it can provide sufficient radio field information for usage in various static and dynamic scenarios. This will ensure development of spectrum-efficient wireless solutions capable of meeting the future networking requirements.

Acknowledgments This work was funded by the EC FP7-248351 FARAMIR project. The authors would like to thank everyone involved.

References

1. Shared Spectrum Company (2013). http://www.sharedspectrum.com/. Accessed 04 June 2013
2. EC FP7-248351 project FARAMIR. http://www.ict-faramir.eu/. Accessed 04 June 2013
3. Shin, K.G., et al.: Cognitive radios for dynamic spectrum access: from concept to reality. IEEE Wirel. Commun. **17**(6), 64–74 (2010)
4. EC FP7-248303 project QUASAR (2013) Deliverable 1.4: Final report on regulatory feasibility. http://quasarspectrum.eu/images/stories/Documents/deliverables/QUASAR_D1.4.pdf. Accessed 04 June 2013
5. Zhang, Z., et al.: Self-organization paradigms and optimization approaches for cognitive radio technologies: a survey. IEEE Wirel. Commun. **20**(2), 36–42 (2013)
6. Wang, J., et al.: Emerging cognitive radio applications: a survey. IEEE Commun. Mag. **49**(3), 74–81 (2011)
7. EC FP7-248303 project QUASAR (2013) Deliverable 2.2: Methodology for assessing secondary spectrum usage opportunities-final report. http://quasarspectrum.eu/images/stories/Documents/deliverables/QUASAR_D2.2.pdf. Accessed 04 June 2013
8. Zhao, Y., et al.: Performance evaluation of radio environment map enabled cognitive spectrum-sharing networks. In: Paper presented at 2007 IEEE Military Communications Conference, pp. 29–31. Orlando, USA (2007)

9. Atanasovski, V., et al.: Constructing Radio Environment Maps with Heterogeneous Spectrum Sensors. In: Demo Presented at 2011 IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks, pp. 3–6. Aachen, Germany (2011)
10. Denkovski, D., et al.: Integration of heterogeneous spectrum sensing devices towards accurate REM construction. In: Paper presented at 2012 IEEE Wireless Communications and Networking Conference, pp. 1–4. France (2012)
11. Denkovski, D., et al.: Reliability of a Radio Environment Map: Case of Spatial Interpolation Techniques. In: Paper Presented at 7th International Conference on Cognitive Radio Oriented Wireless Networks, p. 1820. Stockholm, Sweden (2012)
12. Subramani, S., et al.: Deployment and interface design considerations for radio environment maps. In: Paper presented at 2012 IEEE 8th International Conference on Wireless and Mobile Computing, Networking and Communications, pp. 8–10. Barcelona, Spain (2012)
13. van de Beek, J., et al.: REM-enabled opportunistic LTE in the TV band. In: Demo Presented at 2012 IEEE International Symposium on Dynamic Spectrum Access Networks, Bellevue, pp. 8–10. Washington, USA (2012)
14. Yilmaz, H.B., Tugcu, T.: Location estimation-based radio environment map construction in fading channels. J. Wirel. Commun. Mobile Comput. (2013). doi:10.1002/wcm.2367
15. Wei, Z., et al.: (2013) On the construction of radio environment maps for cognitive radio networks. In: Paper Presented at 2013 IEEE Wireless Communications and Networking Conference, pp. 7–10. Shanghai, China (2013)
16. Universal Terrestrial Radio Access (UTRA) and Evolved Universal Terrestrial Radio Access (E-UTRA) (2011) Radio Measurement Collection for Minimization of Drive Tests (MDT), Overall description, Stage 2, 3GPP Standard TS 37.320, v. 11.0.0, June 2011
17. Axell, E., et al.: Spectrum Sensing for Cognitive Radio: State-of-the-Art and Recent Advances. IEEE Sig. Process. Mag. **29**(3), 101–116 (2012)
18. Chen, Z., et al.: Demonstration on real time spectrum sensing for cognitive radio. IEEE Commun. Lett. **14**(10), 915–917 (2010)
19. Denkovski, D., et al.: HOS nased goodness-of-fit testing signal detection. IEEE Commun. Lett. **16**(3), 310–313 (2012)
20. Federal Communications Commission (FCC) (2010) Second Memorandum Opinion and Order, FCC-10-174, September 2010
21. MS2690A Signal Analyzer Brochure. http://www.anritsu.com/en-AU/Downloads/Brochures-Datasheets-and-Catalogs/Brochure/DWL8558.aspx. Accessed 04 June 2013
22. Universal Software Radio Peripheral 2 (USRP2). http://www.ettus.com. Accessed 04 June 2013
23. Texas Instruments eZ430-RF2500 datasheet. http://focus.ti.com/lit/ug/slau227e/slau227e.pdf. Accessed 04 June 2013
24. Sun SPOT Developers Guide. http://www.sunspotworld.com/Tutorial/index.html. Accessed 04 June 2013
25. Denkovski, D., et al.: Efficient mid-end spectrum sensing implementation for cognitive radio applications based on USRP2 devices. In: Proceedings of the 1st International Conference on Advances in Cognitive Radio 2011, pp. 17–22. Budapest, Hungary (2011)
26. Shepard, D.: A two-dimensional interpolation function for irregularly-spaced data. In: Proceedings of the 23rd ACM National conference, New York, USA (1968)
27. Renka, R.J.: Multivariate interpolation of large sets of scattered data. ACM Trans. Math. Softw. **14**(2), 139148 (1988)
28. Goovaerts, P.: Geostatistics for Natural Resources Evaluation. Oxford, New York (1997)
29. Lin, G., Chen, L.: A spatial interpolation method based on radial basis function networks incorporating a semivariogram model. J. Hydrol. **288**, 288–298 (2004)
30. Hutchinson, M.F.: Interpolating mean rainfall using thin plate smoothing splines. Int. J. Geogr. Inf. Syst. **9**(4), 385–403 (1995)
31. Nalder, I.A., Wein, R.W.: Spatial interpolation of climatic Normals: test of a new method in the Canadian boreal forest. Agric. For. Meteorol. **92**, 211–225 (1998)

32. Denkovski, D., et al.: Practical assessment of RSS-based localization in indoor environments. In: Proceedings of the 2012 IEEE Military Communications Conference, Orlando, USA (2012)
33. Dagres, I., et al.: Algorithms and bounds for energy-based multi-source localization in log-normal fading. In: Paper Presented at 2012 Globecom Workshops, pp. 3–7. Anaheim, USA (2012)

Chapter 10
Cellular Coverage Optimization: A Radio Environment Map for Minimization of Drive Tests

Ana Galindo-Serrano, Berna Sayrac, Sana Ben Jemaa, Janne Riihijärvi and Petri Mähönen

Abstract Coverage is one of the most important targets that has to be achieved by cellular operators. Without coverage provisioning, concepts like service, or Quality of Service (QoS) cannot be considered. Therefore, cellular coverage prediction and enhancement is a basic and prevailing area of research in wireless communications. Our work introduces an automatic and remote self-optimization process based on exploitation of geo-location information for cellular coverage optimization. Specifically, we use Radio Environment Maps (REMs) for cellular network coverage hole detection purposes. We define REM as an intelligent entity which stores incoming radio environmental data and also interpolates this data to benefit from the spatial correlation that exists in it. Furthermore, with the standardization of Minimization of Drive Tests (MDT) in 3GPP, geo-location based solutions/applications are increasingly becoming feasible and popular. The proposed REM-based coverage hole detection approach drastically reduces the required drive tests and enhances the network with self-responsive capabilities to handle key obstacles towards cellular networks autonomy.

A. Galindo-Serrano (✉) · B. Sayrac · S. B. Jemaa
Orange Labs, 38-40 rue du Général Leclerc, 92794 Issy les Moulineaux cédex 9, France
e-mail: anamgase@gmail.com

B. Sayrac
e-mail: berna.sayrac@orange.com

S. B. Jemaa
e-mail: sana.benjemaa@orange.com

J. Riihijärvi · P. Mähönen
Institute for Networked Systems, RWTH Aachen University, Aachen, Germany
e-mail: jar@inets.rwth-aachen.de

P. Mähönen
e-mail: pma@inets.rwth-aachen.de

M.-G. Di Benedetto and F. Bader (eds.), *Cognitive Communication and Cooperative HetNet Coexistence*, Signals and Communication Technology,
DOI: 10.1007/978-3-319-01402-9_10,

10.1 Introduction

Coverage estimation is the first step for cellular network deployments. It consists of the signal strength estimation in the area to be served through sophisticated planning tools. These planning tools apply comprehensive propagation models to the terrain profile, also considering the building layout. Despite the accuracy of these planning tools and the efforts put by the operators in the coverage estimation and deployment phase, the existence of coverage holes is almost impossible to avoid since the network is always subject to unforeseen changes in the propagation environment, or equipment failures. Therefore, the coverage optimization process is required during the operational phase. The coverage optimization has two stages, the coverage hole prediction and the solution deployment, which remedies or removes the coverage problem in the uncovered zones. The solution deployed to solve coverage holes needs to be timely and cost-effective, i.e., providing full coverage without creating excessive interference to the already covered neighboring areas. The success of the given solutions is directly related to the accuracy of coverage hole location and shape information, which comes from the coverage hole prediction approach.

This chapter focuses on coverage hole prediction, and showcases how to use the powerful mathematical tools coming from spatial statistics as a cognitive solution to these problems. So far, operators have adopted the following procedure to deal with coverage prediction: (1) to perform *drive tests*, which consist of geographically measuring different network metrics and indicators with motor vehicles equipped with specialized mobile radio measurement equipments and Global Positioning System (GPS), and (2) to analyze the collected measurements for coverage prediction. On the one hand, drive tests generate a tremendous amount of data to be processed, allowing the operators to get realistic network information close to the actual user experience. This is a very useful and desired information by operators [1]. With the processed drive test measurements, the operator can have a realistic picture of the network in terms of coverage and service quality, and find the right optimization solutions through the modification of one or several network parameters such as the transmission power, antenna locations, antenna orientations and tilts, etc. On the other hand, drive tests are quite an inefficient means to solve the coverage problems since they: (1) imply large Operational Expenditure (OPEX), (2) incur delays in detecting and predicting the coverage holes, (3) are an undesirable source of pollution, and (4) provide an *incomplete* picture of the "ground-truth" since they are limited to roads and other regions accessible by motor vehicles.

All these disadvantages make mandatory for operators to make the most of the information collected through the drive tests, and to minimize the use of them. For this purpose, the 3rd Generation Partnership Project (3GPP) standardization body has been working on the minimization of the use of drive tests for Long Term Evolution (LTE) since Release 9 [2]. In Release 10, a Minimization of Drive Tests (MDT) work item [3, 4] for Universal Terrestrial Radio Access Network (UTRAN) was also included. The main focus of MDT in Release 10 is on coverage optimization. Release 11 focuses on Quality of Service (QoS) verification, further improvements

in coverage optimization, positioning enhancement and the study of other MDT use cases [5].

The key idea of the MDT proposed by 3GPP is to take advantage of the (esp. geo-location) measurement capabilities of the new advanced User Equipments (UEs) as well as of the radio measurements performed as part of the Radio Resource Management (RRM) procedures. The main characteristic of MDT is that the UEs report their *geo-located* measurements to the network upon operator request. The collected MDT measurements are at operator's direct disposition to ease any kind of (automated as well as manual) network operation, management and optimization task. Hence, MDT also acts as a valuable input to the Self-Organized Network (SON) functionalities whose main goals are to decrease operational expenses, increase the autonomy of networks and allow the achievement of an automatic, near optimal network quality in next generation networks [6].

The use of location information for maximizing the efficiency of wireless network resource utilization also appears as a fundamental component, called as *location/environment awareness*, in Cognitive Radio (CR), put forward by Mitola in [7]. The first work where spatial information exploitation appears explicitly in wireless networks was the Available Resource Map (ARM) introduced by Krenik, which is proposed as a real-time map of all radio activities in the network for CR applications in Unlicensed Wide Area Networks (UWAN) [8]. Then, it was extended to the Radio Environment Map (REM) concept by Zhao [9], who defined it as an integrated database for enhancement of CR systems, mainly for dynamic spectrum access purposes (such as TV whitespaces). Such a REM stores environmental information, past experience and radio knowledge with the aim of improving the network performance as well as aiding in self-optimization and management. Shortly after, a more comprehensive/cognitive version of REMs, also called as *Interference Cartography* (*IC*), has been proposed for a mesh-like REM structure where the location points are found on a rectangular grid [10–12], and we refer to the minimum area of the regular grid as a pixel. In our work, we consider this broader view of the concept of REM, where the main idea is to: (1) *spatially interpolate* the collected geo-located measurements (i.e. drive test measurements and MDT measurements) in order to predict the measurement values at those pixels where measurements are not available, and (2) to request additional measurements at intelligently chosen pixels to enhance the quality of those predictions. This chapter falls in this latter line of work and unless otherwise stated, REM will mean IC.

In the literature on spatial statistics, there are several powerful spatial interpolation techniques that can be used. Among them, we have selected the Bayesian kriging interpolation, introduced by Kitanidis in [13], as it takes into account the various uncertainties in the models used, and does not underestimate the standard errors of predictions. Furthermore, it automatically calculates the interpolation model parameters through a process of sub-settings and simulations, all this at the cost of an increment in the computation complexity[1] as the number of measurements

[1] The computational complexity of classical kriging interpolation is roughly $O(n^3)$, where n is the number of available measurements. In practice several thousands of measurements can be processed

increases [14]. Particularly in our problem, this disadvantage does not entail a drawback since we are dealing with a semi-static REM whose construction is an *offline* process that in practice would be activated by the operator once per day or even a week (for comparison, classical drive test-based coverage prediction is often performed just once per year due to the amount of efforts and costs involved).

The work presented in this chapter is the continuation of the work presented in [15, 16], where the Bayesian kriging method was used for the REMs generation. Differently from these works, in this chapter we give a more realistic evaluation of the REM-based coverage hole prediction, since in the present analysis, a more realistic coverage hole definition is considered. A coverage hole frequently consists of *neighboring* uncovered pixels. Thus, a set of neighboring N pixels constitute a single coverage hole, but not N separate coverage holes, as assumed in the pixel-based analysis presented in [16].

First, this chapter presents the mathematical details of Bayesian kriging, applied to the problem of coverage prediction in cellular networks. We analyze the performance of the interpolation process for a LTE network modeled with a grid size of $25\,\text{m} \times 25\,\text{m}$, in a dense urban environment. Results obtained with real network measurements show that the chosen spatial interpolation approach is indeed a very efficient and promising means of processing the MDT data which is expected to overcrowd the operator databases in the near future, in terms of improving the quality of the coverage predictions. Second, we adopt the notion of neighboring pixels in defining a coverage hole and perform the performance analysis. Our aim is to measure the gains introduced by the use of REMs in coverage hole prediction where coverage hole is defined in a more realistic sense, as described above. We first present a *local* coverage analysis where it is assumed that operators have a rough knowledge about the areas with coverage problems where they apply the REM to perform the coverage analysis in these areas. Then, a more comprehensive solution is presented, the *global* coverage analysis, which consists of the construction of a REM over the Base Station (BS) coverage region to automatically determine the potential areas with coverage problems.

To the best of our knowledge, automated coverage analysis based on spatial statistics has been studied extensively so far for sensor networks but not for cellular networks. So our line of work is the first to introduce spatial statistics in cellular coverage studies and to perform realistic performance evaluations. However, the significance of this work reaches beyond: considering that MDT measurements are currently on their way to overcrowd the operator databases, operators are in need of cost effective, feasible and well-performing solutions that allow them make use of the valuable information for high quality network performance, and this chapter provides a cognitive radio-inspired candidate solution.

The remainder of this chapter is organized as follows. Section 10.2 presents in detail the Bayesian kriging interpolation method for coverage prediction. Section 10.3

(Footnote 1 continued)
on a typical computer in the time scales discussed in the text. Further, recent approximation techniques yield very good results with *linear* complexity [27, 28]. We have taken first steps to apply these techniques for REM construction in [29] with very promising initial results.

presents the algorithm used for the REM construction and methodology implemented for the coverage hole detection. The initial analysis is performed *locally* and then the solution is extended to a *global* analysis of the coverage region. Section 10.4 presents results evaluating the interpolation process performance and the local and global coverage hole detection processes. Finally, Sect. 10.5 summarizes our main conclusions.

10.2 Coverage Prediction with Bayesian Kriging

In this section we summarize the steps to be followed for applying Bayesian kriging in cellular coverage prediction. First, we present the used network model and our assumptions. Second, we introduce the prediction method and, finally, we explain how the parameters used in our model are estimated.

10.2.1 Modeling and Assumptions

We consider the Down Link (DL) transmission of a cellular radio access network with a given BS transmitter equipped with an omnidirectional antenna. Let $y(\mathbf{x}_i)$ denote the DL received power (in dBm) at location $\mathbf{x}_i$. Assuming that the fast fading effects are averaged out by the receivers, $y(\mathbf{x}_i)$ can be expressed as

$$y(\mathbf{x}_i) = p_0 - 10\alpha \log_{10} d_i + s(\mathbf{x}_i) + z_i, \tag{10.1}$$

where p_0 is the transmitted power (in dBm), α is the pathloss coefficient, d_i is the distance (in m) between the transmitter and the receiver location $\mathbf{x}_i$, $s(\mathbf{x}_i)$ is the shadow fading factor (in dB), and z_i is the zero-mean additive noise term which incorporates the uncertainties of the measurement process and all other random effects due to the propagation environment. Note that for directional antennas we can use the same but modifying Eq. 10.1 by including the antenna pattern.

Equation (10.1) is the well-known large-scale propagation model in the logarithmic scale, which models the wireless channel as the sum of a deterministic linear pathloss term and two stochastic terms: shadowing and noise. This model is one of the most widely used wireless channel models due to its simplicity and to its overall ability to represent the main characteristics of the wireless channel behavior in a variety of important wireless environments. The random noise process is assumed to consist of independent and identically distributed Gaussian samples, which are also independent of the shadowing term. Shadowing is a zero-mean Gaussian random variable that is spatially correlated according to the exponential correlation model [17]

$$\mathbb{E}\left\{s(\mathbf{x}_i)s(\mathbf{x}_j)\right\} = r_{ij} = \frac{1}{\theta}\exp\left(-\frac{d_{ij}}{\phi}\right), \tag{10.2}$$

where $\frac{1}{\theta}$ is the shadowing variance in dB, d_{ij} is the Euclidean distance between locations $\mathbf{x}_i$ and $\mathbf{x}_j$, and ϕ controls the correlation distance of the shadowing.

We assume that such power measurements are carried out by a set of N receiving terminals, located at the set of locations $\mathbf{x} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$. Arranging these measurements in a $N \times 1$ column vector $\mathbf{y}(\mathbf{x})$, we obtain the vector-matrix relation

$$\mathbf{y} = X\beta + \mathbf{u}, \tag{10.3}$$

where

$$X = \begin{bmatrix} 1 & -10\log_{10}(d_1) \\ \vdots & \vdots \\ 1 & -10\log_{10}(d_N) \end{bmatrix}, \quad \beta = \begin{bmatrix} p_0 \\ \alpha \end{bmatrix}, \text{ and } \mathbf{u} = \begin{bmatrix} s(\mathbf{x}_1) + z_1 \\ \vdots \\ s(\mathbf{x}_N) + z_N \end{bmatrix}. \tag{10.4}$$

Here, X is a $N \times 2$ deterministic matrix of known functions of the measurement locations $\mathbf{x}$, β is the 2×1 parameter vector of the spatial mean and $\mathbf{u}$ is a $N \times 1$ multivariate Gaussian vector whose covariance matrix is $Q_{yy}(\theta, \phi, \tau) = \frac{1}{\theta}(R_{yy}(\phi) + \tau I_N)$, where $\frac{1}{\theta}R_{yy}(\phi)$ is the $N \times N$ covariance matrix of the spatially correlated shadowing term whose (i, j)th entry is equal to r_{ij} of Eq. (10.2), and I_N is the $N \times N$ identity matrix. Note that the variance of the noise process is $\frac{\tau}{\theta}$. Note also that $\mathbf{y}$, X and $\mathbf{u}$ are functions of the locations $\mathbf{x}$.

10.2.2 Prediction

The aim is to predict the received power values at locations where we do not have measurements. Let $\mathbf{x}_0$ denote the $M \times 1$ vector of those locations. The same underlying model is assumed for the predictions, namely

$$\mathbf{y}_0 = X_0\beta + \mathbf{u}_0, \tag{10.5}$$

where $\mathbf{y}_0$ is the $M \times 1$ vector of received power values at locations $\mathbf{x}_0$, X_0 is the $M \times 2$ matrix of deterministic effects for $\mathbf{y}_0$ and $\mathbf{u}_0$ is the $M \times 1$ stochastic vector whose covariance matrix is denoted by $Q_{00}(\theta, \phi, \tau)$. Note that for notational convenience, the dependence of $\mathbf{y}$, X and $\mathbf{u}$ ($\mathbf{y}_0$, X_0 and $\mathbf{u}_0$ resp.) on $\mathbf{x}$ ($\mathbf{x}_0$ resp.) is not shown in Eqs. (10.3–10.5).

Putting the measurements and the predictions together, we obtain the multivariate Gaussian model given by

$$\begin{bmatrix} \mathbf{y}_0 \\ \mathbf{y} \end{bmatrix} \equiv \mathcal{N}\left(\begin{bmatrix} X_0 \\ X \end{bmatrix} \beta, \begin{bmatrix} Q_{00}(\theta, \phi, \tau) & Q_{0y}(\theta, \phi, \tau) \\ Q_{y0}(\theta, \phi, \tau) & Q_{yy}\theta, \phi, \tau \end{bmatrix} \right), \tag{10.6}$$

where $Q_{0y}(\theta, \phi, \tau)$ and $Q_{y0}(\theta, \phi, \tau)$ are the cross-covariance matrices between $\mathbf{y}_0$ and $\mathbf{y}$. Note that $Q_{0y} = Q_{y0}^T$, where $[\,\cdot\,]^T$ denotes the matrix transpose.

The model parameters β, θ, ϕ and τ are unknown. Therefore, they have to be estimated from the existing measurement dataset. However, those parameters are not *completely* unknown to us: we have some prior knowledge on their probable values. For example, we can say that the radiated power p_0 is close to the power at the antenna feeder, the propagation pathloss coefficient α is around 3.5 in urban areas [18] and the shadowing standard deviation $\sqrt{1/\theta}$ typically ranges between 8 and 11 dB for typical outdoor Above RoofTop to Below RoofTop scenarios [19]. Including this prior information into the model helps us to enhance the prediction quality. Besides, the Gaussian assumption of the stochastic component vector, $\mathbf{u}$, in the chosen model, provides tractability in the complex analytical derivations of the Bayesian inference framework. However, we emphasize that the Gaussian nature of the problem is firmly grounded on experiments, and is not simply assumed for mathematical convenience.

The task of predicting the received power values at locations $\mathbf{x}_0$ is equivalent to finding an estimator $\hat{\mathbf{y}}_0$ of the random vector $\mathbf{y}_0$ given measurements $\mathbf{y}$. This estimator is preferably a linear function of measurements $\mathbf{y}$ which minimizes a given loss/risk/cost function. In the Bayesian context, this is equivalent to the *Bayes estimator* which minimizes a given *posterior* expected loss/risk/cost (a.k.a. *Bayes risk*). The most commonly used risk function is the *squared error risk*, or the Mean Squared Error (MSE), resulting in the estimate

$$\hat{\mathbf{y}}_0 = \min_{\mathbf{y}_0^*} \mathbb{E}\left\{(\mathbf{y}_0 - \mathbf{y}_0^*)^T(\mathbf{y}_0 - \mathbf{y}_0^*)|\mathbf{y}\right\} = \mathbb{E}\left\{\mathbf{y}_0|\mathbf{y}\right\}, \tag{10.7}$$

which is equivalent to the posterior mean $\mathbb{E}\{\mathbf{y}_0|\mathbf{y}\}$. For the MSE Bayes estimator of Eq. (10.7), the Bayes risk (or the MSE) is the posterior variance

$$\begin{aligned} MSE &= \frac{1}{M}\mathbb{E}\left\{(\mathbf{y}_0 - \hat{\mathbf{y}}_0)^T(\mathbf{y}_0 - \hat{\mathbf{y}}_0)|\mathbf{y}\right\} \\ &= \frac{1}{M}\mathbb{E}\left\{(\mathbf{y}_0 - \mathbb{E}\{\mathbf{y}_0|\mathbf{y}\})^T(\mathbf{y}_0 - \mathbb{E}\{\mathbf{y}_0|\mathbf{y}\})|\mathbf{y}\right\} \\ &= \frac{1}{M}\,\mathrm{trace}\,(\mathrm{cov}\,\{\mathbf{y}_0|\mathbf{y}\}). \end{aligned} \tag{10.8}$$

For the MSE Bayes estimator, we need to calculate the marginal posterior (or Bayesian) pdf $p(\mathbf{y}_0|\mathbf{y})$, or at least its moments such as mean $\mathbb{E}\{\mathbf{y}_0|\mathbf{y}\}$ and covariance $\mathrm{cov}\,\{\mathbf{y}_0|\mathbf{y}\}$. In the following, we will present a brief derivation of these calculations.

We start by the following expression for the Bayesian pdf:

$$p(\mathbf{y}_0|\mathbf{y}) = \iiiint_{\tau,\phi,\theta,\beta} p(\mathbf{y}_0|\beta, \theta, \phi, \tau, \mathbf{y}) p(\beta, \theta, \phi, \tau|\mathbf{y}). \tag{10.9}$$

The first integrand in Eq. (10.9) is the conditional pdf of $\mathbf{y}_0$ given $\mathbf{y}$ and the model parameters. Assuming that Q_{yy} is non-singular, it can be shown (see Proof 1 in [20])

that this pdf is Gaussian with mean $X_0\beta + Q_{0y}Q_{yy}^{-1}(\mathbf{y} - X\beta)$ and covariance matrix $Q_{00} - Q_{0y}Q_{yy}^{-1}Q_{y0}$. Note that for notational convenience, the dependence of Q_{0y} and Q_{y0} on the model parameters is not shown here, and in integral expressions measures of integration are omitted when there is no danger of confusion.

The second integrand in Eq. (10.9) is the joint posterior pdf of the model parameters which can be decomposed into the product of two joint posterior pdfs:

$$p(\beta, \theta, \phi, \tau|\mathbf{y}) = p(\beta, \theta|\phi, \tau, \mathbf{y})p(\phi, \tau|\mathbf{y}). \tag{10.10}$$

We can express the first joint posterior pdf $p(\beta, \theta|\phi, \tau, \mathbf{y})$ as the product of a likelihood term and a prior pdf given by

$$p(\beta, \theta|\phi, \tau, \mathbf{y}) = c_1(\mathbf{y}, \phi, \tau)\, p(\mathbf{y}|\beta, \theta, \phi, \tau)p(\beta, \theta|\phi, \tau), \tag{10.11}$$

where $c_1(\mathbf{y}, \phi, \tau) = \left[\iint_{\beta,\theta} p(\mathbf{y}|\beta, \theta, \phi, \tau)p(\beta, \theta|\phi, \tau)\right]^{-1}$ is a normalizing constant which can be computed numerically, $p(\beta, \theta|\phi, \tau)$ is the joint prior pdf of β and θ (with ϕ and τ known), and $p(\mathbf{y}|\beta, \theta, \phi, \tau)$ is the likelihood term which is Gaussian with mean $\mathbf{y} - X\beta$ and covariance Q_{yy}:

$$p(\mathbf{y}|\beta, \theta, \phi, \tau) = (2\pi)^{-N/2}\left|Q_{yy}\right|^{-1/2}\exp\left[-\frac{1}{2}(\mathbf{y} - X\beta)^T Q_{yy}^{-1}(\mathbf{y} - X\beta)\right]. \tag{10.12}$$

An important issue in Bayesian inference is the choice of prior distributions. It is common practice to choose functional forms that facilitate the involved analytical treatments. Such functional forms are called as *conjugate priors*, meaning conjugate to the likelihood function, such that the posterior distribution has the same functional form as the prior distribution. Conjugate priors have been identified for the most widely used distribution functions [14].

In order to find the conjugate prior for $p(\beta, \theta|\phi, \tau)$, we rewrite the likelihood $p(\mathbf{y}|\beta, \theta, \phi, \tau)$ of Eq. (10.12) in the following form (see Proof 2 in [20]):

$$\begin{aligned} p(\mathbf{y}|\beta, \theta, \phi, \tau) = (2\pi)^{-N/2}|S_{yy}|^{-1/2}\theta^{\rho/2}\exp\left[-\frac{\theta}{2}(\beta - \mathbf{b})^T H(\beta - \mathbf{b})\right] \\ \times\, \theta^{\kappa/2}\exp\left[-\frac{1}{2}\kappa q\theta\right], \end{aligned} \tag{10.13}$$

where

$$S_{yy} = \theta Q_{yy} = R_{yy} + \tau I_N \tag{10.14}$$

$$H = X^T S_{yy}^{-1} X \tag{10.15}$$

$$H\mathbf{b} = X^T S_{yy}^{-1}\mathbf{y} \tag{10.16}$$

$$\rho = \text{rank}(H) \tag{10.17}$$

$$\kappa = N - \rho \tag{10.18}$$

$$q = \mathbf{y}^T S_{yy}^{-1}(\mathbf{y} - X\mathbf{b})/\kappa. \tag{10.19}$$

Note that the expression in Eq. (10.13) is a proper pdf if $\rho = \text{rank}(X)$. Otherwise, it represents only $\rho < \text{rank}(X)$ linear combinations of β.

The functional form of Eq. (10.13) calls for the *Normal-Gamma-2* density as the conjugate prior for $p(\beta, \theta|\phi, \tau)$ [14], given by

$$p(\beta, \theta|\phi, \tau) = (2\pi)^{-\rho'/2}|H'|^{1/2}\theta^{\rho'/2}\exp\left[-\frac{\theta}{2}(\beta - \mathbf{b}')^T H'(\beta - \mathbf{b}')\right] \times \left(\frac{\kappa' q'}{2}\right)^{\kappa'/2} \Gamma^{-1}\left(\frac{\kappa'}{2}\right)\theta^{\kappa'/2-1}\exp\left[-\frac{1}{2}\kappa' q'\theta\right] \tag{10.20}$$

with $\rho' = \text{rank}(H')$. This pdf represents a multivariate Gaussian random vector β whose covariance matrix is scaled with a random variable whose inverse θ is Gamma distributed.

The joint posterior pdf $p(\beta, \theta|\phi, \tau, \mathbf{y})$ is obtained by replacing Eqs. (10.12–10.20) in Eq. (10.11) and is also a Normal-Gamma-2 density

$$p(\beta, \theta|\phi, \tau, \mathbf{y}) = (2\pi)^{-\rho''/2}|H''|^{1/2}\theta^{\rho''/2}\exp\left[-\frac{\theta}{2}(\beta - \mathbf{b}'')^T H''(\beta - \mathbf{b}'')\right] \times \left(\frac{\kappa'' q''}{2}\right)^{\kappa''/2} \Gamma^{-1}\left(\frac{\kappa''}{2}\right)\theta^{\kappa''/2-1}\exp\left[-\frac{1}{2}\kappa'' q''\theta\right] \tag{10.21}$$

with the following parameters (see Proof 3 in [26]):

$$H'' = H' + H \tag{10.22}$$

$$H''\mathbf{b}'' = H'\mathbf{b}' + H\mathbf{b} \tag{10.23}$$

$$\rho'' = \text{rank}(H'') \tag{10.24}$$

$$\kappa'' = \rho' + \kappa' + \rho + \kappa - \rho'' \tag{10.25}$$

$$q'' = \frac{\kappa' q' + \mathbf{y}^T S_{yy}^{-1}\mathbf{y} + \mathbf{b}'^T H'\mathbf{b}' - \mathbf{b}''^T H''\mathbf{b}''}{\kappa''}. \tag{10.26}$$

The second joint posterior of Eq. (10.10), $p(\phi, \tau|\mathbf{y})$, can be expressed as follows:

$$p(\phi, \tau|\mathbf{y}) = c_2(\mathbf{y})p(\mathbf{y}|\beta, \theta, \phi, \tau)p(\beta, \theta|\phi, \tau)p(\phi, \tau)\frac{1}{p(\beta, \theta|\phi, \tau, \mathbf{y})}. \tag{10.27}$$

Here $c_2(\mathbf{y}) = \left[\iiiint_{\beta,\theta,\phi,\tau} p(\mathbf{y}|\beta, \theta, \phi, \tau)p(\beta, \theta|\phi, \tau)p(\phi, \tau)\right]^{-1}$ is a function of $\mathbf{y}$ only and can be numerically evaluated. Using Eqs. (10.12) and (10.20– 10.21) in the above equation and simplifying the terms we obtain

$$p(\phi,\tau|\mathbf{y}) = c_2(\mathbf{y})(2\pi)^{-(N+\rho'-\rho'')/2}|S_{yy}|^{-1/2}|H'|^{1/2}|H''|^{-1/2} \times \Gamma^{-1}\left(\frac{\kappa'}{2}\right)\Gamma\left(\frac{\kappa''}{2}\right)\left(\frac{\kappa' q'}{2}\right)^{\kappa'/2}\left(\frac{\kappa'' q''}{2}\right)^{-\kappa''/2} p(\phi,\tau). \quad (10.28)$$

The joint prior $p(\phi,\tau)$ can be written as the product of the marginal densities $p(\phi)$ and $p(\tau)$, since the noise process is assumed independent of the shadowing process. Furthermore, we assume discrete pdfs for $p(\phi)$ and $p(\tau)$. Therefore,

$$p(\phi,\tau) = \sum_k \sum_l \phi_k \tau_l \delta(\phi-\phi_k)\delta(\tau-\tau_l), \quad (10.29)$$

where $\delta(\cdot)$ is the Dirac delta function. Then, Eq. (10.28) takes the following form:

$$p(\phi,\tau|\mathbf{y}) = c_2(\mathbf{y})(2\pi)^{-(N+\rho'-\rho'')/2}|H'|^{1/2}\Gamma^{-1}\left(\frac{\kappa'}{2}\right)\Gamma\left(\frac{\kappa''}{2}\right)\left(\frac{\kappa' q'}{2}\right)^{\kappa'/2} \times \sum_k \sum_l \phi_k \tau_l |S_{yy}(\phi_k,\tau_l)|^{-1/2}|H''(\phi_k,\tau_l)|^{-1/2} \times \left(\frac{\kappa'' q''(\phi_k,\tau_l)}{2}\right)^{-\kappa''/2} \delta(\phi-\phi_k)\delta(\tau-\tau_l). \quad (10.30)$$

Note that in the above equation, S_{yy}, H'' and q'' are functions of ϕ and τ.

Now, we can return to the Bayesian pdf of Eq. (10.9) and rewrite it as follows:

$$p(\mathbf{y}_0|\mathbf{y}) = \iiint_{\theta,\phi,\tau}\left[\int_\beta p(\mathbf{y}_0|\beta,\theta,\phi,\tau,\mathbf{y})p(\beta|\theta,\phi,\tau,\mathbf{y})\,d\beta\right] \times p(\theta|\phi,\tau,\mathbf{y})p(\phi,\tau|\mathbf{y})\,d\theta\,d\phi\,d\tau. \quad (10.31)$$

The inner integral, $p(\mathbf{y}_0|\theta,\phi,\tau,\mathbf{y})$, is the marginalization (w.r.t. β) of a joint Gaussian pdf, and therefore is Gaussian with the following mean and variance [13]:

$$\begin{aligned}\mathbb{E}\{\mathbf{y}_0|\theta,\phi,\tau,\mathbf{y}\} &= (X_0 - Q_{0y}Q_{yy}^{-1}X)(H' + X^T Q_{yy}^{-1}X)^{-1} \\ &\quad H'\mathbf{b}' + [Q_{0y}Q_{yy}^{-1} + (X_0 - Q_{0y}Q_{yy}^{-1}X) \\ &\quad (H' + X^T Q_{yy}^{-1}X)^{-1}X^T Q_{yy}^{-1}]\mathbf{y} \quad (10.32)\\ \mathrm{var}\{\mathbf{y}_0|\theta,\phi,\tau,\mathbf{y}\} &= (Q_{00} - Q_{0y}Q_{yy}^{-1}Q_{y0}) + \\ &\quad (X_0 - Q_{0y}Q_{yy}^{-1}X)(H' + X^T Q_{yy}^{-1}X)^{-1} \\ &\quad (X_0 - Q_{0y}Q_{yy}^{-1}X)^T. \quad (10.33)\end{aligned}$$

Then, the Bayesian pdf $p(\mathbf{y}_0|\mathbf{y})$ can be rewritten as:

$$p(\mathbf{y}_0|\mathbf{y}) = \iint_{\phi,\tau}\left[\int_\theta p(\mathbf{y}_0|\phi,\theta,\tau,\mathbf{y})p(\theta|\phi,\tau,\mathbf{y})d\theta\right]p(\phi,\tau|\mathbf{y})\,d\phi\,d\tau. \quad (10.34)$$

The inner integral is recognized as a Student distribution [14] with mean and covariance matrix of

$$\begin{aligned}\mathbb{E}\{\mathbf{y}_0|\phi,\tau,\mathbf{y}\} &= (X_0 - S_{0y}S_{yy}^{-1}X)H''^{-1}H'\mathbf{b}' \\ &\quad + [S_{0y}S_{yy}^{-1} + (X_0 - S_{0y}S_{yy}^{-1}X)H''^{-1}X^T S_{yy}^{-1}]\mathbf{y} \qquad (10.35)\\ \operatorname{cov}\{\mathbf{y}_0|\phi,\tau,\mathbf{y}\} &= \frac{\kappa''q''}{\kappa''-2}[S_{00} - S_{0y}S_{yy}^{-1}S_{y0} \\ &\quad + (X_0 - S_{0y}S_{yy}^{-1}X)H''^{-1}(X_0 - S_{0y}S_{yy}^{-1}X)^T], \qquad (10.36)\end{aligned}$$

where $S_{00} = \theta Q_{00}$, $S_{0y} = \theta Q_{0y}$ and $S_{y0} = \theta Q_{y0}$ are the scaled covariance and cross-covariance matrices respectively.

Finally, using Eqs. (10.30 and 10.35–10.36), the posterior mean MSE Bayes estimator can be calculated as

$$\mathbb{E}\{\mathbf{y}_0|\mathbf{y}\} = \sum_{\phi,\tau} p(\phi,\tau|\mathbf{y})\mathbb{E}\{\mathbf{y}_0|\phi,\tau,\mathbf{y}\} \qquad (10.37)$$

and the covariance is given by

$$\begin{aligned}\operatorname{cov}\{\mathbf{y}_0|\mathbf{y}\} &= \sum_{\phi,\tau} p(\phi,\tau|\mathbf{y})[\operatorname{cov}\{\mathbf{y}_0|\phi,\tau,\mathbf{y}\} + (\mathbb{E}\{\mathbf{y}_0|\phi,\tau,\mathbf{y}\} \\ &\quad - \mathbb{E}\{\mathbf{y}_0|y\})(\mathbb{E}\{\mathbf{y}_0|\phi,\tau,\mathbf{y}\} - \mathbb{E}\{\mathbf{y}_0|\mathbf{y}\})^T]. \qquad (10.38)\end{aligned}$$

10.2.3 Model Parameter Estimation

Estimation of the model parameters is carried out by calculating their posterior expectations, i.e. evaluating

$$\begin{aligned}\mathbb{E}\{\beta|\mathbf{y}\} &= \int_\beta \beta p(\beta|\mathbf{y})\,d\beta \\ &= \int_\beta \beta \iint_{\phi,\tau} p(\beta|\phi,\tau,\mathbf{y})p(\phi,\tau|\mathbf{y})\,d\phi\,d\tau\,d\beta \\ &= \iint_{\phi,\tau}\left[\int_\beta \beta p(\beta|\phi,\tau,\mathbf{y})\,d\beta\right]p(\phi,\tau|\mathbf{y})\,d\phi\,d\tau \\ &= \iint_{\phi,\tau} \mathbb{E}\{\beta|\phi,\tau,\mathbf{y}\}\,p(\phi,\tau|\mathbf{y})\,d\phi\,d\tau \qquad (10.39)\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}\{\theta|\mathbf{y}\} &= \int_{\theta} \theta p(\theta|\mathbf{y})\, d\theta \\ &= \int_{\theta} \theta \iint_{\phi,\tau} p(\theta|\phi,\tau,\mathbf{y}) p(\phi,\tau|\mathbf{y})\, d\phi\, d\tau d\theta \\ &= \iint_{\phi,\tau} \left[\int_{\theta} \theta p(\theta|\phi,\tau,\mathbf{y})\, d\theta\right] p(\phi,\tau|\mathbf{y})\, d\phi\, d\tau \\ &= \iint_{\phi,\tau} \mathbb{E}\{\theta|\phi,\tau,\mathbf{y}\}\, p(\phi,\tau|\mathbf{y})\, d\phi\, d\tau. \end{aligned} \tag{10.40}$$

To calculate the above expressions, we need the marginal posteriors for θ and β, which become

$$\begin{aligned} p(\beta|\phi,\tau,\mathbf{y}) = \pi^{-\frac{\rho''}{2}} \left|\frac{H''}{q''}\right|^{\frac{1}{2}} \kappa''^{-\frac{\rho''}{2}} \Gamma^{-1}\left(\frac{\kappa''}{2}\right) \Gamma\left(\frac{\rho''+\kappa''}{2}\right) \\ \times \left(1 + \frac{1}{\kappa''}(\beta - \mathbf{b}'')^T \frac{H''}{q''}(\beta - \mathbf{b}'')\right)^{-\frac{\rho''+\kappa''}{2}} \end{aligned} \tag{10.41}$$

and

$$p(\theta|\phi,\tau,\mathbf{y}) = \left(\frac{\kappa'' q''}{2}\right)^{\frac{\kappa''}{2}} \Gamma^{-1}\left(\frac{\kappa''}{2}\right) \theta^{\frac{\kappa''}{2}-1} \exp\left[-\frac{\theta}{2}\kappa'' q''\right]. \tag{10.42}$$

Equations (10.41) and (10.42) are obtained by integrating (10.21) w.r.t β and θ respectively. Equation (10.41) is a ρ''-multivariate Student distribution with κ'' degrees of freedom, scaling matrix $q'' H''^{-1}$ and mean $\mathbf{b}''$ while (10.42) is a Gamma-2 distribution with mean q''^{-1} and variance $2/(\kappa'' q''^2)$. Then we can rewrite Eqs. (10.39–10.40) as

$$\mathbb{E}\{\beta|\mathbf{y}\} = \iint_{\phi,\tau} \mathbf{b}'' p(\phi,\tau,\mathbf{y}) p(\phi,\tau|\mathbf{y})\, d\phi\, d\tau \tag{10.43}$$

and

$$\mathbb{E}\{\theta|\mathbf{y}\} = \iint_{\phi,\tau} \frac{1}{q''(\phi,\tau,\mathbf{y}) p(\phi,\tau|\mathbf{y})}\, d\phi\, d\tau. \tag{10.44}$$

Considering the discrete nature of $p(\phi,\tau|\mathbf{y})$, the two integrals of Eqs. (10.43) and (10.44) become weighted sums of pdfs, hence the posterior parameter expectations become weighted averages of the conditional parameter expectations:

$$\mathbb{E}\{\beta|\mathbf{y}\} = \sum_{\phi_k,\tau_k} p(\phi_k,\tau_k|\mathbf{y}) \mathbf{b}''(\phi_k,\tau_k,\mathbf{y}) \tag{10.45}$$

$$\mathbb{E}\{\theta|\mathbf{y}\} = \sum_{\phi_k,\tau_k} p(\phi_k,\tau_k|\mathbf{y}) \frac{1}{q''(\phi_k,\tau_k,\mathbf{y})}. \tag{10.46}$$

These equations form the foundation for our coverage prediction algorithm introduced in the following.

10.3 REM Construction

Based on partial information regarding the network coverage, the operator can use the Bayesian kriging technique, described above, to construct a full map over the area under analysis. To do so, we assume the existence of a REM manager, a software framework for carrying out such spatial data interpolation [21], located at the network's Operation and Maintenance (O&M) center. The REM manager has two main tasks, namely (1) to collect the received power information in the analyzed area, making use of the reports stored by the Trace Collection Entity (TCE) as well as other network performance estimation tools such as drive tests and measurement request to specific UEs; and (2) to perform the coverage estimation. With the Bayesian kriging technique, the operator will have an automated and remote representation of the network coverage, avoiding or minimizing drive tests and the expends and delays they imply. The coverage estimation algorithm proposed in this work assumes that the area of interest is spatially discretized. We consider a regular grid and we refer to the minimum area of the grid as a *pixel*. We assume that the measurements performed by any UE inside this pixel corresponds to only one location, for instance the center of the pixel. We define p as the minimum percentage of required measurements, which corresponds to the minimum percentage of pixels with available measurements. The coverage estimation algorithm can be summarized then in the following steps:

1. The REM manager sends measurement requests to the UEs until it collects a required percentage of pixels with available measurements, p, in order to perform the interpolation process. To achieve p, two options are possible: (1) the REM manager broadcasts the locations where the measurements are required, and only the UEs in these locations report their measurements; or (2) the REM manager requests that all the UEs report their locations with additional information such as velocity or battery life, and then chooses the UEs that should perform and report the measurement. While in the first option, the signaling overhead is minimized, an algorithm that compares the requested location to the current UE locations and makes the decision of performing and reporting the measurement, needs to be implemented by the UEs. The second option is simpler, from an implementation point of view, as the UEs are only asked to report their location and measurements. Besides, it is worth pointing out that the delay in waiting for the required UEs measurements will be always lower than the required time to perform and process the drive tests. Therefore, we can affirm that the delays associated with the automated methodology are not significant in comparison with those in manual coverage analysis.
2. Once the minimum required measurements are gathered, the REM manager performs the Bayesian kriging interpolation, presented in Sect. 10.2, to estimate

the signal power values in those points where it lacks information. Finally, the REM manager constructs the REM by overlapping the available real measurements with the interpolated ones.

The details of the REM functional architecture can be found in [22]. In [16], an initial application of the Bayesian kriging to coverage REM construction is presented, where a preliminary performance analysis on pixel-level coverage is also given. The proposed methodology does not imply any modification in the existing network entities. This makes its deployment very straightforward and cost-effective.

As discussed in the introduction, the work we present in this chapter is a continuation of the work presented in [16]. Differently from [16], we give here a more realistic evaluation of the REM-based coverage hole prediction, since in the present analysis a more realistic coverage hole definition is considered. In [16], a non-covered pixel was considered as a coverage hole, independently from the fact that its neighboring pixels were covered or not. However, in reality, a coverage hole frequently consists of *neighboring* uncovered pixels. Two pixels are considered neighbor pixels when there is at least a common edge between them. Thus, a set of neighboring N pixels constitute a single coverage hole, and not N separate coverage holes. Therefore, we define a coverage hole as N *neighboring* uncovered pixels, and we use this definition in performance evaluations. An important parameter to define is the minimum number of neighboring uncovered pixels which an operator considers as an area with coverage problems where some actions have to be taken.

10.3.1 Local Coverage Hole Analysis

In this section we present the local coverage hole analysis, the introduced solution for operator to analyze specific regions with coverage problems. A more detailed work in this subject can be found in [23].

The data we have used for the local coverage analysis consist of 3G received pilot powers, i.e. *Received Signal Code Power (RSCP)* values. The geo-located RSCP measurements are obtained with a very accurate planning tool which uses a sophisticated ray-tracing propagation model developed and used for operational network planning [24]. The propagation model uses specific environment information such as terrain profile, height, clutter, building data, etc. and is calibrated through repeated drive tests. Therefore, the RSCP data obtained from this tool and used in this chapter can be considered as real measurements reflecting the "ground-truth" on the coverage situation in the area of interest.

We consider an urban area of 2000 m $\times$ 2000 m in the south west of Paris, whose received signal power map is presented in Fig. 10.1. In what follows we refer to the map obtained from this RSCP data as the *real coverage map*. This map has a grid granularity of 25 m $\times$ 25 m.

We define a minimum RSCP coverage threshold of $\delta = -123$ dBm and those pixels where the received RSCP is below this threshold are considered to be uncovered.

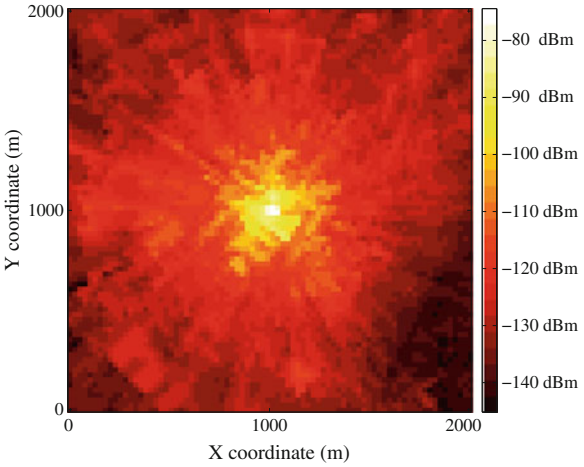

Fig. 10.1 RSCP real coverage map for a grid size of 25 m × 25 m

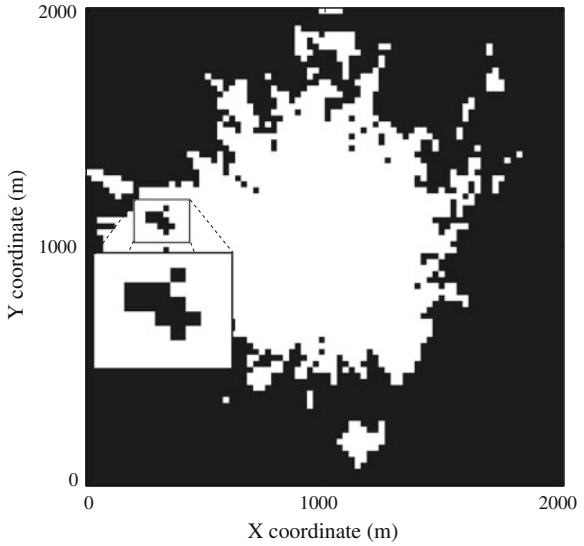

Fig. 10.2 RSCP coverage map used as the ground-truth

This threshold is chosen for specific operator coverage requirements for 3G networks. Figure 10.2 represents the binary map of the real coverage map, where the uncovered pixels are represented in black and the covered pixels are represented in white. Another important parameter to define is the minimum number of neighboring uncovered pixels which an operator considers as an area with coverage problems where some corrective actions have to be taken. Considering the environment (urban) and the grid size (25 m × 25 m), we define a *coverage hole* to be formed by the typical value of $N = 4$ neighboring pixels.

In dealing with coverage holes over a geographical zone of operation, usually a practical approach taken by operators is to prioritize (groups of) coverage holes rather than handling all the coverage holes at once. Therefore, those coverage holes that have the potential to have a larger (negative) impact on operator's prioritized targets are dealt with higher priority than the others. In this chapter, we adopt a similar *local* approach and consider a typical prioritization aspect by focusing on the *largest* coverage hole (i.e. coverage hole with the largest geographical area). Therefore, the proposed analysis is carried out on the neighboring area around the largest coverage hole, as can be seen Fig. 10.2.

Operators already have an idea on such neighboring areas thanks to traditional network diagnostics based on human-expert processing of alarm tickets, customer complaints and routine drive-test measurements collected from the neighborhood. When such an area is suspected to have a coverage hole, the traditional method to deal with this situation is to send out drive test equipment and experts specifically to that area to perform detailed/thorough drive tests and analyze the obtained measurements in order to: (1) detect the presence and (2) accurately identify the shape of the coverage hole for further (corrective) actions.

The REM-based analysis proposed in this chapter is meant to replace this phase of detailed/local drive test measurement collection and analysis, by using geo-located measurements reported by UEs present around the coverage hole. Although UEs cannot report *immediate* measurements when they are *within* a coverage hole, mechanisms such as *logged* Radio Link Failure (RLF) reporting exist in the MDT framework which allow the UEs to log the measurements when they lose the network connection and report the logged measurements as soon as they get reconnected [3]. Furthermore, measurements could be requested to UEs inside the coverage hole, attached to other technologies. In this way, valuable geo-located data related to the coverage loss can be used by the network operator. Thus, instead of performing a second round of (local and dedicated) drive test measurements and manually processing/analyzing those measurements, the operator constructs a remote representation of the local network coverage (i.e. REM) over the suspected area, hence minimizing the overall number of drive tests and the expenses/delays they imply.

In Fig. 10.3a, we can see the binary map of the larger coverage hole found in the real coverage map after applying the minimum RSCP coverage threshold of δ, and

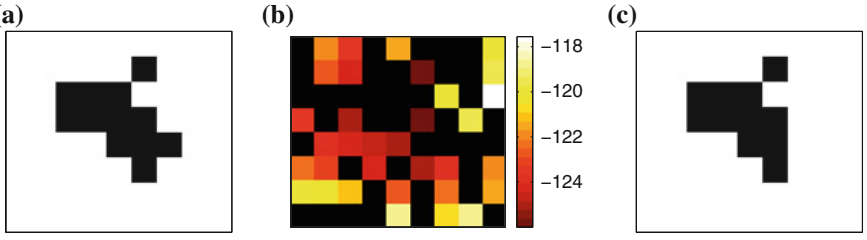

Fig. 10.3 Real and interpolated ground-truth maps. **a** Larger coverage hole. **b** Available measurements for interpolation. **c** Resulting map after interpolation

zoomed in Fig. 10.2. To obtain this map, let $\mathscr{M}$ denote the matrix representation of the real coverage map of the region under analysis with $\mathscr{M}(r, c)$ as the RSCP value at the rth row and the cth column ($r = 1, ..., R$; $c = 1, ..., C$). $\mathscr{P}$, with $\mathscr{P}(r, c)$ as its entry at the rth row and cth column ($r = 1, ..., R$; $c = 1, ..., C$), which is constructed as follows:

- Each value in $\mathscr{M}$, $\mathscr{M}(r, c)$, where $r \in R$ and $c \in C$ is compared to δ
- If $\mathscr{M}(r, c) < \delta$ we set $\mathscr{P}(r, c) = 0$
- If $\mathscr{M}(r, c) \geq \delta$ we set $\mathscr{P}(r, c) = 1$

Figure 10.3b represents the operator's (partial) view on the zoomed area: colored pixels show locations where the operator has network measurements and black pixels show locations where there are no measurements. The automated construction of REM uses spatial statistics, the Bayesian kriging presented in Sect. 10.2, to have a realistic representation of the ground-truth. Thus, the REM manager estimates the RSCP values in the black pixels and obtain the ground-truth map by adding them with the network measurements. Then, the minimum RSCP coverage threshold is applied to the REM, obtaining a map as the one presented in Fig. 10.3c, which shows the coverage hole estimated after one realization of the proposed local coverage analysis algorithm.

10.3.2 Global Coverage Estimation

In this section we explain the global coverage analysis, where the existent areas with coverage problems are determined remotely and automatically. In reference [25] a more comprehensive analysis is presented, including results for a rural scenario.

Considering that we are currently in the deployment phase of LTE technology, it is more relevant and timely to address coverage issues in LTE rather than in 3G, which has been in use since nearly 10 years now. Therefore, for the global coverage analysis, we align our coverage detection and prediction work with LTE. For this purpose, we translate/map the RSCP measurements in 3G, presented in the previous section, to their equivalent metric in LTE, namely the *Reference Signal Received Power* (*RSRP*).

RSCP is the received power on the Common Pilot Channel (CPICH) measured over the 5 MHz 3G carrier bandwidth, and the RSRP is the linear average of the received powers on the time-frequency resource elements that carry cell-specific reference signals over the 15 kHz LTE carrier bandwidth. Assuming that we use the same overall bandwidth (5 MHz) for both systems, we can apply the solution presented in [26] to obtain the wideband RSRP, which is the sum of the RSRP values measured over all reference subcarriers in the bandwidth. Due to the similarity between RSRP and RSCP, authors in [26] propose to compute the wideband RSRP based on the measured RSCP values through the following equation:

$$RSRP_{5\,\mathrm{MHz}} = RSCP + 10\log\left(\frac{RS_{\mathrm{Tx\ power}}}{CPICH_{\mathrm{Tx\ power}}}\right) + \Delta_{pl} + \Delta_{lb} \tag{10.47}$$

where Δ_{pl} is the over-the-air path loss difference between the carrier frequencies of the two technologies and Δ_{lb} is the link budget difference between LTE and 3G systems. Δ_{lb} includes Node B and device antenna gains, receiver's noise figure and feeder loss differences.

In the solution presented in [26], the authors assume the usual cell planning strategy for a new technology roll-out deployment, where the existent sites are reused, antennas have the same coverage size and beamwidths, with the same antenna azimuths and tilt settings. Therefore, it is assumed that transmitter and receiver antenna gains are the same for both technologies. The extension of results given by Eq. (10.47) to a 20 MHz LTE system is straightforward: we change the considered bias, which would imply also a change in the considered RSRP threshold.

For the global coverage hole analysis we consider the obtained RSRP values after applying the previous solutions for the same BS as for the local coverage analysis case. We consider a 1000 m × 1000 m region, whose received signal power map is presented in Fig. 10.4. We reduce the size of the map regarding the local coverage hole analysis, presented in previous section, to reduce the computational complexity of the algorithm, since the interpolation process is performed in the whole coverage area rather that in a particular region with coverage problems, as it was the case in the local analysis. Therefore, in the global analysis we focus on the coverage holes at the cell-edge.

Operators do not have the whole information on the ground-truth as in Fig. 10.4. Instead, they have a partial map where they lack coverage information in some pixels. Figure 10.5 represents an example of such a partial coverage map that the

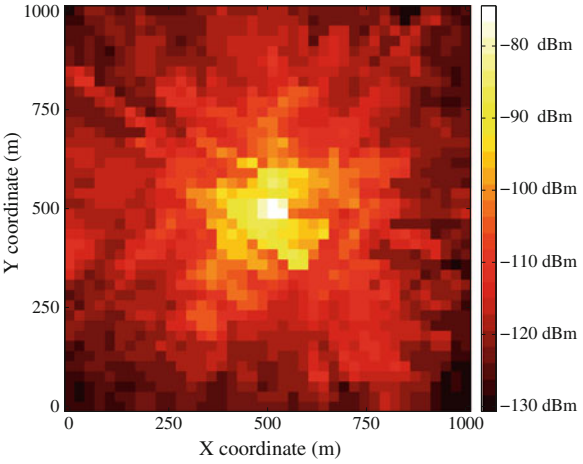

Fig. 10.4 RSRP real coverage map for a grid size of 25 m × 25 m

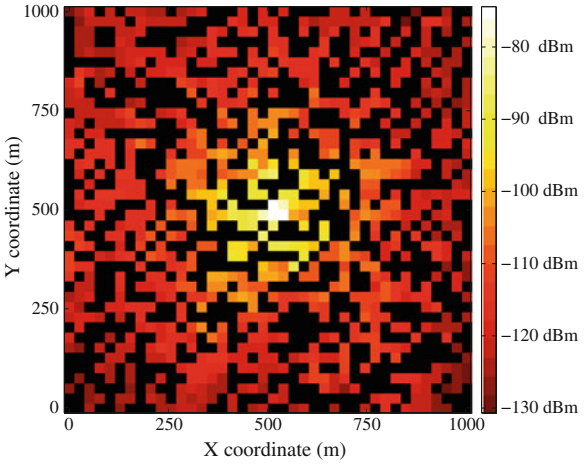

Fig. 10.5 RSRP measurements available for interpolation in the urban scenario

operator could potentially have, for $p = 50\,\%$ of pixels with available information. In Fig. 10.5, black pixels represent those grid areas where operator lacks the RSRP information. With such a map at hand, the operators carry out detailed analysis to perform coverage detection and prediction. These analysis are performed manually by human experts, combining information coming from other sources such as alarm tickets, customer complaints etc. Needless to say, this process is long, expensive and cumbersome. To add to these drawbacks, additional drive tests may be needed in cases/areas where the available information is deemed insufficient and the coverage problem too important. What we propose instead, is to replace this manual process by an efficient automated process that uses the available measurements, avoiding or minimizing drive tests together with the expenses and delays they imply. The proposed automated process applies Bayesian kriging to the partial map of Fig. 10.5 to obtain a REM which provides RSRP predictions in the black pixels. The obtained REM is then used for the global coverage analysis.

We define a minimum RSRP threshold $\delta = -124\,\text{dBm}$ and those pixels where the received RSRP is below this threshold are considered to be out of coverage. Similarly as for the 3G network case presented in the previous section, this threshold is chosen for specific operator network coverage requirements for a 4G network but it is not related to Eq. 10.47. Applying this threshold to the real coverage map we obtain a binary map, $\mathscr{P}$, of Fig. 10.6, where the uncovered pixels are represented in black, and the covered pixels are represented in white, is obtained following the same algorithm as explained in previous section. As for the local analysis, in this case we also consider that a coverage hole is formed by the minimum typical value of $N = 4$ neighboring pixels.

To determine the existent coverage holes we construct a new matrix $\mathscr{N}$, where at the end all the values will be zeros, except those determined as belonging to a coverage

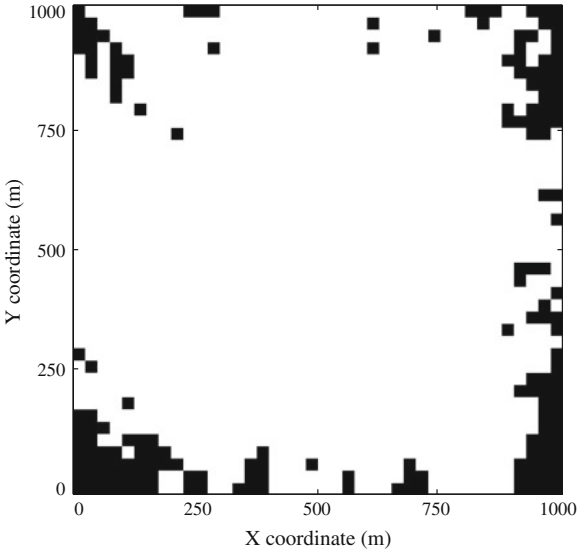

Fig. 10.6 Binary coverage map for δ for the urban scenario

hole. The algorithm we have developed to analyze the global binary coverage matrix, $\mathscr{P}$, and generate matrix $\mathscr{N}$ consist in the following steps.

- The analysis is performed row by row. Those uncovered pixels without an left-hand side neighbor uncovered pixel, are consecutively numerated.
- In that row, those pixels in the right-hand side of an uncovered pixel take the right-hand side pixel number.
- The indexes of the uncovered pixels are stored in *current row* variable.
- Once the analysis of a row is finished, we compare the positions of the pixels detected in that row, stored in *current row* variable, with the values in *previous row* variable, which stores the positions of the uncovered pixels detected in the previous row, in order to determine which pixels belong to the same coverage hole. Finally, the *current row* indexes become the *previous row* and the process is repeated for the next row in the matrix.

10.4 Coverage Analysis and Simulation Results

The performance evaluation presented in the section, first corresponds to the interpolation process performance, second to the coverage hole detection for the local analysis and third to the coverage hole detection for the global analysis. In particular, the results presented in this chapter are obtained for a random selection of the pixels where the operator lacks information for reasons of simplicity. In reality, the measurement locations are expected to follow certain mobility patterns. The incorporation of

realistic mobility patterns in measurement locations is part of our ongoing research work. In what follows we first introduce the performance evaluation metrics used.

10.4.1 Performance Evaluation

To present the gains obtained by the REM in coverage analysis, we define the following metrics, where predicted coverage hole refers to a coverage hole detected in the REM, and real coverage hole refers to those coverage holes in the real map. The following metrics consider the coverage hole concept, of N uncovered neighboring pixels.

- *Coverage hole detection probability* is the probability for a real coverage hole to be correctly detected, as a function of c. This probability is evaluated by dividing at each snapshot, the number of detected coverage holes by the total number of real coverage holes, and averaging the resulting ratio over the iterations.
- *Pixel detection probability* is the detection probability of the pixels forming the coverage holes. We estimate the amount of pixels forming a coverage hole that are detected when considering only network measurements and the increment introduced by the use of REM.
- *The average number of detected coverage holes* is the number of the real coverage holes that are correctly detected, averaged over the iterations. A coverage hole is considered as correctly detected if $c\,\%$ of the pixels of this coverage hole belong to a predicted coverage hole (so called in the reminder the "corresponding" real coverage hole).

Also, to present the results we use the following definitions:

- Average percentage of measured pixels, which is equal to the number of coverage hole pixels measured by the operator, divided by the number of pixels forming the coverage hole.
- Average percentage of detected pixels with REM which is equal to the number of coverage hole pixels measured by the operator plus those (uncovered) pixels correctly estimated by the REM, divided by the number of pixels forming the coverage hole.

These average percentage values are obtained over a statistically significant number of independent "snapshots" where at each snapshot, the available measurement pixels are uniformly chosen. Results presented for the local and the global coverage analysis have been calculated for different percentages of available measurements, $p = \{50, 60, 70, 80, 90\}\,\%$ used in the interpolation process. Having 80–90 % of measurements may seem excessive. However, when we translate p into a more tangible metric, such as the number of measurements per square meter, we obtain $q = \{0.08, 0.096, 0.112, 0.128, 0.144\}$ UE measurements per square meter, which is a small amount due to the static nature of the problem and the size of the grid.

Table 10.1 Accuracy of predictions

Measures (dB)	$p = 50\,\%$	$p = 80\,\%$
80 % errors	−5 to 3	−2 to 4
MAE range	2.55 to 3.25	2.00 to 3.05
MAE peak	2.75	2.45

10.4.2 Bayesian Kriging Accuracy Evaluation

We evaluate the Mean Absolute Error (MAE) obtained when performing the interpolation process in those pixels without signal power information. The MAE measures the average magnitude of the errors in a set of forecasts, without considering their direction. We compare the obtained MAE when $p = \{50, 80\}\,\%$ measurements and we summarize the obtained results in Table 10.1.

As expected, when "low" amount of measurements are available, i.e. $p = 50\,\%$, the obtained errors are higher than for the case of "high" amount of available measurements, i.e. $p = 80\,\%$, since the interpolation process is more accurate. In particular, when $p = 50\,\%$, the MAE goes from around 2.55 to 3.25 dB with a peak in 2.75 dB, and when there is a higher number of available measurements, $p = 80\,\%$, the MAE goes from 2 to 3.05 dB with a peak in 2.45 dB. Therefore, when p is increased form 50 to 80 %, the histogram moves to the left. Thus, we can affirm that in average the accuracy of predictions improves by 0.3 dB. When analyzing the distribution of the error for $p = \{50, 80\}\,\%$ measurements in a typical realization we obtain an 80 % of the errors distributed between −5 and 3 dB for $p = 50\,\%$ and between −2 and 4 dB for $p = 80\,\%$.

10.4.3 Coverage Hole Detection: Local Analysis

In this section we evaluate the accuracy of the proposed pixel-wise methodology for realistic coverage hole detection. Due to the uniform distribution followed to select the pixels with available measurements, the average percentage of measured pixel values correspond almost exactly to the percentage of available measurement values. As the number p of measurements available for the interpolation process increases, the average number of the coverage hole pixels known by the operator also increases, as presented in Table 10.2. The REM introduces a 20 % of additional knowledge on the coverage hole pixels to the network measurements, when $p = 50\,\%$. The knowledge on the coverage hole pixels reaches an approximate value of 94 % for $p = 90\,\%$ when REM is used.

Comparing the probability of detecting a coverage hole for the cases when the REM is constructed and when only network measurements are used, we can conclude that the detection probability significantly increases with the use of REM. Table 10.3 summarizes the coverage hole detection probability for both cases for the highly

Table 10.2 Average known coverage hole pixels for only network measurements and REM

p (%)	Av. percentage of measured pixels	Av. percentage of detected pixels with REM
50	50.1	71.1
60	59.9	76.6
70	70	82.4
80	80	88.2
90	90.1	93.7

Table 10.3 Average probability of coverage hole detection for only network measurements and REM

p (%)	$c = 90\,\%$ network measured	$c = 90\,\%$ REM
50	10.7	44.7
60	23.4	56.8
70	41.5	70
80	65.5	83.6
90	88.7	94.8

demanding requirement of coverage hole detected pixels $c = 90\,\%$. For a low amount of measurements, $p = 50\,\%$, the coverage hole detection probability increases 34 % when REM is used and for high amount of available measurements, $p = 90\,\%$, the coverage hole detection probability still increases by about 6 % with the REM. Furthermore, a very striking fact is that, for achieving a 70 % coverage hole detection probability, the REM requires 70 % of available measurements, whereas without the REM, we need more than 80 % of measurements to achieve the same detection probability. Thus, REM saves up more than 10 % of measurements for the same performance level of coverage hole detection.

10.4.4 Coverage Hole Detection: Global Analysis

In this section we present the results obtained for the global coverage analysis introduced in Sect. 10.3.2. Table 10.4 presents the average number of detected coverage holes, when using only network measurements and when using the REM. The number of real coverage holes for the analyzed scenario is 10. Results presented in Table 10.4 were obtained for $c = 70\,\%$. As it was expected, when the REM is used, the predicted coverage holes increase with the number of measurements used in the interpolation process, p. These results give the notion of the average misdetection in the coverage hole detection, since the difference between the number of real coverage holes and the average number of detected coverage holes are the average real coverage holes misdetected.

Table 10.4 Comparison of existent and predicted coverage holes with and without the REM in the urban scenario

p (%)	Network measurements	REM
50	3.93	5.55
60	5.6	7.36
70	8.11	8.97
80	9.4	9.66
90	9.96	10

Table 10.5 Average probability of coverage hole detection for different percentages c of pixels for which measurements are available (see end of Sect. 10.4.1), when using REM in the urban scenario

p (%)	REM coverage hole detection (%)		
	$c = 50\,\%$	$c = 70\,\%$	$c = 90\,\%$
50	83.5	55.5	24.1
60	91.2	73.6	32.1
70	97.8	89.7	47.2
80	99.1	96.6	66
90	100	100	89.9

Specifically, in Table 10.4, it can be observed that in average, more than the half of the coverage holes are predicted when the REM is applied, even for the case of low amount of available measurements, $p = 50\,\%$. For this case, the prediction probability increases in more than 15 % when the REM is applied, in comparison to the case when only measurements are used.

Table 10.5 presents the coverage hole detection probability for $c = \{50, 70, 90\}\%$ when the REM is used. It can be observed that, when half of the coverage hole pixels are required to be detected, $c = 50\,\%$, the coverage hole is detected in more than 80 %, even for low amounts of available measurements, i.e. $p = 50\,\%$. It can be observed that for the highly demanding case of $c = 90\,\%$, the coverage hole is detected in more than 45 % of the snapshots when p is above 70 %, and the detection probability doubles for $p = 90\,\%$.

10.5 Conclusions

In this chapter we have presented a cognitive tool, the REM, for automated and remote coverage hole prediction. This cognitive tool uses geolocated measurement data and exploits techniques from spatial statistics, specifically, the Bayesian kriging interpolation method. We have presented detailed mathematical derivations for the developed interpolation method and we have evaluated the prediction performance of the interpolation process for cellular coverage prediction. Then, we have introduced

the concept of coverage hole to perform the coverage analysis. We presented two coverage estimation algorithms, one for local analysis and one for the global coverage estimation. The proposed REM-based solutions have been shown to improve the coverage hole detection probability with regards to the case when only network measurements are considered. We therefore can affirm that the proposed solutions allow a remote and automated cellular coverage prediction, that enhances the accuracy of the coverage hole detection with a limited required number of measurements.

The proposed automated solution replaces the long and expensive task of manual coverage hole analysis and it allows the operators to rapidly deploy solutions which overcome the coverage problem. The obtained results demonstrate that the REM-based automated coverage detection and prediction is a promising approach for future cognitive cellular networks, which is definitely worth further investigation.

References

1. Tomala, M., Keskitalo, I., Bodog, G., Sartori, C.: Supporting Function: Minimisation of Drive Tests (MDT), pp. 267–310. Wiley, Chichester (2011)
2. 3GPP TR 36.805 v1.3.0 1 study on minimization of drive-tests in next generation networks; (release 9). Technical report, 3GPP organization (2009)
3. 3GPP TS 37.320 v11.1.0 universal terrestrial radio access (UTRA) and evolved universal terrestrial radio access (E-UTRA); radio measurement collection for minimization of drive tests (MDT); overall description; stage 2 (release 11). Technical report, 3GPP organization (2012)
4. Hapsari, W.A., Umesh, A., Iwamura, M., Tomala, M., Gyula, B., Sebire, B.: Minimization of drive tests solution in 3GPP. Commun. Mag., IEEE **50**(6), 28–36 (2012)
5. Holma, H., Toskala, A.: LTE Advanced: 3GPP Solution for IMT-Advanced. West Sussex, UK (2012)
6. NGNM Alliance: NGMN recommendation on SON and O&M requirements. White paper, Next Generation Mobile Networks (2008). http://www.ngmn.org/uploads/media/NGMN_Recommendation_on_SON_and_O_M_Requirements.pdf. Available online (53 pages)
7. Mitola, J.: Cognitive radio: an integrated agent architecture for software defined radio. Doctoral dissertation, Royal Institute Technology (KTH), Stockholm (2000)
8. Krenik, W., Batra, A.: Cognitive radio techniques for wide area networks. In: Proceedings of the 42nd annual Design Automation Conference, (DAC '05), pp. 409–412. ACM, New York (2005). http://doi.acm.org/10.1145/1065579.1065688
9. Zhao, Y., Le, B., Reed, J.H.: Network support—the radio environment maping. In: Fette B.A. (ed.) Cognitive Radio Technology. Elsevier, Amsterdam (2006)
10. Alaya-Feki, A.B.H., Sayrac, B., Jemaa, S.B., Moulines, E.: Interference cartography for hierarchical dynamic spectrum access. In: Proceedings of the 3rd IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN) October (2008)
11. Alaya-Feki, A., Ben Jemaa, S., Sayrac, B., Houze, P., Moulines, E.: Informed spectrum usage in cognitive radio networks: interference cartography. In: Proceedingsof the IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), September (2008)
12. Riihijärvi, J., Mähönen, P., Wellens, M., Gordziel, M.: Characterization and modelling of spectrum for dynamic spectrum access with spatial statistics and random fields. In: Proceedings of the IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC) (2008)

13. Kitanidis, P.: Parameter uncertainty in estimation of spatial function: bayesian analysis. Water Resour. Res. **22**(4), 507 (1986)
14. Raiffa, H., Schlaifer, R.: Applied Statistical Decision Theory. New york, USA (2000)
15. Grimoud, S., Sayrac, B., Ben Jemaa, S., Moulines, E.: An algorithm for fast REM construction. In: Proceedings of the 6th International ICST Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM), pp. 251–255 (June 2011)
16. Sayrac, B., Riihijärvi, J., Mähönen, P., Ben Jemaa, S., Moulines, E., Grimoud, S.: Improving coverage estimation for cellular networks with spatial bayesian prediction based on measurements. In: Proceedings of the 2012 ACM SIGCOMM Workshop on Cellular Networks: Operations, Challenges, and Future Design, Cell Net '12, pp. 43–48 (2012)
17. Gudmundson, M.: Correlation model for shadow fading in mobile radio systems. Electron. Lett. **27**, 2145–2146 (1991)
18. Jo, H.-S., Yook, J.-G.: Path loss characteristics for IMT-advanced systems in residential and street environments. IEEE Antennas Wirel. Propag. Lett. **9** 867–871 (2010)
19. Senarath, G., et al.: Multi-hop relay system evaluation, methodology (2007)
20. Sayrac B., Galindo-Serrano A., Ben Jemaa S., Riihijärvi J., Mähönen P.: Bayesian spatialinterpolation as an emerging cognitive radio application for coverage analysis in cellularnetworks. Transactions on Emerging Telecommunications Technologies (Submitted)
21. Final system architecture.: Technical report deliverable D2.4, FARAMIR Project (2011). http://www.ict-faramir.eu/fileadmin/user_upload/deliverables/FARAMIR-D2.4-Final.pdf
22. Van de Beek, J. et al., How a layered REM architecture brings cognition to todays mobilenetworks. IEEE Wirel. Commun. **19**(4), 17–24 (2012)
23. Galindo-Serrano, A., Sayrac, B., Ben Jemaa, S., Riihijärvi, J., Mähönen, P.: Automated coverage hole detection for cellular networks using radio environment maps. In: Proceedings of the 9th International Workshop on Wireless Network Measurement (WiNMee 2013), in conjuction with WiOPT (2013)
24. ASSET: Website. http://www.aircominternational.com/Products/planning/asset.aspx. Accessed 05 June (2013)
25. Galindo-Serrano A., Sayrac B., Ben Jemaa S., Riihijärvi J., Mähönen P.: Harvesting MDTdata: radio environment maps for coverage analysis in cellular networks. In: Proceedings ofthe 8th International Conference on Cognitive Radio Oriented Wireless Networks (CROWNCOM) (2013)
26. Landre, J.-B., Rawas, Z.E., Visoz, R., Bouguermouh, S.: Realistic performance of LTE: In a macro-cell environment. In: Proceedings of the IEEE 75th Vehicular Technology Conference (VTC Spring), May (2012)
27. Cressie N., Johannesson G.: Fixed rank kriging for very large spatial data sets. J. R. Stat. Soc. Series B (Statistical Methodology) **70**(1), 209–226 (2008)
28. Cressie N., Shi T., Kang E.: Fixed rank filtering for spatio-temporal data. J.Comput. Graph. Stat. **19**(3), 724–745 (2010)
29. Riihijärvi J., Nasreddine J., Mähönen P.: Demonstrating radio environment map construction-from massive data sets. In: Proceedings of IEEE DySPAN 2012 (demonstrations track)(2012)

Chapter 11
Use of Learning, Game Theory and Optimization as Biomimetic Approaches for Self-Organization in Heterogeneous Networks

Ali Imran and Lorenza Giupponi

Abstract In this chapter, we present the use of several bio inspired approaches called biomimetics for the design of Self-organization (SO) in heterogeneous network scenarios. Mainly these approaches are further categorized in indirect biomimetics and direct biomimetics depending on whether the inspiration from the biological systems is used indirectly or it is applied as it is to design SO. Under the umbrella of indirect biomimetics we discuss in detail the emerging paradigms in learning theory that have been recently shown to have strong potential for designing SO solution in heterogeneous networks. In the second part of the chapter, we investigate a rather under explored paradigm of direct biomimetic. Building on a case study of a self-organising systems in nature we extract generic SO design principles that can be used as a direct biomimetic approach for designing distributed, scalable and agile solutions to many problems in complex heterogamous networks. We demonstrate this direct biomimetic approach through a use case of a heterogeneous network scenario with outdoor fixed relays. By exploiting one to one mapping between a natural SO system and our system model we systematically apply the bio-inspired design principle directly and obtain a distributed SO solution to the problem under consideration. The performance of this solution is evaluated through numerical results and substantial gains are observed. Finally we conclude this chapter with remarks on some important considerations and limitations of the use of biomimetic approaches.

A. Imran (✉)
Qatar Mobility Innovations Center, Tech 2, Qatar Science and Technology Park, Doha, Qatar
e-mail: alii@qmic.com

L. Giupponi
Centre Tecnolgic de Telecommunicacions de Catalunya (CTTC), Barcelona, Spain
e-mail: lorenza.giupponi@cttc.es

M.-G. Di Benedetto and F. Bader (eds.), *Cognitive Communication and Cooperative HetNet Coexistence*, Signals and Communication Technology,
DOI: 10.1007/978-3-319-01402-9_11,

Acronyms

BS	Base Station
BSOF	Biomimetic Self Organisation Framework
CCDF	Complementary Cumulative Distribution Function
ITW	Iterative Waterfilling
LTE	Long Term Evolution
ML	Machine Learning
MDP	Markov Decision Process
OFR	Outdoor Fixed Relay
OFDMA	Orthogonal Frequency Division Multiple Access
RB	Resource Block
RRM	Radio Resource Management
RL	Reinforcement Learning
RNC	Radio Network Controller
SIR	Signal to Interference Ratio
SINR	Signal to Noise and Interference Ratio
TD	Time Difference
UTRAN	UMTS Terrestrial Radio Access Network
SO	Self-organization

11.1 Introduction to Biomimetics: A Bio Inspired Design Approach

Nature contains plethora of enormously complex systems that are quite perfect in their design and operation. Biomimetic is a recently evolved branch of science that investigates such natural systems with aim to exploit their working principles for improvement in design and operation of man made systems.

In nature, there are myriad of examples of self-organizing behavior, for example, ant colonies finding shortest routes to food sources, termites collectively building complex constructions without using a blueprint, fish schools organizing themselves without a leader, and swarms of fireflies in south-east Asia synchronously emitting light flashes. The fact the SO is originally a bio inspired phenomenon and the abundance of SO in biological systems makes Biomimetics a perfect paradigm for investigating the constituents and working principles of SO with aim to design SO in engineering systems through these inspirations. There are two different approaches in Biomimetics: the direct and the indirect approach. We use both of them in this chapter and therefore explain them briefly below:

- *Direct Biomimetics*: In the direct (or top-down) approach, an engineering problem is tackled by looking for natural systems solving an equivalent problem. The biological solution and its principles are then analyzed and re-built in a technical application. Examples of the direct approach are the design of airplane wings that

directly copy the gliding flight of birds or design of camera that copies design of human eye etc.

- *Indirect Biomimetics*: In contrast, the indirect (or bottom-up) approach of bio-inspired design involves, first, the derivation of principles by analyzing natural systems. The principle is then abstracted from its biological context and used in the technical applications where it could be suitable. Examples of such indirect approach are the concept of artificial intelligence, which attempts to exploit the human learning behavior, or the concept of game theory, which aims to exploit the findings from dynamics of a free economy in various engineering applications.

The rest of this chapter is organized as follows. In Sect. 11.2, we discuss a key indirect biomimetic approaches, i.e., learning and discuss some useful taxonomy, besides some interesting result. In learning theory, we particularly focus on an emerging paradigm of docitive learning and demonstrate how it is more promising in heterogeneous network scenario compared to conventional learning techniques. While the indirect approaches are currently more popular in the literature for developing solutions for wireless systems, in Sect. 11.3, we present a direct biomimetic approach for developing SO solutions for heterogeneous networks. To this end, we establish one to one analogy of our system model with a SO system in nature that addresses a problem analogous to ours. Based on this analogy we demonstrate the novel approach of developing SO solution for heterogeneous network scenario, by use of direct biomimetics. The presented solution is developed for capacity enhancement on the backhaul access links for the Outdoor Fixed Relay (OFR) scenario, by SO of macro Base Station (BS) antenna tilts. Finally, Sect. 11.4 summarizes the main conclusions of this chapter.

11.2 Indirect Biomimetic Approaches

We introduce in this section a set of algorithms that can be used to solve RRM and self-organizing problems, they are all based on learning theory. We discuss the appropriateness of using them for different problems, we introduce some basic technical details and we introduce a useful taxonomy.

11.2.1 Overview of Learning Theory Based Approaches

A particularly interesting framework in realistic decentralized wireless networks is the literature of Reinforced Learning (RL). The reason is that RL provides model free and online learning features, which makes it suitable for taking decisions in realistic wireless settings characterized by a high degree of dynamism due to e.g., lognormal shadowing, fast fading, mobility of users, multiuser scheduling, random femtocell nodes activity patterns, etc. The majority of studies that can be encountered in RRM

literature, applying RL techniques, are formulated for centralized settings where all decisions are taken by a single entity e.g., Radio Network Controller (RNC) in UTRAN systems. RL in single agent systems possesses a firm foundation in the theory of Markov Decision Processes (MDP) and can be shown to optimally perform. Small cell systems, however, cannot be formulated through a centralized learning process due to their deployment model, scalability and signaling overhead constraints. We therefore focus on decentralized learning processes based on RL. In this case, (1) the intelligent decisions are made by multiple intelligent and uncoordinated nodes; (2) the nodes partially observe the overall scenario; and (3) their inputs to the intelligent decision process are different from node to node since they come from spatially distributed sources of information. This distributed system can be mapped onto a multi-agent system, which consists of multiple nodes who are similarly and simultaneously adapting. The topic of learning in distributed systems has actually been studied in game theory since 1951 when Brown proposed the fictitious play algorithm [1]. The underlying assumption of fictitious play is that an agent assumes that its opponents sample the actions from some fixed distribution, i.e. at each time step opponents use a stationary mixed strategy. Then, each agent in the game estimates its opponents strategies by keeping a score of the appearance frequencies of the different actions. This means that, each agent needs to know the strategies followed by the other players in the game.

In Machine Learning (ML), the literature of single agent learning is extremely rich, while it is only in recent years that attention has been focused on distributed learning aspects, in the context of multiagent learning. It has been yielding some enticing results, being arguably a truly interdisciplinary area and the most significant interaction point between computer science and game theory communities. The theoretical framework to formulate RL problems can be found in MDP for the single agent system, and in stochastic games, for a multiagent system. In what follows, we give a brief introduction of learning in single and multiagent systems.

11.2.2 Learning in Single-Agent Systems

A MDP provides a mathematical framework for modeling decision-making processes in situations where outcomes are partly random and partly under the control of the decision maker. A MDP is a discrete time stochastic optimal control problem. Here, operators take the form of actions, i.e. inputs to a dynamic system, which probabilistically determine successor states. A MDP is defined in terms of a discrete-time stochastic dynamic system with finite state set $\mathscr{S} = \{s_1, \ldots, s_k\}$. Time is represented by a sequence of time steps, $t = 0, 1, \ldots, \infty$. At each time step, a controller observes the system's current state and selects an action, which is executed by being applied as input to the system. Let us assume that s is the observed state, and that the action is selected from a finite set of admissible actions $\mathscr{A} = \{a_1, \ldots, a_l\}$. When the controller executes action $a \in \mathscr{A}$, the system state at the next step changes from s to v, with a state transition probability $P_{s,v}$. We further assume that the

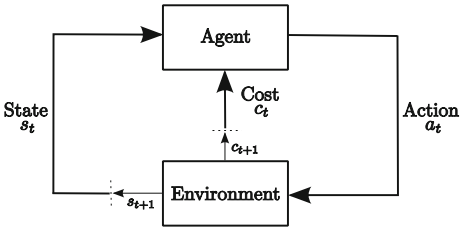

Fig. 11.1 Learner-environment interaction

application of action a in state s incurs an immediate cost $c(s, a)$. When necessary, we refer to states, actions, and immediate costs by the time steps at which they occur, by using s_t, a_t and c_t, where $a_t \in \mathscr{A}$, $s_t \in \mathscr{S}$ and $c_t = c(s_t, a_t)$ are, respectively, the state, action and cost at time step t. A graphic representation of the learner-environment interaction is shown in Fig. 11.1, here by the use of c_{t+1} and s_{t+1} we aim to emphasize that, as consequence of the performed action at time t, a_t, in the next time step, the agent receives a cost c_{t+1} and finds itself in a new state $v = s_{t+1}$. To sum up, a MDP consists of:

- a set of states $\mathscr{S}$.
- a set of actions $\mathscr{A}$.
- a cost function $C : \mathscr{S} \times \mathscr{A} \to \mathbb{R}$.
- a state transition function $P : \mathscr{S} \times \mathscr{A} \to \Pi(\mathscr{S})$, where a member of $\Pi(\mathscr{S})$ is a probability distribution over the set $\mathscr{S}$ (i.e. it maps states to probabilities).

The state transition function probabilistically specifies the next state of the environment as a function of its current state and the agent's action. The cost function specifies expected instantaneous cost as a function of current state and action. The model is a *Markov* model if the state transitions are independent of any previous environment states or agent actions. The objective of the MDP is to find a policy that minimizes the cost of each state s_t. As a result, the aim is to find an optimal policy for the infinite-horizon discounted model, relying on the result that, in this case, there exists an optimal deterministic stationary policy [2].

RL problems model the world using MDP formulism. In the literature, three ways have been identified to solve RL problems. The first one consists of the knowledge of the state transition probability function from state s to state v, $P_{s,v}(a)$, and is based on dynamic programming. The second and third forms to solve RL problems, on the other hand, do not rely on this previous knowledge and are based on Monte Carlo and Time Difference (TD) methods. As a result, Monte Carlo and TD are primarily concerned with how an agent ought to take actions in an environment so as to minimize the notion of long-term cost, that is, so as to obtain the optimal policy, when the state transition probabilities are not known in advance. When state transition probability is not known, but a sample transition model of states, actions and costs can be built, Monte Carlo methods can be applied to solve the MDP problem. On the other hand, if the only way to collect information about the environment is to interact with it, TD methods have to be applied. TD methods combine elements of

dynamic programming and Monte Carlo ideas, they learn directly from experience which is a characteristic of Monte Carlo methods and they gradually update prior estimate values, which is common of dynamic programming. TD methods allow an online learning which is crucial for long-term/continuos applications.

RL algorithms are based on the computation of *value functions*, i.e. the state-value function, $V(s)$, or the state-action value function, $Q(s, a)$, which measure how good, based on the future expected cost, is for an agent to be in a given state or to execute an action in a given state, respectively. The expected costs for the agent in the future are given by the actions it will take and therefore, the value functions depend on the policies being followed. The state-value of state s is defined as the expected infinite discounted sum of costs that the agent gains if it starts in state s and then executes the complete decision policy π,

$$V^{\pi}(s) = E_{\pi}\left\{\sum_{t=0}^{\infty} \gamma^t c_t \mid s_t = s\right\} \tag{11.1}$$

where $0 \leq \gamma < 1$ is a discount factor which determines how much expected future costs affect decisions made now.

Similarly, the Q-value $Q(s, a)$ represents the expected decreased cost for executing action a at state s and then following policy π thereafter.

$$Q^{\pi}(s, a) = E_{\pi}\left\{\sum_{t=0}^{\infty} \gamma^t c_t \mid s_t = s, a_t = a\right\} \tag{11.2}$$

Solving a RL problem means to find the best return in the long term. This is defined as finding an optimal policy, which is the one giving minimum expected return. We define the optimal value of state s as:

$$V^*(s) = \min_{\pi} V^{\pi}(s) \tag{11.3}$$

According to the principle of Bellman's optimality [2], the optimal value function is unique and can be defined as the solution to the equation:

$$V^*(s) = \min_{a}\left(C(s, a) + \gamma \sum_{v \in \mathscr{S}} P_{s,v}(a) V^*(v)\right) \tag{11.4}$$

which asserts that the value of state s is the expected cost $C(s, a) = \mathbb{E}\{c(s, a)\}$, plus the expected discounted value of the next state, v, using the best available action. Given the optimal value function, we can specify the optimal policy as:

$$\pi^*(s) = \arg\min_{a}\left(C(s, a) + \gamma \sum_{v \in \mathscr{S}} P_{s,v}(a) V^*(v)\right) \tag{11.5}$$

Applying the Bellman's criterion in the action-value function, first we have to find an intermediate minimum of $Q(s,a)$, denoted by $Q^*(s,a)$, where the intermediate evaluation function for every possible next state-action pair (v,a') is minimized, and the optimal action is performed with respect to each next state v. $Q^*(s,a)$ is:

$$Q^*(s,a) = C(s,a) + \gamma \sum_{v \in \mathscr{S}} P_{s,v}(a) \min_{a' \in \mathscr{A}} Q^*(v,a') \tag{11.6}$$

Then, we can determine the optimal action a^* with respect to the current state s. In other words, we can determine π^*. Therefore, $Q^*(s,a^*)$ is minimum, and can be expressed as:

$$Q^*(s,a^*) = \min_{a \in \mathscr{A}} Q^*(s,a) \tag{11.7}$$

11.2.3 Learning for Multiagent Systems

The characteristics of the distributed learning systems, as mentioned before, are as follows: (i) the intelligent decisions are made by multiple intelligent and uncoordinated nodes; (ii) the nodes partially observe the overall scenario; and (iii) their inputs to the intelligent decision process are different from node to node since they come from spatially distributed sources of information. These characteristics can be easily mapped onto a multiagent system, where each node is an independent intelligent agent. The theoretical framework is found in stochastic games [3] described by the five-tuple $\{\mathscr{N}; \mathscr{S}; \mathscr{A}; P; C\}$. Here, $|\mathscr{N}| = N$ is the set of agents, indexed $1, 2, \ldots, N$; $\mathscr{S} = \{s_1, s_2, \ldots, s_k\}$ is the set of possible states, or equivalently, a set of N-agent stage games; $\mathscr{A}$ is the joint action space defined by the product set $\mathscr{A}^1 \times \mathscr{A}^2 \times \cdots \times \mathscr{A}^N$, where $\mathscr{A}^f = \{a_1^f, a_2^f, \ldots, a_l^f\}$ is the set of actions (or pure strategies) available to the fth agent; P is a probabilistic transition function defining the probability of migrating from one state to another provided the execution of a certain joint action or, equivalently, it specifies the probability of the next stage game to be played based on the game just played and the actions taken in it; $C = \{c^1 \times c^2 \times \cdots \times c^N\}$, where c^f is the immediate cost of the fth agent in a certain stage of the game, which is a function of the joint actions of all N nodes.

In multiagent systems, the distributed decisions made by the multiple nodes strongly interact among each other. These kind of problems are usually modeled as non-cooperative games. The simplest and most common interpretation of a non-cooperative game is that there is a single interaction among players ("one-shot"), after which the payoffs are decided and the game ends. However, many, if not all strategic endeavors occur over time, and in a state dependant manner. That is, the games, and so the environment in which the nodes make decisions progress over time, passing through an infinite number of states, and the current game is decided based on the history of the interactions. Stochastic games form a natural model for such interactions. A stochastic game is played over a state space, and is played in rounds. In each round, each player chooses an available action simultaneously with

and independently from all other players, and the game moves to a new state under a possible probabilistic transition relation based on the current state and the joint actions [3]. We distinguish in this context two different forms of learning. On the one hand, the agent can learn the opponent's strategies, so that it can then devise a best response. Alternatively, the agent can learn a strategy of his own that does well against the opponents, without explicitly learning the opponent's strategies. The first approach is sometimes referred to as model-based learning, and it requires at least some partial information of the other players strategies. The second approach is referred to as model-free learning, and it does not necessarily require to learn a model of the strategies played by the other players.

We will discuss in the following a very partial sample of multiagent learning techniques, which we consider representative for the aim of this taxonomy:

- *Model-based approaches*: This approach, generally adopted in game theory literature, is based on building some model of the other agents strategies, following which, the node can compute and play the best response strategy. This model is then updated based on the observations of their actions. As a result, these approaches require knowledge or observability of the other agents' strategies, which may pose severe limits from the feasibility point of view in terms of information availability and signalling overhead. The best known instance of this scheme is fictitious play [1], which is a static game that simply counts the plays of the other agents in the past. Different variations of the original schemes exist, for example those considering that the agent does not play the exact best response, but assigns a probability of playing each action. Other algorithms in literature that can be classified into this group are the Metastrategy [4] and the Hyper-Q algorithms [5].
- *Model-free approaches*: A completely different approach, commonly considered by the AI literature, is the model-free approach, also known as TD learning, which avoids building explicit models of other agents' strategies. Instead, over time, each agent learns how properly the various available actions work in the different states. TD methods typically keep memory of the appropriateness of playing each action in a given state by means of some representation mechanism, e.g., lookup tables, neural networks, etc. This approach follows the general framework of RL and has its roots in the Bellman equations [2]. TD methods can be roughly classified into two groups, i.e. on-policy and off-policy methods. On-policy methods learn the value of the policy that is used to make decisions and off-policy methods can learn about policies other than that currently followed by the agent [6].

In Fig. 11.2, we present a useful taxonomy summarizing the discussion of this section. We consider the femtocell system as a distributed system given its deployment model. Distributed systems are commonly modeled by means of stochastic games, where players select their actions independently from the other agents in the system. Since we aim to perform an online learning in such a way that agents can adapt to the environmental changes automatically, we formulate the problem through RL. In the problem we are considering, building explicit models of other agents' strategies is highly complex. The solution is then found through model-free techniques, also known as TD learning methods, which construct the knowledge based on experience.

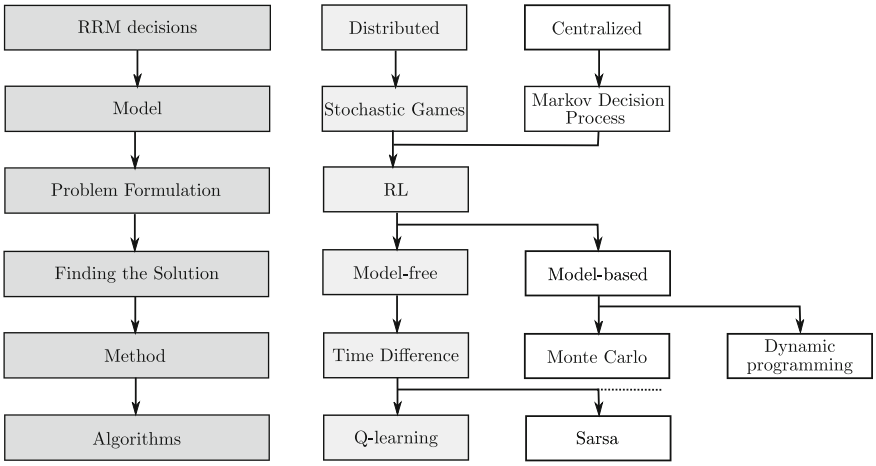

Fig. 11.2 Taxonomy of the formulation of the given problem

Among TD learning methods, typical learning approaches are Q-Learning, Sarsa, Actor-Critic, etc.

11.2.4 Improving the Learning Process

In decentralized multiagent systems, the environment perceived by a given agent is not stationary, as in turn it would be in the case of single-agent scenarios, since it consists of other nodes who are similarly adapting. This may generate may generate oscillating behaviors that not always reach an equilibrium. The dynamics of learning may thus be long and complex, with complexity increasing with an increasing observation space. A possible solution to speed up the learning process and to create rules for unseen situations, is to facilitate expert knowledge exchange among learners [7, 8]. We introduced then in [9] an emerging framework for femtocells, referred to as docition, from "docere" = "to teach" in Latin, which relates to nodes teaching other nodes. This concept perfectly fits a femtocell network scenario, where a femtocell is active only when the users are at home, so that it can take advantage of the decision policies learnt by the neighbor femtocells, which have been active during a longer time. Depending on the degree of docition among nodes, the following cases can be distinguished:

- *Start-up Docition*: Docitive radios teach their policies to any newcomers joining the network. In this case, again, each node learns independently; however, when a new node joins the network, instead of learning from scratch how to act in the surrounding environment, it learns the policies already acquired by more expert

neighbours. Gains are due to a high correlation in the environments of adjacent expert and newcomer nodes. Policies are shared by exchanging Q-tables.
- *IQ-Driven Docition*: Docitive radios periodically share part of their policies with less expert nodes, based on the degree and reliability of their expert knowledge. Policies are shared by exchanging (a weighted version) of the entire Q-table or rows thereof, corresponding to states that have been previously visited.
- *Performance-Driven Docition*: Docitive radios share part or the entirety of their policies with less expert nodes, based on their ability to meet prior set performance targets. Example targets are maximum created interference, achieved capacity.
- *Perfect Docition*: The multi-user system can be regarded as an intelligent system in which each joint action is represented as a single action. The optimal Q-values for the joint actions can be learned using standard centralized Q-learning. In order to apply this approach, a central controller should model the Markov decision process MDP and communicate to each node its individual actions. Alternatively, all nodes should model the complete MDP separately and select their individual actions; whilst no communication is needed here, they all have to observe the joint actions and individual rewards. Due to an exponential growth of the states, this approach is typically not feasible.

11.2.5 Results

Results presented in this chapter considers a single Long term Evolution (LTE) heterogeneous scenario where femtocells coexist with the macro deployment. Docition, implemented over a Q-Learning algorithm is compared to two benchmarks:

- *Distance-Based Non-Cognitive*. The rationale behind this reference algorithm is that femtocell f selects the transmission power of Resource Block (RB) r based on its distance from the macrouser using that RB. Notice that this reference algorithm is only proposed as a non-cognitive benchmark for comparison purposes, and for its implementation we make the hypothesis that the femto network has at least some approximate knowledge of the position of the macrousers, which is a quite difficult hypothesis in a realistic cellular network.
- *Iterative Waterfilling (ITW)*. It is a non-cooperative game where agents are selfish and compete against each other by choosing their transmit power to maximize their own capacity, subject to a total power constraint.

To evaluate the proposed docitive approaches we divide the femtocells in the scenario into two groups. In the first group, we have the docitive or teaching entities and in the second group we have the learning entities, which start their learning process 500,000 learning iterations later than the docitive ones. The given results were obtained for the second group of femtocells in order to show the improvement achieved when agents take advantage of acquired knowledge of other comparable entities in the system. In the startup case, learning entities update their Q-tables at

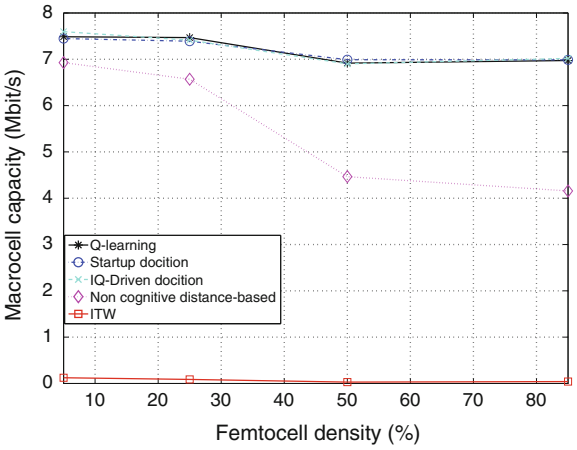

Fig. 11.3 Macrocell capacity as a function of femtocell density

the beginning of the learning process based on the selected docitive entity policies. In the IQ-Driven case, learning entities update every 10,000 learning iterations the Q-values of those rows with lower knowledge, i.e. higher Q-values.

Figure 11.3 shows the macrocell capacity as a function of the femtocell density. It can be observed that learning techniques do not jeopardize the macrocell capacity, maintaining it at a desired level independently of the number of femtocells. On the other hand, with the distance-based reference algorithm, the macrocell capacity decreases when the number of femtocells increases, since the reference algorithm does not adaptively consider the aggregated interference coming from the multiple femtocells in the power allocation process. Furthermore, the ITW algorithm dramatically reduces the macrocell capacity due to its selfish power allocation policy. Finally, with respect to the implementation, it is worth mentioning that the Q-learning approaches only need feedback from the macro network about the Signal to Noise and Interference Ratio (SINR) at the macrousers. However, the non-cognitive distance-based approach relies on stronger hypotheses, such as the positions of the macrousers.

As for the performance of docition, Fig. 11.4 shows performances in terms of precision, i.e. oscillations around the target SINR. We assumed a 50 % femtocell occupation ratio, composed of the probability that a femtocell is present and that it is switched on. In particular, it represents the Complementary Cumulative Distribution Function (CCDF) of the variance of the average SINR at the control point with respect to the set target of $SINR_{Th} = 20$ dB. It can be observed that due to the distribution of intelligence among interactive learners the paradigm of docition stabilizes the oscillations by reducing the variance of the SINR with respect to the specified target. More precisely, at a target outage of 1 %, we observe that the IQ-Driven docition outperforms the startup docition by a factor of two, and the Q-learning algorithm by about an order of magnitude.

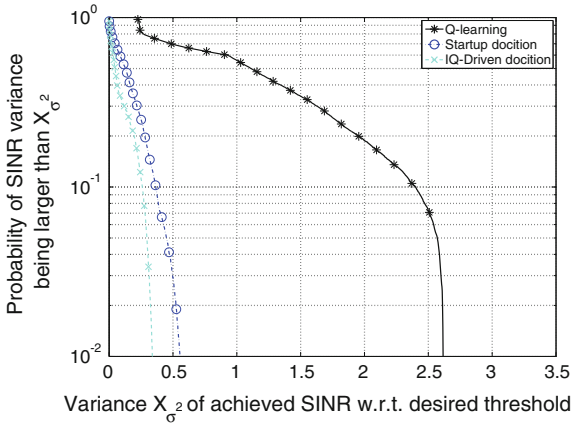

Fig. 11.4 CCDF of the average SINR at macrouser for a femtocell occupation ratio $p_{\text{oc}} = 50\,\%$

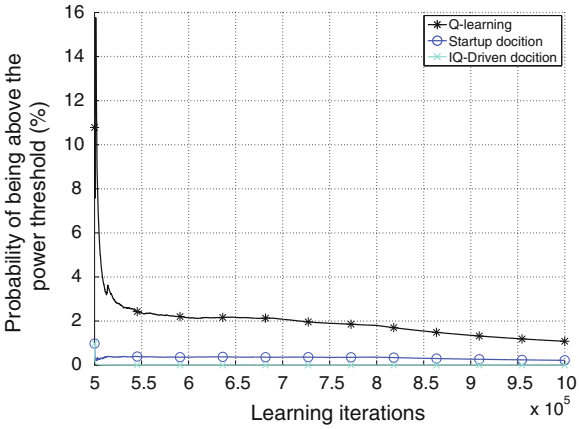

Fig. 11.5 Probability of being above the power threshold as a function of the learning iterations for different docitive cases

Figure 11.5 shows the average probability that the total power at femtocells is higher than P_{max}^{F} as a function of the learning time when docition is applied. It can be observed that the docitive approaches better satisfy the constraint in terms of total transmission power since the early stages of the learning process. More in particular, for the startup docition case, after docition, the femtocell continues with its learning process adapting the knowledge to its own situation. On the other hand, for the IQ-Driven docition case, since the learner agent periodically updates the knowledge corresponding to states of the environment where it performs poorly, the agent presents a very accurate behavior during all the learning process. This accurate behavior is achieved thanks to the continuous adjustment in the agent policy.

11.3 Direct Biomimetic Approach

In this section we present the use of direct biomimetic approach to develop a SO framework for spectral efficiency enhancement on the access link between OFR and their donor BS through adaptation of system wide BS antenna tilts in a distributed manner.

The rest of this section is organized as follows: First a case study of SO system in nature is presented to analyse how a SO solution to a complex optimisation problem is achieved in the natural system. A generic Biomimetic Self Organisation Framework (BSOF) is extracted from this analysis. Then we present the system model and problem formulation. The BSOF is applied to our problem by exploiting a one to one mapping between the two systems. Finally, selected numerical results are also presented to demonstrate the gain the SO solution can achieve without relying on global coordination.

11.3.1 A Biomimetic Self Organization Framework

As explained earlier, myriad of systems in nature exhibit perfect self organization e.g. school of shoaling fish, swarming insects, herding sheep, synchronously flashing fire flies, and flocking birds to name a few. This provides us with an opportunity to devise and extract the generic principles of SO from nature. Here we take one specific case study to investigate the underlying principles of self organization in such system with aim to come up with a generic design and operational framework for self organization.

11.3.2 Revisiting SO in Nature: Flock of Common Cranes

A flock of common cranes is one of the myriad of perfect examples of SO system in nature. We consider the case study of flock of common cranes to further delve into their SO group flight phenomenon as shown in Fig. 11.6. A flock of common cranes *adapts* its flight attributes and flies such that the average flight efficiency of the whole flock is maximised by minimising the average air drag each bird faces by up to 70 % compared to individual bird flight efficiency [10] by dynamically maintaining group flight optimal delta formation during flight. Most importantly, it does so:

- without a leader and without global information exchange or coordination among *all* birds of the flock i.e. no explicit global signalling and no central control (Scalability)
- without swaying from the long term direction of flight or breaking apart even in face of changing wind and weather conditions (Stability)
- without loosing their ability of acutely execute individual as well as collective maneuvering to avoid predator attacks and large hurdles (Agility)

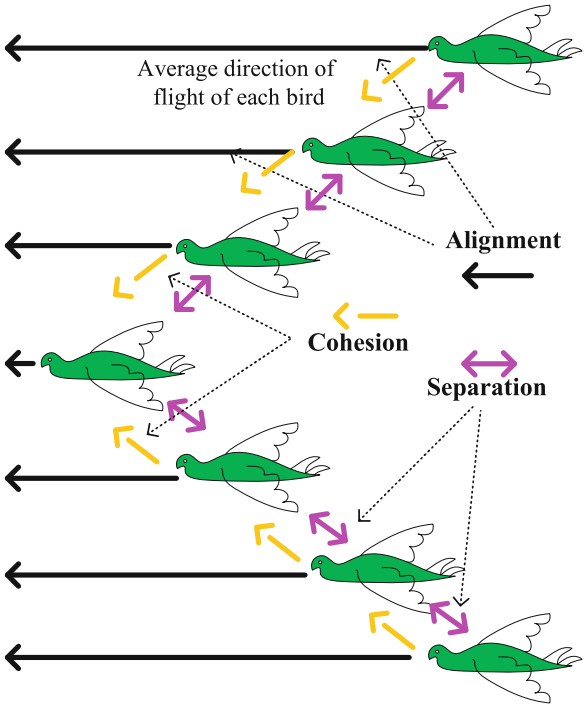

Fig. 11.6 A flock of common cranes optimises its flight efficiency by nearly 70 % through self organization i.e. by maintaining near V-formation through simple individual actions of cohesion, separation and alignment executed by each bird

It has been investigated that each birds individually executes and maintains a certain set of simple flight attributes such that the flock is always in near V-formation that happens to be the optimal formation for group flight efficiency [11]. For common cranes, nature has solved the complex problem of group flight efficiency optimisation to a much simpler problem of maintaining a V-formation. Nurture on the other hand has taught them, how to control their own flight attributes with reference to their immediate i.e. line of sight neighbours to maintain the V-formation while flying. This attributes have been identified to be barely three i.e. *cohesion*, *separation* and *alignment* [12].

1. Separation: Each bird tries to avoid crowding neighbours
2. Alignment: Each bird tries to steer towards average heading of neighbours
3. Cohesion: Each bird tries to steer towards average position of neighbours

Figures 11.6 and 11.7 explains how cohesion, separation and alignment executed by each bird during flight results into an over all V-formation and hence flight efficiency optimization.

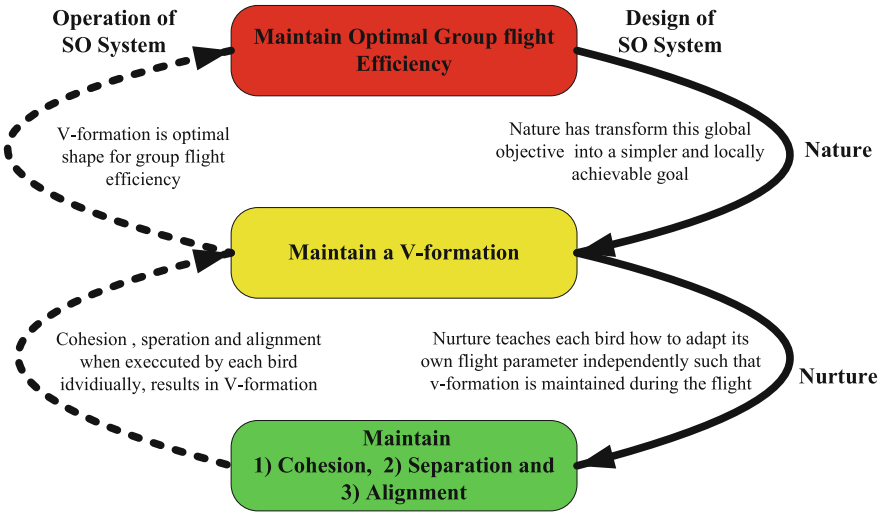

Fig. 11.7 Main elements of self organization working in flock of common cranes

11.3.3 Constituents of SO

By inspection of the above example, three main components of SO can be identified:

1. A specific objective e.g. maximization of flight efficiency.
2. A goal which is effectively a much simpler manifestation of the same objective e.g. formation of V shape.
3. A small set of simple functions which are performed by the entities of the system to achieve that goal e.g. separation, alignment and cohesion in this case.

In order to put our future discussion in a consistent context we will refer to these three identified components of a SO as SO-Objective, SO-Goal and SO-Function respectively.

11.3.3.1 Relationship Between SO-Objective and SO-Goal

A pivotal observation to be made here is that in a system with SO, the complex SO-Objective (e.g. the maximisation of flight efficiency in this case study) is not dealt with, as it is, by each bird, rather it is first mapped to a much simpler goal having equivalent semantics (i.e. flying in V shape), as it is an optimal formation to minimize aerodynamic drag and hence maximise group flight efficiency [10]. Thus a crucial step to achieve SO is translating the complex SO-Objective into a simpler SO-Goal. Next equally important step in design of SO is design of such simple SO-Functions that can achieve the SO-Goal(s).

11.3.3.2 Designing SO-Functions

Simplicity of SO-Functions is one of the basic properties for achieving SO and its associated benefits. Other then that following general characteristics of SO-Functions are worth noting in all natural systems with SO:

1. SO-Functions do not require any discriminative abilities e.g. each bird in the flock can perform all the three SO-Functions i.e. separation, alignment and cohesion.
2. SO-Functions do not require explicit or global co-ordination among entities of system e.g. in bird flock example separation, alignment and cohesion are performed by each bird by relying on its local observation only.

So far, we have explicated three basic constituents of SO i.e. SO-Objective, SO-Goal, and SO-Functions and have figured out the characteristics of each of them. Next, we need to determine the key operational principle of SO i.e. the nature of interaction of self organizing entities in order to target a common SO-Goal. The full understanding of this operational principle is crucial in designing appropriate SO-Functions in a real world engineering system. Following section investigates the type of mutual behavior among entities of the natural system with SO.

11.3.3.3 Coopetition: An Interesting Principle Behind SO Functions

A probe into the mutual behavior among the birds, in the example of bird flock, deciphers that SO-Functions executed by each bird are neither purely cooperative towards other birds nor purely competitive. A pure cooperative behavior among birds will require explicit communication which is not the case in the bird flock. On the other hand, a pure competitive behavior among birds will result in a conflict of interests and will not result in common SO-Goal e.g. instead of single well maintained V-shape multiple independent subgroups with different shape can result or all birds might flight in straight line in parallel to each other as result of pure competition. This implies that a system cannot evolve to SO if the individual entities of the system behave either in pure cooperative fashion or pure competitive fashion.

A cautious inquisition of the above case study shows that the three SO-Functions enacted by each bird are a well composed combination of both cooperative and competitive behavior towards other birds. In fact SO-Function: '*Separation*,' exhibits a kind of cooperative behavior towards other birds as by executing it each bird is being friendly to other by giving it more space. The SO-Function: '*Alignment*,' has dominantly competitive nature as by executing it each bird is trying to reach the same destination as its peers in a rush to avail the resources (food etc) available at that destination. Whereas, SO-Function: '*Cohesion*,' is an intricate amalgam of both cooperation and competition as by executing it birds tend to push into their neighbor birds but simultaneously maintaining a threshold separation. In essence, SO-Functions executed by the entities of a SO system are such that their mutual behavior is neither pure competition nor cooperation. Rather, it is a judicious combination of both of these extreme attitudes. The most suitable term to embody this

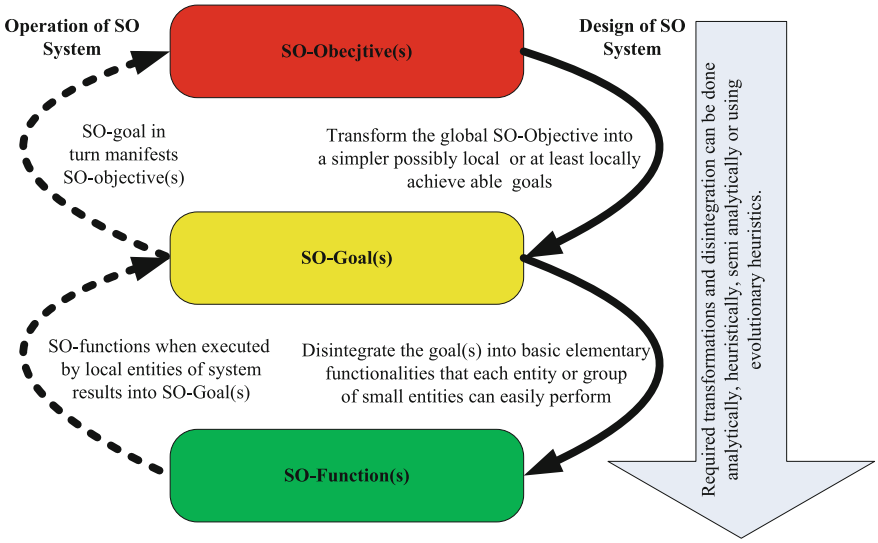

Fig. 11.8 A Biomimmetic Self Organising Framework (BSOF) inspired by case studies of self organization in nature

type of behavior is coopetition: a neologism introduced in the field of economics to represent cooperative competition [13]. Hence, it is neither competition nor cooperation but coopetition among system's entities which is more suitable for emergence of SO behavior on system level.

11.3.4 A Generic Biomimetic Self Organization Framework

Building on the above findings through analysis of SO in natural systems, we can infer a general framework to design SO in a system which is based on three steps:

1. Identification of SO-Objective.
2. Mapping of SO-Objective into simple SO-Goal.
3. Identification of functionalities bearing the characteristics of SO-Functions that are executable under the principle of coopetition to achieve the SO-Goal. These three steps are illustrated in Fig. 11.8. We call this generic frame work BSOF i.e. Biomimetic Self Organization Framework.

11.3.5 An Application of BSOF in Heterogeneous Networks

11.3.5.1 System Model and Assumptions

Assumptions and Nomenclature: We consider multicell cellular relay enhanced cellular system. We assume a frequency reuse of one with interference limited scenario. BSs and RSs are multiplexed in time (or frequency) such that there is no cross interference among the backhaul (BS-R) and coverage (BS-user and RS-user) links. It is assumed that all user devices have omnidirectional antennas with 0 dB gain. Similarly, RSs are also equipped with omnidirectional antennas with a constant gain in all directions. We use spectral efficiency in b/s/Hz as optimization metric and we define it as the long term average bandwidth normalized throughput on link given by $\log_2(1 + SIR)$, where SIR stands for Signal to Interference Ratio. Due to the geometrical context of the following analysis, by referring to BS, RS and users we mean the location of their antennas unless specified otherwise. Symbol *tilde* e.g. $\tilde{x}$ is used to denote optimal value of variable x and symbol *hat* e.g. $\hat{x}$ is used to denote an approximation of a variable x.

System Model: We consider the downlink scenario of a sectorised multi cellular network as shown in Fig. 11.9. Each BS has three cells (sectors) and each cell has at most one RS station placed at an arbitrary location, to cover random hotspots of users. Let $\mathscr{B}$ denote the set of points corresponding to the transmission antenna location of all BS cells, $\mathscr{R}$ the set of points representing the locations of the RSs antennas in the system and $\mathscr{U}$ the set of points representing the antennas of all the user devices randomly located in the system. The geometric SIR on the backhaul link of a RS located at point $r \in \mathscr{R}$ associated with bth cell, can be written as:

$$\gamma_r^b = \frac{P^b G_r^b G_r \alpha \left(d_r^b\right)^{-\beta}}{\sum_{\forall \acute{b} \in \mathscr{B} \backslash b} \left(P^{\acute{b}} G_r^{\acute{b}} G_r \alpha \left(d_r^{\acute{b}}\right)^{-\beta} \right)} \quad b, \acute{b} \in \mathscr{B}, r \in \mathscr{R} \tag{11.8}$$

where P^b is the transmission power of the bth cell, d_r^b and $d_r^{\acute{b}}$ are the distances between the b and $\acute{b}$ transmitting cell antenna locations and receiving RS antenna location r. α and β are the pathloss model coefficient and exponent, respectively. The operator '\' in $\mathscr{B}\backslash b$ means all elements of $\mathscr{B}$ excluding b. G_r^b and $G_r^{\acute{b}}$ are the antenna gains perceived at RS r, from BS b and $\acute{b}$, respectively. For 3GPP LTE and LTE-A the three dimensional antenna pattern can be modelled as proposed in [14], and with the simplifications introduced in[15]. Using the geometry in Fig. 11.9, the perceived antenna gain from a bth BS, at location r, of a RS can be written in dBs as follows:

$$G_r^b = 10^{-1.2\left(\lambda_v \left(\frac{\psi_r^b - \psi_{tilt}^b}{B_v}\right)^2 + \lambda_h \left(\frac{\phi_r^b - \phi_a^b}{B_h}\right)^2\right)} \tag{11.9}$$

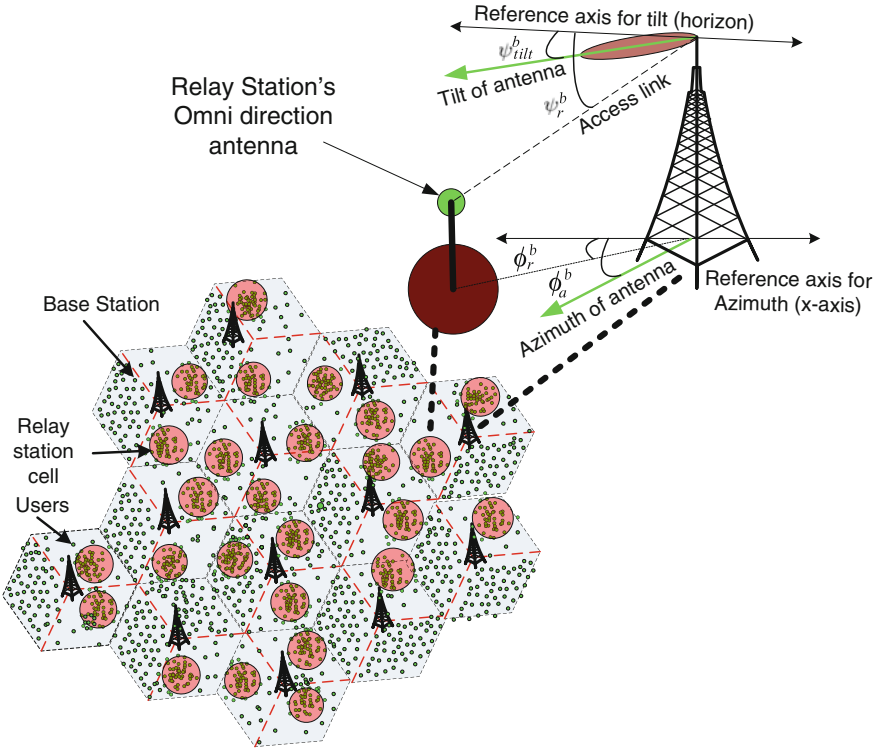

Fig. 11.9 Illustration of geometrical background of the analysis

where ψ_r^b is the vertical angle at the bth BS, in degrees from reference axis (horizon) to the rth RS. ψ_{tilt}^b is the tilt angle of the bth cell, as shown in Fig. 11.9. The ϕ_a^b is the angle of the azimuth orientation of the antenna with respect to the horizontal reference axis, i.e. positive x-axis. ϕ_r^b is the angle of location r of the RS from the horizontal reference axis, at BS b. Subscripts h, a and v denote horizontal, azimuth and vertical, respectively. Thus B_h and B_v represent the horizontal and vertical beamwidths of the BS antenna, respectively, and λ_h and λ_v represent the weighting factors for the horizontal and vertical beam patterns of the antenna in the 3D antenna model [14], respectively.

We assume that all the BSs transmit with the same power and all RS antennas have unity gain i.e. $G_r = 1$. For the sake of simplicity, we use the following substitutions:

$$c_k^b = \frac{B_v^2 \lambda_h}{\lambda_v} \left(\frac{\phi_r^b - \phi_a^b}{B_h} \right)^2 ; \quad c_k^{\acute{b}} = \frac{B_v^2 \lambda_h}{\lambda_v} \left(\frac{\phi_r^{\acute{b}} - \phi_a^{\acute{b}}}{B_h} \right)^2 \tag{11.10}$$

$$h_r^b = \alpha \left(d_r^b \right)^{-\beta} \quad ; \quad h_r^{\acute{b}} = \alpha \left(d_r^{\acute{b}} \right)^{-\beta} ; \quad \mu = \frac{-1.2\lambda_v}{B_v^2} \tag{11.11}$$

Using the substitutions in (11.8)–(11.11) the SIR on the backhaul link of the rth RS can be determined as:

$$\gamma_r^b = \frac{h_r^b 10^{\mu\left(\left(\psi_r^b - \psi_{tilt}^b\right)^2 + c_r^b\right)}}{\sum_{\forall b \in \mathscr{B} \setminus \acute{b}} \left(h_r^{\acute{b}} 10^{\mu\left(\left(\psi_r^{\acute{b}} - \psi_{tilt}^{\acute{b}}\right)^2 + c_r^{\acute{b}}\right)} \right)} \tag{11.12}$$

Note that γ_r^b is a function of $\psi_{tilt}^B = \left[\psi_{tilt}^1, \psi_{tilt}^2, \psi_{tilt}^3 ... \psi_{tilt}^B\right]$, where $B = |\mathscr{B}|$, but for the sake of simplicity, we will show this dependency only where necessary. Similarly, the geometric SIR perceived by a user at a location u, being served by the bth BS cell is given by:

$$\gamma_u^b = \frac{P^b G_u^b \alpha \left(d_u^b\right)^{-\beta}}{\sum_{\forall \acute{b} \in \mathscr{B} \setminus b} \left(P^{\acute{b}} G_u^{\acute{b}} \alpha \left(d_u^{\acute{b}}\right)^{-\beta} \right)} \qquad b, \acute{b} \in \mathscr{B}, u \in \mathscr{U} \tag{11.13}$$

where d_u^b and $d_u^{\acute{b}}$ are distances between the b and $\acute{b}$, and user location u. Following the same steps as above, the SIR for the BS-user link can be written as:

$$\gamma_u^b = \frac{h_u^b 10^{\mu\left(\left(\psi_u^b - \psi_{tilt}^b\right)^2 + c_u^b\right)}}{\sum_{\forall \acute{b} \in \mathscr{B} \setminus b} \left(h_u^{\acute{b}} 10^{\mu\left(\left(\psi_u^{\acute{b}} - \psi_{tilt}^{\acute{b}}\right)^2 + c_u^{\acute{b}}\right)} \right)} \tag{11.14}$$

11.3.6 Designing a Self Organising Solution

11.3.6.1 Problem Formulation: Identifying SO-Objective

This section describes the application of first step of BSOF i.e. identifying the SO-objective, that is formulating the actual problem.

We aim to minimise the radio resources required by the backhaul link in order to maximise the net gain of RSs in terms of system wide capacity. To achieve this objective, we propose to optimise system-wide BS antenna tilts such that it maximize the SE all the backhaul links in the system. The tilt optimisation is carried out on a long time scale of hours to day, so that only the path loss remains as the dominant factor in the determination of the channel gain, as the fast fading and the shadowing are averaged out and consequently can be neglected. Since the adaptation of the BS antenna tilts while it aims to maximise the SE of the BS-RS backhaul links, has an impact on the BS-user links, we also have to take into account the BS-user links in the optimisation problem. Let $|R|$ denote the cardinality of the set $\mathscr{R}$, i.e. $|\mathscr{R}| = R$,

and $\acute{\mathscr{U}}$ the users located in the sectors that contain a RS, where $\acute{\mathscr{U}} \subset \mathscr{U}$. The rest of the users are located in sectors without RS and their locations will be given by the set $\mathscr{U} \setminus \acute{\mathscr{U}}$. The average SE of users in the sectors with RS can be anticipated to be not *critically* affected by the BSs tilt adaptations, as a significant proportion of them is served by the RSs. Furthermore, since the RSs are usually located to cover hotspot areas in the cells, the tilt optimization with respect to RS locations will naturally optimize tilts with respect to majority of users in those cells. On the other hand, the users in cells that are only served by the BS are the users who has to be exclusively taken into consideration in the optimisation problem as they can be significantly affected by the antenna tilt configurations of their cell and neighbor cells, while optimising tilts with respect to backhaul links. By taking these users into account, our problem of maximising system wide BS-RS and BS-user link average spectral efficiency ζ through optimisation of system-wide BS antenna tilts can be mathematically written as (11.15). In (11.15) the second summation term represents the users in sectors without RS deployment.

$$\max_{\psi^B_{tilt}} \left(\frac{1}{|\mathscr{R}|} \sum_{\forall r \in \mathscr{R}} \log_2 \left(1 + \gamma^b_r \left(\psi^B_{tilt}\right)\right) + \frac{1}{\left|\mathscr{U} \setminus \acute{\mathscr{U}}\right|} \sum_{\forall u \in \mathscr{U} \setminus \acute{\mathscr{U}}} \log_2 \left(1 + \gamma^b_u \left(\psi^B_{tilt}\right)\right) \right) \tag{11.15}$$

Note that (11.15) is a nonlinear multi variable optimisation problem. Its solution would require global cooperation among all cells in the system, which would make it not distributed and consequently not in line with the basic idea of online local self-organisation [16, 17]. Furthermore, as we will see in subsequent sections, the objective function in (11.15) is non-convex and characterised by a large number of the optimisation variables, i.e. $\psi^B_{tilt} = \left[\psi^1_{tilt}, \psi^2_{tilt}, \psi^3_{tilt} \ldots \psi^B_{tilt}\right]$, meaning that we are dealing with a large scale optimisation problem. Therefore, numerical or exhaustive search based heuristics are also not a practically feasible approach either. In the following section we present a novel biologically inspired approach to solve this problem in order to develop a pragmatic distributed self-organising solution.

11.3.6.2 Transforming SO-Objective into SO-Goals

This section describes the second step of BSOF i.e. simplifying the SO-objective to SO-Goal. We present the following theorem that paves the way to determine the SO-Goal i.e. a simpler and decomposable manifestation of (11.15).

Theorem 11.1 *For given antenna tilts of neighbouring cells the antenna tilt $\tilde{\psi}^b_{tilt}$ of a cell without RS is optimal for maximising the SE, if it satisfies the following condition:*

$$\sum_{\forall u \in \mathscr{U}^b} \left((\psi^b_u - \tilde{\psi}^b_{tilt}) \frac{\tilde{\gamma}^b_u}{1 + \tilde{\gamma}^b_u} \right) = 0, \quad \mathscr{U}^b \subseteq \left|\mathscr{U} \setminus \acute{\mathscr{U}}\right| \tag{11.16}$$

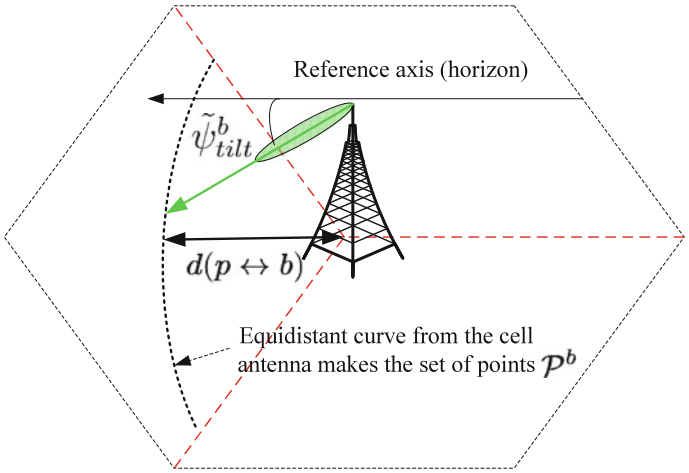

Fig. 11.10 The optimal tilt $\hat{\psi}^b_{tilt}$ cab be mapped to a set of points equidistant distant from the BS. Any of the points on this equidistant curve can be taken as the Center of Gravity (CG) of the user geographical distribution for which the optimal tilt angle is calculated

where $\tilde{\gamma}^b_u$ *is the SIR perceived by user u in cell b, when its antenna is tilted by* $\tilde{\psi}^b_{tilt}$ *degrees.*

Summation in (11.16) *sums over all user in the cell.*

Proof The SE in (11.15) is a twice differentiable function of the tilt, therefore, the proof of Theorem 11.1 can be easily obtained by finding the optimality conditions through the first derivative of the sum of the SE at all user locations, and by means of the second derivative test confirming that this condition provides a maximum point. Details are omitted for space limit.

As a result, Theorem 11.1 provides a method to calculate the optimal tilts that maximize the BS-user link SE, in cells without RS. The following corollary can be directly deduced from Theorem 11.1:

Corollary 11.1 *For given tilt angles of neighbouring cells, the optimal tilt angle* $\tilde{\psi}^b_{tilt}$ *of cell b is the tilt angle that optimizes the SE at any point p. Where p belongs to a set of points* $\mathscr{P}^b$ *in that cell such that* $\mathscr{P}^b = \{p|, d(p \leftrightarrow b) = d^b\}$*, and where* $d^b = \left(H^b - H^p\right) / \tan(\tilde{\psi}^b_{tilt})$*.*$H^b$ *and* H^p *are the heights of the bth cell antenna and point p, respectively.* $d(p \leftrightarrow b)$ *denotes the distance between the location of cell b antenna and the user location p.*

Proof This corollary follows from Theorem 11.1, from the fact that the optimal tilt angle $\tilde{\psi}^b_{tilt}$ given by Theorem 11.1 can be transformed into a set of certain points $\mathscr{P}^b$, *which lie at distance* d^b from the cell antenna b. This is illustrated in Fig. 11.10.

Notice that, according to Theorem 11.1 and its subsequent corollary, the tilt angle of *b*th cell, optimized for any of the points in set $\mathscr{P}^b$, optimizes the average spectral

efficiency in that cell. Thus, taking advantage of Theorem 11.1 and its corollaries, the users in cells without RS can be represented by a single point in each cell which cam act as focal point of user geographical distribution in a cell (see Fig. 11.10) for tilt optimisation. Such focal point of each cell can be used to constitute a set of points $\mathscr{V}$ across the cells defined as:

$$\mathscr{V} = \bigcup_{b=1}^{\mathscr{B}\backslash\acute{\mathscr{B}}} p^b, \qquad p^b \in \mathscr{P}^b \tag{11.17}$$

where set $\mathscr{B}\backslash\acute{\mathscr{B}}$ denote cells that do not contain RS, by using definition of $\mathscr{V}$ in (11.17), in conjunction with Theorem 11.1, the second term of the right hand side in (11.15) can be written as:

$$\sum_{\forall u \in \mathscr{U}\backslash\acute{\mathscr{U}}} \log_2\left(1+\gamma_u^b\left(\psi_{tilt}^B\right)\right) = \sum_{\forall v \in \mathscr{V}} \log_2\left(1+\gamma_v^b\left(\psi_{tilt}^B\right)\right) \tag{11.18}$$

Substituting (11.18) in (11.15), the optimization problem can be written as (11.19).

$$\max_{\psi_{tilt}^B}\left(\frac{1}{|\mathscr{R}|}\sum_{\forall r\in\mathscr{R}}\log_2\left(1+\gamma_r^b\left(\psi_{tilt}^B\right)\right)+\frac{1}{|\mathscr{V}|}\sum_{\forall v\in\mathscr{V}}\log_2\left(1+\gamma_v^b\left(\psi_{tilt}^B\right)\right)\right) \tag{11.19}$$

Note that, since $|\mathscr{V}|$ = number of cells that do not have a RS , $|\mathscr{V}| < |\mathscr{B}|$, which further implies that$|\mathscr{V}| \ll \left|\mathscr{U}\backslash\acute{\mathscr{U}}\right|$. This means that the second summation in (11.19), is a summation over a much smaller number of terms, compared to (11.15), which significantly simplifies the problem. For further simplification, if we define a set $\mathscr{S} = \{\mathscr{R}\cup\mathscr{V}\}$, based on arguments presented above through (11.16)−(11.18) the optimization problem in (11.19) can be written as:

$$\max_{\psi_{tilt}^B}\zeta\left(\psi_{tilt}^B\right) = \max_{\psi_{tilt}^B}\frac{1}{|\mathscr{B}|}\sum_{\forall s\in\mathscr{S}}\log_2\left(1+\gamma_s^b\left(\psi_{tilt}^B\right)\right) \tag{11.20}$$

The points in set $\mathscr{S}$ are shown in Fig. 11.11, where circles represent RSs, i.e. points in set $\mathscr{R}$; and stars represent the focal point of users' geographical distribution in each cell with no RS, i.e. points in set $\mathscr{V}$. For ease of discussion, we will refer to points in set $\mathscr{S}$ as Centre of Gravity (CGs) of the cells. Note that as highlighted in Sect. 11.3.3.1, (11.20) is the required simplified manifestation of the original problem in (11.15).

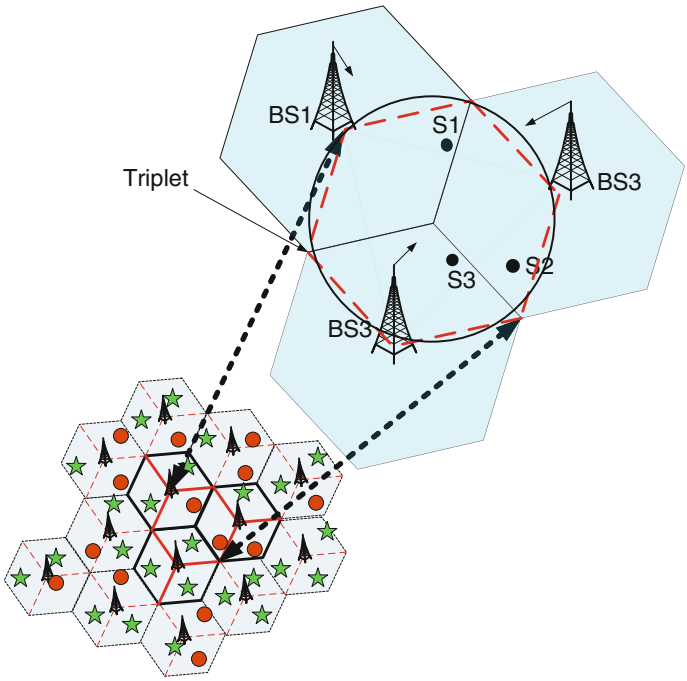

Fig. 11.11 *Circles* represent points in set $\mathscr{R}$ i.e. RS locations and *stars* represent points in set $\mathscr{V}$ i.e. focal points of user distributions in a cell determined through Theorem 11.1 and its corollaries. *Stars* and *circles* together make set $\mathscr{S}$

11.3.6.3 Decomposing SO-Goal to Design SO-Functions

This section describes the third step of BSOF i.e. decomposing the SO-Goal into SO-Functions.

While the complexity of the problem in (11.15) has been significantly reduced in (11.20), still we are dealing with a large scale optimization problem over a vector of variables ψ^B_{tilt}. The resulting dependency of the potential solution over all the BS tilts in the system, prevents it from being a practical solution from implementation point of view. Due to the dependency of the SIR in (11.20) on all the antennas in the system, any direct solution (11.20) would require a global cooperation among all the cells of the system. The system wide signalling required to achieve this cooperation will make such solution too slow and complex to cope with the on-the-run deployment or switching on and off of the RSs. As discussed above, a SO solution will require a further decomposition of (11.20) into local subproblems. Such decomposition is common in SO systems in nature, as it is explained for the case study of flock of common cranes, in [10] and [18]. We refer to the same case study and more in particular to the result, discussed in the above references, according to which, for achieving the flock-wide objective of flying in V-formation, each crane merely

relies on the observation of its immediate two neighbours, on its two sides. Thus, although cranes do not achieve the perfect V shape, they can still achieve up to 70 % gain in group flight efficiency [10]. To exploit the same principle in our problem, we compromise on the global optimisation perspective and we propose the novel concept of *triplet* to enable a local problem decomposition. A *triplet* consists of three immediate neighbour cells as it is illustrated in the enlarged part of Fig. 11.11. The key idea is that, as it happens for the cranes, each cell observes (tilts and CG locations) of its immediate two neighbours cells, when optimising its own tilts. In this way, tilts are optimised within each of the $N = \frac{B}{\hat{B}}$ triplets independently, where $\hat{B}$ is the size of the triplet, i.e. three in this case. As a result, the problem in (11.20) can be approximated as:

$$\max_{\psi_{tilt}^{B}} \hat{\zeta}\left(\psi_{tilt}^{B}\right) = \max_{\psi_{tilt}^{B}} \frac{1}{|\mathscr{B}|} \sum_{\forall s \in \mathscr{S}} \log_2(1+\hat{\gamma}_s^b) \tag{11.21}$$

where $\hat{\gamma}_s^b$ is the approximate SIR at point s (CG) that takes into account the observations from the only two other members of the triplet, and can be rewritten as:

$$\hat{\gamma}_s^b\left(\psi_{tilt}^{\hat{B}}\right) = \frac{h_s^b 10^{\mu\left(\left(\psi_s^b - \psi_{tilt}^b\right)^2 + c_s^b\right)}}{\sum_{\forall \check{b} \in \check{\mathscr{B}} \setminus b}\left(h_s^{\check{b}} 10^{\mu\left(\left(\psi_s^{\check{b}} - \psi_{tilt}^{\check{b}}\right)^2 + c_s^{\check{b}}\right)}\right)} \tag{11.22}$$

where b represents the antenna location of the cell which point s lies. $\hat{\mathscr{B}}$ represents a triplet, such that $\left|\hat{\mathscr{B}}\right| = \hat{B} = 3$. $\psi_{tilt}^{\hat{B}}$ is the vector of tilt angles of the $\hat{B}$ sectors within the triplet. Now consider the following propositions:

Proposition 11.1 *As β and the cell radius grows large, $\hat{\zeta}$ becomes a closer approximation of ζ.*

Proof Proposition 11.1 can be easily proved by putting large values of β and d in (11.14) and (11.22).

Proposition 11.2 *If the SIR is given by $\hat{\gamma}_s^b$, the maximum aggregate throughput achieved in the system by optimizing the tilts within each triplet independently, is the same as the throughput achieved by optimizing the system-wide tilts. Mathematically, $\hat{\zeta}_{N,\max} = \hat{\zeta}_{max}$, where ζ_{max} is the maximum average SE that can be achieved by solving the optimization problem in* (11.21) *and*

$$\hat{\zeta}_{N,\max} = \frac{1}{|\mathscr{N}|} \sum_{\forall n \in \mathscr{N}} \left\{ \max_{\psi_{tilt}^{Tn}} \frac{1}{|\mathscr{T}_n|} \sum_{\forall s \in \mathscr{S}_n} \log_2\left(1+\hat{\gamma}_s^b\left(\psi_{tilt}^{Tn}\right)\right) \right\} \tag{11.23}$$

where $\mathscr{S}_n \subset \mathscr{S}$, $\mathscr{T}_n$ *is the nth triplet and* $|\mathscr{S}_n| = |\mathscr{T}_n| = T_n = 3, \forall n \in \mathscr{N}$, $\psi_{tilt}^{T_n}$ *is the vector of tilt angles of sectors within nth triplet such that*

$$\mathscr{S}_n \cap \mathscr{S}_{n'} = \Phi \text{ and } \mathscr{T}_n \cap \mathscr{T}_{n'} = \Phi, \forall \neq n' \text{ where } n, n' \in \mathscr{N} \tag{11.24}$$

$\mathscr{N}$ *is set of all the triplets, such that* $|\mathscr{N}| = \frac{|B|}{|\mathscr{T}_n|} = N$ *is the total number of triplets in the system.*

Proof Since $|\mathscr{N}| \times |\mathscr{T}_n| = |\mathscr{N}| \times |\mathscr{S}_n| = |\mathscr{B}| = |\mathscr{S}|$ and in right hand side of (11.23) all the terms are mutually exclusive hence the proposition.

Note that each term in the summation in (11.23) is now a very small scale optimization problem over three tilt angles within each triplet. Next we present a methodology to solve this subproblem.

11.3.6.4 Solving the SO-Function

This section describes the final step of BSOF i.e. solving the local sub-problem to enable the execution of SO-Functions.

To enable execution of SO-Function, the following subproblem needs to be solved for each of the N triplets locally and independently:

$$\begin{aligned}\hat{\zeta} = &\log_2\left(1 + \left(\frac{h_1^1 10^{-1.2\mu\left(\left(\psi_1^1 - \psi_{tilt}^1\right)^2 + c_1^1\right)}}{\left(h_1^2 10^{-1.2\mu\left(\left(\psi_1^2 - \psi_{tilt}^2\right)^2 + c_1^2\right)}\right) + \left(h_1^3 10^{-1.2\mu\left(\left(\psi_1^3 - \psi_{tilt}^3\right)^2 + c_1^3\right)}\right)}\right)\right) \\ &+ \log_2\left(1 + \left(\frac{h_2^2 10^{-1.2\mu\left(\left(\psi_2^2 - \psi_{tilt}^2\right)^2 + c_2^2\right)}}{\left(h_2^1 10^{-1.2\mu\left(\left(\psi_2^1 - \psi_{tilt}^1\right)^2 + c_2^1\right)}\right) + \left(h_2^3 10^{-1.2\mu\left(\left(\psi_2^3 - \psi_{tilt}^3\right)^2 + c_2^3\right)}\right)}\right)\right) \\ &+ \log_2\left(1 + \left(\frac{h_3^3 10^{-1.2\mu\left(\left(\psi_3^3 - \psi_{tilt}^2\right)^2 + c_3^3\right)}}{\left(h_3^1 10^{-1.2\mu\left(\left(\psi_3^1 - \psi_{tilt}^1\right)^2 + c_3^1\right)}\right) + \left(h_3^2 10^{-1.2\mu\left(\left(\psi_3^2 - \psi_{tilt}^2\right)^2 + c_3^2\right)}\right)}\right)\right)\end{aligned} \tag{11.25}$$

$$\max_{\psi_{tilt}^1, \psi_{tilt}^2, \psi_{tilt}^3} \hat{\zeta}\left(\psi_{tilt}^1, \psi_{tilt}^2, \psi_{tilt}^3\right) \tag{11.26}$$

subject to: $\psi_{tilt}^1, \psi_{tilt}^2, \psi_{tilt}^3 < \frac{\pi}{2}$

where $\hat{\zeta}$ is given by (11.25). We drop the subscript n to simplify the notation and since the following analysis is valid for any triplet. Notice that (11.26) is still a non convex optimization problem. However, compared to our original problem in (11.15),

the problem in (11.26) is now a small scale optimization problem, as the number of optimization parameters is only three with limited range of $0° < \psi < 90°$. Since the search space of this problem is now reasonably small ($\approx 90 \times 90 \times 90 = 729{,}000$), any of exhaustive search based evolutionary heuristics listed in [17] can be used to find the solution of (11.26). Alternatively, a solution can also be determined using a non linear optimization techniques that can tackle a non-convex optimization objective. For example, since the objective function is twice differentiable and the constraint is also differentiable, an option could be to solve (11.26) through Sequential Quadratic Programming SQP. To this end, the problem can be written in the standard form as:

$$\min_{\psi} \quad -\hat{\zeta}\,(\psi) \tag{11.27}$$

subject to: $g_j\left(\psi_j\right) < 0, \quad j = 1, 2, 3$
where $\psi = [\psi_1, \psi_2, \psi_3]$ and $g_j\left(\psi_j\right) = \psi_j - \frac{\pi}{2}$.
The Lagrangian of the problem in (11.27) is given by:

$$\mathscr{L}(\psi\,, \lambda) = \hat{\zeta}\,(\psi) - \sum_{j=1}^{3} \lambda_j (\psi_j - \frac{\pi}{2}) \tag{11.28}$$

If $\hat{H}$ denotes the approximation of the Hessian matrix $\mathbf{H}$, then we can define a quadratic subproblem to be solved at the ith iteration of SQP as follows:

$$\min_{w \in \Re^J} \quad \frac{1}{2} w^T \hat{H}\,(\mathscr{L}\,(\psi\,, \lambda)\,)_i\, w + \nabla \hat{\zeta}\,(\psi)_i w \tag{11.29}$$

subject to: $w_j + \psi_{j_i} - \frac{\pi}{2} < 0 \quad j = 1, 2, 3$

At each iteration the value of $\hat{H}$ can be updated using the Broyden-Fletcher-Goldfarb-Shannon BFGS) approximation method. Once the Hessian is known the problem in (11.29) is a quadratic programming problem that can be solved using standard methods such as the gradient projection [19].

Based on the above steps of the SQP, the problem in (11.26) can be solved within each triplet independently. The solution provides the optimal tilt angles to be maintained by each of the three cells in the triplet for given locations of CGs, within that triplet. The execution of these local solutions in each triplet results in the achievement of the system-wide objective in (11.20), which was a close manifestation of the original system wide objective in (11.15). Since the solution is distributed, i.e. executable in each triplet independently and autonomously, the near optimal tilt angles can always be maintained locally to maximise system-wide SE on the backhaul links, as well as on the coverage links, despite the impromptu deployment or removal of RSs. In the following we refer to the developed framework as SOT (Self-Organisation of Tilts).

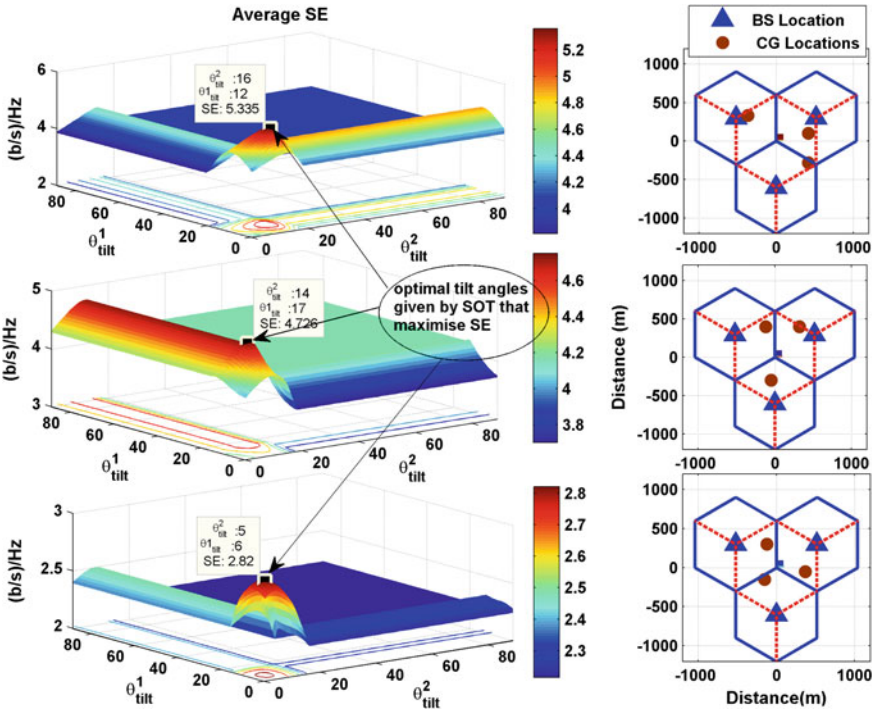

Fig. 11.12 $\frac{\hat{\xi}}{3}$ plotted for a stand alone triplet against tilts of two sectors while third is fixed at 0° for three different CG locations within the triplet. It can be seen that optimal tilt angles for maximum spectral efficiency change as locations of CG's change

11.3.7 Results

In this section, first we present the numerical results for a single triplet cellular system, which can be readily obtained from the analysis presented above. These results indicate the gains theoretically achievable by the proposed solution. This is followed by performance results of the SOT evaluated by implementing it in a full scale system level simulator. These results evaluate the gain that can be obtained by the proposed SOT in more practical and realistic scenarios, with shadowing and interference from multiple tiers of cells.

11.3.7.1 Numerical Results

Numerical results for three different sets of locations of CGs are plotted in Fig. 11.12. These results can be obtained by plotting (11.25) with $\beta = 4$, $B_v = 10°$, $B_h = 70°$, a cell radius of 600 m, BS and CG height of 30 and 10 m respectively. In addition, we

Table 11.1 System level simulation parameters

Parameters	Values
System topology	19*3 sector/cells
BS transmission power	39 dB m
BS Inter site distance	1,200 m
BS height	32 m
RS height	10 m
User height	1.5 m
User antenna	0 dB (Omini directional)
RS antenna	0 dB (Omini directional)
BS antenna horizontal beamwidth, B_h	70°
BS antenna vertical beamwidth, B_v	10°
BS antenna vertical Gain Weight, λ_v	0.5
BS antenna vertical Gain Weight, λ_h	0.5
BS antenna maximum gain, G_{max}	18 dB
BS antenna maximum attenuation, A_{max}	20 dB
Frequency	2 GHz
Pathloss model	Urban Macro [20]
Shadowing STD for BS-user links	8 dB
Shadowing STD for BS-RS links	4 dB

normalise $\hat{\zeta}$ by 3, i.e. the number of cells in the triplet. $\frac{\hat{\zeta}}{3}$ is plotted in Fig. 11.12 and it represents the average SE in per cell in a triplet. It can be seen that by optimising the antenna tilts the average SE at the CGs can change from 3.9 to 5.3, 3.7 to 4.7 and 2.1 to 2.8 b/s/Hz, respectively (in figures from top to bottom), depending on the location of CGs that represent either RS or focal points of user distribution. By determining the optimal tilt angles for given locations of CGs in each triplet, the SOT can always set the antenna tilts for maximum SE. The values of SE achieved by optimal tilts determined through the SOT, compared to the SE achievable with the wide range of other tilts, imply that the SOT in general can maintain a substantial gain in SE, compared to a large range of arbitrary tilting.

11.3.7.2 System Level Simulation Results

For system level proof of concept we use an Orthogonal Frequency Division Multiple Access (OFDMA) based generic cellular system where some cells contain randomly located RS and other cells are served by the BS only as would be the case with LTE-A. Table 11.1 shows the summary of simulation parameters. We compare the performance of the SOT against three fixed antenna tilting benchmarks of low, medium and high tilts, i.e. 0°, 15°, 30°. Performances are evaluated for both backhaul and coverage links on the downlink, in terms of SE (b/s/Hz). Figure 11.13 plots the average SE with the fixed tilts and with the SOT implementation. It can be seen

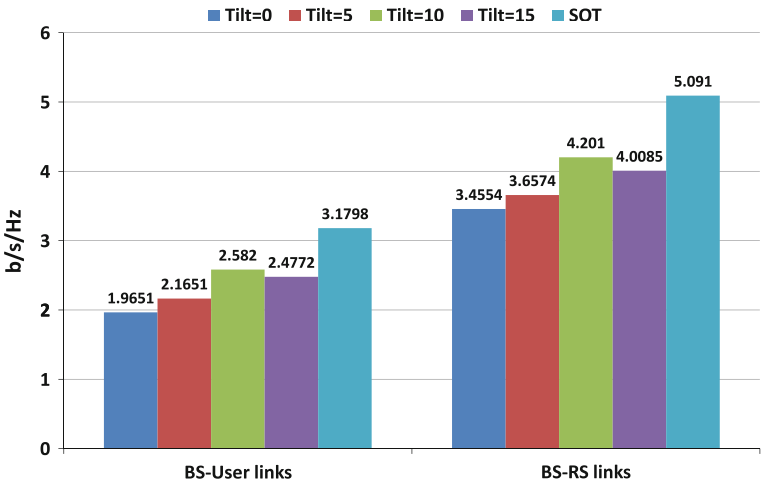

Fig. 11.13 Average spectral efficiency on BS-RS and BS-user links

that with $Tilt = 0°$, the SE is worst and with $Tilt = 15°$ the SE in general improves. However, with $Tilt = 30°$, the SE starts decreasing again. This is due to the fact that with high tilting values, although the interference decreases, the coverage on the cell edge starts decreasing as well. On the other hand, the SOT provides a substantial boost in average spectral efficiency. This is because, the SOT allows tilts to be optimised locally, based on the location of RS and CG of user geographical distributions, in individual cells. It can be observed that for BS-user links, the SOT yields a gain of $3.1798 - 1.9651 \approx 1.2$ b/s/Hz compared to the worst fixed titling ($Tilt = 0$) and $3.1798 - 2.4772 = 0.7$ b/s/Hz compared to the best fixed tilting options. On the other hand, on the backhaul links the same gains are observed to be 1.64 and 1.1 b/s/Hz, respectively, which are larger than that achievable on BS-user links because of obvious reasons that RS can be better represented with single point compared to a user distribution. However, it can be inferred from both numerical and simulation results that the SOT provides significant gains in SE, and the exact gain in each triplet depends on the RS and users' geographical distribution.

11.4 Conclusions and Considerations in Use of Biomimetics

In this chapter, use of directly and indirectly bio inspired techniques has been presented for enabling self-organization for macrocell and femtocell coexistence. Indirect biomimetics, techniques inspired by learning theory are explored and their performance is compared via simulation results. Another contribution of the chapter is a novel demonstration of use of direct biomimetic approaches for developing SO for a system-wide antenna tilt adaption problem which is a large scale non

linear optimization problem. The target scenario considers outdoor fixed relay based femtocells coexisting with macro system. Results show that proposed biomimetic solution can yield substantial gain in spectral efficiency without resorting to central control or system wide signalling.

Although, this chapter demonstrated biomimetics as a rich and promising paradigm for design of SO, but there are some consideration and implications that must be apprehended while applying Biomimetics, particulary the direct approach of Biomimetics to engineering systems. These limitations mainly arise due to the notable differences between biological and engineering systems. Due to these differences some functionalities of biological systems are hard to mimic and it is not technically feasible or even possible to achieve such functionality in engineering system may be because of lack of their understanding or because of technological limitations. Examples of such not mimicable functionalities so far are the intelligence of human brain, or 100 % efficiency of conversion process of chemical energy to electromagnetic energy i.e. light in fire flies.

Thus, Biomimetics instead of all its potential might not be an appropriate approach when: (i) the biological solution is too difficult to rebuild with technical means or (ii) a technical solution can be more efficient by taking advantage of mechanisms that cannot be found in the biological paradigm. A mistake to be avoided is to stick to biological solutions just because of their seeming elegance.

References

1. Brown, G.: Iterative solution of games by fictitious play. Activity Analysis of Production and Allocation. Wiley, New York (1951)
2. Bellman, R.: Dynamic Programming. Princeton University Press, Princeton, NJ (1957)
3. Fudenberg, D., Levine, D.K.: The Theory of Learning in Games. MIT Press, Cambridge, MA (1998)
4. Powers, R., Shoham, Y.: New criteria and a new algorithm for learning in multi-agent systems. In: Proceedings of Advances in Neural Information Processing Systems (NISP2004), pp. 1089–1096, 13–18 Dec (2004)
5. Tesauro, G.: Extending Q-learning to general adaptive multi-agent systems. In: Proceedings of Advances in Neural Information Processing Systems (NISP2003), pp. 1089–1096, 13–18 Dec (2004)
6. Sutton, R.S., Barto, A.G.: Reinforcement Learning: An Introduction. The MIT Press, Cambridge, MA (1998)
7. Panait, L., Luke, S.: Cooperative multi-agent learning: the state of the art. Auton. Agent. Multi-Agent Syst. **3**(11), 383–434 (Nov 2005)
8. Ahmadabadi, M.N., Asadpour, M.: Expertness based cooperative Q-learning. IEEE Trans. Syst. Man. Cyber. B **32**(1), 66–76 (Feb 2002)
9. Giupponi, L., Galindo, A., Blasco, P., Dohler, M.: Docitive networks—an emerging paradigm for dynamic spectrum management. IEEE Wireless. Comm. Magazine **17**(4), 47–54 (2010)
10. Lissaman, P.B.S., Shollenberger, C.A.: Formation flight of birds. Science **168**(3934), 1003–1005 (1970) http://www.sciencemag.org/cgi/content/abstract/168/3934/1003
11. Potts, W.: The chorus line hypothesis of maneuver coordination in avian flocks. Nature **309**, 344–345 (1984)

12. Reynolds, C.W.: Flocks, herds, and schools: A distributed behavioral model, in computer graphics. In: Proceedings of the SIGGRAPH, 87 Conference, vol. 21, No. 4, pp. 25–34 (1987). http://www.red3d.com/cwr/boids/
13. Brandenburger, A., Nalebuff, B.: Co-Opetition: A Revolution Mindset That Combines Competition and Cooperation: The Game Theory Strategy That's Changing the Game of Business. Currency Doubleday, New York (1997)
14. Viering, I., Dottling, M., Lobinger, A.: A mathematical perspective of self-optimizing wireless networks. In: IEEE International Conference on Communications (ICC '09), pp. 1–6, June 2009
15. Imran, A., Imran, M., Tafazolli, R.: Relay station access link spectral efficiency optimization through SO of macro BS tilts. IEEE Commun. Lett. **15**, 1326–1328 (2011)
16. Prehofer, C., Bettstetter, C.: Self-organization in communication networks: principles and design paradigms. IEEE Commun. Mag. **43**(7), 78–85 (July 2005)
17. Aliu, O., Imran, A., Imran, M., Evans, B.: A survey of self organisation in future cellular networks. IEEE Commun. Surv. Tutor. **99**, 1–26 (2012)
18. Imran, A., Bennis, M., Giupponi, L.: Use of learning, game theory and optimization as bio-mimetic approaches for self-organization in macro-femtocell coexistence. In: IEEE Wireless Communications and Networking Conference, pp. 103–108, April 2012
19. Gill, P., Murray, W., Wright, M.H.: Practical Optimization. Academic Press, London (1981)
20. Baum, D., Hansen, J., Salo, J.: An interim channel model for beyond-3G systems: extending the 3GPP spatial channel model (scm). In: IEEE 61st Vehicular Technology Conference (VTC'05-Spring), vol. 5, pp. 3132–3136, May-1 June 2005

Chapter 12
Cooperation and Competition for Spectrum Sharing in Cognitive Radio Networks: The Practical Perspective

Hanna Bogucka and Marcin Parzy

Abstract The goal of the flexible, efficient and fair spectrum allocation is to increase the spectrum utilization of radio resources in future wireless systems. According to the cognitive radio (CR) concept, the nodes are expected to sense their radio environment, take decisions on their operation in the network and learn from their past actions to better adjust to the network dynamics process in various unplanned situations. The CR node can take actions resulting from the input information processing, although in most cases this information is incomplete or inaccurate. CR node may acquire the necessary information needed for its efficient operation by accessing the control or management channel(s) or by interaction with other nodes. Unfortunately, control channels may not always be available and the neighbouring nodes may not be interested in cooperation due to the cost of the spectrum and energy resources. Therefore, the major challenge for a CR node is to operate efficiently with incomplete or limited knowledge on the network and cooperate with its competitors. For CR, game theory provides interesting tools to study competition and cooperation among rational and intelligent players taking decisions with limited or incomplete information. CR nodes can exchange information, cooperate or learn because they were programmed to perform such tasks. Such processes have the associated cost, usually expressed in consumed energy, time or spectrum. These costs must be balanced with benefits. Therefore, the GT models must be carefully selected and evaluated for the application in resource sharing to comprise with practical limitations of the dynamic CR networks. In this work the practical issues of cooperation among cognitive radio nodes competing for available resources in the decentralized networks are considered. It is pondered how the theory of competition and cooperation (game theory) meet the practice, by discussing the quantitative metrics of the cost of avoiding cooperation

H. Bogucka (✉) · M. Parzy
Chair of Wireless Communications, Poznan University of Technology, Polanka 3,
60-965 Poznan, Poland
e-mail: hbogucka@et.put.poznan.pl

M. Parzy
e-mail: mparzy@et.put.poznan.pl

M.-G. Di Benedetto and F. Bader (eds.), *Cognitive Communication and Cooperative HetNet Coexistence*, Signals and Communication Technology,
DOI: 10.1007/978-3-319-01402-9_12,

(the Price of Anarchy—PoA), of having limited knowledge of the competitors (the Price of Ignorance—PoI). Some practical approaches to the spectrum sharing and allocation problem are also presented, which make use of representative, intentionally reduced information that the CR nodes have to exchange. One of the presented methods is based on the repeated game against the network-nodes community using the aggregated knowledge of its possible behavior. The other one is based on the coopetition methodology, which combines the advantages of both cooperative and competitive approaches. It is shown that the problem of radio resource allocation in wireless systems can be solved efficiently by using these not-optimal but practical approaches, by presenting some indicative results: the information-data sum-throughput, Jain's fairness index, PoA, PoI, and the network welfare function equal to the sum-throughput net.

12.1 Introduction

Opportunistic spectrum access as well as flexible and efficient spectrum allocation procedures are considered as measures to increase the utilization of the scarce radio resources in future wireless communication networks. Apart from the spectral efficiency, fairness in resources distribution is also in the focus of research towards the cognitive, opportunistic and dynamic spectrum access. For the future communication concepts, such as Cognitive Radio (CR), the nodes are expected to sense their radio environment, take decisions on their operation in the network and learn from their past actions to better adjust to the network dynamics. Thus, CR nodes are expected to have a certain degree of intelligence.[1] The study on the definitions of intelligence and the so-called cognitive intelligence direct us to the observation that the ability of active *information processing* to better adjust to the varying environment, as well as effectiveness of the decision making process in various unplanned situations are the indicators of intelligence and are used for its testing [1]. The intelligence of a living being as well as of a machine (a CR node), can be evaluated based on its actions resulting from the input information processing, although in most cases this information is incomplete or inaccurate. In order to acquire the information necessary for its efficient operation, a CR node has to have an access to the control (management) channels or to interact and cooperate with other nodes. On the other hand, these control channels and centralized manager may not always be available, and the neighboring nodes in the CR network area are often competitors in acquiring resources necessary for their operation, namely the spectrum and energy resources. Thus, the major challenge for an intelligent CR node is to operate efficiently while

[1] Note, that these mentioned features of a CR are a subset of abilities mentioned in the definition of intelligence as "a very general mental capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly, and learn from experience" [1]. The quoted document [1] was a public editorial statement signed by fifty-two researchers in fields allied to intelligence testing that claimed to present those findings widely accepted in the expert community.

possessing an incomplete knowledge on the network environment and to cooperate with its competitors.

The centralized spectrum allocation procedures usually apply some optimization procedures that require the Channel State Information (CSI) of all links in the network, and involve a significant amount of the overhead traffic, which in turn occupies the scarce radio resources. These procedures are not in our focus, since we concentrate on CR concept, which incorporates discussed intelligence of devices, not solely on the opportunistic, but centrally managed spectrum access. Moreover, as said before, optimization theory may not be the right tool when the initial conditions and boundary values are unknown or uncertain. Interestingly, game theory provides tools to study competition and cooperation among rational (intelligent) players taking decisions based on limited or incomplete information, and possibly learning from their past decisions outcomes [2–6]. Note that our CR nodes are such intelligent and highly-rational players, because contrary to the living beings, they are machines that can be programmed in such a way, to behave and respond rationally to the outside world stimuli. They can be also programmed to exchange information, cooperate and learn, however in practice, the volume of exchanged information, the range of cooperation and the time needed to learn may be restricted. This is because these processes have the associated cost, usually expressed in the amount of consumed resources (energy, time and spectrum), that must be balanced against the benefits. There exist metrics that allow to study the benefit of the network operation in terms of some predefined welfare function resulting from a certain degree of cooperation among the players, from the learning process as well as from handling the incompleteness of the information. These metrics must be addressed to find the trade-offs between efficiency of the network operation and its cost. For example, the cooperation among all the mobile CR nodes in a network would result in their optimized operation, however the necessary exchange of the information would require a lot of control traffic comparable in volume to the actual information-data traffic. Competitive behavior and distributed decision making could also converge to some optimal operation point, if the learning process was properly designed, however every learning algorithm requires time to converge, which may be not acceptable in highly-dynamic radio environment. Thus, the game-theory models must be carefully selected and evaluated for the application in CR resource sharing to comprise with practical limitations of the dynamic CR networks.

12.2 Typical Game Models and Their Inappropriateness for Cognitive Radio Resource Sharing

Engineers tend to see some contradiction between theory and practice. On the other hand, they say that if the theory does not apply in practice, it means that the theory is wrong. It can be observed that recently, application of game theory for considering and solving the practical problems of wireless networks has attracted a considerable attention (see [5, 6] and the references therein). It allows for studying competition

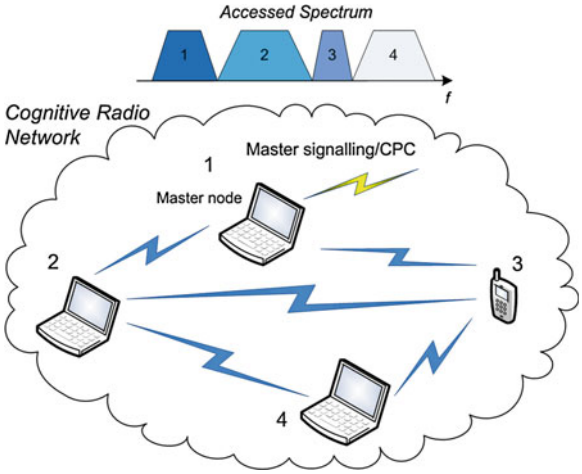

Fig. 12.1 Distributed cognitive radio network scenario

between the nodes, as well as possible cooperation for their broadly understood common communication benefit. Theories, however, even if attractive for providing useful solutions to the stated problems, may turn out impractical, if their basic assumptions are difficult to be met. Let us discuss, what practical game models could be applied for solving the problem of spectrum sharing among the nodes in CR networks. For this purpose, we consider some major game types that result from their typical categorization. It is not our goal to provide a short course on game theory, rather we aim at indicating most popular game models and discuss their appropriateness for the application in our general scenario of spectrum sharing.

Our considered distributed CR network is presented in Fig. 12.1, in which a number (K) of intelligent wireless devices are competitors in sharing available spectrum resources, i.e. a number (N) of spectrum units, called orthogonal frequency channels. It is assumed that one channel can be only assigned to one player, namely the CR node. A spectrum unit has the bandwidth narrow enough to be considered a flat-fading channel for each player. An example of the multiple access technique using such orthogonal channels is the well-known Orthogonal Frequency Division Multiple Access (OFDMA), in which case the network nodes share a set of accessible subcarriers. Every player's strategy is a combination of channels to be used for her communication. The players which are able to sense the resources availability, can communicate and exchange some information. There is no centralized management, however the CR nodes may have an access to some signaling channels, such as the master-node signaling or the so-called Cognitive Pilot Channel (CPC), in which some necessary signaling data are transmitted. The goal of each player is to acquire as much spectrum bandwidth as possible to maximize its transmit-data throughput, that is also understood as its own benefit. The CR network performance as a whole is usually evaluated according to the efficiency of the spectrum-resources usage (sum-throughput net) as well as to the fairness of these resources distribution.

12.2.1 Cooperative Versus Non-cooperative Games and the Price of Anarchy

The first categorization that is usually made in game theory is to divide games as cooperative and non-cooperative ones. The analysis of non-cooperative games is usually focused on searching the equilibrium point, i.e. the strategy set of the players, such that any deviation by a player from her respective strategy could only result in a worse payoff for this player. It is worth mentioning that equilibrium can be defined in a number of ways (the Nash Equilibrium—NE is the most known), and proving of its existence and uniqueness is often a difficult analytical problem. In cooperative games, the players are allowed to cooperate, bargain their assets and form coalitions to increase their benefit, or to increase the broadly defined *welfare*. The solution of a cooperative game is called bargaining solution, an example of which is Nash Bargaining Solution (NBS). The popular metric that measures the cost of lack of cooperation between the players is the *Price of Anarchy* (PoA), defined as the ratio between the worst possible Nash equilibrium (the worst-case outcome at any equilibrium in a non-cooperative game) and the social optimum, as a measure of the performance of the system, as it was defined in [7]:

$$\mathrm{PoA} = \frac{\max_{\mathbf{s}\in\mathbf{S}} W(\mathbf{s})}{\min_{\mathbf{s}\in\mathbf{E}} W(\mathbf{s})}, \tag{12.1}$$

where $W(\mathbf{s})$ is the value of the welfare function resulting from the adopted vector of strategies $\mathbf{s}$, $\mathbf{S}$ is the Cartesian product of the players' sets of strategies, and $\mathbf{E}$ is a subset of $\mathbf{S}$ containing vectors of strategies that are in equilibria. The natural candidate for $W(\mathbf{s})$ is the sum of players utilities $u_i(\mathbf{s})$:

$$W(\mathbf{s}) = \sum_i u_i(\mathbf{s}) \tag{12.2}$$

The PoA reflects the players' common gain (welfare), if they play the cooperative game versus non-cooperative equilibrium for the same number of players K, their strategies and individual outcomes.[2] To be precise in some cases the 'min' and 'max' in (12.1) can be inversed due to the goal of the optimization of $W(\mathbf{s})$. In such defined PoA, the worst equilibrium utility may be 0 as social utility; thus the denominator could be 0 and the PoA becomes infinite.

The *welfare* of the community of the CR network nodes is often defined as the network sum-throughput net (the control-traffic necessary for successful communication is subtracted from the total traffic) over the available spectrum bandwidth. Note, that for the CR nodes to cooperate and maximize the sum-throughput, the

[2] Note, that sometimes PoA is defined using the cost function rather than welfare, i.e. the maximal cost resulting from the non-cooperative game equilibrium over the minimal cost resulting from the cooperation of the players.

Channel State Information (CSI) of all links must be exchanged between them. The CSI for each possible link consists of the quality indicators, e.g. Signal-to-Noise power Ratios (SNRs), for each available frequency channel. Given the nodes mobility, and the channels dynamics in the coherence time, the volume of this information to be exchanged can be comparable to the actual information-data traffic. Alternatively, the CSI of the nodes links can be delivered to a master node or other central element that can solve the cooperative game and inform the players on the assigned spectrum. If we apply the complete-information non-cooperative game model to our scenario, again the all-players CSI has to be made known to all other players, what is completely impractical. Thus, even if the PoA is moderate in our CR network scenario, the achieved common welfare (sum-throughput net) is very poor, because of the communication cost of providing complete information to the players.

12.2.2 Complete versus Incomplete Information Games and the Price of Ignorance

Another typical distinction is made between the complete-information and incomplete -information games. In the first case, it is assumed that the players have complete information on the game rules, every player's set of strategies and payoffs resulting from the combination of the chosen strategies. Incomplete information means that a player does not know the other players' payoff functions, i.e. either their mathematical definitions (which translate to the players' goals)[3] or the parameters of the defined functions (which reflect how much the other players value their goals). The incomplete-information games can also refer to the uncertainty concerning the players' behavior and associated set of strategies, and are treated as Bayesian games, i.e. using some probabilities of the possible sets of strategies for the players, it is assumed with some probability, that a particular player will behave in a way to take a particular set of options. A metric describing the cost of not having the complete information is the *Price of Ignorance* (PoI), introduced in [8], and defined as the change in the network performance under partial knowledge as compared to that of the network under global knowledge:

$$\mathrm{PoI} = \max_{\mathbf{s}\in\mathbf{E},\mathbf{s}'\in\mathbf{E}'} \frac{W_{\text{knowledge}}(\mathbf{s}) - W_{\text{ignorance}}(\mathbf{s}')}{W_{\text{knowledge}}(\mathbf{s})}, \tag{12.3}$$

where $\mathbf{E}$ and $\mathbf{E}'$ are sets of strategies in equilibrium for the case of played complete and incomplete information game respectively. The performance W can be understood as the network achieved welfare, and can be defined in a number of ways, e.g. the total

[3] Note that imperfect information means something else, i.e. the information about the other players' payoffs or actions is known but may contain an error, that usually has to be accepted if other actions to minimize or eliminate this error are not undertaken.

network power saving or its spectral efficiency (sum-throughput over the available bandwidth) as in [8], or sum-throughput net. Moreover, in [8] the authors showed that while local knowledge has little effect on the maximum transmission power used by the network, it has a significant effect on the spectral performance. The ignorance can be understood as either uncertainty of the information or as possessing complete (certain) information, which is however reduced representation of the information describing the players' environmental conditions and options in detail.

As mentioned in the previous section, providing the complete information on all the channels quality of all players to all other players in the considered CR network is associated with a huge communication cost. Applying the Bayesian game model to the considered problem is even more impractical, because the fading statistics of all channels for all players are required to consider every player behavior with a given probability. In a dynamic radio environment these statistics change with time. Moreover, it seems impractical to consider the channel gains probability density functions with high granularity because it exponentially increases the computational complexity of calculating the equilibrium point.

12.2.3 Single Stage Versus Repeated Games

Another distinction is often made between single-stage games (when a game is played just once) and repeated (multi-stage) games. In repeated games, the players choose their actions at every stage taking their overall payoff into account over a certain horizon of time (multiple stages of the game). It has been found that the optimal method of playing a repeated game is to play a socially optimum strategy. In wireless communication networks, due to the lack of information of the other players payoff function parameters (the other players' CSI), repeated games are considered, which usually serve determination of these parameters by the rest of the players. At each game stage, the players take their past experience into account, and apply a learning mechanism to get a complete-information on the other players and approach the social optimum, depending on the applied model. A popular example of such games application is distributed power control in wireless networks. Practical application of such a game for power control is possible due to the fact, that the players do not have to know all other players complete CSI or transmit powers for calculating their payoff, but the aggregated result of the actions taken by them, i.e. the interference power. In the spectrum sharing games, in general this is not possible. We will come back to the issue of aggregating the other player's behavior in the next section.

It is important to note that repeated games with learning typically assume infinite horizon, i.e. infinite number of game repetitions. In practice, the game is stopped at a certain moment, when the social goal of the network is achieved, however the players do not know when. Learning in repeated games requires a number of iterations to converge. In dynamically changing radio environment, the CR nodes may not have enough time to learn and adjust their strategies to achieve the global optimum, i.e. the network maximum welfare. Limited computational resources may also lead to the decision of stopping the game before the learning convergence result is achieved.

Apart from social optimum in games with infinite time horizon the players may play any Nash equilibrium which weakly dominates the min–max payoff profile of the constituent stage game. When Nash equilibrium dominates the min-max payoff the players has no incentives to deviate by simply play the min–max strategy at every stage. According to the folk theorem [9] the Nash equilibrium to be feasible it must lie in the convex hull of the set of possible payoff profiles of the stage game.

12.2.4 Fairness

Let us note, that apart from the network welfare defined as spectral efficiency or sum-throughput fairness should be considered in spectrum sharing algorithms, because it translates to the nodes perceived quality of experience. The Jain's fairness index [10] is broadly adopted as a metric to measure fairness of resource allocation in wireless networks. Its value ranges from $1/K$ (when one player is allocated all resources) to 1 (when all players achieve the same throughput, regardless of their links quality). In this work the cooperation is considered to improve the fairness and the resource distribution process especially in cases when the CR nodes compete for the resources.

Many fairness metrics can be used in resource distribution such as max-min rule or a generalized proportional fairness based on NBS and coalitions which was presented in [11]. Such an approach is quite interesting and similar to the coopetition approach presented in Sect. 12.3.2. But the coopetition model is more flexible and allows for controling the trade–off between the spectral efficiency and fairness. Thus the nodes may choose which metric is more important for them in the current game. In our opinion such algorithms should be strongly desired. The trade-off between the efficiency and fairness was considered deeply last years, especially in [12] and in [13]. In [12] the authors addressed this problem by managing the system fairness index. They proposed two adaptive utility-based resource allocation frameworks for subcarriers assignment and power allocation for non–real–time and real–time services. In [13] the authors presented the comprehensive tutorial about the downlink packet scheduling in LTE (Long Therm Evolution). In LTE it is important to provide fair and efficient transmission using the time and frequency resources, channel condition and QoS requirements. The authors presented the state of the art and key design issues for new packet scheduling algorithms.

12.3 Practical Approaches to Resource Sharing with Reduced Information

As discussed in the previous section, the major problem encountered when applying game theory models to resource sharing in the CR network is a huge (impractical) communication cost of exchanging information necessary for the nodes cooperation and for the application of complete information game models.

Below, we consider two promising practical approaches to solve this problem. The first one is based on narrowing the required information for non-cooperative game by treating all other players as one, namely the Network Nodes Community (NNC), and by limiting the players in resources they can acquire at a time as a remedy for greedy behavior of the players. It has been shown in [14, 15], that such a limitation may be advantageous for achieving high network spectral efficiency. The second approach is also based on reducing the information required in the game, however the social behavior is coerced by the coopetition model [16], which combines competition and cooperation between the network nodes. In both cases the key is the required information reduction, i.e. in a spectrum sharing game, the other players' CSI is reduced to a representative value, which can be made completely known to every player. It allows to consider these games as complete-information games, and to apply them in practical systems.

12.3.1 Game Against the Community

As mentioned above, the complete-information non-cooperative game models have been formulated, and have practical application for distributed interference management, due to the fact that the necessary complete information on the interference level actually aggregates the power levels of all players and can be available for each player, since the nodes can measure it locally. In [15], distributed subcarriers allocation method has been presented for a network of the OFDMA-based opportunistic radios. Definition of this game involves aggregation of the players, in such a way that each player (the network node) can view all other players as one, namely the NNC. The complete information required in this game does not include the individual CSI of the other network nodes, but only the local (single-link) CSI. This way non-cooperative game with full information is reasonable and practically applicable in the dynamically changing network scenarios.[4] The complete-information game is defined as the extensive-form game, i.e. the players take decisions one after another, as they appear in the network, what requires an admission control or collision avoidance mechanism, which however maybe decentralized as in stochastic medium-access control.[5] The individual players' set of strategies consist of possible numbers i of spectrum units (orthogonal channels) they can acquire at the game respective stages when they take their decisions. The NNC set of strategies $\mathbf{J} = \{j\}$ consist of numbers j of spectrum units it can occupy at a particular game stage, given the fact that after every stage the available resources shrink, i.e. once a player acquires

[4] An observation can be made that also in every day life, whenever individuals share limited common resources, e.g. parking places in a city, they act by viewing the rest of the community as a whole, and play against this community, e.g. against other drivers willing to occupy available parking lots, and do not consider each individual other player and her possible payoff.

[5] The users can access the channel randomly, use the mechanism similar to 802.11 or a mechanism based on the token exchange. Please, note that it is assumed that users can detect collisions and react for them.

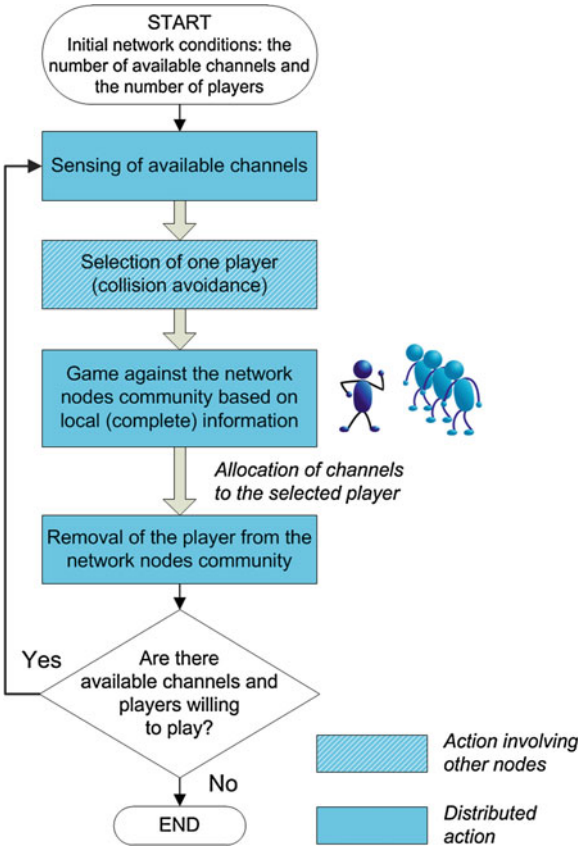

Fig. 12.2 An algorithm for the game against the Network Nodes Community

some spectrum resources, these resources are not available for the NNC in the next stage and this player is removed from, the NCC. The players are also limited in the maximum number of spectrum units, they can take at a time ($i \leq I$). The algorithm of the game is presented in Fig. 12.2. The game is fully distributed. First, selection of a player is done based on the random access with collision avoidance. Although it is a distributed action, it involves other players. Decision is made by each player individually. A player removal from the NNC can also be done in a distributed fasion, simply by tagging the players who have already made their decisions in the game. The minimum signalling is required in a control channel to start and finish the game. Here, we do not specify the details of these mechanisms.

The payoff function $p_{i,j}^{(k)}$ applied to assist player k at the game stage, in which she takes her decision is defined as follows:

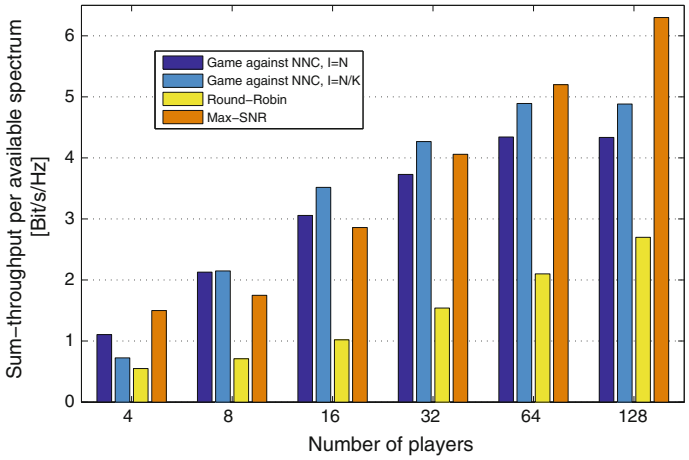

Fig. 12.3 Data sum-throughput averaged over the available bandwidth versus the number of CR nodes for the repeated game against the network nodes community

$$p_{i,j}^{(k)} = \left\{ \frac{1}{N} \sum_{n \in \mathbf{S}_i^{(k)}} \log_2(1 + \alpha^{(k)} \gamma_n^{(k)}) \right\} \cdot \{N^{(k)} - i - j\}, \tag{12.4}$$

where $p_{i,j}^{(k)}$ is the number of channels available at the kth stage of the game (note that $N = N^{(1)}$), $\gamma_n^{(k)}$ is the kth player's (estimated) SNR in channel n, whose index n belongs to the set $\mathbf{S}_i^{(k)}$ of indices of the kth CR node's i estimated highest-SNR channels (the cardinality of $\mathbf{S}_i^{(k)}$ is i, and $i \leq I$, since we restrict each player from occupying more than I spectrum units at a time), and $\alpha^{(k)}$ is the so-called SNR-gap, depending on the kth player's target Bit Error Probability (BEP) P_e. One may interpret Eq. (12.4) as the total normalized throughput (throughput divided by the spectrum unit bandwidth), which could be obtained by the NNC in the next stage, in case it occupied the remaining channels and had the same average spectral efficiency as the considered player. The first factor in Eq. (12.4) is the contribution of player k to the network spectral efficiency, while the second represents the network potential to serve other nodes. This way, in the decision-making on how many SCs to occupy, the players factor the social aspect of the network (to serve multiple nodes) and not just their own benefit.[6]

In Fig. 12.3, simulation results of the applied game against the NNC for spectrum sharing are presented as the sum-throughput in the example OFDMA system. The assumed number of CR nodes (players) K ranges from 4 to 128, and $N = 256$

[6] Note that in [15], additionally pricing component has been included in formula (12.4), however when such a Social-Behaviour Model is used, and when I is properly chosen, optimal pricing parameter can be close to zero.

orthogonal subcarriers are available. Rayleigh fading outdoor rural channel with exponentially decaying power profile is assumed. The total transmit power in the system is limited and fixed to $K \cdot P_{\text{node}}$, where P_{node} is the average transmit power per the CR node. The players have random SNR of uniform distribution ranging from 10 to 35 dB. The target BEP is assumed to be 10^{-4}, the same for all users. The results are compared against classical centralized approaches: max-SNR, which provides the highest throughput, and Round-Robin algorithm, which is considered to provide good fairness. Note that, the game against the NNC is efficient when the number of players in high enough to be able to occupy all available resources (higher than 16 in our case) and when the maximum number of channels that can be acquired at a time is well adjusted (in our case $I = N/K$, and thus, it is possible that all players get their fair share of the resources. When I increases the efficiency increases while the fairness decreases and vice versa).

12.3.2 Coopetition

Coopetition is a neologism combining the notion of cooperation and competition. Originally coopetition was defined in [16]. This approach is based on the fact that cooperation creates an added value to the resulting product (outcome) of this cooperation and that this added value can be distributed among competitors who have created it. According to the popular definition, coopetition is the behavior of two or more competitors cooperating in some areas of their usual activities or business. They ussually cooperate because they want to decrease common costs.

The idea of applying coopetition in resource sharing is to let the players compete for resources using reduced and representative information about their CSI, and then to form coalitions and cooperate to refine the resulting spectrum allocation. The applied algorithm is presented in Fig. 12.4.

In the first phase, each CR node of the network calculates a compact metric, which represents their spectrum demands, and their CSI. A good example of such metric representative for the CSI is the effective SNR used as channel quality indicator (CQI). The effective SNR [17, 18] may be calculated for the whole available frequency band. Such compact (reduced) information can be exchanged by the nodes or transmitted to a master node at relatively low cost, expressed in the volume of traffic. Such metrics allow for signalling reduction and provides necessary information which can improve the system performance and fairness than in system without the knowledge about other nodes. In the competition phase, non-cooperative game model with complete information is applied, e.g. the Cournot oligopoly model [3, 4]. The Cournot competition has been considered in [19–21] for radio resource allocation in flat fading channels or in the channels described by the CQI. The game outcome is the number of the spectrum units "won" (acquired) by each player, however the exact location of these units on the frequency axis is not resolved at this stage, yet. These acquired numbers of spectrum units constitute the players' assets reflecting their input to possible coalitions in the next phase of the algorithm.

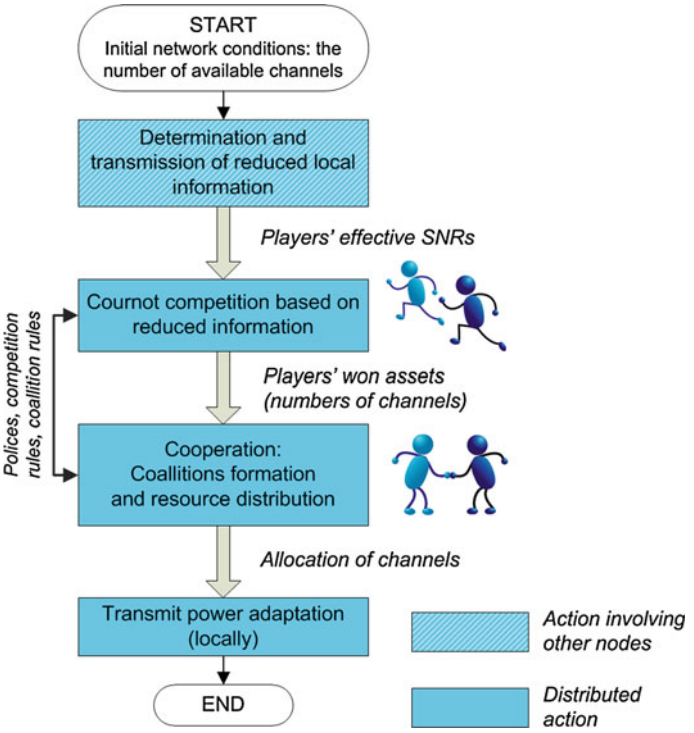

Fig. 12.4 Coopetition algorithm for spectrum sharing

In the cooperation phase, the coalition games can be used to decide how the final resource allocation will be. The coalitions are formed just to define the order of spectrum allocation. The stronger the coalition in terms of their assets (the total number of acquired channels), the higher priority in the order of acquiring the resources. A good option is to form a coalition that can obtain the majority of resources, yet having enough flexibility in sharing them between the coalitionists. Flexibility of a coalition c is defined as the number of possible combinations of choosing n_c channels from a set of $N^{(c)}$ channels available for this coalition. In such a case, the strongest player with the highest assets would form a coalition with possibly the weakest player, who would assure the least majority of gained assets (e.g. 51 % of remaining available resources), because weak players (with a small number assets) in the coalition allow for higher flexibility in resource distribution.[7] The strongest coalition (with the highest number of channels) is the first one for the particular channels selection. The second strongest coalition has to sense the spectrum again and selects a subset of remaining channels, etc. As a result of the cooperation phase the concrete allocation of the spectrum units (frequency channels) to the players is done. In the considered

[7] This approach is similar to Parliamentary Games, in which after elections, the players have some assets (sits in the parliament), and form coalitions to distribute resources (government positions).

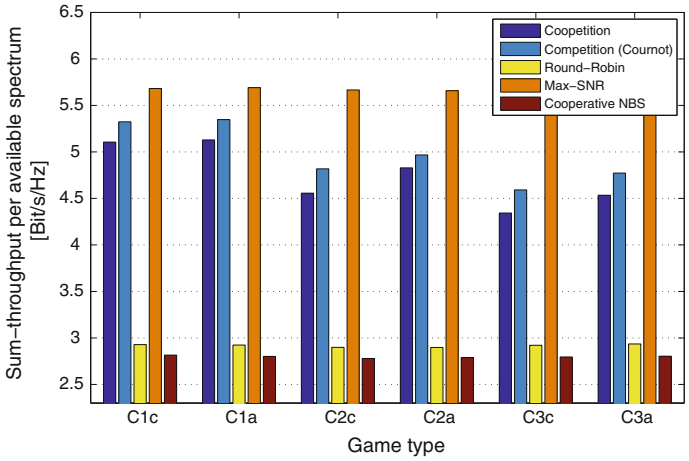

Fig. 12.5 Data sum-throughput averaged over the available bandwidth for various coopetition scenarios

option of coalition's formation, the players with the higher number of resources are prioritized. The main goal for this phase is to mix the "strongest" players with the highest number of resources, in average having high link quality with the weak players having worse channels quality. This improves fairness in resource sharing. Finally, the transmit power is adjusted by each player. For this purpose the classic water-filling can be applied, fixed power allocation per subcarrier or the so-called *game against the nature* in case of the CSI uncertainty (imperfect CSI) [22].

In Figs. 12.5 and 12.6, simulation results of the applied coopetition methodology for spectrum sharing are presented as the sum-throughput and Jain's fairness index in the example OFDMA system. The assumed system parameters are the same as in the previous subsection, however a fixed number of players $K = 8$ has been adopted. Different game types refer to the parameters of the Cournot game, defined in [21], i.e. C1, C2 and C3 respectively refer to the game outcome in case of NE without any minimum resource assignment, of NE with guaranteed minimum resource assignment, and of the behavior of players maximizing their payoff. The letters 'c' and 'a' in the game-type refer to constant values of the cost-function parameters in the Cournot formulation (fixed cost equal to 0, and the spectrum unit cost equal to 1), and to adapted valued of these parameters (For details of the Cournot game formulation, refer to [19] and [21]). The coopetition results presented in Figs. 12.5 and 12.6 are compared with the results obtained for pure competition using the Cournot game only (without the cooperation phase), in which the the players choose their resources according to their strength starting from the strongest one, with a pure cooperative complete-information game NBS (as defined in [23]), and with the Round-Robin and max-SNR algorithms outcomes.

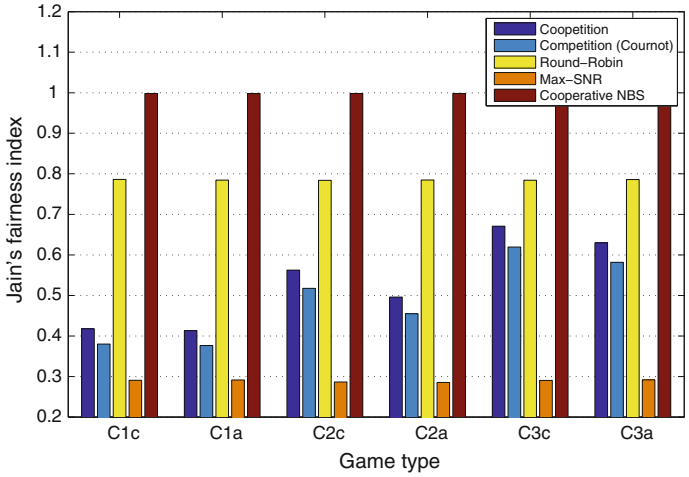

Fig. 12.6 Jain's fairness index for various coopetition scenarios

Table 12.1 Indicative network quality metrics

Model	PoA	PoI	Fairness	Net-throughput
Game against the community	0.36	0.43	0.82	2.1
Coopetition (C1c-type)	0.15	−0.38	0.42	5.1
Cournot competition	0.14	−0.43	0.39	5.3
Cooperation (NBS)	1	0.78	1	2.8

Note, that coopetition significantly increases spectral efficiency of the network in comparison with cooperative NBS and Round-Robin solutions, and its fairness is improved when compared with pure Cournot competition and max-SNR algorithms.

Here above, we have presented the game models outcomes—the network efficiency expressed in terms of the information-data sum-throughput. As mentioned above, the key idea in the presented games is to reduce the information that must be exchanged by the nodes. In Table 12.1, we present the values of some considered metrics that reflect the cost of this reduction for the considered examples of game models (game against the NNC with $I = N/K$, coopetition of C1c type, pure Cournot competition and cooperative complete-information game defined in [23] for number of players $K = 8$. The PoA and PoI are calculated for the reduced complete-information games outcomes with respect to unreduced (full CSI) complete information games. The welfare considered for PoA and PoI calculation is the network sum-throughput-net (maximum data throughput minus the control traffic necessary for required information exchange), i.e. the effective spectral efficiency. It is assumed that the channel coherence time allows to represent either the complete CSI per channel or the reduced information (effective SNR) in average of 2 bits per an OFDM symbol.

12.4 Conclusion

The question of how the theory of competition and cooperation applies to practical spectrum sharing in wireless cognitive radio networks is often posed by wireless communication engineers and researchers, who understand practical limitations of these networks. These limitations refer to limited knowledge of the wireless environment of the nodes, and thus limited possibilities of cooperation. Gathering the complete information comes on a significant communication cost of the order of actual data traffic. We have presented two practical approaches to this problem, which make use of intentionally reduced information that the CR nodes have to exchange. In particular the coopetition methodology, which combines the advantages of healthy competition and limited cooperation seems to be a good idea compromising the benefit (network welfare) versus the above mentioned cost. The fairness of resources distribution using this methodology is also increased when compared to pure competitive behavior of nodes.

Acknowledgments The research leading to these results has received funding from the Polish Ministry of Science and Higher Education, under the grant No. 779/N-COST2010/0 which supports participation in the European COST Action IC0902.

References

1. Editorial: Mainstream science on intelligence. Intelligence **2**4(1), 13–23 1997 [also in Wall Street Journal on December 13, 1994]
2. Straffin, P.: Game Theory and Strategy. The Mathematical Association of America, Washington (2002)
3. Rasmusen, E.: Games and Information: An Introduction to Game Theory. Blackwell Publishers, Oxford (2006)
4. Watson, J.: Strategy Introduction to Game Theory. W.W. Norton & Company, New York (2002)
5. MacKenzie, A., DaSaliva, L.: Game Theory for Wireless Engineers. Morgan & Claypool Publishers, California (2006)
6. Lasaulce, S., Tembine, H.: Game Theory and Learning for Wireless Networks: Fundamentals and Applications. Academic Press, Boston (2011)
7. Koutsoupias, E., Papadimitriou, C.: Worst-case equilibria. In: Proceedings of the 16th Symposium on Theoretical Aspects of Computer Science, 1999
8. Komali, R.S., Thomas, R.W., DaSiva, L.A.: The price of ignorance: distributed topology control in cognitive networks. IEEE Trans. Wireless Commun. **9**(4), 1434–1445 (2010)
9. Ratliff, J.: A folk theorem sampler. Great Introductory Notes to the Folk Theorem (1996)
10. Jain, R., Chiu, D.M., Hawe, W.: A Quantitative Measure of Fairness and Discrimination for Resource Allocation in Shared Computer Systems. DEC Research, Report TR-301 (1984)
11. Han, Z., Ji, Z., Liu, K.J.R.: Fair multiuser channel allocation for OFDMA networks using nash bargaining solutions and coalitions. IEEE Trans. Commun. **53**(8), 1366–1376 (2005)
12. Rodrigues, E.B., Casadevall, F.: Control of the trade-off between resource efficiency and user fairness in wireless networks using utility-based adaptive resource allocation. IEEE Commun. Mag. **9**, 90–98 (2011)
13. Capozzi, F., Piro, G., Grieco, L., Boggia, G., Camarda, P.: Downlink packet scheduling in LTE cellular networks: key design issues and a survey. IEEE Commun. Surv. Tutorials **99**, 1–23 (2012)

14. Perlaza, S.M., Debbah, M., Lasaulce, S., Bogucka, H.: On the benefits of bandwidth limiting in decentralized vector multiple access channels. In: Proceedings of 4th International Conference on Cognitive Radio Oriented Wireless Networks and Communications, Germany (2009)
15. Bogucka, H.: Optimal resource pricing coercing social behavior in wireless networks. In: Proceedings of the IEEE International Communications Conference (ICC), Kyoto (2011)
16. Brandenburger, A.M., Nalebuff, B.J.: Co-opetition: A Revolutionary Mindset That Combines Competition and Co-operation: The Game Theory Strategy Thats Changing the Game of Business. Doubleday Publication Press, New York (1996)
17. Nortel: Effective SIR computation for OFDM system-level simulations. TSG-RAN WG1 35, R03–1370 (2003)
18. Kliks, A., Zalonis, A., Dagres, I., Polydoros, A., Bogucka, H.: PHY abstraction methods for OFDM and NOFDM systems. J.Telecommun. Inf. Technol. 3, 116–122 (2009)
19. Niyato, D., Hossain, E.: Microeconomic models for dynamic spectrum management in cognitive radio networks. Microeconomic models for dynamic spectrum management in cognitive radio networks. Springer, New York (2007)
20. Niyato, D., Hossain, E.: A game-theoretic approach to competitive spectrum sharing in cognitive radio networks. In: Proccedings of the IEEE Wireless Communications and Networking Conference, Hong Kong (2007)
21. Parzy, M., Bogucka, H.: Coopetition practical methodology for efficient sharing of radio resources in wireless networks. In: Proceedings of the 4th International ICST Workshop on Game Theory in Communication Networks (Gamecomm), Paris, France (2011)
22. Bogucka, H.: Game theoretic model for the OFDM water-filling algorithm with imperfect channel state information. In: IEEE International Communications Conference (ICC) (2008)
23. Chee, T.K., Lim, C.-C., Choi, J.: A cooperative game theoretic framework for resource allocation in OFDMA systems. In: 10th IEEE Singapore International Conference on Communication systems (ICCS), Singapore (2006)

Chapter 13
Cooperative Detection of PUE Attacks in CRNs

Olga León, Juan Hernández-Serrano and Miguel Soriano

Abstract Cognitive radio networks (CRNs) act as secondary users of the spectrum left unused by licensed services or primary users thus improving spectrum usage. However, the specific features of CRNs give rise to new security threats, such as the primary user emulation (PUE) attack, in which a malicious user impersonates a primary transmission to prevent the CRN from using a vacant band. With the aim of detecting such kind of attacks, in this chapter we describe a cooperative localization method specifically suited to CRNs. Localization is a powerful PUE-detection technique in several scenarios where the position of the primary users is well known, e.g. TV towers in IEEE 802.22 standard. Simulations results show the goodness of the proposed method and its suitability to typical CRN scenarios.

13.1 Introduction

Cognitive Radio Networks (CRNs) [4] are regarded to be a possible solution to the current underutilization of the spectrum by allowing Cognitive Radios (CRs) to act as secondary users of the spectrum left unused by licensed services or primary users. Thus, spectrum sensing is a crucial task in order to detect vacant bands or white spaces and avoid interfering primary transmissions. If a primary signal is detected in the operation channel, the CRN must switch to another band (a process known as spectrum handoff). Besides, if another secondary user is already operating in such band, self-coexistence mechanisms are needed to share the spectrum fairly.

O. León (✉) · J. Hernández-Serrano · M. Soriano
Universitat Politècnica de Catalunya, Barcelona, Spain
e-mail: olga@entel.upc.edu

J. Hernández-Serrano
e-mail: jserrano@entel.upc.edu

M. Soriano
Centre Tecnològic de Telecomunicacions de Catalunya, Barcelona, Spain
e-mail: soriano@entel.upc.edu

M.-G. Di Benedetto and F. Bader (eds.), *Cognitive Communication and Cooperative HetNet Coexistence*, Signals and Communication Technology,
DOI: 10.1007/978-3-319-01402-9_13,

Due to the specific characteristics of CRNs, new threats have arisen [13, 25]. In particular, the community research has paid special attention to the Primary User Emulation (PUE) attack, which can severely undermine the primary detection process. In the PUE attack [10], an attacker emulates a primary transmission in order to prevent secondary users from using a vacant band. Consequently, there is a need for providing effective methods so as to distinguish between legitimate primary transmissions and fake ones (PUE attacks).

Research on this topic has been generally based on the recently approved standard IEEE 802.22 WRAN Wireless Regional Area Networks (WRANs) [1], that defines a centralized network composed by a Base Station (BS) and a set of CRs. In such kind of networks, two different types of primary users are defined: TV emitters and wireless microphones. Most of the existing countermeasures for the PUE attack are based on energy measurements of the received signal that are provided by members of the CRN. These measurements are combined in a fusion center, typically the BS, in order to detect whether a given emitter is a legitimate primary user or not. However, these approaches can be easily overcome by an attacker by selecting an optimal position and properly adjusting its transmission power.

Location of the transmission source can be a valuable tool to detect such attacks whenever the position of true primary transmitters is known, as it is the case of TV towers in IEEE 802.22 WRAN networks. In [26] we presented a cooperative location method to effectively deal with PUE attacks in such scenarios. In this method, the position of a potential attacker is estimated based on a set of time measures, which are derived from the feedback provided by members of the CRN. Besides, a heuristic approach is adopted to improve the accuracy of the estimation. The decision about the existence of a primary user or an attacker is then performed by comparing the position estimate with the known positions of true primary users.

The work presented in this chapter is based on [26]; work that is complemented with background knowledge that makes the proposal easier to understand for a non-specific audience. The chapter is structured as follows. Section 13.2 provides an overview of the threats to CRNs. Section 13.3 is devoted to describe the PUE attack, its variants and the proposed countermeasures. Section 13.4 analyzes several location techniques used in wireless networks and their suitability for CRNs, and describes the localization method to detect PUE attacks. In Sect. 13.5 the goodness of the proposed method is proved via simulation and finally, in Sect. 13.6 we provide the conclusions of this work.

13.2 Security Threats in CRNs

Wireless networks are growing in popularity due to its easy deployment and their ability to provide high-speed access to portable devices, and to areas where running cable is not an option. However, malicious users can also take advantage of such ease of access and mobility to attack from any location. As wireless communications use the air as the physical media, they are more easily accessed by an attacker and

Table 13.1 Attacks to CRNs (new specific ones in bold)

Physical	*Link*	*Network*	*Transport*
Spoofing* / sybil* [15, 22]			
Jamming*	Packet injection		
PUE* [11]	Selfish [22]		Jellyfish [2]
Learning engine* [13]	Selective forwading [22]		**lion*** [24]
Common control channel* [35]	**False feedback** [22, 28]		Key depletion* [28]
	Worm/sink-hole [22]		

thus intrinsically more vulnerable than its wired counterparts to attacks such as eavesdropping, data modification, impersonation or Denial of Service (DoS).

Due to its wireless nature, CRNs inherit most of the threats already reported in the literature in the context of wireless networks. However, the flexibility and reconfigurability capabilities of these networks can make them even more sensitive to conventional attacks but also expose them to new security implications [7, 13, 25].

In the context of CRNs, we define an attack as an action that achieves at least one of the following goals:

- **Unacceptable interference to primary users**. Because of the attack, the communication channel of the primary/licensed users is diminished or becomes unusable, i.e., a DoS attack.
- **Missed opportunities for secondary users**. An attacker could prevent secondary users from using available spectrum bands, by reducing the channel performance or denying service to secondary users, among others.
- **Access to private data**. An attacker could try to access data without authorization. As a consequence data must be secured by cryptographic primitives.
- **Modification of data**. An attacker could try to modify the data exchanged between several entities to its own advantage. Thus, integrity of data must be assured.
- **Injection of false data**. Injection of false data could lead the CRN to behave in an unpredictable way or to follow the attacker guidelines. Therefore, authentication of information sources must be guaranteed.

As in any other type of network, the last three threats may be overcome by providing basic security services such as confidentiality, integrity and authentication. However, the remaining threats are specific to CRNs and constitute the base for several attacks that will be described throughout this section.

Table 13.1 depicts a taxonomy of the attacks to CRNs according to the layer that is the target of the attack. Specific attacks to CRNs appear in bold. Next we provide an overview of these attacks based not only on the target layer but also on whether the attacker is a member of the CRN (inside attacks) or not (outside attacks). While any attack in the table can be carried out by the actual CRN members (often with an increased scope), some attacks can also be performed by outsiders; such attacks are marked with an asterisk in the table.

13.2.1 Outside Attacks

Outside attacks are those carried by an entity from outside the victim CRN, i.e., attacks that can be executed by a non-authorized entity. As a result, we assume the attacker cannot log in (spoof) the victim network. Obviously, when the attacker spoofs an authorized identity can also execute insider attacks (those performed by authorized entities). For this reason, authentication/authorization at the link layer should be mandatory to secure the network.

The Sybil attack is a form of spoofing attack in which the attacker spoofs multiple identities, and hence it can gain a larger influence than a simple spoofing attack. This attack, which was originally intended to attack the redundancy mechanisms of peer-to-peer storing systems, is also commonly used to attack routing protocols, data aggregation mechanisms, voting systems, avoid detection of malicious users, etc.

Notice that it cannot be assumed that an outside attacker, although it is not authorized, has no knowledge about the victim network. Quite the opposite, the attacker may know the sensing protocols, the potential primary users, etc. and hence execute more specific attacks.

One of the most known attacks to wireless networks at the physical layer is jamming. Within the simplest form, a jamming attack involves the radiation of radio signals that intentionally[1] disrupt communications in the victim network. If the generated interferences are big enough, they can substantially decrease the performance of communications or completely interrupt them, thus implementing a DoS attack. In CRNs, however, interferences can be also created by fake primary users, leading to the PUE attack, which will be describe more in detail in Sect. 13.3, or the learning engine attack. Learning engine attacks are based on disrupting the learning mechanisms used by CRs by altering the sense medium, thus leading CRs to wrong decisions with regard to network configuration [13].

A jamming attack can also lead to a common control channel attack when the target channel is the one used by members of the CRN to share information about spectrum availability/usage. Note that the lack of knowledge about available bands may prevent the CRN from operating, i.e., leads to a DoS attack. Besides, because these channels carry sensitive information, they must be secured in order to guarantee confidentiality, authentication and integrity of the data by means of cryptographic primitives, and thus prevent an eavesdropper from obtaining information about the operation of the CRN. As we will see in Sect. 13.3, an attacker may use this information to perform a more damaging attack, such as the PUE attack.

Finally, CRNs are also exposed to cross-layer attacks, which exploit the vulnerability of a given layer but are targeted to disrupt an operation at a different layer. Because these attacks involve different layers they are especially difficult to detect. However, the amount of cross-layer attacks that can be performed by outsiders is somehow

[1] The term *jamming* is used for intentionally disruption of communications while the term *interferences* usually refers to unintentionally one.

limited, since they can only access the physical or the link layer. As an example, an outsider could perform a key depletion attack [28] or the Lion attack [20, 24], which is based on the PUE attack and will be described in Sect. 13.3.

13.2.2 Inside Attacks

An inside attacker can perform any of the outside attacks presented in the previous section but also others that take advantage of its authorization to use the network. In the following, we present an overview of such attacks and classify them according to the TCP/IP model layer in which they are performed.

Attacks to CRNs at the link layer are mainly targeted to the cooperative sensing mechanisms. It is well known that cooperation among CRs improves spectrum sensing, but it is always assumed that secondary users are honest and willing to cooperate. However, a given CR may report false measurements either because it has been compromised or due to malfunctioning; or it may decide to not cooperate in order to increase its benefit (for example, saving energy) sinking in selfish behavior. Both type of behaviors can increase the probability of wrong decisions regarding spectrum opportunities [9].

At the network layer, most attacks focus on the routing protocols, either producing a DoS or modifying the protocol behavior to acquire a profit. Note that these attacks may only affect to distributed CRNs in which, typically, the used routing protocols are the same as in other existing wireless networks. Thus, well-known attacks to routing, especially the ones targeted to ad-hoc routing such as the selective forwarding attack, the sink hole attack or the worm-hole attack [23], can also be executed against CRNs.

Finally, most attacks at the transport layer are based on disrupting TCP connections. As an example, packet injection aims at producing a DoS or gaining access to protected resources by injecting forged packets into an existing connection. Another example are cross-layer attacks, such as the Jellyfish attack [2], which is performed at the network layer by means of packet disordering, packet dropping and packet delay but targeted to degrade the throughput of TCP connections.

13.3 The Primary User Emulation (PUE) Attack

13.3.1 Description of the Attack

CRNs act as secondary users of the licensed spectrum and must not interfere with primary transmissions. An adversary could take advantage of this feature in order to disrupt communications within a CRN by performing a PUE attack. In a PUE attack, first coined in [10], an attacker pretends to be a primary user or incumbent by transmitting a signal with similar characteristics to a primary signal or replying a real one. If the attack succeeds, it prevents the CRN from using a vacant band.

Moreover, with previous knowledge on the CRN operation, the attacker can force PUE attacks whenever the CRN switches from one channel to another, leading to a DoS. The attacker can gather information about the channel being used by the CRN in the following ways:

- By performing spectrum sensing until finding the new channel of operation of the CRN. In order to minimize the search time, the attacker may discard some channels directly, e.g., channels already in use by primary users. Moreover, if the attacker lies in the CRN area, it can estimate the most likely CRN channel by means of its own sensing measures.
- By eavesdropping the common control data of the CRN (if exists). This threat can be easily overcome by securing the channel by means of cryptographic primitives, as it is already recommended by the IEEE 802.22 WRAN standard.

The impact of the PUE attack depends on several factors, such as the location of the attacker or the sensing mechanism used by the CRN. Selecting an optimal position to perform the attack will cause many secondary users reporting the existence of a primary transmission, and therefore will lead the CRN to look for another portion of the spectrum. Moreover, if the sensing mechanism used by the CRN looks for specific characteristics of the signal, the fake signal should fulfill several requirements with regard frequency, code, modulation, etc., in order to appear as a legitimate one. Although in this case the PUE attack is harder to perform, it is still quite feasible because the attacker can program its CR device in order to match the transmission parameters of a primary user, or even transmit a real primary signal previously recorded.

The most widespread sensing mechanism used for primary detection is energy detection [8], which simply relies on an energy threshold in order to take a decision about the existence of a primary. Because energy detection is unable to discriminate between primary and secondary signals, the 802.22 standard suggests the use of quiet periods in the CRN in which transmissions are not allowed. This allows performing spectrum sensing while avoiding the potential interferences that secondary users could produced. Thus, any transmission detected during that period may be considered as a primary signal if the received power at CRs is above a given threshold. This threshold plays an essential role since the lower the threshold, the higher the detection probability but also the easier to perform a PUE attack.

Although frequency handoffs could also be forced by means of jamming, there are fundamental differences which may motivate an attacker to perform specifically a PUE and not simply jam the channel. First, a CRN is required to perform a frequency handoff upon detection of a primary transmission, even if the remaining channels offer worse transmission conditions. If the channel is jammed, the victim CRN may just perform the handoff if the overall transmission conditions are below a certain threshold and a better channel is available. Moreover, note that the cost of a PUE attack comes down to transmit a signal similar to a primary signal, e.g., TV or wireless microphones signals, or replay a real one. Besides, with the same effort or amount of resources the scope of a PUE attack can be much larger. Even if the fake primary transmission is only detected in a small area of the CRN, it can force a

frequency handoff and affect the whole CRN. On the contrary, the fact of degrading the communication channel only in a small area should not be enough to force the CRN to perform such handoff.

13.3.2 Variants of the Attack and Their Scope

The basic PUE attack is targeted to force a frequency handoff in a CRN. However, it must be noted that an attacker may also perform a PUE attack aiming at disrupting the learning process of CRs, i.e., to perform a learning engine attack (see Sect. 13.2.1). If a successful PUE attack is performed on a given channel during the learning phase of the CRN, that channel may be considered to be occupied during the whole operation of the CRN. Thus, the effect of this variant lasts for a longer period of time than in the basic attack.

PUE attacks can also be used to perform cross-layer attacks, such as the Lion Attack [20, 24]. The Lion attack is performed at the physical-layer and targeted to the transport layer, aiming to degrade the throughput of TCP connections established in CRN. It is based on performing a PUE attack in order to force a frequency handoff in the CRN, which leads to the interruption of all communications in the network for a given period of time. This interruption can have a harmful impact over the throughput of TCP connections due to the interaction with its congestion control mechanisms [5].

A smart version of the attack could be implemented based on the knowledge of the value of the retransmission timer of the TCP connection. In typical CRNs such as WRAN 802.22 networks [14], the Round-Trip-Time (RTT) value for in-network communications is around some hundreds of microseconds. Although the value of the retransmission timer, from now on RTO (Retransmission TimeOut), is variable and depends on the RTT estimations, most implementations round off its RTO to a minimum value of typically 100 or 200 ms, much higher than the real RTT. This fact will lead the TCP sender to make use of a fixed value for the RTO, which will be doubled for each unsuccessful attempt. The attacker can take advantage of this information to force handoffs at the specific instants in which retransmission attempts are performed and produce a DoS [20].

13.3.3 Countermeasures

Protecting CRNs from PUE attacks is indispensable and requires devising robust techniques for verifying the authenticity of primary signals, such as TV broadcast systems or wireless microphones. The simplest way to achieve it would be to embed a signature in an incumbent signal or to use an authentication protocol between primary and secondary users. However, these approaches do not conform to the requirement established by the Federal Communications Commission (FCC) [17], which states

that *no modification to the incumbent system should be required to accommodate opportunistic use of the spectrum by secondary users*.

Under such demand, most proposals appeared in the literature are based on energy measurements and usually make use of hypothesis testing to reach a decision about the existence of an attacker [11, 12, 21]. As energy detection is typically used in CRNs with spectrum sensing purposes, it is also often the preferred mechanism to detect PUE attacks, as it does not require extra hardware in CRs.

Energy-based approaches can generally deal with PUE attacks, assuming primary users with known and fixed locations, and with transmission powers considerably higher than an attacker [11]. Despite it, in network environments where primary users are mobile and transmit with low power transmission, i.e. wireless microphones, they are prone to fail in detecting such attacks.

There exist alternative countermeasures such as Radio Frequency Fingerprinting (RFF) [3], which is based on the fact that the radio signal emitted by a radio transmitter when it is activated exhibits a transient behavior with respect to instantaneous frequency and amplitude, known as the fingerprint of a radio device. This method requires recording a wide variety of signals emitted by both CRs and incumbents, extracting the transient portion that reflects the unique attributes of each device for each signal, and storing the different patterns or fingerprints. Another disadvantage is that, as the transient behavior of emissions vary due to aging and degradation of the devices, the set of fingerprints must be periodically updated. In spite of its complexity, the advantage of this technique is that it can be applied to identify any transmitter, no matter whether it has a fixed location or it is mobile.

In IEEE 802.22 WRAN networks two types of primary users are defined: TV and wireless microphone transmissions. When a PUE attack is based on wireless microphones, energy-based techniques are not applicable because wireless microphones are low power devices with a transmission range of at most 100–150 m. Anti-PUE schemes based on cooperative energy sensing, where decisions are taken based on measurements performed by many CRs, would probably fail on detecting such attack: due to the large dimension of CRNs, most of the CRs would not be in the range of the emitter. Moreover, wireless microphones are usually mobile, a fact that makes even more difficult to determine whether a given source is legitimate or not.

Being aware of the difficulty of detecting such attacks and against the requirement established by the FCC, the 802.22 Working Group (WG) proposed the use of a beacon protocol [6]. In this approach, wireless microphones should send a special signal before starting transmission for signaling their presence. It could be embedded additional information, such as digital signatures, in the beacon signal, thus allowing to easily authenticate the device.

In [27], the authors proposed an approach similar to random frequency hopping, where secondary users randomly select a channel to transmit, avoiding PUE attacks. Although it represents an effective countermeasure to avoid PUE attacks and can also deal with traditional jamming attacks, it leads the CRN to continuously perform frequency handoffs, implying the interruption of all communications until the CRN is completely operating at a new frequency. Moreover, the success of this

countermeasure strongly depends on the number of available channels when the attack is performed: lack of alternative channels can lead to a DoS.

Finally, an effective countermeasure to mitigate the impact of the PUE attack on TCP connections, i.e., the Lion attack, could be to make TCP aware of the interruptions caused by frequency handoffs. In [24], a cross-layer solution is proposed. This proposal is a modified version of Freeze-TCP in which the TCP sender gets information physical disconnections and freezes its transmission and congestion control parameters during the frequency handoff.

13.4 Detection of PUE Attacks Based on Cooperative Localization

In this section, we describe a cooperative location method to estimate the position of an emitter and detect PUE attacks in CRNs [26]. The proposed method is intended to be applied to infrastructure-based CRNs based on the IEEE 802.22 WRAN standard, where primary users are TV towers.

In order to detect primary emitters in the CRN channel of operation, either legitimate or fake, all CRs perform spectrum sensing and report their measurements to the BS. The BS acts as a data fusion center and takes a decision about the existence of a primary transmission based on these reports. If there is evidence of the presence of a primary user, the CRN applies a localization method to estimate the position of the emitter. The transmission is considered to be legitimate whenever its position matches any of the known TV transmitters' positions. Otherwise, it is assumed that a PUE attack is being performed.

The location method is based on the Time Difference of Arrival (TDoA) technique [31] and applies multilateration in order to estimate the position of the emitter. In particular, Series-Taylor estimation is used to solve the system of equations derived from the set of TDoA measures. On the one hand, TDoA does not require the collaboration of the node to be located and provides higher accuracy than other techniques such as RSS (Received Strength Signal). Series-Taylor estimation is a commonly used approach for node's location because it provides higher accuracy than most of the non-iterative methods. Besides, it is preferable to filter schemes such as the Extended Kalman's filter, which exhibit higher complexity and provide no significantly improvement when the node to be located is static.

In the following we provide a detailed description of this method. We first describe the existing location techniques and justify the use of TDoA in our proposal. Then, we present the assumptions that have been considered for its design. Next, we describe the steps to follow in order to obtain TDoA measurements and to apply Series-Taylor estimation in order to derive the position of the emitter.

13.4.1 Background on Location Techniques

Physical location of RF transmission sources has been a hot topic for many years in wireless applications, especially for those related to military and safety purposes, but recently its interest has even increased due to its application to a wide variety of civil services such as E-911 telephone service, traffic routing, fleet management, etc.

Wireless location techniques fall into two main categories: mobile based and network based. In mobile-based location systems, a given mobile node determines its position by means of signals received from some BSs or from the global positioning system (GPS). Network-based technologies rely on measurements of certain distance-dependent parameters performed at the BS or at nodes whose position is well known. Typically, these parameters are Received Signal Strength (RSS), Angle of Arrival (AoA) and Time of Arrival (ToA).

GPS-based techniques rely on devices that report their position which have been computed by precisely timing the signals sent by GPS satellites. Obviously, this will not be the case of an attacker device, and as a consequence GPS-based techniques are not suitable for locating an attacker but an actual legitimate node of the network.

RSS-based techniques rely on the fact that the signal strength varies inversely with the square of the distance in free space. Therefore, assuming that the path loss model is known, it is possible to locate a transmission source by measuring the signal strength received at several nodes. Although RSS measurements are relatively inexpensive and simple to implement in hardware [32], they are susceptible of high errors due to the dynamics of indoor/outdoor environments, mainly due to multipath signals and shadowing. The effect of shadowing is usually modeled as log-normal and leads to RSS-based estimates with variance proportional to its range. Therefore, although there have been some successful attempts in order to apply RSS-based location techniques to CRNs [11], they are limited to ranges of at most 2 km and a great number of sensor nodes, from hundreds to thousands, cooperating in the localization process. As a result, RSS-based techniques may not be suited for IEEE 802.22 WRAN networks: first, in these networks the BS range is up to 100 km; and second, since these networks are targeted to rural environments, one cannot expect cooperation of such a great amount of nodes during the location process.

In ToA, distances from the node to be located to some reference nodes are computed by measuring the signal propagation time. Usually [33], measurements are obtained by transmitting a signal to the node we wish to locate, which immediately replies with its own signal. The time elapsed since the transmission of the signal until the reception of the reply is then used to estimate the distance between both nodes. As in mobile-based approaches, it requires the collaboration of the node to be located and therefore we cannot either apply ToA in CRNs to obtain the position of a potential attacker. Time Difference of Arrival (TDoA) techniques bypass the need of cooperation by computing the differences in ToA measurements for a given signal's source at two or more pairs of nodes. However, TDoA requires a tight synchronization among the nodes that cooperate in the location process.

Finally, the angle of arrival (AoA) method can be used to locate an emitter by measuring the direction of the received signal at different reference nodes and applying simple triangulation. However, this technique is extremely affected by multipath and requires every reference node to have a directional antenna or antenna arrays for locating purposes [34]; an strong requirement in CRNs.

From the previous reasoning, only TDoA can be used in IEEE 802.22 WRAN networks for locating a PUE attacker. RSS-based techniques are not accurate enough for long-range links, and GPS an AoA-based require new specific hardware. In the following subsections we go into detail in how TDoA ideally works and how to apply it to non-ideal (noisy) real channels.

13.4.1.1 TDoA

TDoA utilizes cross-correlation to calculate the difference in the ToA of an emitter signal at two or more pairs of nodes, thus requiring at least two TDoA measurements (three nodes or more) to locate an emitter on a surface, and three measurements (at least four nodes) to locate the emitter on the 3-dimensional space. As a drawback, it requires a tight synchronization among the nodes. Next, we describe how to apply TDoA in a 2-dimensional space.

Let (x, y) be the position of the emitter to be located. Such position is within the range of n receivers at positions (x_i, y_i) with $i \in [1, n]$, and a reference receiver node (typically the BS) that actually knows the position of all receivers. For the sake of simplicity, we will assume that the reference node is located at position $(0, 0)$ and the other nodes' positions are relative to the former. From the previous assumptions, the TDoA measurement between the pair made up of receiver i and the reference node, when the propagation velocity is v_p, can be expressed as in (13.1) as a function of (x, y).

$$f_i(x, y) = \frac{\sqrt{(x - x_i)^2 + (y - y_i)^2} - \sqrt{(x)^2 + (y)^2}}{v_p} \tag{13.1}$$

Ideally, the intersection of 2 or more different of the above hyperbolas unequivocally determines the emitter position (x, y). If we apply the same reasoning to a 3-dimensional space, with two receivers at known locations, an emitter can be located onto a hyperboloid. A third receiver provides then a second TDoA measurement and hence locates the emitter on a second hyperboloid. The intersection of both hyperboloids describes a curve on which the emitter lies. If a fourth receiver is now introduced, a third TDoA measurement is available and the intersection of the resulting third hyperboloid with the curve already found with the other three receivers defines a unique point in space. This process is known as multilateration or hyperbolic positioning.

However, in practice measurements are subjected to errors and the surfaces rarely intersect. In this case, the TDoA measurement may be expressed as in (13.2) with $\Delta_i(x, y)$ the error performed in the ith TDoA measurement; and the Euclidean

distance error $e_i(x, y)$ derived from measurement τ_i as in (13.3) as a function of the emitter position (x, y).

$$\tau_i = f_i(x, y) + \Delta_i(x, y) \tag{13.2}$$

$$e_i(x, y) = v_p \Delta_i(x, y) = v_p \tau_i - v_p \cdot f_i(x, y) \tag{13.3}$$

The fact of minimizing the localization error can be posed as an optimization problem and solved using, for example, a Least Squares (LS) method or an extended Kalman-Bucy filter. Since Kalman-Bucy provides no significant improvement in accuracy when the emitter to be located is usually stationary [18], we have decided to apply a LS method based on Taylor-Series Estimation, an iterative scheme which starts with a rough initial guess and improves the guess at each step by determining the local linear least-sum squared-error correction.

By applying a LS method, we aim at minimizing the sum of squared errors performed in the set of TDoA measurements. Nevertheless, since every node has different environment conditions (e.g. different SNRs), not all the measurements have the same reliability. Therefore, assuming that the measurements performed by each node are statistically independent, a Weighted Least Squares (WLS) approach is preferred. In the WLS technique a different weight is assigned to each measurement according to its reliability, and then the target is to minimize the expression in (13.4).

$$\sum_{i=1}^{n} w_i e_i^2(x, y) \tag{13.4}$$

This problem is usually simplified by finding the offset in the x and y axes (δ_x, δ_y) that minimizes the quadratic sum of a linearized approximation of $e_i(x, y)$ around a guess (x_v, y_v) of the true emitter's position. This linearized form, denoted as $\hat{e}_i(x, y)$, is obtained as in (13.5) by expanding $e_i(x, y)$ in Taylor's series around the guess (x_v, y_v) keeping only the first order coefficients a_{ix} and a_{iy}, defined as in (13.6).

$$\hat{e}_i(x_v + \delta_x, y_v + \delta_y) = e_i(x_v, y_v) + a_{ix}\delta_x + a_{iy}\delta_y \tag{13.5}$$

$$a_{ix} = \left.\frac{\partial e_i(x, y)}{\partial x}\right|_{x=x_v} \qquad a_{iy} = \left.\frac{\partial e_i(x, y)}{\partial y}\right|_{y=y_v} \tag{13.6}$$

Upon (δ_x, δ_y) is obtained, the guess of the emitter's position is updated as $(x'_v, y'_v) = (x_v + \delta_x, y_v \delta_y)$ and, under convergence conditions (which clearly depend on the initial guess), the guess is being refined by successive iterations until (δ_x, δ_y) are below a given threshold.

The process of obtaining the value of (δ_x, δ_y) that minimizes (13.4) is easily obtained by operating with the matrix representation of the linearized forms of the error as in (13.7) with vectors and matrices defined as in (13.8).

$$\hat{\mathbf{e}} = \mathbf{A} \times \delta + \mathbf{e} \tag{13.7}$$

$$\hat{\mathbf{e}} = \begin{pmatrix} \hat{e}_1 \\ \hat{e}_2 \\ \vdots \\ \hat{e}_n \end{pmatrix} \mathbf{A} = \begin{pmatrix} a_{1x} & a_{1y} \\ a_{2x} & a_{2y} \\ \vdots & \vdots \\ a_{nx} & a_{ny} \end{pmatrix} \delta = \begin{pmatrix} \delta_x \\ \delta_y \end{pmatrix} \mathbf{e} = \begin{pmatrix} v_p\tau_1 - f_1(x_v, y_v) \\ v_p\tau_2 - f_2(x_v, y_v) \\ \vdots \\ v_p\tau_n - f_n(x_v, y_v) \end{pmatrix} \tag{13.8}$$

Then, assuming that $\hat{\mathbf{A}}$ is full rank, the value of δ that minimizes expression in (13.9) can be obtained as in (13.10), with $\mathbf{W}$ a diagonal $n \times n$ matrix where each diagonal element w_{ii} represents the weight assigned to measurements of node i.

$$\sum_{i=1}^{n} w_{ii}\hat{e}_i^2(x_v + \delta_x, y_v + \delta_y) = \hat{\mathbf{e}}^{\mathrm{T}}\mathbf{W}\hat{\mathbf{e}} = (\mathbf{A}\delta + \mathbf{e})^{\mathrm{T}}\mathbf{W}(\mathbf{A}\delta + \mathbf{e}) \tag{13.9}$$

$$\delta = -[\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A}]^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{e} \tag{13.10}$$

13.4.2 Assumptions

In order to develop the proposed method, the following assumptions have been adopted:

- CRs are randomly located and their position is fixed and known to the BS.
- The BS has a database with the true position of TV primary emitters and their respective frequencies of operation.
- The PUE attacker remains at a fixed position and can have different capabilities regarding power transmission.
- The BS is responsible for scheduling quiet periods with sensing purposes, in which no station is allowed to transmit in the channel to be sensed. As recommended by the IEEE 802.22 WRAN standard, two different types of sensing mechanisms are used: fast sensing, which can be performed frequently and generally relies on energy-based sensing techniques, and fine sensing, which is based on more complex techniques that allow obtaining specific features of the signal but require a longer observation time. As we will further explain, the proposed localization method will be embedded in the fine sensing process.
- The measurements performed by each node are normally distributed with zero mean [19] and statistically independent. The latter can be assumed if CRs are distanced more than a few hundred of meters [30].

13.4.3 Acquisition of Valid TDoA Measures

In order to detect primary transmitters (no matter whether they are real or fake ones), all the anchor nodes within the CRN perform sensing and report their measurements to the BS, which is responsible for making a decision about the existence of a primary transmission in the CRN band of operation. Once a potential primary transmission is detected, the CRN performs a localization method to know whether this source of primary transmissions is legitimate (by comparing its position with known TV transmitters' positions) or fake (a PUE attack). In order to apply the localization technique explained in Sect. 13.4.1.1, a set of at least two TDoA measurements are required. However, as mentioned before, the accuracy of these measurements has a strong impact on the position estimation and highly depends on a tight synchronization between the CRN nodes and the BS. Many synchronization proposals on the state of the art [31] can be applied to CRNs but their accuracy is at most in the order of microseconds which could lead to errors of hundreds of meters. For this reason, in the following we propose a method to synchronize reports from several anchor nodes in order to obtain accurate TDoA measurements. The chosen solution is fairly straight forward and can be summarized in five steps.

1. When the localization process starts, the BS first requests all the CRN anchor nodes to start recording the primary signal and then starts its own recording.
2. During the recording, the BS sends a marker signal through the CRN operation antennas and every anchor node will add this marker to its record at the precise instant it has received it. The marker signal is a predefined signal that any anchor node can recognize.
3. Every anchor node sends the recording with the marker to the BS. Since the BS knows the physical position of every anchor node, it can compute the elapsed time since the BS has sent the marker until it has been received. Consequently, the BS is able to synchronize the recordings. Assuming that the BS is placed on $(0, 0)$, the elapsed time can be defined as in expression (13.11) with i the anchor node receiving the marker, (x, y) its position and v_p the propagation velocity.

$$\Delta t_i = \frac{\sqrt{x_i^2 + y_i^2}}{v_p} \tag{13.11}$$

4. Once the records have been synchronized, the BS derives the set of TDoA measurements by computing the delays between the recordings of every anchor node and its own one.
5. Finally, by applying the least squares method described in 13.4.1.1, the BS makes an estimation of the position of the emitter.

Figure 13.1 summarizes the first four steps. The signal on the top is the one recorded by the BS just after the synchronization marker is sent (step 2). The other two signals represent the recordings of two different anchor nodes. Since both nodes send the recording with the marker, the BS can synchronize the recordings: the time

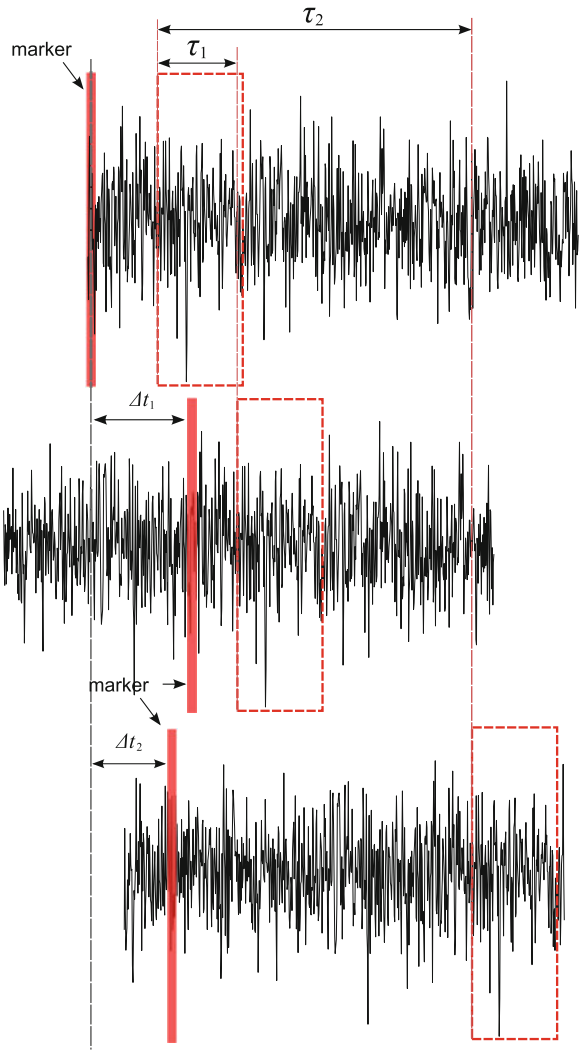

Fig. 13.1 Obtaining TDoA measurements from synchronized recordings

when the BS sent the marker is exactly the time when the anchor node i received the marker minus Δt_i. Once the recordings are synchronized, the BS just needs to compute the delay between recordings that actually corresponds to the TDoA measurements τ_1 and τ_2.

13.4.4 Applying WLS Position Estimation to Guess the Emitter's Position

As explained in Sect. 13.4.1.1, the set of TDoA measurements leads to a non-linear system of equations with no unique solution; an optimization problem that we have decided to solve by means Taylor-Series estimation. Therefore, it is required to select an initial estimation of the position of the emitter (x_v, y_v) and the weights assigned to each measurement in the WLS method.

Assuming that the measurements performed by each node are statistically independent, the weight w_i assigned to the TDoA measurement obtained from the pair node i and the BS can be computed as the inverse of the error variance of such measure.

The error in ToA measurements is modeled as a normally-distributed random variable whose variance is derived by means of the Cramér-Rao Lower Bound (CRLB) [19], which provides a lower bound on the variance in a multipath-free channel. In the case of TDoA, as any observation is computed as the difference of two ToA observations, any observation between a node i and the BS can be modeled as $\tau_i \approx \mathcal{N}\left(f_i(x, y), \sigma_i^2 + \sigma_0^2\right)$, with σ_i as in (13.12) the variance of the measurements from anchor node i, σ_0 the variance of the BS's measures, B the signal bandwidth and SNR_i the signal-to-noise relation at anchor node i.

$$\sigma_i^2 \geq \frac{1}{8\pi^2 \cdot B^2 \cdot SNR_i} \tag{13.12}$$

In the IEEE 802.22 WRAN standard, the Hata model for suburban areas has been proposed for the path loss calculation [16]. As a result, assuming that the CRN nodes have isotropic antennas, the SNR_i at a given anchor node i can be expressed as in (13.13), with ΔL_p as in (13.14), SNR_0 the signal-to-noise ratio at the BS, d_i and d_0 the distances from the emitter to the anchor node i and the BS respectively, and h_b the current altitude of the emitter to be located.

$$SNR_i(dB) = SNR_0(dB) - \Delta L_p(dB) \tag{13.13}$$

$$\Delta L_p(dB) = \left[44.9 - 6.55 \cdot \log_{10}(h_b)\right] \log \frac{d_i}{d_0} \tag{13.14}$$

The main disadvantage of iterative WLS estimators is that the convergence depends on the goodness of the initial guess. As a result, the choice of the initial guess (x_v, y_v) plays a crucial role. In order to improve the choice of the initial guess, we propose to proceed as follows.

The BS defines an m-by-n matrix with m the number of potential initial guesses and n the number of anchor nodes. This matrix stores precomputed values of TDoA measurements for the m initial guesses. p out of the m positions match the position of

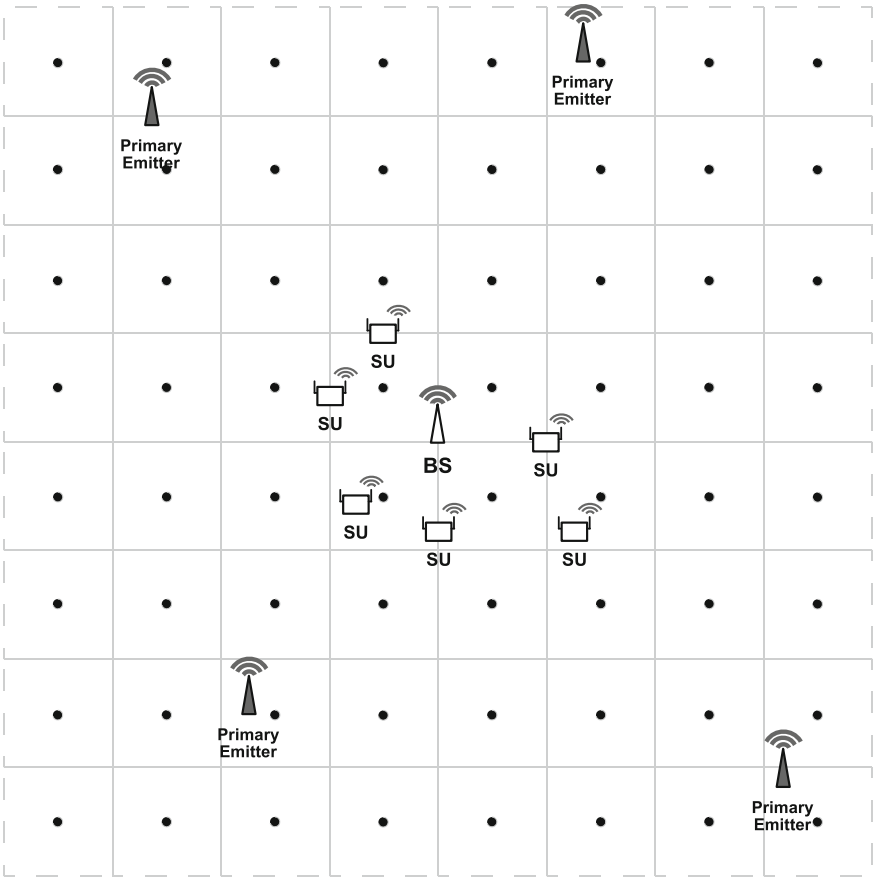

Fig. 13.2 Positions for TDoA measurements in database

real primaries (already known to the BS), the remaining ones are obtained by drawing a grid on the area and taking the central point of every single cell (see Fig. 13.2).

When the localization process starts, the BS first collects the recordings and the following steps are performed:

- The BS computes the Euclidean distances between the vector containing the TDoA measurements performed by the n anchor nodes and each of the m rows of the matrix containing the TDoA measurements stored at the BS's database. These distances are sorted in ascending order and stored in a vector V.
- The BS selects the first element of V as the initial guess and applies the iterative WLS algorithm.
  - If the algorithm does not converge, the BS selects the following element of V as a new initial guess and restarts the iterative WLS algorithm. If there are no more initial guesses, the algorithm ends without converge.

– If the algorithm converges, the output of the WLS method is the estimation of the emitter's position.

Intuitively, the first element of V, which denotes the position that minimizes the distance between the TDoA measurements and its precomputed TDoA measurements, should be the closest one to the real position of the emitter. However, small differences in any of the axes may mix up the elements of V. In any case, the closest precomputed solution will always be within the first positions of V, so we change the initial guess whenever the algorithm does not converge by selecting the next element in V.

The initial error estimation is therefore the Euclidean distance in the coordinate system between the chosen element of V (precomputed position) and the emitter's real position. Obviously, a more accurate estimation can be obtained by reducing the cell size at the expense of a higher cost in terms of allocated memory and computation time.

We set a maximum value for the error ($\delta_{x_{max}}$, $\delta_{y_{max}}$) and perform as many iterations as needed until the error in both axes is below these thresholds. The decision about the existence of a primary user or an attacker will then be made according to the value estimated in the last iteration.

13.5 Performance Evaluation

In this section we evaluate the goodness of the proposed method which has been programmed in Matlab [29]. Simulation results for different scenarios are provided.

13.5.1 Simulation Scenario

In order to evaluate the performance of the proposed location method, we have considered a CRN following the IEEE 802.22 WRAN standard, which is composed by a BS located at the origin and a set of CR nodes uniformly distributed within a square area of $30 \times 30\,\text{km}^2$, as shown in Fig. 13.2, and the following assumptions have been made:

- The BS is located at the origin (see Fig. 13.2)
- 10–100 anchor nodes are randomly placed in the area.
- Primary TV transmitters are located outside the CRN perimeter at well-known positions, at distances ranging from 30 to 150 km far away from the BS.
- The sensing area is divided into a grid of 256 square cells of $\frac{150000}{\sqrt{256}} = 9375\,\text{m}$ side. As explained in 13.4.4, we assume that the BS stores a database with the potential TDoA measurements for transmission sources located at the center of every cell.

- The Hata model for suburban areas is used for modeling the radio frequency propagation. The choice relies on the fact that IEEE 802.22 WRAN networks are intended to be deployed over rural areas.
- The primary signal is received at the BS with different SNR values ranging from −10 to 10 dB (worst to best conditions).
- All the provided results are obtained from 10,000 experiments of every different case.

We have also considered a malicious user performing a PUE attack from different positions both inside and outside the CRN area.

13.5.2 Methodology

The proposed method may introduce a significant overhead in terms of control data transmission and lost of capacity of the CRN. Because of this, there is a need for analyzing the amount of extra time needed to carry on the location process. The results of this analysis are shown in Sect. 13.5.4. Besides, the efficiency of this method in detecting PUE attacks relies on the accuracy of the estimated position of the emitter. Thus, in order to prove the goodness of our proposal, in Sect. 13.5.3 we evaluate it in terms of location error. In particular, we provide elliptical confidence intervals of the error that are derived from a set of samples as follows.

Figure 13.3 represents the normalized cumulative histogram of the position estimation error obtained in 1000 simulations, when an attacker is located at (8000 m, 1000 m), 100 anchor nodes are cooperating in the location process and assuming a SNR = −10 dB at the BS. The confidence levels of 50, 95 and 99 % are represented in the figure with red lines and they correspond to distance errors of approximately 8.66, 20.76 and 27.56 m respectively. Notice that under these conditions, with probability 0.99 the error represents at most just a $\frac{27.56}{\sqrt{(8000)^2+1000^2}} \approx 0.34$ % of the distance between the attacker and the BS.

Figure 13.4 represents the mean of the distance error and the confidence intervals for a varying number of anchor nodes cooperating in the localization process. The figure clearly shows that cooperation actually improves the estimation of the emitter's position, leading in the worst case ($n = 10$) to an error around 450.7 m, and in the best case ($n = 100$) to an error of approximately 9.07 m with a 99 % confidence interval.

By plotting the error samples on a 2-dimensional Cartesian coordinate system as in Fig. 13.5, it can be seen that these errors spread over an elliptical area with the real emitter's position in the center of the ellipse and an angle as in (13.15) between the abscissa and the ellipse's major axis. Therefore, the confidence levels can be represented as ellipses and thus they provide an improved idea about the error dispersion.

$$\varphi = \arctan\frac{y}{x} + k \qquad k = \begin{cases} \pi \text{ if } x < 0 \\ 0 \text{ if } x \geq 0 \end{cases} \tag{13.15}$$

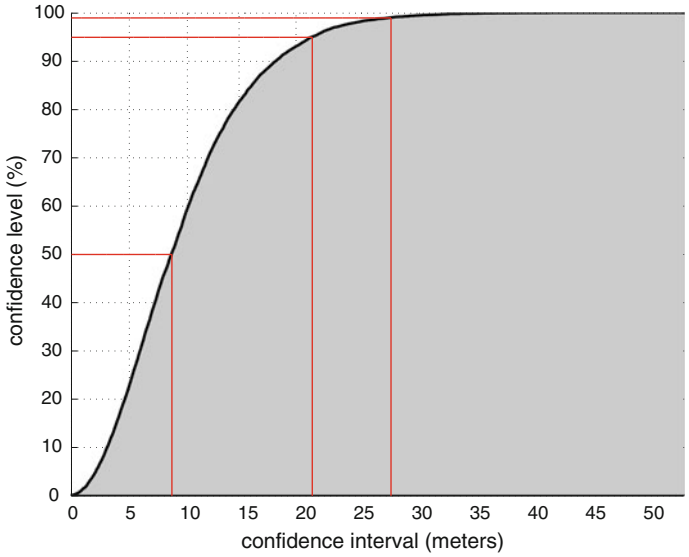

Fig. 13.3 Normalized cumulative histogram of the estimator error. Attacker at (8000, 1000 m), SNR = −10 dB, $n = 100$

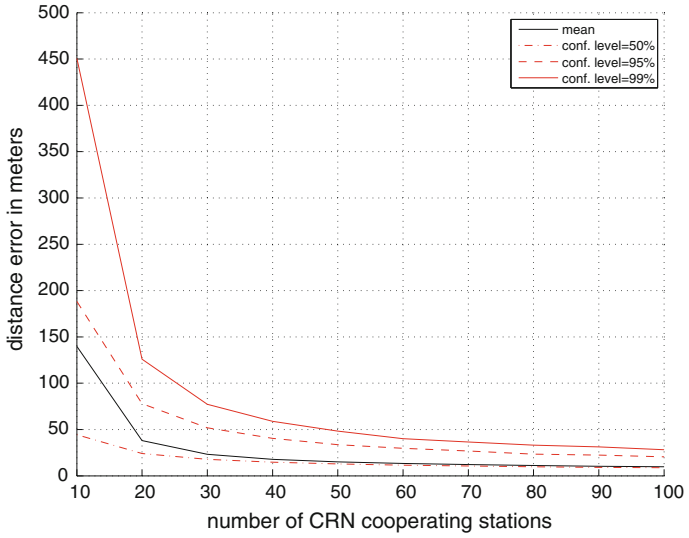

Fig. 13.4 Confidence interval of error. Attacker at (8000, 1000 m), SNR = −10 dB, $n = 100$

In order to obtain the eccentricity of the ellipse we project every error ($e_{x_i} = x_i - x, e_{y_i} = y_i - y$) to the major and minor ellipse axes as in (13.16) with E'_x the vector with the projected error values e'_{x_i} over the major axis of every iteration

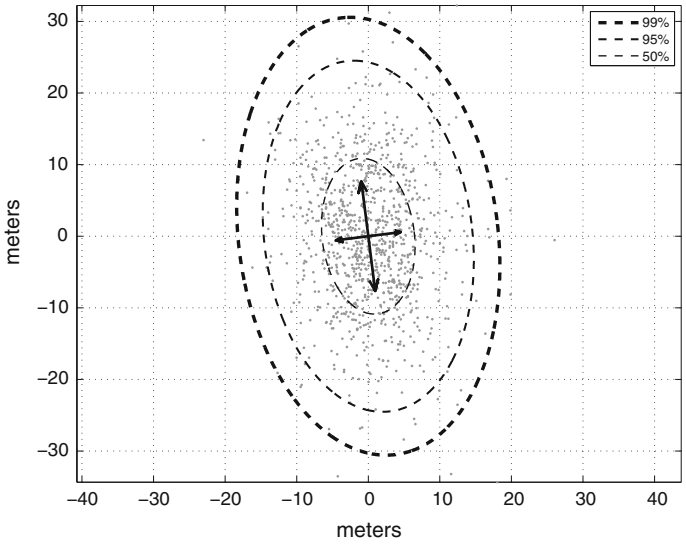

Fig. 13.5 Elliptical confidence interval. Attacker at (8000, 1000 m), SNR = −10 dB, $n = 100$

and E'_y the vector with the projections on the minor axis e'_{y_i}. Then, we can compute the eccentricity as in (13.17) with MAD(E'_x) and MAD(E'_y) the mean absolute deviation of the error in each of the projected axis obtained as in (13.18).

$$\begin{gathered} e'_{x_i} = e_i \cos(\theta) \qquad e'_{y_i} = e_i \sin(\theta) \\ \theta = \arctan\left(\tfrac{y-y_v}{x-x_v}\right) - \varphi + k \qquad k = \begin{cases} \pi \text{ if } x - x_v < 0 \\ 0 \text{ if } x - x_v \geq 0 \end{cases} \end{gathered} \tag{13.16}$$

$$\varepsilon = \sqrt{1 - \left(\frac{\text{MAD}(E'_y)}{\text{MAD}(E'_x)}\right)^2} \tag{13.17}$$

$$\text{MAD}(E'_x) = \frac{1}{n}\sum_{i=1}^{n} |e'_{x_i}| \qquad \text{MAD}(E'_y) = \frac{1}{n}\sum_{i=1}^{n} |e'_{y_i}| \tag{13.18}$$

Once the eccentricity of the ellipse has been obtained, plotting the one that contains a given confidence level (%) of the samples is fairly straight forward. Ellipses in Fig. 13.5 delimit the elliptical confidence interval related to confidence levels of 50, 95 and 99 %. MAD(E'_x) ≈ 4.54 m and MAD(E'_y) ≈ 7.7 m are presented as arrows with origin at the emitter position. As it can be seen, under these conditions, in the 99 % of the cases the estimation error is limited to just 30.55 m on the major axis (worst case) which represents a position error of just $\frac{30.55}{\sqrt{8000^2+1000^2}} \approx 0.37$ % of the distance between the emitter and the BS.

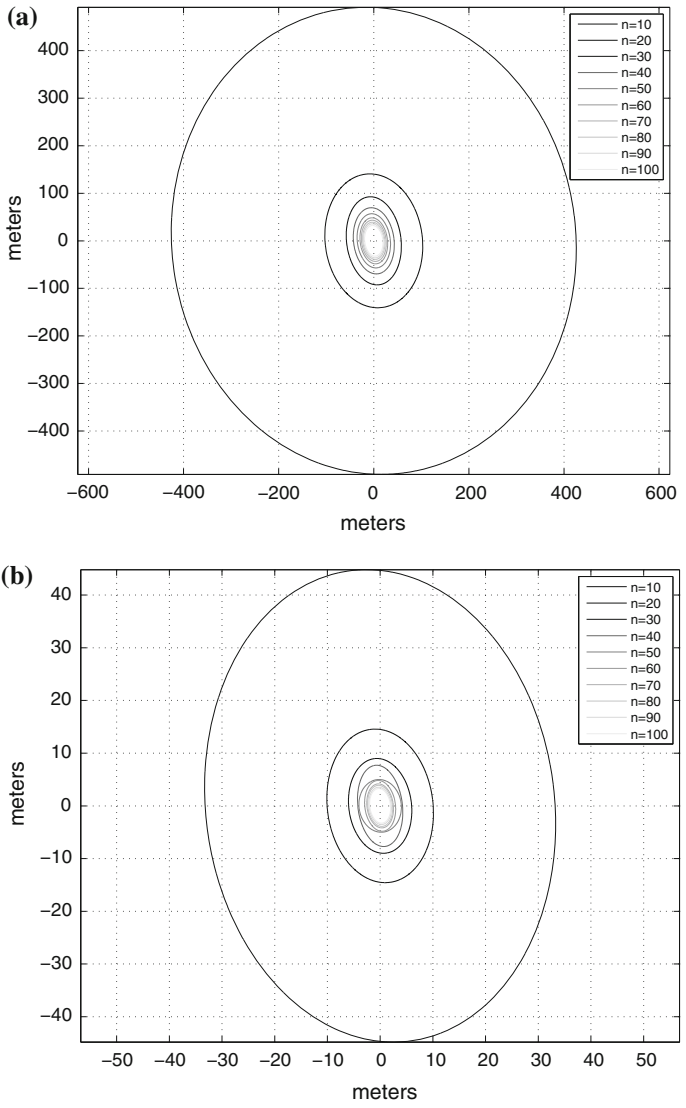

Fig. 13.6 Elliptical confidence interval. Attacker at (8000, 1000 m). **a** SNR = −10 dB. **b** SNR = 10 dB

13.5.3 Evaluation of the Location Error

In this section we show the simulation results with regard location error by means of elliptical confidence intervals, which have been obtained as explained in Sect. 13.5.2.

Figure 13.6a and b show the 99 % elliptical confidence intervals for a reception SNR of −10 and 10 dB respectively and a varying number of anchor nodes. Under

the worst conditions (SNR = −10 dB) the error in the major axis ranges from approximately 30 m, when 100 anchor nodes are cooperating, to around 491 m when there are only 10. Under good reception conditions (SNR = 10 dB) the error ranges just from roughly 3.3 m with 100 anchor nodes to 40 m with 10 nodes. In any case, even in the worst conditions, these errors are acceptable considering a network deployment over an area of 30-by-30 km^2.

Notice that, as the variance of the TDoA measurements is inversely proportional to the (linear) value of the SNR, the estimation error increases as the SNR decreases proportionally to the square of the variance. In our simulations we have considered $SNR_1 = 10$ dB and $SNR_2 = -10$ dB. As a consequence, the linear relation between the respective variances will be $\sigma_2^2 = 100\sigma_1^2$, meaning that, for a given number of cooperating nodes, the error will be approximately ten times bigger when the SNR is −10 dB than when it is 10 dB.

We have seen until now how the SNR and the number of anchor nodes could affect the accuracy of the estimator when the attacker is within the BS range. Now we evaluate an example of attacker placed outside the CRN, specifically at (−75000 m, 50000 m).

Figure 13.7a and b show the 99 % elliptical confidence intervals for reception SNR −10 and 10 dB respectively and a varying number of anchor nodes. The error is now much greater than when the attacker was inside the CRN and, in the worst case (SNR = −10 dB), the 99 % confidence interval ranges in the ellipse's major axis from values around 1.39 km with 10 anchor nodes to 137 m with 100. Despite the error increasing due to the distance to the CRN, once again, the achieved accuracy is enough in order to detect a PUE attack. Notice that the attacker could succeed if it was at 1.39 km far away or nearer from a real TV emitter. In the improbable case of an attacker having such great transmission capabilities and being so close to a legitimate TV transmitter, the attacker would represent an even higher threat to the primary network. Obviously, such an unlikely situation is out of the scope of the CRN prevention mechanisms.

When a primary source is located, its estimated position is checked to be within the chosen elliptical confidence interval α of the known legitimate primary transmitters. As a result, the probability of false negative, that is to say that the attacker succeeds, depends on two factors: (1) a probability of $1 - \alpha$ of error when the localization method succeeds to obtain an estimation; and (2) the probability that the localization method fails or does not converge. During the proposal evaluation, we checked that the probability of non-convergence of our proposed method is always less than 10^{-4} regardless of the position of the attacker and the reception SNR. As a result, the probability of false negative can be approximated as just the former.

Anyhow, after detection of a PUE attack, the CRN should try to counteract the attack, e.g. by reporting the attack to the authorities or ignoring the primary signal and continuing interfering with it. Consequently, a false positive, that is to say, a legitimate emitter that is considered as an attacker, could lead to an unnecessary waste of resources and to interfere with the primary networks,which is strictly forbidden by

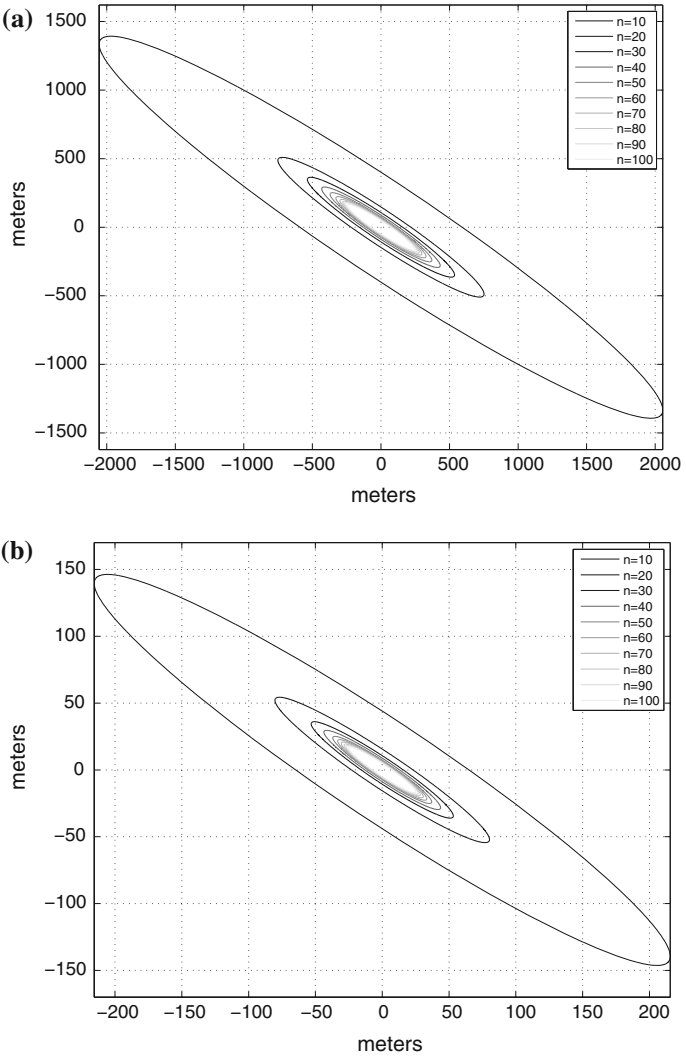

Fig. 13.7 Elliptical confidence interval. Attacker at (−75000, 50000 m). **a** SNR = −10 dB. **b** SNR = 10 dB

CRN regulations [17]. According to this, minimizing the probability of false positive is much more important than minimizing the probability of false negatives.

In the case of our proposal, the probability of false positive can be just expressed as $1 - \alpha$. Consequently, if the assumed error is obtained from the 99 % confidence level, the probability of false positive is 1 %. However, a probability of 1 % is likely to be much higher than the CRN requirements. Anyway, the solution to minimize such probability is rather simple: as mentioned before, the probability of m esti-

mations being out of the confidence interval centered at the primary emitter position (already known) is just $P_f = (1-\alpha)^m$ and, therefore, for a given P_f at least $m_{min} = \lceil \log_{1-\alpha} P_f \rceil$ estimations should be carried out. For example, if we want a probability of false positive of 10^{-6} with a confidence level of 99 %, 3 different estimations (with different TDoA measures) are required every time the localization process is started.

13.5.4 Evaluation of the Location Process Time

The aim of this section is to show that the duration of the localization process conforms to the specifications of the IEEE 802.22 WRAN standard. This standard establishes that the total amount of time required for detecting a primary user in a given channel, i.e., the channel detection time, must be bounded to 2 s [1].

The location-process time depends on (1) the time needed by the CRN anchor nodes to record the primary signal and transmit their recordings to the BS; and (2) the WLS computational time at the BS. The WLS method has a polynomial complexity of $O(n^3)$ that can be solved in nowadays machines in the order of milliseconds or at most hundreds of milliseconds ($\approx$150 ms in a Pentium 4, 32-bits, 2.4 GHz CPU, 1 GB RAM). Consequently, in the following we focus on obtaining the time needed by the anchor nodes in order to record the potential primary transmission and send it back to the BS.

As shown in Fig. 13.8, an IEEE 802.22 WRAN frame is divided into a downstream (DS) subframe and an upstream (US) subframe. The DS subframe is used to send data from the BS to the nodes, and contains the necessary information for nodes to synchronize (frame preamble) and about the bandwidth allocation (US-MAP and DS-MAP). The US subframe is used to transmit data from the nodes to the BS according to the bandwidth allocation established in the DL subframe. The anchor nodes will then use the UL subframe in order to send their recordings to the BS.

In the IEEE 802.22 WRAN standard every frame is made of 26–42 symbols, each one are divided into 60 subchannels, and each subchannel into 28 sub-carriers, 4 of them are pilots and then the remaining 24 are addressed for data transmission. As a result, the number of data bits per subchannel and per OFDMA symbol is $24 \cdot \log_2 M \cdot C$, with M the number of symbols of the modulation and C the coding rate.

Let us denote L_{record}, the amount of bits to be recorded by every anchor node; S as in (13.19) the number of slots that should be assigned to every anchor node in order to transmit the recorded data; and S_{ul} the number of symbols per subchannel assigned to the US subframe, i.e., for data transmission from the CRs to the BS. Then, considering the 58 available subchannels for transmitting data on the US frame, the number of frames F_l needed to transmit the recordings from n anchor nodes can be computed as in (13.20).

$$S = \left\lceil \frac{L_{record}}{24 \cdot \log_2 M \cdot C} \right\rceil \tag{13.19}$$

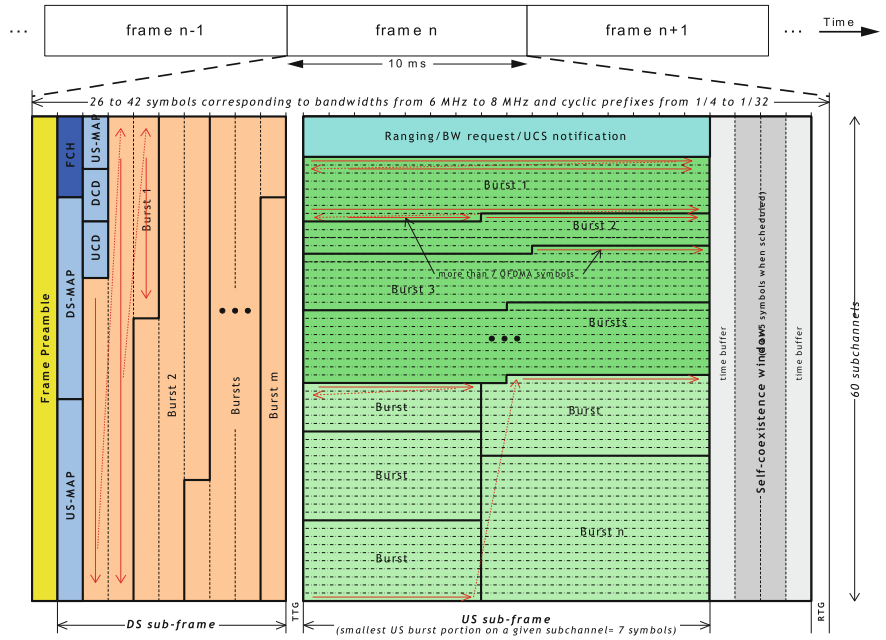

Fig. 13.8 IEEE 802.22 WRAN frame structure

$$F_l = \left\lceil n \cdot \frac{S}{58 \cdot S_{ul}} \right\rceil \tag{13.20}$$

In order to correlate the recording made by a given node and the BS, each recording should be at least as long as the maximum propagation delay within the network. Considering a typical IEEE 802.22 WRAN scenario with a BS range of 30 Km, the maximum propagation delay is approximately 100 μs. Therefore, we have considered that every anchor records 2 complete ATSC DTV frames, that is $2 \cdot 77\,\mu s = 144\,\mu s$. Assuming the worst scenario, when the primary signal comes from a DTV source transmitting at the maximum data rate of 19.4 Mbps, every anchor node should then record $L_{record} = 144\,\mu s \cdot 19.4\,\text{Mbps} = 3348\,\text{bits}$.

Table 13.2 shows the number of frames and the corresponding time needed for transmitting the anchors' recordings to the BS in the previous scenario. We have chosen representative values for the number of cooperating anchor nodes n, modulation and coding rate, and symbols per US subframe S_{ul}. It can be observed that the minimum amount of time needed for transmitting the anchors' recordings to the BS is one frame or 10 ms for 64-QAM $\frac{5}{6}$ modulation and 10 anchor nodes. However, in order to obtain accurate estimations of the emitter's position, a greater number of cooperating anchor nodes is needed. Table 13.2 clearly shows that the time needed to transmit the recordings, despite it grows with the number of anchor nodes, is limited to a few hundred of milliseconds even for 100 anchor nodes.

Table 13.2 Frames/time for localization process

Mod. (CR)	S_{ul}	F_l (time) $n = 10$	$n = 100$
QPSK $\left(\frac{1}{2}\right)$	10	3 (30 ms)	25 (250 ms)
	20	2 (20 ms)	13 (130 ms)
16-QAM $\left(\frac{2}{3}\right)$	10	1 (10 ms)	10 (100 ms)
	20	1 (10 ms)	3 (30 ms)
64-QAM $\left(\frac{5}{6}\right)$	10	1 (10 ms)	5 (50 ms)
	20	1 (10 ms)	3 (30 ms)

From the previous reasoning, we can conclude that the location process time, including both the transmission and computational times, is in the order of hundreds of milliseconds. This is a very promising result that actually meets FCC requirements regarding primary-signal detection time: the detection of a primary transmission and the necessary handoff must be done in less than 2 s [1].

13.6 Conclusions

CRNs rely on spectrum sensing mechanisms to identify white spaces in the spectrum left unused by primary users. A malicious user can take advantage of this feature by emulating a primary transmission, i.e., perform a PUE attack, and prevent the CRN from using available bands.

In this chapter we have described a cooperative method to effectively detect PUE attacks based on TDoA location. The method can be applied to infrastructure-based networks where the location of primary users is known, as it is the case of TV emitters in WRAN 802.22 networks. The rationale behind it is to estimate the location of the emitter and compare it with the known positions of primary users so as to identify potential attacks.

During the location process, CRs are requested to record a fragment of the received signal and send it to the BS, which derives a set of TDoA measurements by correlating the signal received by itself with those coming from the CRs. The non-linear system of equations given by this measurements is then solved by applying Taylor series estimation, which applies a WLS criterion in order to obtain a position estimate while minimizing the sum of squares error. When applying a WLS method, estimation accuracy can be improved with respect to the basic LS mechanism by assigning more weight to those measurements that are expected to be more reliable.

The method poses several challenges that we have addressed as follows. First, TDoA measurement require a tight synchronization between the BS and the CRs. In order to overcome this problem, the BS transmits a marker signal which is recorded by the CRNs together with the emitter's signal, and allows to synchronize every pair of recordings. Second, Taylor series estimation requires an initial guess of the emitter position and the convergence of the algorithm strongly depends on it. In order

to obtain this initial guess, we have proposed to use a database with pre-stored TDoA measures.

Our proposal has been evaluated via simulation under a set of scenarios with different environment conditions and different positions for the attacker. The goodness of the method has been proved based on two different metrics: location accuracy and duration of the location process. The simulation results show that an attacker can be easily identified whenever the PUE attacker is located inside the CRN or at a position relatively far away from true primary users. As the bigger the distance between the emitter and the CRN, the lower the location accuracy, the attack could be performed by selecting a position near a primary user, thus increasing the probability of a successful PUE attack. However, this would require a big amount of resources with regard to power transmission and this probably wouldn't be available to most attackers. Besides, such attack would probably interfere primary transmissions, pushing the responsibility of detecting it and providing the corresponding countermeasures to the primary network. Finally, the amount of time required to perform the location process has been shown to be bounded to reasonable values, meaning that the overhead introduced by the method does not represent a significant decrease of the CRN capacity.

We can conclude that the proposed method can effectively deal with PUE attacks and prevent CRNs from losing spectrum opportunities. However, further work is needed in order to refine the method and improve its performance. On the one hand, position accuracy could be improved by applying hybrid-mechanisms which rely on different types of measures. Given that the inherent spectrum sensing mechanisms of CRNs already require RSS measures, an interesting approach would be to evaluate the performance of location methods by combining RSS and TDoA measures. Besides, a more realistic model where the effect of multipath is taken into account should be evaluated. Multipath is one of the major sources of TDoA measurements errors, but it is also difficult to model because it depends on the particular characteristics of the terrain where the network has been deployed.

References

1. 802.22 WG: IEEE standard for information technology-telecommunications and information exchange between systems wireless regional area networks (wran)-specific requirements part 22: Cognitive wireless ran medium access control (mac) and physical layer (phy) specifications: Policies and procedures for operation in the tv bands. IEEE Std 802.22-2011 (2011). doi:10.1109/IEEESTD.2011.5951707
2. Aad, I., Hubaux, J.P., Knightly, E.W.: Denial of service resilience in ad hoc networks. In: MobiCom '04: Proceedings of the 10th Annual International Conference on Mobile Computing and Networking, pp. 202–215. ACM, New York (2004). URL http://doi.acm.org/10.1145/1023720.1023741
3. Afolabi, O., Kim, K., Ahmad, A.: On secure spectrum sensing in cognitive radio networks using emitters electromagnetic signature. In: Computer Communications and Networks, 2009. ICCCN 2009. Proceedings of 18th Internatonal Conference on, pp. 1–5. IEEE (2009)

4. 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). URL http://dx.doi.org/10.1016/j.comnet.2006.05.001
5. Allman, M., Paxson, V., Stevens, W.: RFC 2581: TCP congestion, control (1999)
6. Buchwald, G., Kuffner, S., Ecklund, L., Brown, M., Callaway, E.: The design and operation of the ieee 802.22. 1 disabling beacon for the protection of tv whitespace incumbents. In: New Frontiers in Dynamic Spectrum Access Networks, 2008. DySPAN 2008. 3rd IEEE Symposium on, pp. 1–6. IEEE (2008)
7. Burbank, J.: Security in cognitive radio networks: the required evolution in approaches to wireless network security. In: 3rd International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CrownCom), pp. 1–7 (2008). doi:10.1109/CROWNCOM.2008.4562536
8. 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 (2004). doi:10.1109/ACSSC.2004.1399240
9. Chen, R., Park, J., Hou, Y., Reed, J.: Toward secure distributed spectrum sensing in cognitive radio networks. IEEE Commun. Mag. **46**(4), 50–55 (2008)
10. Chen, R., Park, J.M.: Ensuring trustworthy spectrum sensing in cognitive radio networks. In: 1st IEEE Workshop on Networking Technologies for Software Defined Radio Networks (SDR), pp. 110–119 (2006). doi:10.1109/SDR.2006.4286333
11. Chen, R., Park, J.M., Reed, J.: Defense against primary user emulation attacks in cognitive radio networks. IEEE J. Sel. Areas Commun. **26**(1), 25–37 (2008). doi:10.1109/JSAC.2008.080104
12. Chen, Z., Cooklev, T., Chen, C., Pomalaza-Raez, C.: Modeling primary user emulation attacks and defenses in cognitive radio networks. In: Performance Computing and Communications Conference (IPCCC), 2009 IEEE 28th International, pp. 208–215 (2009). doi:10.1109/PCCC.2009.5403815
13. Clancy, T., Goergen, N.: Security in cognitive radio networks: threats and mitigation. In: 3rd International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CrownCom), pp. 1–8 (2008). doi:10.1109/CROWNCOM.2008.4562534
14. Cordeiro, C., Challapali, K., Birru, D., Shankar Sai, N.: Ieee 802.22: an introduction to the first wireless standard based on cognitive radios. J. Commun. **1**(1), 38–47 (2006)
15. Douceur, J.R.: The sybil attack. In: IPTPS '01: Revised Papers from the First International Workshop on Peer-to-Peer Systems, pp. 251–260. Springer, London (2002)
16. Eksim, A., Kulac, S., Sazli, M.: Effective cooperative spectrum sensing in ieee 802.22 standard with time diversity. In: International Conference on Advances in Computational Tools for Engineering Applications, ACTEA'09, pp. 528–531 (2009). doi:10.1109/ACTEA.2009.5227849
17. FCC: Federal Communications Commission. ET Docket No. 03–322. Notice of proposed rule making and order (2003)
18. Foy, W.: Position-location solutions by taylor-series estimation. IEEE Trans. Aerosp. Electron. Syst. **AES-12**(2), 187–194 (1976). doi:10.1109/TAES.1976.308294
19. Gezici, S., Tian, Z., Giannakis, G., Kobayashi, H., Molisch, A., Poor, H., Sahinoglu, Z.: Localization via ultra-wideband radios: a look at positioning aspects for future sensor networks. IEEE Signal Process. Mag. **22**(4), 70–84 (2005). doi:10.1109/MSP.2005.1458289
20. Hernández-Serrano, J., León, O., Soriano, M.: Modeling the Lion attack in cognitive radio networks. EURASIP J. Wireless Commun. Netw. **2011**, 633–652 (2010)
21. Jin, Z., Anand, S., Subbalakshmi, K.P.: Mitigating primary user emulation attacks in dynamic spectrum access networks using hypothesis testing. SIGMOBILE Mob. Comput. Commun. Rev. **13**(2), 74–85 (2009). URL http://doi.acm.org/10.1145/1621076.1621084
22. Karlof, C., Wagner, D.: Secure routing in wireless sensor networks: attacks and countermeasures. Ad Hoc Netw. (Sens. Netw. Protoc. Appl.) **1**(2–3), 293–315 (2003). doi:10.1016/S1570-8705(03)00008-8

23. Karlof, C., Wagner, D.: Secure routing in wireless sensor networks: attacks and countermeasures. Ad Hoc Netw. **1**(2), 293–315 (2003)
24. León, O., Hernández Serrano, J., Soriano, M.: A new cross-layer attack to tcp in cognitive radio networks. In: Second International Workshop on Cross Layer Design (IWCLD), pp. 1–5 (2009). doi:10.1109/IWCLD.2009.5156526
25. León, O., Hernández-Serrano, J., Soriano, M.: Securing cognitive radio networks. Int. J. Commun. Syst. **23**(5), 633–652 (2010). doi:10.1002/dac.1102
26. León, O., Hernández-Serrano, J., Soriano, M.: Cooperative detection of primary user emulation attacks in crns. Comput. Netw. **56**(14), 3374–3384 (2012)
27. Li, H., Han, Z.: Dogfight in spectrum: combating primary user emulation attacks in cognitive radio systems, part i: known channel statistics. IEEE Trans. Wireless Commun. **9**(11), 3566–3577 (2010). doi:10.1109/TWC.2010.091510.100629
28. Mathur, C.N., Subbalakshmi, K.P.: Security Issues in Cognitive Radio Networks. In: Mahmoud, Q.H. (ed.) Cognitive Networks: Towards Self-Aware Networks, pp. 284–293. John Wiley, Chichester (2007)
29. MATLAB: version 7.9 (r2009b) (2009). URL http://www.mathworks.com
30. Min, A., Shin, K., Hu, X.: Secure cooperative sensing in ieee 802.22 wrans using shadow fading correlation. IEEE Trans. Mobile Comput. **10**(10), 1434–1447 (2010)
31. Pandey, S., Agrawal, P.: A survey on localization techniques for wireless networks. Chin. Inst. Eng. **29**(7), 1125 (2006)
32. Patwari, N., Ash, J., Kyperountas, S., Hero A.O., I., Moses, R., Correal, N.: Locating the nodes: cooperative localization in wireless sensor networks. IEEE IEEE Signal Process. Mag. **22**(4), 54–69 (2005). doi:10.1109/MSP.2005.1458287
33. Patwari, N., Hero A.O., I., Perkins, M., Correal, N., O'Dea, R.: Relative location estimation in wireless sensor networks. IEEE Trans. Signal Process. **51**(8), 2137–2148 (2003). doi:10.1109/TSP.2003.814469
34. Rong, P., Sichitiu, M.: Angle of arrival localization for wireless sensor networks. In: Sensor and Ad Hoc Communications and Networks, 2006. SECON '06. 2006 3rd Annual IEEE Communications Society on, vol. 1, pp. 374–382 (2006). doi:10.1109/SAHCN.2006.288442
35. Zhang, Y., Xu, G., Geng, X.: Security threats in cognitive radio networks. In: High Performance Computing and Communications, 2008. HPCC '08. 10th IEEE International Conference on, pp. 1036–1041 (2008). doi:10.1109/HPCC.2008.21

Chapter 14
Cognition as a Tool for Green Next Generation Networks

Fabrizio Granelli, Oliver Holland and Nelson L.S. da Fonseca

Abstract The chapter discusses issues related to the implementation of the different steps of the cognitive cycle, especially focusing on reasoning, and applies this to energy saving for green networking. The application of cognition to networking and communications can be readily implemented into current TCP/IP networks. Indeed, the use of the cognitive paradigm represents a way: (i) to address the multiple temporal and spatial fluctuations in the operation of a network, and (ii) to gain and take advantage of additional causal information related to the network configuration and its performance. Network performance is a multi-faceted concept, including simple measures such as throughput as well as far more complicated or subjective measures such as user-level QoS. Recently, an additional parameter has been added to this equation: energy consumption. The need for identifying suitable methodologies to optimize performance from the above viewpoints, also including the contradictory requirement to save energy, is driving research interests towards the emergence of "green networks". Green networking represents an appropriate scenario where cognition and associated radio adaptation can immensely contribute to the given objectives. This chapter describes how cognitive networking can be implemented to support green network operation, proposing a test case demonstrating its potential in a 3G cellular context. Experimental results based on real traffic data demonstrate the capability of a 3G base station to implement cognition to the purpose to save energy without any a-priori information.

F. Granelli (✉)
Dept. of Information Engineering and Computer Science (DISI),
University of Trento, Trento, Italy
e-mail: granelli@disi.unitn.it

O. Holland
King's College London, London, UK
e-mail: oliver.holland@kcl.ac.uk

Nelson L.S. da Fonseca
Institute of Computing, State University of Campinas, São Paulo, Brazil
e-mail: nfonseca@ic.unicamp.br

M.-G. Di Benedetto and F. Bader (eds.), *Cognitive Communication and Cooperative HetNet Coexistence*, Signals and Communication Technology,
DOI: 10.1007/978-3-319-01402-9_14,

14.1 Introduction

As communication networks and the Internet become pervasive and exceed the threshold of a billion connected devices, and, therefore, their management becomes increasingly complex. Moreover, the plethora of applications and services run over the global communications highway are generating novel, diverse and fast-changing requirements. In coping with the complexity in managing future communication networks and to make them flexible to services' requests, the paradigm of cognitive communications emerged.

In its early stages the concept of cognitive communications was often applied primarily to the enablement of better spectrum utilization, giving rise to "spectrum sensing cognitive radio" and similar embodiments. It has been more recently extended to encompass whole networks, leading to the "cognitive networking" paradigm. However, there's still a significant gap between theoretical/simulation studies and the actual implementation of these concepts, such that significant steps forward in introducing cognitive functionalities towards autonomic network management can be realized.

Autonomic network management is indeed extremely relevant in current and future networks, given the short timescales and spatial densities at which management actions need to be performed, and the complex relationships among the operating parameters of networks. Such decision making becomes increasingly difficult when you have conflicting goals in network management such as to provide target QoS guarantees while minimizing energy consumption. Based on such considerations, this chapter describes how the cognitive paradigm can represent a useful tool for the design of green next-generation networks.

The structure of the chapter is as follows. First, Sect. 14.2 surveys the state-of-the-art and research challenges in cognitive networking, while Sect. 14.3 focuses on the key functionality of automated reasoning. Section 14.4 discusses how cognitive networking can be used in the framework of green networking, and Sect. 14.5 provides a test case related to 3G cellular networks. Finally, Sect. 14.6 concludes this chapter with some final remarks and identification of future research challenges related to this subject.

14.2 Research Challenges in Cognitive Networking

14.2.1 Guiding Principles

A pioneering work envisioning cognition-enhanced networks has been undertaken by Clark et al. [1], focusing primarily on the ability of a network to selfrecover. Motivated by the fact that further improvement of existing works was not likely to reach the target goal, they introduced a key component defined 3 Knowledge Plane (KP), basically a distributed cognitive system equipped with reasoning and learning capabilities and

designed as a closed-loop control system. According to the researchers view, the KP spans both vertically through the protocol stack and horizontally across the network nodes, and should be designed taking into consideration some precise goals:

- take advantage of the different observations that can be made in different points of the network, that is, it should exploit observation diversity;
- to implement a unified approach to solve problems, avoiding ad-hoc solutions;
- to include network edges, in order to exploit their knowledge;
- to be able to function in dynamical, continuously changing environments, also in the presence of misleading and/or incomplete information, and under conflicting high-level goals.

It is important to note that these basic principles are shared by virtually all other cognitive networking-related work in the literature. In fact, Clarks KP represents a possible instantiation of the cognitive loop. However, there seems to be no agreement in the literature on which cognitive network model should be taken as a reference, but a proliferation of architectures based on the KP concept, with minor adjustments.

The first formal research about cognitive networks has been performed by Thomas et al. [2], proposing the following features for a cognitive network: (i) extensibility, flexibility, and proactivity; (ii) the capability by the decision process to use network metrics as input and provide actions as output; (iii) the capability to achieve higher performance levels with respect to traditional networks. Besides offering a definition for this new paradigm, the work also delineates a possible framework for the development of cognitive networks, logically divided into three layers, mirroring the defining aspects of a generic cognitive entity (behavioral, computational, neurophysical). Specifically, the guiding end-to-end objectives are specified at the top layer, where they are also redefined in terms of local objectives. Such local objectives are successively handed to the middle (cognitive) layer, which reasons about them and selects the appropriate actions to take. According to the actions selected, the bottom layer sets the tunable elements it controls. It also senses the environment and reports to the cognitive layer, in order to make it learn, thereby closing the cognitive loop. A remarkable characteristic of this cognitive network model is that no strict rules are defined as to how the three layers should be mapped with respect to the actual network nodes. This means that, for instance, the cognitive process may be performed by just one cognitive element, comprising all the devices of a network, or by multiple elements, each one in charge of controlling just a part of the ensemble. In other words, no design limits are implied.

Several architectures are built on the belief that a more holistic approach is needed. Mahonen et al. [3], for instance, acknowledge that information has to be extracted from all layers but assert that so far no attempts towards the joint analysis of the behavior of a communication system and its crosslayer interactions have been made. To the authors, the cognitive networking paradigm is particularly appealing as it represents a valid alternative to basic crosslayer optimization. Emphasis is also given on the collaborative aspect: according to the researchers, a local view of the network is simply not sufficient to perform global optimization. Collaboration is, therefore,

fundamental to perform distributed information gathering with the objective being to build a comprehensive knowledge of the environment.

The thesis of holism is supported also by the work by Sutton et al. [4], underlining that the limited scope of communication protocols prevents a node from performing a global optimization. In this sense, the authors propose an architecture characterized by a network-wide scope, similar to the KP, extending over both the protocol stack and the network itself. An interesting characteristic of such an architecture is that the cognitive process is explicitly split between two distinct entities: a reconfigurable entity is in charge of manipulating sensors and actuators and a separate cognitive entity takes care of reasoning and learning tasks. This separation can rise to concerns regarding how commands and measurements flow between the cognitive entity and the reconfigurable entity, and how much should one expect communication overhead to be. However, no specific details are discussed. According to their vision, the cognitive entity can comprise all the layers of a node or even just a subset of them, whereas the reconfigurable entity involves reconfigurations not only inside a single node but possibly also in each node of the network.

14.2.2 Research Challenges

14.2.2.1 Reasoning and Learning

Reasoning exploits the knowledge that an entity has been able to build by learning, in order to draw new inferences and beliefs. A similar definition states that to reason is to infer conclusions starting from beliefs that are already present in an agents mind. Therefore, a necessary condition is for a cognitive entity is to possess the capability of drawing relationships among such beliefs, and using them to derive some conclusions.

Despite the wide consensus in the literature about the need for an holistic approach, not many works in the literature provide concrete solutions able to support such vision. Even more important, to the best of the authors knowledge, no research has specifically taken into consideration cross-layer interactions in the reasoning process, a key aspect to fully enable a truly holistic approach. Indeed, such interactions concur in defining the overall performance of a network and therefore should be jointly considered when performing the cognitive process. Therefore, to cope with these challenges, an appropriate reasoning technique has to be chosen. However, in the networking architectures proposed so far only a few five reasoning techniques are employed and often their use is not completely justified over the available alternatives.

Thomas et al. [2] suggest thinking of cognition as a machine learning problem. Formalisms like neural networks, genetic algorithms, expert systems, and control theory (e.g., Kalman filters and learning automata) can therefore be employed. However, rather than analyzing the theoretic aspects of cognition, they focus more on providing practical guidelines for bringing intelligence into communication networks by arguing that a cognitive process need not involve exactly one reasoning strategy. On the

contrary, depending on the type of problem being solved and the objectives strived to achieve, more than one technique can be used, possibly creating hybrids.

Along the same line, Mahonen et al. propose to equip cognitive network nodes with an ensemble of tools to perform reasoning tasks [3]. Besides the already mentioned neural and Bayesian networks, such a toolbox may also include pattern recognition and classification techniques, and multidimensional optimization algorithms like simulated annealing or genetic algorithms. The authors though give no indication of which technique should be preferred in a given situation or upon which basis it should be selected. What seems to really matter, however, is the convergence time. Whatever the solution chosen, it is of utmost importance that its convergence time be appropriately less than the typical period of environmental changes.

Another key research topic is learning, which can be seen as a planned and ordered gathering of knowledge, and can happen in several ways. Being strictly connected to reasoning, it has been given little attention as a topic on its own. Besides, learning techniques are usually straightforward and reflect those techniques actually used by living organisms. According to [1], knowledge can be built thanks to external input, the same way children learn at school thanks to their teachers, or by trial and error, as happens with people facing new problems when no one is available to give indications. Other techniques to boost learning often-cited are analogy and generalization.

14.2.2.2 Adaptation

Adaptation, which is the process of turning reasoned decisions into actions, is complementary to reasoning. More specifically, after reasoning the node must adapt itself so that the decisions taken by the cognitive entity. However, being a cognitive network characterized by a global scope, it is possible that more nodes need to adapt themselves in a coordinated manner, in order for the effect to take place. Adaptation therefore implies that nodes have to coordinate themselves to carry out a common goal, or in more general terms, to extend the cognitive process to the whole network.

Network-wide adaptation in cognitive networks, though, seems not to have received much attention in the literature so far. Instead, proposals have focused more on the need for inter-node communication (or lack thereof).

14.2.2.3 Information Representation

The topic of information representation is in part related to both learning and adaptation. Many researchers agree on the need for a proper representation of objectives and knowledge. By pursuing the vision of a knowledge-centric network, knowledge representation becomes fairly important. Information representation plays a key role also in optimization processes: as optimization can be performed with respect to different, possibly contrasting objectives (e.g., performance, reliability), representation is necessary to establish priorities and decide upon which dimension that optimization

should aim to. Additionally, the goals of the system should be clearly expressed so that no misinterpretation can take place. However, even when goals are not correctly specified, the system should be able to act in a reasonable manner.

Information propagation, both within nodes and across the network is linked to information representation. As a general guideline, a desirable feature of the chosen information propagation scheme is that it should be independent from the underlying communication technology.

As for the implementation of information representation, it is noted that it can be analyzed from both the semantic and the syntactic points of view. The semantic vision concerns which information we want network nodes to exchange. Updating all the nodes in a network with all the network knowledge can not be the best option, mainly for two reasons: first, the waste of resources may be too high, and second, an information overload can be burdensome, not to say misleading, for the reasoning process. To this end, it is imperative to understand which nodes need which kind of information.

Syntax is about how messages should be exchanged and how they should be defined. This translates into the definition of standardized protocols that have to be efficient from the point of view of resource consumption, and explicitly or implicitly negotiate and ask for resource reservation.

14.2.3 State of the Art

14.2.3.1 Reasoning and Learning

A popular technique used to infer general conclusions is represented by first-order logic and it is typically used by agents who can rely on certain knowledge. As a consequence, employing first-order logic schemes in cognitive nodes may not appear to be the best choice. This seems to be acknowledged in the literature, since to the best of the authors knowledge, there is no cognitive network architecture based on such a paradigm. Some researchers have introduced expert systems, i.e., systems aiming to store human experts knowledge in a specific field, useful to perform reasoning in cognitive networks provided the problem to be solved is characterized by a limited number of variables. However, the potentially narrow domain of application typical of expert systems clashes with the concept of a cognitive network architecture, which should aim to reason across a variety of diverse domains.

Structural equation modeling, although not a proper artificial intelligence technique, has been mentioned as a potential formalism for reasoning purposes in [5]. However, the same authors admit that such a formalism is more suitable for confirming already-defined causal structures (hypothesis testing), rather than discovering them. This could place a limit on cognitive entities, which can be prevented from adapting to new situations thus not evolving over time. Heuristic optimization algorithms like simulated annealing, genetic algorithms or swarm intelligence, are used to automatically find optimal solutions, and could be employed as alternative

reasoning methods. Genetic algorithms have been used for taking measurement of cognitive radio networks. However, some researchers assert that, although applicable in a variety of contexts, such techniques should be preferred when the environment is well-known and the problem is centralized [2]. For this reason, in distributed scenarios like wireless ad-hoc networks where the nodes may be highly scattered over the environment, such techniques are generally less favorable with respect to others.

Neural networks are often considered as a standard artificial intelligence technique and have been applied to a wide range of applications including cognitive networks. Their main drawback lies in the fact that they are black boxes: once a neural network reaches a solution, its inner structure does not necessarily reflect the motivation behind that outcome. In another sense, the actual relationships that exist among the variables of a system are not reflected by the configuration of the neural network that led to the solution. Therefore, if the purpose is to gain some insights into a network internals, neural networks can hardly represent the optimal solution.

Bayesian networks are another reasoning tool traditionally associated with artificial intelligence, with the capability of representing causal relationships among variables of a given problem and of being applied where knowledge is not certain. As they are based on directed acyclic graphs, their major limitation lies in their impossibility to deal with causality loops. Bayesian networks was suggested by Clark et al. in their pioneering work on cognitive networking [1]. Strictly related to Bayesian networks are Markov random fields. Similarly to Bayesian networks, they are generative models, i.e., they represent a probability model for all the variables of a problem. However, unlike Bayesian networks, they need not represent causal relationships. Other formalisms derived from Markov random fields are Markov logic networks, which combine Markov random fields, first-order logic, and conditional random fields. They are a discriminative type of model, i.e., they model the dependency of unobserved variables on observed variables. Markov random fields (and all models based on them) suffer less from the limitation peculiar to Bayesian networks about loop-free networks. It is also worth noting that the undirected nature of such structures prevents them from handling induced dependencies.

Fuzzy Cognitive Maps (FCMs) are mathematical structures for modeling dynamical systems. They emphasize the causal relationships among the variables of a system and upon those they base reasoning [6]. FCMs are described in details in a separate section of the chapter.

With respect to learning, the techniques used so far in the cognitive networking field are usually straightforward, such as the use of a knowledge base to be updated as new situations are experienced. Specific techniques are employed to update connections in neural networks and FCMs. Popular algorithms commonly leveraged in neural networks are back-propagation learning, in which error with respect to a target outcome is repeatedly back-propagated through the network, and reinforcement learning, in which a reinforcement signal is a measure of the performance level achieved after the system has performed a set of actions. As for FCMs, updating techniques are based on Hebbian learning, according to which connections between concepts that activated together should be given more weight [7].

14.2.3.2 Adaptation

Papers in the literature provide hints at techniques that could be used to accomplish network-wide adaptation, but examples of implementations of this are limited. Clark et al. have proposed a distributed and decentralized model so that different parts of the network can work toward the achievement of divergent goals [1]. However, they acknowledge that classic artificial-intelligence techniques are not suitable for employment in distributed environments and suggest that robust, highly dynamic algorithms should be employed. They also mention multi-agent systems, a tool to model and design systems populated with intelligent agents, each capable of (and responsible for) its own actions, as a framework that could potentially represent a base for building such algorithms.

Friend et al. state their reference for a distributed implementation of the cognitive process, claiming it is more convenient when compared to the centralized counterpart [8]. Their preference stems not only from the typical advantages that such a model offers, but also from the benefits it presents in the specific case of cognitive networking. First, a centralized architecture would lead to a greater communication overhead, and second, in environments populated by nodes that already have cognitive capabilities (e.g., where there are some cognitive radios), the centralization of cognition would inevitably lead to a greater waste of channel resources.

According to this point of view, Thomas proposes several formalisms to study the behavior of a network [9], introducing the use of game theory with the aim of analyzing the interactions occurring among a population of rational decision-makers which have strategic behavior. One potential drawback of game theory is that, usually, assumptions are made in order to make game analysis tractable: common assumptions in this sense are of a homogeneous player set (which means nodes have to share similar characteristic), or players having complete information (that is, nodes exactly know what other node wants and what they can do). Were such assumptions not acceptable, a game would likely become more complicated and, possibly, would not present equilibrium points.

Thomas proposes another tool to investigate interacting elements which is the set of so-called interaction models, in order to identify fixed points of operation of complex systems. Some of the models listed are infinite particle systems and Petri nets, to model general distributed discrete systems.

14.2.3.3 Information Representation

Most works offering practical insights on information representation, resort to the use of markup languages, such as the Extensible Markup Language (XML). For example, an interesting solution comes from Kephart and Chess [10], for whom autonomic elements should register the services they offer in a public registry, such as the Universal Description, Discovery and Integration (UDDI). Such a registry will contain high-level descriptions of the objectives and policies offered by autonomic elements, and how they should be invoked.

Another suggestion is to use the Radio Knowledge Representation Language (RKRL), first promoted by Mitola [11] with the objective of representing radio knowledge through the use of structured natural language. The final scheme for the representation of knowledge worth mentioning is the DARPA Agent Markup Language (DAML), conceived by the Defense Advanced Research Projects Agency (DARPA), which capitalizes on both XML and the Resource Description Framework (RDF) to support ontologies for web objects.

As for information propagation, it is acknowledged that in order to foster cross-network communication a signaling architecture must first be devised. However, to the best of the authors knowledge, no mechanism has been proposed so far specifically for cognitive networks but only general guidelines have been discussed. Regardless of the signaling method, two broad categories of signaling can be distinguished: (i) an in-band signaling scheme, where signaling messages are mixed with data messages, and (ii) an out-of-band signaling scheme, in which messages constitute a separate communication channel. A signaling scheme belonging to the first category may be realized by embedding information in header fields that are normally not used by a protocol. Among the advantages of such solution is that can be mentioned is that it incurs in no overhead, moreover, it is possible to reuse already existing protocols. However, nodes along a path may misclassify such enhanced packets as malicious or malformed, and discard them. Other drawbacks are represented by the limited room to allocate information, and the fact that information passing becomes dependent on the communication technology used (i.e., the protocols). An out-of-band signaling scheme needs a separate communication setup (and possibly the design of a new protocol). The quantity of information that can be sent is greater than in the previous scheme, but complexity and overhead are also greater.

Depending on the architecture chosen for the cognitive process, different message passing schemes can be deployed. In a centralized network, a natural option would be a master-slave communication scheme, but direct communications among slaves may be desirable as well. In a distributed network, flooding could represent a simple solution, even though it could increase communication overhead and may be too general. For instance, two neighboring nodes may harbor the same beliefs about the environment: as a consequence neither of the two will benefit from the others information. More elaborate protocols can represent more viable alternatives: as an example, a seemingly popular choice in distributed scenarios is represented by epidemic protocols, i.e., protocols that imitate the diffusion typical of diseases to spread information. They are generally scalable and resistant to failure.

14.3 Quantitative Reasoning in Cognitive Networks

14.3.1 Introduction

The cognition loop is central to any cognitive architecture. According to this cycle, during the sensing stage a cognitive entity acquires knowledge about the environment and pre-processes it. Afterward, the reasoning, driven by end-to-end principles, takes place during the planning stage while the actual decision is taken in the decision stage. Finally, in the acting stage, actions are executed and their effect evaluated, by sensing the environment again. The learning stage is another key element building knowledge on causality that can feed reasoning.

Cognitive networking needs cross-layering to operate. In this framework, cross-layering represents the means to provide optimization, while a cognitive engine represents the learning, adaptation and decision processes which drive it to achieve end-to-end goals. Several cognitive network architectures are available in the literature, dealing with different stages of the cognition loop.

As correctly summarized in [12], the techniques used in cognitive networks for reasoning purposes are limited in number and often not justified over other available alternatives. In addition, it appears that the cognitive networks proposed in the literature, when addressing the reasoning problem, do not explicitly consider cross-layer relations. Given this, the aim of this section is to introduce the usage of the graph-like structures commonly known as Fuzzy Cognitive Maps (FCMs), to explicitly represent cross-layer and network-wide interactions, and to use such information as a base for the reasoning process.

14.3.2 Fuzzy Cognitive Maps

FCMs are mathematical structures conceived in 1986 by Kosko [6] as a tool for modeling dynamical systems through the causal relationships that characterize them. A pictorial representation of a hypothetical FCM is shown in Fig. 14.1a. Figure 14.1b shows the adjacency matrix of the graph, which is an alternative widely used representation for FCMs.

Graphically, an FCM is rendered as a direct graph, in which a node represents a generic concept (e.g., an event, a process, a variable or other generic entities) and edges between any two concepts mean that there is a causal relation between them, the cause being the node from which the arrow starts.

In the simplest case, the domains of nodes and the weights of edges are discrete. In more complex FCMs, however, nodes can be mapped on larger sets depending on the detail the designer wants to achieve. Indeed, nodes and edges can in principle be fuzzy, in that they may take any value in the continuous sets $[0, 1]$ or $[-1, 1]$. Though the use of larger sets generally results in a greater flexibility of a model, it is often the case that FCMs concepts are mapped on the discrete set 0, 1 and edge

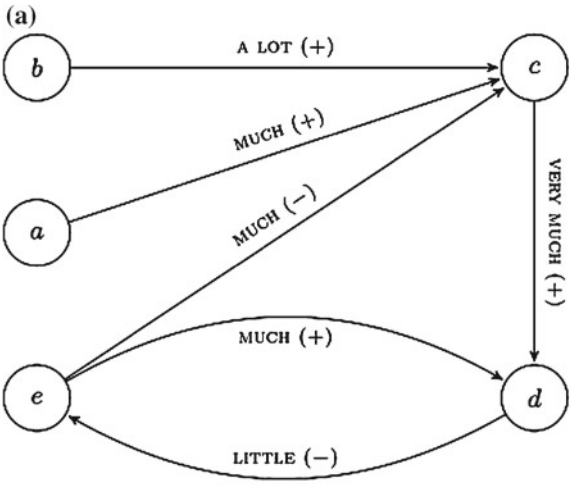

Fig. 14.1 An example of a fuzzy cognitive map. **a** Graphical representation. **b** Mathematical representation (adjacency matrix)

(b)

$$F = \begin{pmatrix} 0 & 0 & 0.6 & 0 & 0 \\ 0 & 0 & 0.9 & 0 & 0 \\ 0 & 0 & 0 & 0.8 & 0 \\ 0 & 0 & 0 & 0 & -0.2 \\ 0 & 0 & -0.6 & 0.6 & 0 \end{pmatrix}$$

labels mapped on $-1, 0, 1$. Such maps are better known as 12 simple FCMs and are particularly suited to obtain a preliminary model of a problem.

A zero-valued concept denotes that the concept is off, inactive, irrelevant, in a low-state or it can even mean that it is not considered at all. Conversely, a concept set to one means it is regarded as high, relevant or active. Edge labels measure the degree of causality; values of either $+1$ or -1 denote a strong causal relationship, positive in the former case and negative in the latter. A zero-valued label means that the two concepts are not causally related to each other. No concept can cause itself, hence edges leaving and entering the same node cannot exist. Equivalently, the trace of the adjacency matrix of any FCM yields zero.

The state of a system having n distinct concepts is a vector of dimensions $1-by-n$. In the inference process, this vector is repeatedly multiplied by the FCM matrix and the result thresholded each time, until it converges either to a fixed point or to a limit cycle. This complete procedure is shown using a toy example in Fig. 14.2.

FCMs offer some advantages over other potential reasoning techniques (neural networks, Bayesian and Markov networks). Unlike both Bayesian and Markov networks, the inference procedure used in FCMs, since it involves exclusively vector-by-matrix multiplications and thresholding operations, has a low computational footprint.

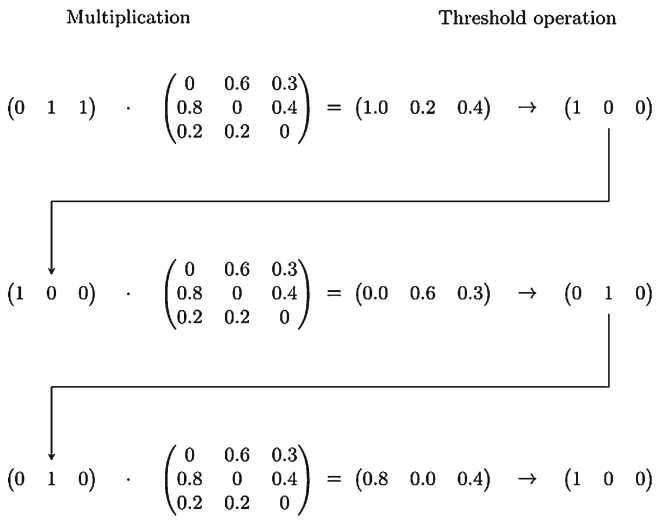

Fig. 14.2 The inference process for a FCM

Moreover, FCMs are more powerful when there are causality loops in the problem; for such problems, Bayesian networks cannot be applied. In Markov networks, loopy belief propagation algorithms can be applied, but convergence is not guaranteed. Additionally, the FCM inference process can be applied to both loop-free and non-loop-free problems. Moreover, inference is guaranteed to converge to either a fixed point or a limit cycle, provided that concepts take their values in any finite discrete set.

Another considerable advantage is the possibility of exchanging and merging together multiple FCMs, resembling the operations of people exchanging opinions. This aspect has its roots in the primary purpose for which FCMs were created, i.e., to allow experts to represent their causal knowledge about some situation. Different experts may have different opinions about the same matter, and may encode differently their beliefs, hence drawing conflicting FCMs. Merging helps to smooth (possibly divergent) beliefs and biases, thereby reducing the possibility of biased reasoning.

However, there are some disadvantages. One major drawback concerns the automatic synthesis of FCMs: FCMs were not conceived for being constructed starting from observational data. They were initially devised in the social science field as a tool to be used by experts in order for them to formally express their beliefs about a given matter. For this reason, auto-synthesis of FCMs is difficult, mostly because, for non-humans, causal relationships between variables are generally more complex to detect than simple correlations.

Another limitation of FCMs is their restricted ability to deal with the complementary process of inference, i.e., abductive reasoning. Abduction is the process of stating which causes are responsible for a given effect: in case of FCMs, as well as in Bayesian and Markov Networks, it requires the solution of an NP-hard problem.

Despite such problems, the benefits deriving from the use of FCMs in cognitive networking appear to outweigh their drawbacks. The interested reader can find a thorough survey of FCMs in [13].

14.3.3 Applying FCMs to Cognitive Networking

The proposed approach to implement FCMs in cognitive nodes can be described as a three-step process:

1. To identify which variables of the different communication protocols could be of interest for the intended application. In particular, concepts should be categorized according to what they represent: everything that can be tuned should belong to the action class, everything that cannot be controlled directly and conveys a QoS-related meaning should belong to the QoS class, and everything else (i.e. variables that cannot be controlled and are not QoS-related) should belong to the environment class.
2. To define the domain of each variable, avoiding continuous sets and keeping discrete sets as small as possible (or else being prepared to increased computational complexity). In finding the right domain for boundless concepts, it could be of help to think of a threshold, so that greater (or lower) values entail the same causality as the threshold value. An error probability could be mapped to the discrete set 0, 1, indicating absence or presence of errors; however, a proper threshold has to be chosen (for instance, according to the transmission modulation used), so to distinguish values of the error rate that affect the system behavior from values that do not affect it. Finally, the same domain could be apt for modeling the throughput: the lowest value can be used to represent non-satisfactory situations, the highest to represent favorable situations.
3. To design and implement the algorithm for building and updating the FCM, embedding into the matrix any available a-priori knowledge. The choice and configuration of the algorithm is critical for achieving satisfactory results. For instance, it is important to introduce zeros in the FCM matrix representation corresponding to absence of relationship between two concepts.

It should be noted that the first two points resemble a pre-processing stage, rather than the reasoning stage itself. However, the two stages cannot be separated from one another and have to be accomplished in a jointly fashion.

14.4 Application of Cognition to Green Networking

14.4.1 Introduction

Global emissions of greenhouse gases (GHGs) represent a footprint on the world of the development of humanity, especially after industrialization. The ICT sector itself roughly accounts for 2 % of todays global carbon footprint, but figures are expected to significantly increase in forthcoming years, with forecasts predicting levels in 2020 around three times what they were in 2002. However, it is also forecast that it will contribute, both directly and indirectly, to reducing global emissions of about five times its own footprint, potentially leading to approximately 600 billion savings for energy costs.

The most significant direct effect is that the telecom infrastructure is expected to grow significantly, ultimately being responsible for 13 % of the total sector footprint. Considering also that power generation in ICT is acknowledged to be one of the main causes behind the increase of man-made greenhouse gases, it is evident the importance of energy optimization in the telecom infrastructure.

14.4.2 Motivation

In current cellular networks, base stations are usually kept powered on and operating all day long, pursuing the vision of an always-on network. As power consumption in such networks is mainly due to base stations, which account for almost 80 % of the total, it is no wonder that several research efforts have tried to intervene directly on the functioning of transceivers, at different levels of details. Most of the works in the literature address the problem of energy consumption from a static point of view, by applying static optimization algorithms. While this approach works fine in quasi-static or highly predictable scenarios, the same may not hold in cases where the environment is subject to dynamic changes or in cases of emerging and differentiated usage of the communication infrastructure. For example, even if it is known that the average number of users in a cell follows a predictable pattern, the actual number may be markedly different from the mean, thereby reducing the effectiveness of energy saving mechanisms implementing a static approach. In line with such considerations, this chapter aims to explore the possibility of developing dynamic energy-reduction schemes.

In this scenario, the cognitive networking paradigm represents a potential approach to pursuing the objective of energy efficiency. Cognitive networking is a relatively recent research field, stemming from cognitive radio technology. It aims to extend the principles underlying cognitive radios and apply them to the whole communication protocol stack according to a network-wide perspective.

Works in this area are traditionally targeted at reducing management complexity or optimizing Quality-of-Service-related metrics. In this chapter, we propose to adapt

cognitive networking principles to address the problem of energy saving in cellular networks. More precisely, we employ the Fuzzy Cognitive Maps (FCMs) reasoning tool introduced earlier to analyze causal relationships between energy consumption in base stations and other variables characterizing the cell, in order to identify the most appropriate run-time decisions to reduce energy consumption while maintaining a suitable performance level. It should be noted that a dynamic approach, in principle, substitute static schemes, but can also be used in conjunction to them.

14.4.3 Related Works

The literature on green wireless networks is huge, and it is not possible to provide a comprehensive survey in this section. For reasons of ease of presentation we focus our attention on network deployment strategies, and provide a quick review of the most promising technologies to build energy-efficient wireless cellular networks.

14.4.3.1 Cell Deployment Strategies

Energy saving through cell deployment strategies can be achieved by implementing denser networks, consisting of smaller cells and higher number of base stations. With this approach it is possible to counteract the high energy consumption caused by high path losses during transmission. On the other hand, a higher number of base stations in the system implies on greater power consumption, especially during low traffic mode where most active stations are in the idle state, thus consuming power during the listening process. As a consequence, the energy consumption for the overall network operation is a trade-off and the question is to identify the optimal inter-site-distance (ISD) among cells [14]. For each deployment strategy there is an optimal inter site distance among cells that can be deployed in order to achieve minimum power consumption. For example, optimal inter site distance for urban and dense urban environments (when considering homogenous macro-stations deployment) is typically between 500 and 1500 m. Another important issue is to design scalable networks, with over-provisioning, since it is not possible to change deployment after rollout. Over-provisioning provides possibilities for densification of network with additional macro base stations and/or micro cells [15], so that the inter-site distance in case of increasing traffic demand can be reduced. It is worth nothing that over-provisioning influences the total power consumption, while denser networks reduces power consumption during transmission states [16]. Furthermore, additional power saving could be achieved by adapting deployment to switch off the cells or to modify some configuration parameters [17, 18].

It is also necessary to analyze the trade-off between deployment costs and the gain of energy efficiency. Cell deployment has to take into account given requirements of coverage and capacity. More precisely, a proper deployment has to fulfill specified conditions in terms of coverage area and available capacity. If some energy-efficient

deployments do not provide the required coverage level by operator, they should not be considered. Likewise, the deployment should also be discarded in case the network is not able to satisfy the capacity demand in a covered area.

Besides macro cells which are designed to cover larger areas, micro cells are designed for relatively short distances, utilizing the good channel conditions for transmission and using much less transmit power, reducing, power consumption. The energy consumption of deployed networks is mostly connected to the high density of base stations. Higher densification is employed to increase the capability of data transmission and the level of coverage. Thus, the optimization of a cell size is a very important factor, which highly influences energy saving. The optimal cell size when considering energy efficiency in wireless architectures is investigated in [19]. The cell size can be optimized dynamically, according to temporal traffic conditions. For very high local traffic demand an additional capacity can be provided by small underlay cells or by the densification of macro cells that can be switched off during low traffic times. However, there is no guarantee that cells in sleeping mode will not consume energy at all, but the amount of power is very small.

14.4.3.2 Energy Efficiency Through Relay Stations

The introduction of relay nodes (RNs) in the infrastructure network architecture is a novel approach that aims both to extend coverage and increase capacity [20], hence allowing more flexible and cost effective tradeoffs. RN covers a much smaller area than the macro cells described in the previous section. Therefore, they consume less transmission power. Considering this fact, it is expected that relay nodes built for small transmission ranges will consume less power.

Different types of relay nodes are introduced in 3GPP. In-band relays share the same frequency band for the base station relay node (BS-RN) and relay node user equipment (RN-UE) links. Two cases are distinguished in 3GPP:

- Type 1 relays in 3GPP notation are RN operating on Layer 3, i.e. the protocol layers up to Layer 3 for user data packets is available at the RN. Such a L3 relay has all functions that an BS has, and can receive and forward IP packets (PDCP SDUs);
- Type 2 relays may be either Layer 3 or Layer 2 relay nodes, depending on the particular solution/implementation.

Coherently with 3GPP terminology, Type 1 relays are considered visible to the mobile devices, while Type 2 relays are transparent relay nodes. The performance evaluation of these two schemes showed that the deployment of Multicast Cooperative schemes improve more the capacity, while the deployment of two-hop schemes provides higher energy efficiency.

14.4.3.3 Co-Deployment of Different Radio Access Technologies

The co-deployment of different radio access technologies (RAT) aims to provide the so-called Multi-RAT systems, which can achieve higher performance, especially in terms of energy efficiency. Next-generation wireless networks (e.g. 4G) will integrate heterogeneous technologies with the purpose of achieving enhanced performance. This implies necessary support for co-deployment with other wireless networks based on different RATs. To provide co-existence of such mixed networks, and to improve spectrum and bandwidth utilization efficiency, many different co-deployment mechanisms were proposed. In the last few years a significant number of studies have addressed various approaches for improving bandwidth efficiency. A well-known technique in this group is the MIMO technology, mainly used to increase system capacity, which leads to enhancements of bandwidth utilization. Spectrum utilization efficiency can increase through a distributed spatial division multiple access (SDMA) technique, so that mutual interference among coexisting peer systems in areas of overlap can be prevented.

The 3GPP long term evolution (LTE) technology [20] provides increased capacity and transmission rate. Interesting scenarios of network co-deployment are proposed, where conventional RATs are combined to cooperate together with LTE to provide mobile services. When high capacity is important, standalone LTE have shown better results. Deployment in lower frequency bands could also represent a suitable option while when coverage capacity is important LTE has to be deployed in higher frequency bands, combined LTE and legacy RAT solutions have shown up to 50 % of gain, with respect to standalone LTE [21].

14.4.3.4 Integration of Different Radio Access Technologies

The combined operation of LTE and Wi-Fi represents a promising solution, aiming to support rapidly increasing demand for wireless services and bandwidth. Wireless operators are realizing that additional 4G bandwidth will not cope with the increasing requirements of the emerging media-rich devices and social media applications. Hence, Wi-Fi represents a potential offload technology to convey traffic from the overloaded 3G and 4G networks. The synthesis of Wi-Fi and LTE will represent an essential combination to deliver the necessary bandwidth required by the industry for emerging video-heavy, media-savvy applications. However, co-deployment of LTE and various RATs should be very carefully performed, and trade-offs should be carefully estimated and optimized, in order to preserve power consumption on reasonable level, while providing highquality services in terms of capacity and coverage [22].

14.5 Test Case: Green Self-Configuration of 3G Base Stations

14.5.1 Embedding FCMs into Radio Network Controllers

To the purpose of presenting a test case on the usage of cognitive network paradigm in green networking, we propose to equip 3G base stations with such cognitive capabilities, in order to learn how to save energy while adapting to the changing environment.

According to the procedure described in Sect. 14.3.3, we will define in the following sections a proper FCM for the problem of energy saving in cellular networks.

14.5.1.1 Identification of the Concepts Characterizing the Problem

The first step towards the definition of a FCM involves the identification of the concepts that will compose the system state. We can define a set C of all the concepts that characterize the system under study. We can think of such concepts as belonging to different sets:

- Set A comprises all concepts on which the reasoning entity has direct control;
- Set Q collects all concepts that the reasoning entity cannot control directly but that are interesting because they give feedback on the achieved performance;
- Set E collects all concepts on which the reasoning entity has no direct control nor carry relevant information regarding the performance.

Once concepts are found and classified, it is possible to create the system state vector, $s = (a, q, e)$, where: $v = (vi, ..., v_n V) vi \in V \forall (v, V) \in (a, A), (q, Q), (e, E)$. The FCM needs to converge to a solution state $s* = (a*, q, e)$ by finding a vector a* such that the constraints expressed by q are satisfied before environmental conditions e change. It should be noted that elements in set E are important, although they cannot be directly controlled by the system nor do they provide any information directly related to the performance of the system. In fact, they might have relationships with concepts in the other sets, which must be taken into account to be, eventually, exploited. For the considered test case, the following concepts are selected:

- Concepts in $S \cap A$: the use of higher frequencies (hi), the use of trisectorized operational mode (tri);
- Concepts in $S \cap Q$: the energy consumption (en), the blocking rate (br), the Signal to Interference-plus-Noise Ratio (snr);
- Concepts in $S \cap E$: the amount of voice users (v), the amount of users browsing the web (h), the amount of users that transfer data (f).

14.5.1.2 Definition of Concept Domains

The second step in defining the FCM involves the identification of concept domains and the pre-processing operations needed to perform the mapping operation. The number of steps needed to make the reasoning process converge depends on the domains on which the concepts are mapped and the number of concepts themselves. More precisely, the inference process reaches a solution within lc steps, l being the number of levels of concept domains and c being the number of concepts. For this reason, we considered appropriate to adopt binary sets as the domains for all concepts. In particular, we have opted for the $\{-1, 1\}$ domain as the domain for all concepts except the blocking rate, for which we used the binary set $\{0, 1\}$. This choice has been made to highlight the fact that a low blocking rate (0) should not be able to influence or, better, cause any other concept. Avoiding the use of zero in all other cases means that any variation can, at least in principle, entail some change in the other concepts. It should be noted that only concepts in $S \cap A$ can be naturally mapped on such interval.

Pre-processing operations are needed in order to map all other concepts. The identification of optimal values for the pre-processing stage depends on the problem under analysis and is a step that must be taken into account for the correct outcome of the operations. More information on the subject is provided by [23].

14.5.1.3 Definition and Update of the Fuzzy Cognitive Map

The third and last step is about embedding of any a priori knowledge of the problem to the FCM. Let us denote by f_{ij} the edge of the FCM that departs from i and arrives to j, i, and j being generic concepts in S.

1. We assume that concepts in the same set are causally independent from one another. Considering the set $S \cup Q$ as an example, this means that, for instance, the variation of the number of users that browse the web has no causal implication to (and from) the variation of the number of the users that place voice calls. This means that no edges arrive or depart from concepts that belong to the same class: $f_{i,j} = f_{j,i} = 0 \forall i, j \in S \cap V, V \in A, Q, E$
2. Concepts in the action set are not directly caused by any other concept. Instead, they are triggered by the reasoning process. This translates into the fact that no edges point to any action concept, that is: $f_{i,a} = 0 \forall a \in S \cap A, \forall i \in S \cap V, V \in Q, E$
3. Similarly, concepts related to quality metrics do not cause any variation in the concepts related to the environment. As an example, users will decide to call, browse the web and download files ignoring channel conditions (Signal to Interference and Noise Ratio (SINR), blocking rate and energy consumed by the base station). Mathematically: $f_{qe} = 0 \forall q \in \{S \cap Q\}, \forall e \in \{S \cap E\}$ (7)

4. We also know that both actions increase the number of frequency slots available, and, as a direct consequence, reduce the blocking rate. Therefore we may want to embed such information, by properly setting $f_{hi,br}$ and $f_{tri,br}$.

The resulting FCM is as follows:

$$F = \begin{pmatrix} 0 & 0 & 0 & 0\,0 & f_{v,en} & f_{v,br} & f_{v,snr} \\ 0 & 0 & 0 & 0\,0 & f_{h,en} & f_{h,br} & f_{h,snr} \\ 0 & 0 & 0 & 0\,0 & f_{f,en} & f_{f,br} & f_{f,snr} \\ 0 & 0 & 0 & 0\,0 & f_{hi,en} & f_{hi,br} & f_{hi,snr} \\ f_{hi,v} & f_{hi,h} & f_{hi,f} & 0\,0 & f_{tri,en} & f_{tri,br} & f_{tri,snr} \\ f_{tri,v} & f_{tri,h} & f_{tri,f} & 0\,0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0\,0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0\,0 & 0 & 0 & 0 \end{pmatrix}.$$

In order to maintain the FCM updated, a popular learning rule known as Differential Hebbian Learning (DHL) is used. It updates the edges in an FCM proportionally to the value of variations of the concepts. Mathematically, if we denote by $f_{i,j}^t$ a generic edge at time t and by $\dot{C}$ the variation of concept i at time t, the DHL rule states that:

$$f_{i,j}^t = f_{i,j}^{t-1} + \eta(-f_{i,j}^{t-1} + \dot{C}_i^t \dot{C}_j^t)$$

The parameter $\eta \in (0; 1]$ is known as learning rate and its purpose is to lower the responsiveness of the algorithm, which otherwise could produce too abrupt updates.

In order to consider all variations of the same importance, we modified the DHL rule to devise a Linear Learning (LL) rule. According to such rule, edges are updated based only on the polarity of the variation, that is:

$$f_{i,j}^t = f_{i,j}^{t-1} + \eta sgn(\dot{C}_i^t \dot{C}_j^t)$$

where *sgn* denotes the sign operator. Clipping is done to prevent edge values fall off of the $[-1; 1]$ interval.

14.5.2 Simulation Scenario

The simulation scenario is populated by seven base stations arranged according to the traditional honeycomb structure, and we focus our attention on the central base station, which is served by a base station that is equipped with cognitive capabilities. Such cognitive capabilities are based on the FCM described in the previous Section, and allow the base station to reason about the environment in order to reduce energy consumption while monitoring the blocking rate. Other base stations in the network

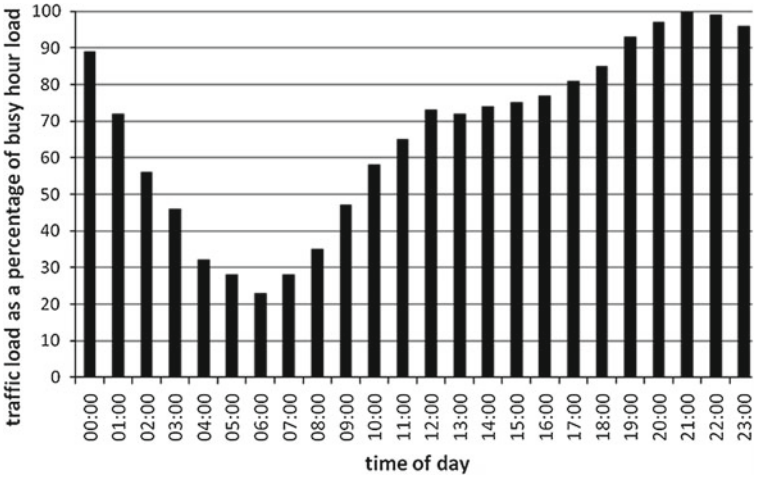

Fig. 14.3 Hourly variation of traffic load as a percentage of busy hour load over a typical day for a mobile network operator in London, UK

do not employ any cognitive scheme and maintain all radio modules enabled at all times.

The simulating platform focuses on the periods of active communications between the base stations and the user terminals associated to it, so that it is possible to monitor energy consumption. Terminals are static during their communications and are distributed over the coverage area of a base station following a uniform random distribution. Associations and de-associations to/from a base station follow a Poisson process, with parameters and, respectively. Both λ and μ depend on the type of communication occurring: voice call, web browsing, data transfer, indicared as λ_v, μ_v, λ_h, μ_h,λ_f, μ_f, respectively.

Each base station has a peak busy load of 50 users, weighted by real data traces in order to reflect real-world situations. Weights reflect the actual hourly load measured in a Vodafone 3G cell in London and were obtained via internal communication within the UKs Mobile VCE Core 5 Green Radio research program (see Fig. 14.3).

Voice traffic is modeled after the well-known Brady six-state model. We assumed an average duration of calls of one minute, resulting in a μv of 1/60. Web traffic has been modeled as a continuous repetition of two states: a downloading period to retrieve a page from the web, and a waiting period, to parse and read the page. The download time depends on the size of the web page and eventual embedded objects. Object sizes follow a truncated lognormal distribution, while the number of the embedded objects in a web page follows a truncated Pareto distribution. Reading and parsing times can be found by sampling an exponential distribution. Assuming that a web session for a mobile user lasts, on average, five minutes, we fixed μ_h to a value of 1/300.

Table 14.1 Simulation parameters. d is the distance in km

Classes	Action mechanism
System configuration	Broadly reflecting HSDPA Rel. 5
Spectral efficiency	0.8 b/s Hz
Bandwidth per HSDPA band	5 Mhz
Channel path loss models	2 GHz: $128.1 + 37.6 * log(d)$
	5 GHz: $141.52 + 28 * log(d)$
HSDPA pilot power	20 % of cell power budget

The model for FTP sessions is similar to that for web sessions, except for the fact that there is no parsing time. The download time exclusively depends on the size of the object to be transferred, which follows a Pareto distribution. Reading time is again modeled by an exponential random variable. Considering that a mobile user is not likely to make extensive use of FTP, we hypothesize an average session duration of two minutes, corresponding to α_f equal to 1/120. The composition of traffic is fixed as follows: 50 % voice traffic (α_v), 40 % web browsing (α_h) and 10 % FTP traffic (α_f). Traffic categories are independent from one another.

The environment represents a typical HSDPA network. The main parameters that characterize the system are reported in Table 14.1. Base stations all operate on two bands, centered at 2 and 5 GHz, and are characterized by a coverage radius of 600 m. According to internal documentation within the Mobile VCE Green Radio research program, it has been shown that an HSDPA base stations consumes 857 W at 100 % transmission power and 561 W at 20 % transmission power.

Regarding user capacity, a base station can accommodate at most 22 users per band when operating in omnidirectional mode and up to 15 users per band per sector when operating in tri-sectorized mode.

14.5.3 Results and Discussion

Simulation time covers three days, starting from midnight. Hourly variation is considered the same from day to day. The dashed line in Fig. 14.4 represents the energy consumption by a traditional base station, i.e. when all six modules are always on. As can be expected, the curve is a linear function of the time, showing no change in the behavior of the base station. The behavior of the cognitive base station is represented by the solid line, which resembles a piecewise linear function. The curve reveals that at times when the user load is low, it is possible to save energy by switching off part of the radio modules. In such quiet periods, it is possible to employ power saving modes, with a saving around 50 % of the total energy consumption of the traditional system (6.25 MJ/h or 1.7 kW). Conversely, when there is a high user load all radio modules must be kept active and no energy saving is possible.

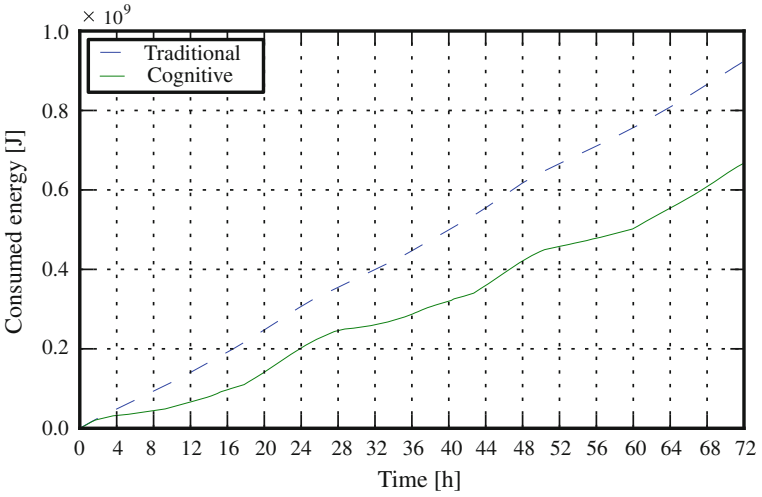

Fig. 14.4 Energy consumption as a function of time for the legacy and cognitive approaches

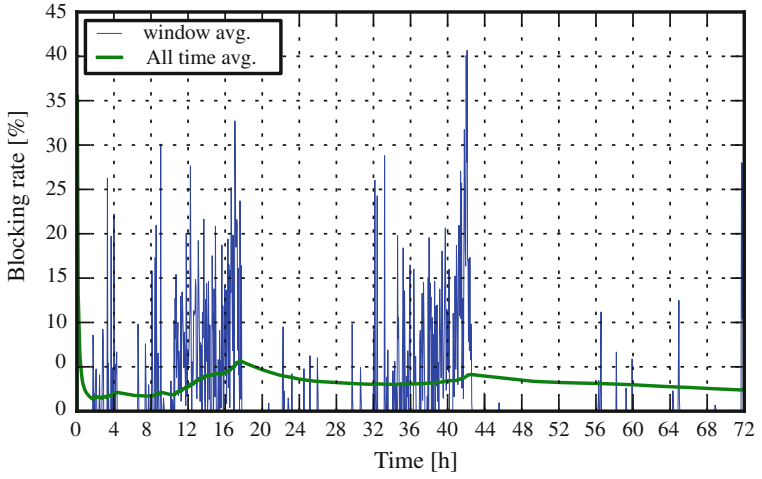

Fig. 14.5 Evolution of the blocking rate for the cognitive approach

Clearly, there is a trade-off between energy saving and blocking rate. As the thin line in Fig. 14.5 indicates, the blocking rate tends to increase when a subset of the transceivers is turned off. The thick curve represents the all-time average, in which the blocking rate remains below 5 % throughout the simulation. As it can be noticed, the cognitive base station learns over time and improves blocking day by day.

The causal relationships between the action concepts and the qualityrelated concepts are reported in Fig. 14.6. It can be noticed that the cause-effect relationship

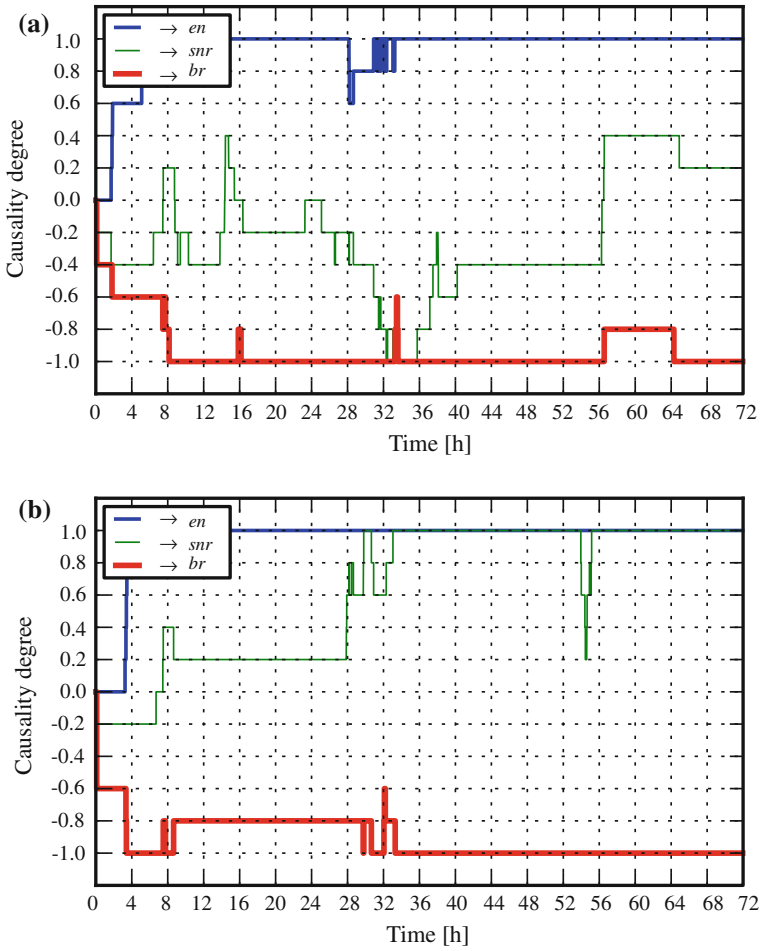

Fig. 14.6 Example of evolution of the causal relationships: (**a**) between *hi* (use of higher frequencies) and the quality-related concepts, and (**b**) between *tri* (use of tri-sectorized mode) and the quality-related concepts

between any action concept and the blocking rate is negative and approaches the lower bound (1). This means that using the higher band in conjunction with the lower band and using the tri-sectorized mode causes the blocking rate to decrease. A similar effect happens with energy: turning on the radio modules causes the consumption of energy to rise (positive cause-effect relation).

14.6 Conclusions and Final Remarks

This chapter has illustrated the application of cognitive networking paradigms to reducing energy consumption in networks, focusing on the usage of FCMs for reasoning. The rationale on the usage of cognition in such a field is related to the complex relationships among the parameters that control power consumption in todays networks, a complexity that is evident even in the relatively simple scenario presented as a proof-of-concept. Yet, such experimental results (based on real traffic data) demonstrate the capability of a 3G base station to implement cognition and to achieve energy saving without any a-priori information but only fucusing on understanding the cause-effect relationships among the parameters and environment variables.

In summary, cognitive networking has been shown in this chapter to enable us to face the problem of designing and managing green networks. Nevertheless, several related research issues are yet to be fully developed, including network-wide cooperation, signaling schemes, as well as the learning and reasoning paradigms for such networks. In addition, proper balance between learning from own actions and a-priori information is needed to improve the efficiency in reasoning and the overall performance of the system.

Acknowledgments The authors wish to express their gratitude to Dr. Christian Facchini for his contribution in studying the problem of reasoning in cognitive networks. Some findings and examples presented in this chapter were studied in his PhD thesis [24].

The authors would also like to acknowledge the interactions and groundwork of the Mobile VCE Core 5 "Green Radio" research programme and the ICT-ACROPOLIS Network of Excellence, www.ict-acropolis.eu, which have in some aspects contributed to the realisation of this work as well as CNPq Science without Border project 402480/2012-0

References

1. Clark, D.D., Partridge, C., Ramming, J.C., Wroclawski, J.T.: A knowledge plane for the internet. In: Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications, (Karlsruhe, Germany), pp. 3–10, ACM (2003)
2. Thomas, R.W., DaSilva, L.A., MacKenzie, A.B.: Cognitive networks. In: 1st IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DyS-PAN), pp. 352–360 (2005)
3. Mahonen, P., Petrova, M., Riihijarvi, J., Wellens, M.: Cognitive wireless networks: Your network just became a teenager. In: 25th Conference on Computer Communications. Barcelona (2006)
4. Sutton, P., Doyle, L.E., Nolan, K.E.: A reconfigurable platform for cognitive networks. In: 1st International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM), pp. 1–5 (2006)
5. Fletcher, D., Nguyen, D., Cios, K.: Autonomous synthesis of fuzzy cognitive maps from observational data: preliminaries. In: IEEE Aerospace Conference, pp. 1–9 (2005)
6. Kosko, B.: Fuzzy cognitive maps. Int. J. Man-Mach. Stud. **24**(1), 65–75 (1986)
7. Kosko, B.: Neural networks and fuzzy systems: a dynamical systems approach to machine intelligence. Prentice-Hall, New Jersey (1992)

8. Friend, D.H., Thomas, R.W., MacKenzie, A.B., DaSilva, L.A.: Distributed learning and reasoning in cognitive networks: methods and design decisions, ch. 9, pp. 223–246. Wiley-Interscience, New York (2007)
9. Thomas, R.W.: Cognitive networks. Ph.D. dissertation, Virginia Tech, Virginia (2007)
10. Kephart, J.O., Chess, D.M.: The vision of autonomic computing. IEEE Comput. **36**(1), 41–50 (2003)
11. Mitola, J.I.: Cognitive radio for flexible mobile multimedia communications. IEEE International Workshop on Mobile Multimedia Communications (MoMuC-99), San Diego, pp. 3–10 (1999)
12. Fortuna, C., Mohorcic, M.: Trends in the development of communication networks: cognitive networks. Comput. Netw. **53**(9), 1354–1376 (2009)
13. Aguilar, J.: A survey about fuzzy cognitive maps papers. Int. J. Comput. Cogn. **3**, 27–33 (2005)
14. Hu, L., Kovacs, I.Z., Mogensen, P., Klein, O., Stormer, W.: Optimal new site deployment algorithm for heterogeneous cellular networks. In: IEEE Vehicular Technology Conference (VTC), Fall 2011
15. Fehske, A.J., Ritcher, F., Fettweis, G.P.: Energy efficiency improvement through micro sites in cellular mobile radio networks. In: IEEE Global Communications Conference (GLOBECOM) Workshops, (2009)
16. Ritcher, F., Fettweis, G.: Cellular mobile network densification utilizing micro base stations. In: IEEE International Conference on Communications (ICC) (2010)
17. Bousia, A., Antonopoulos, A., Alonso, L., Verikoukis, C.: Green distance-aware base station sleeping algorithm in LTE-Advanced. In: IEEE International Conference on Communications (ICC) (2012)
18. Bousia, A., Kartsakli, E., Antonopoulos, A., Alonso, L., Verikoukis, C.: Game theoretic approach for switching off base stations in multi-operator environments. In: IEEE International Conference on Communications (ICC) (2013)
19. Badic, B. et al.: Energy efficient radio access architectures for green radio: large versus small cell size deployment. In: IEEE 70th Vehicular Technology Conference Fall (VTC 2009-Fall), pp. 1–5 (2009)
20. 3GPP TR 36.814 V9.0.0, 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Further advancements for E-UTRA physical layer aspects (Release 9), March 2010
21. Wang, L., Aghvami, H., Nafisi, N., Sallent, O., Perez-Romero, J.: Voice capacity with coverage-based CRRM in a heterogeneous UMTS/GSM environment. In: Second International Conference on Communications and Networking in China, CHINACOM (2007)
22. Rodriguez, J., Marques, P., Radwan, A., Moessner, K., Tafazolli, R., Raspopoulos, M., Stavrou, S., Trapps, P., Noquet, D., Sithamparanathan, K., Gomes, A., Piesiewicz, R., Mokrani, H., Foglar, A., Verikoukis, Ch.: Cognitive radio and cooperative strategies for power saving in multi-standard wireless devices, Future Network and Mobile Summit (2010)
23. Facchini, C., Holland, O., Granelli, F., da Fonseca, N.L.S., Aghvami, H.: Dynamic green self-configuration of 3G base stations using fuzzy cognitive maps. Comput. Netw. (2013)
24. Facchini, C.: Bridging the gap between theory and implementation in cognitive networks: developing reasoning in today's networks. Ph.D. Thesis, University of Trento, Italy, Dec 2011

Chapter 15
Testbeds and Implementation Issues

K. Katzis, A. Perotti and L. De Nardis

15.1 Introduction

A Cognitive Radio (CR) system can be visualised in Fig. 15.1 as a complicated wireless communication system that involves a virtual engine (soul) and a platform (body). The engine is implemented based on a number of logical arguments in order to reason and negotiate with other wireless systems, aiming towards the optimum utilisation of the spectrum while ensuring minimum disruption to existing wireless systems.

Current research and development on CR has been pushing towards the development of different versions of engines running on different types of platforms. This is to address one of the main concerns of CR, which is to ensure that it does not disturb any primary users from existing wireless systems as well as to prove its potential in terms of performance. This cannot be resolved through theoretical models and simulations but it requires to be addressed through a working system which demonstrates that the amount of interference is sufficiently low to justify its operation. The practical demonstration of a CR is essential for bringing such systems in real life by convincing local regulatory bodies of their non-disruptive operation.

Based on the above argument, a number of CR testbeds have been implemented for evaluation purposes on the currently available CR platforms such as USRP2, BEE2, VESNA or any other hardware and software platforms. The testbeds must be designed based on well-defined metrics and test cases in order to generate useful results. Within

K. Katzis (✉)
European University Cyprus, Egkomi, Cyprus
e-mail: K.Katzis@euc.ac.cy

A. Perotti
CSP-Innovation in ICT, Torino, Italy
e-mail: alberto.perotti@ieee.org

L. De Nardis
Sapienza University of Rome, Rome, Italy
e-mail: lucadn@newyork.ing.uniroma1.it

M.-G. Di Benedetto and F. Bader (eds.), *Cognitive Communication and Cooperative HetNet Coexistence*, Signals and Communication Technology,
DOI: 10.1007/978-3-319-01402-9_15,

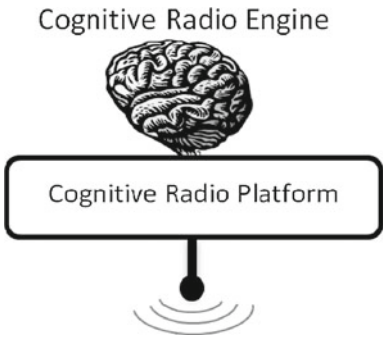

Fig. 15.1 CR engine running on a CR platform

this context, this chapter aims to present testbeds and their implementation issues in four subsections. The first section introduces the concept of CR engine, emphasizing its advantages and disadvantages when compared to legacy communication systems. The second section presents the basics of CR platforms regarding their software and hardware parts, giving an overview of the various available technologies. It then presents in detail all major testbeds that are currently available, commenting on their operation characteristics and their current applications along with some of their breaking field trials. Last, the fourth section discusses known implementation issues of the CR engines, looking as well at related standardisation activities. An example on how to implement a CR testbed is also presented. Finally, a paragraph with the main conclusions regarding the available test beds closes the chapter.

15.2 Cognitive Radio Engines

Cognitive radio (CR) engines in literature are presented as the encephalon, the decision-making part of a cognitive radio system. They are often described as a multiple system of parameters that require delicate tuning to achieve optimum performance. Performance in CR networks is defined in terms of multiple elements, such as bit error rate (BER), bandwidth, throughput, and transmit power. There are numerous CR engines in the research arena that vary in many aspects depending primarily on the system that are to be employed while looking at network architecture, applications utilized, level of security required, hardware employed, radio environment, etc. Nevertheless, all different instances of CR engines share the common goal of optimizing the performance of the system while ensuring coexistence with the existing wireless systems.

Looking back in literature, a significant number of CR engines have been inspired by Genetic Algorithms (GA) [19, 34, 38, 47]. These algorithms are requiring modelling of the physical (PHY) layer traits of the radio within the context of a genetic chromosome. Biologically inspired CR engines are capable of intelligently adapting a radio's physical and MAC behaviour on constantly varying network conditions.

An example where a GA driven CR decision engine is developed is presented in [31]. The GA based engine determines the optimal radio transmitter parameters for single and multicarrier systems. The GA is based on a fitness function that directs evolution of the GA parameters to their optimal values. Further to the GA based CR engines, a biologically inspired cognitive engine with dynamic spectrum access has been presented in [27]. Initial results indicate that the performance of the proposed engine achieved a 20 dB signal to interference plus noise ratio (SINR) increase when compared with the traditional IEEE 802.11 physical layer standard. The wireless GA proposed in [27] is a multi-objective GA designed for the control of a radio by modelling the physical radio system as a biological organism and optimizing its performance through genetic and evolutionary processes. In [26], a prototype smart receiver has been presented including a General Purpose Processor (GPP) based software defined radio platform, signal classification capability and PHY-MAC reconfigurability with hardware-independent radio interface. Another approach based on GA driven CR decision engines is outlined in [1]. This work explores the use of Taguchi method and orthogonal arrays (OA) as a tool for identifying favourable GA parameter settings. The strategies developed here limit the number of required tests needed to identify acceptable parameter values as opposed to the well-known methods such as design of experiment (DOE) and response surface methodology (RSM). As the number of configuration variables grows, DOE and RSM formalisations diminish due to the significant number of test cases that require full factorial designs. Results presented in [1] indicate that the Taguchi method analysis yields a predicted best combination of GA parameters from nine test cases. It utilises an efficient selection of testing configurations based on the concept of OA.

In [35] the theory and the prototypical implementation of the CE is presented. Furthermore a list of cognitive components is presented along with various issues related to developing algorithms for CR behaviour. These components are perception, conception and execution. Figure 15.2 presents a generic architecture of a cognitive radio [35] while showing how a cognitive engine interacts with the rest of the components of a CR architecture. Here there are three input domains that concern the cognitive radio. The user domain informs the cognitive engine about performance requirements of services and applications to guarantee the minimum acceptable quality of service (QoS). Based on the different quality of service levels, the user domain effectively sets the performance goals of the radio. Furthermore, information regarding the changes in performance of waveforms based on different propagation environments is collected, to determine the effect of the external environment and RF channel. The last part of this generalized CR architecture is the policy domain that, as mentioned before, guides the system to perform within the boundaries and limitations set by the local regulatory bodies as interpreted by the CE.

A generalized Cognitive Engine design is also presented in [35]. The engine presented includes a central component called the cognitive controller that acts as the system kernel and scheduler to handle the input / output and timing of the other attached components. The major components listed in [35] and illustrated in Fig. 15.3 as part of a CE are: cognitive controller, sensors, optimisers, decision makers, policy engine, radio framework and user interface.

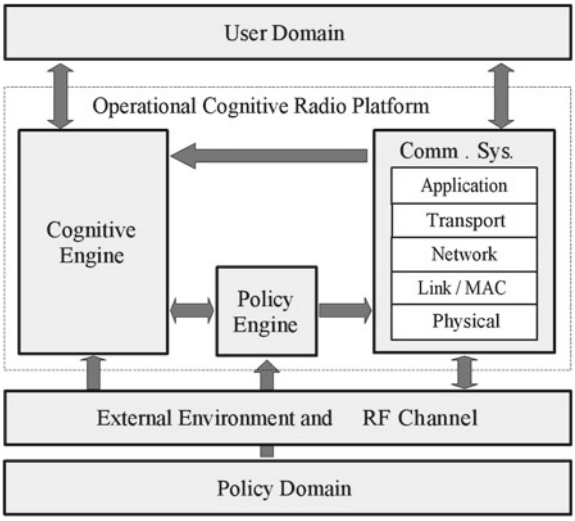

Fig. 15.2 CR architecture [35]

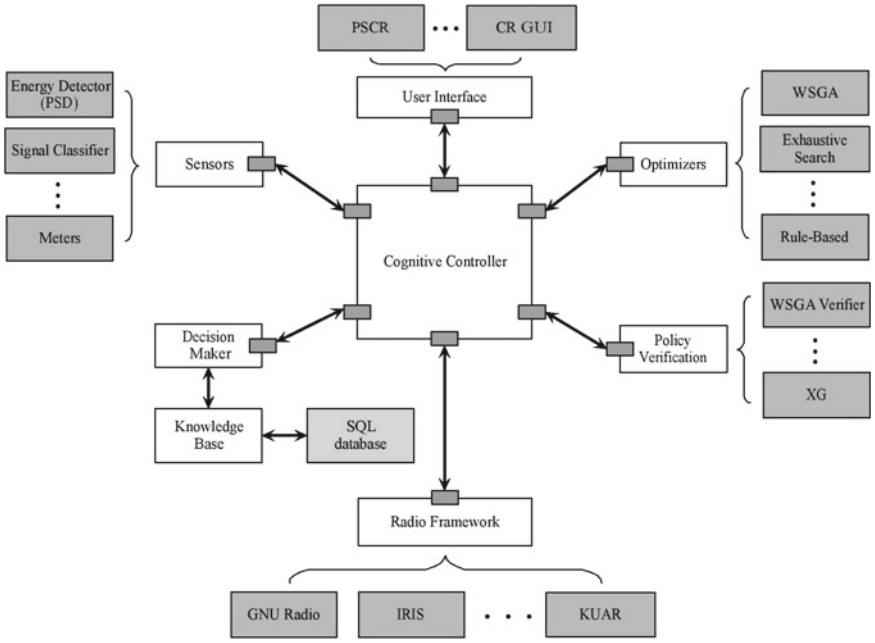

Fig. 15.3 Cognitive engine [35]

Each component is launched as a separate process that interfaces and exchanges data between processes. Components can be designed, implemented and tested in isolation for low coupling between processes, thus enabling distributed processing where different components can reside on different processors or hosts. The proposed architecture also enables testing of different types of algorithms and processes to create new components.

An important part of the design and development of CR engines are the tools employed to achieve the best possible result. In [14], a Policy Reasoner (PR) is proposed based on a language used for expressing policies that allow opportunistic spectrum access. Using a Policy Reasoner in CR engines, it is possible to guarantee policy-specified behaviours while allowing spectrum sharing. The proposed engine performs its processing at a software level rather than hardwired in the system like in legacy radios, thus achieving a device independent policy reasoning. The flexible mechanism coordinates, considers and processes a large number of operating dimensions while adhering to all regulation policies and maintain optimum spectrum sharing. The implementation of the Policy Reasoner in [14] was carried out using Cognitive (Policy) Radio Language (CORAL). CORAL has been first introduced in the neXt Generation (XG) programme funded by DARPA [12, 13]. CORAL is a language for policy specification that was devised to encode the policies in a straightforward, "natural" way in something close to first-order logic. The advantage of combining CORAL language with a device independent platform proposed in [14] was the capability to easily update the operating software based on various types of regulatory policies that exist. This would have an immediate effect on the operation of the engine in respect to:

1. Sensing frequencies (unrestricted bands, identification of primary users).
2. Characterisation of opportunities.
3. Distributed or centralized coordination of resources based on communication with other devices for the identification of resource availability.
4. Enforcing behaviour consistent with policies.

In [9] a cognitive radio architecture is presented while the authors discuss reasoning and learning engines as parts of a CE. Reasoning and planning here are subjected to various radio parameters. With these parameters a knowledge-base is formed just like an expert system in Artificial Intelligence (AI) systems. The proposed engine is implemented based on a generic cognitive engine [40] which combines OSSIE [33], an open source software communications architecture SCA, along with Soar Cognitive Engine. Efficient design of CR engines requires the capability of experimentally verifying the proposed solutions and the identification of engine components and of corresponding implementation choices is a fundamental step towards this direction. For this reason, tools, platforms have been combined together to form an evaluation testbed for CR engines.

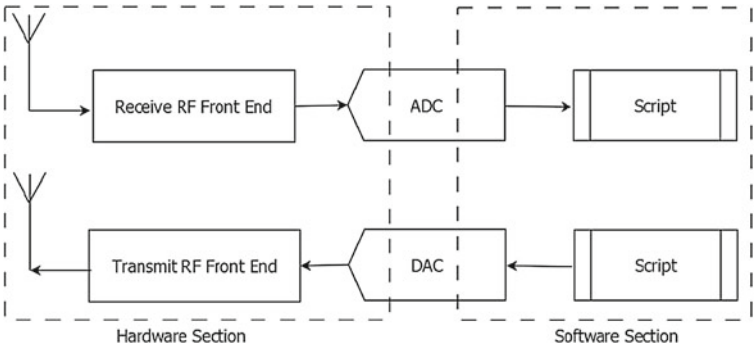

Fig. 15.4 Generic block diagram for SDR

15.3 Cognitive Radio Testbeds

Several testbeds have been designed and deployed in the last few years to evaluate and improve solutions for cognitive radio networks, including algorithms and strategies specific for Cognitive Engine functions. Cognitive radio research requires that these testbed facilities are equipped with appropriate capabilities because of complex interaction between physical and network layers that must be addressed jointly. A set of requirements that a cognitive radio testbed should employ are [5]:

1. Real-time baseband processing for spectrum sensing.
2. Agile transmission with high computational throughput and low latency.
3. Integration of physical and network layers on embedded processors.
4. Sufficiently wide bandwidth radio front end with spatial processing capabilities.
5. Central processing of information exchange between multiple radios for controlled physical and network layer development and analysis.
6. Ability to perform controlled experiments in different propagation environments such as indoors or outdoors.

This section describes the most relevant testbeds deployed in the Cognitive Radio community, moving from the introduction of major Cognitive Radio Platforms, that constitute the building blocks of testbeds. CR platforms provide in fact the means for designing testbeds in order to carry out experiments and thus support design, implementation, testing and performance evaluation of CR algorithms and protocols, including those related to CR engines.

CR platforms are typically the result of the combination of hardware and software components, often developed independently and later combined to build a complete CR system to be deployed within a testbed. The general idea behind a complete CR device, resulting from the combination of hardware and software CR platforms is depicted below in Fig. 15.4.

The front end found in a CR platform is designed to implement the front end of a radio transceiver. The less the CR system relies on its hardware part the better

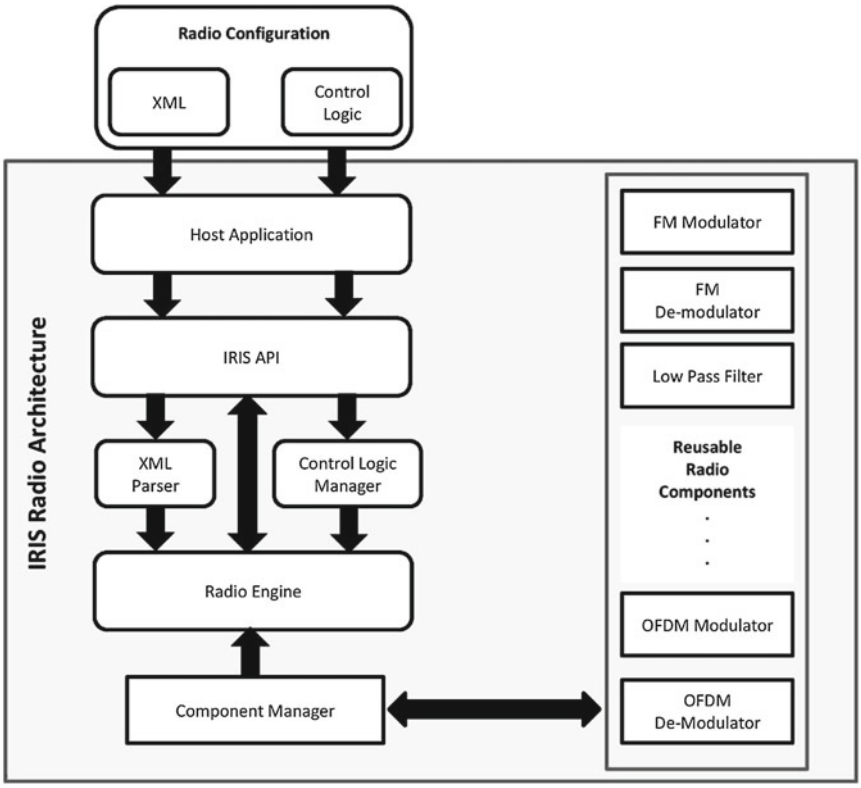

Fig. 15.5 IRIS layers of components

thus allowing more flexibility, re-configurability and upgradeability of the system through its software part.

In the following subsections hardware and software CR platforms are introduced, and later most relevant testbeds built on various combinations of such components are described, identifying their capabilities (in particular related to Cognitive Engine aspects) and known experimental campaigns completed or ongoing using them (Fig. 15.5).

15.3.1 Software CR Platforms

Software CR platforms were in most cases defined as the evolution of Software Defined Radios (SDRs). SDRs have been around for more than 20 years starting in the analogue modem industry where manufacturers implemented the modulating and de-modulating algorithms in software. This allowed them to upgrade / change the

communication standards without having to change any of the hardware components. Based on this concept, SDRs have been further evolved to better utilise the radio spectrum by allowing real-time reconfigurability and improve compatibility and coexistence with different wireless standards. This has been achieved by employing software to perform the modulation and demodulation of the radio signals.

Building on such experience several advanced SDRs, that can be adopted as software CR platforms in the deployment of a CR testbed, were proposed in the last years:

- *GNU Radio* is a widely popular open source implementation of SDR. GNU Radio is a software library written in a combination of C++ and Python that enable the implementation of various signal and data processing blocks to perform complex signal processing operations, by providing simple signal processing primitives written in C++. The code implemented in C++ is then integrated in scripts using SWIG, an interface compiler. Python scripts can be then used to implement specific applications that rely on underlying signal processing blocks written in C++. The advantage of GNU radio is that all these blocks run in their native speed without having to be interpreted by a compiler thus achieving optimum processing speed. Systems that have been implemented in GNU radio include HDTV video signal decoders, AM/FM transmitters and receivers, as well as a simple packet radio system using GMSK modulation and demodulation to transmit packets. It is worth mentioning that GNU Radio is a platform mainly focusing on physical layer and signal processing aspects, with very few examples of projects developed on top of it that focus on network and higher layers.
  Implementation of algorithms in GNU Radio typically requires programming skills in C++ and Python, although software tools such as the GNU Radio Companion can provide a more intuitive interface to GNU Radio components. GNU Radio is released under the GNU Public License (GPL) v3, and can be freely downloaded from the GNU radio website [20].
- *IRIS*, standing for Implementing Radios In Software, is a software radio engine designed to operate on a general-purpose processor, and to interact with a hardware frontend [18]. It is used to create software radios that are reconfigurable in real-time. The IRIS radio architecture allows to create a radio from the existing suite of components or by creating new ones. IRIS relies on an architecture based on plug-ins, that are dynamically loaded by the main program when required by the desired SDR architecture. Each plug-in implements in fact a radio component that implements a specific radio functionality allowing full flexibility and reconfigurability in the design of the overall system.
  The Radio Component is one of the building blocks in the IRIS architecture, that foresees six major entities: Radio Component, XML parser, component manager, radio engine, control logic manager and API [18]. As already explained, the Radio Component implements an individual stage in the signal processing chain of an IRIS reconfigurable radio. The XML parser is used to interpret the radio design that is stored in an external XML configuration file, listing the components required to build the desired radio architecture, and how these components will interact.

Information in this file is converted into a form that can be later processed by the component manager and radio engine. The Component Manager has in fact the role of loading the radio components specified in the XML configuration file. These components define the radio implementation and are located either on the host PC or in a network-accessible remote location. The radio engine is the key element in the implementation of a reconfigurable radio IRIS SDR architecture and is supported by the Component Manager and Control Logic Manager. The control logic manager's main purpose is to reconfigure the existing radio into a new one. It has the ability to access the structure and parameters of all the components in real time and modify them in order to implement the new radio layout. Finally, the IRIS API is the component that allows integration of the reconfigurable radio application with other applications running on the host PC.

IRIS is proposed as a solution for the implementation of both physical layer applications, thanks to the definition of all key signal processing blocks required for modulation/demodulation and coding of signals, and of higher layer applications, involving Medium Access Control, network and above layers. For this reason, IRIS might be a better starting point compared to GNU Radio when targeting the design and implementation of higher layer protocols for cognitive radio. Other significant differences between IRIS and GNU Radio include the fact that IRIS employs a controller that allows to dynamically reconfiguring the CR mechanism when compared to GNU radio. It is also easier to integrate the stack engine with the physical layer. In addition, GNU radio can process threads either in a single block (one thread that runs in the processor) or if the multithread mechanism is employed, each block is considered as a different thread which effectively overloads the processor. In IRIS, each component can be defined with a different number of threads.

IRIS can be downloaded by a dedicated github repository, as explained in [25].

- *ASGARD*, standing for Application-Oriented Software on General Purpose Processors for Advanced Radio Development, is another platform that recently gained attention [2]. ASGARD is a software tool utilised for the implementation of Cognitive Radio communication systems over Software Defined Radio equipment. Its development started at Aalborg University within the framework of the EU FP7 SAMURAI project [7], the main idea being to design a Software CR Platform that provides the necessary degree of flexibility for the implementation of reconfigurable, multi-layered communication systems. The platform architecture is component based and supports a customized thread of execution architecture thus allowing modularity, flexibility and domain driven development. The software architecture of ASGARD is based on application programming interfaces (API) written in C++, while the platform relies on XML for configuration files. In ASGARD the following objects can be defined as APIs: (1) the Application, encompassing all other objects, (2) Modules, (3) Components typically grouped in Modules, and (4) Communication Objects enabling communications between components belonging to the same or different modules. In order to ensure testbed evolution and expansion, compatibility with Off-The-Shelf hardware, modern PCs and integration within the networking stack of recent Linux OS releases has been

taken into account when developing this platform. The software is in fact designed to run in the Linux (Ubuntu) OS free user space.
Differently from the GNU Radio platform introduced earlier in this section ASGARD focuses on the support of implementation and experimentation of protocols and algorithms for higher layers.
Aasgard is released under the Aalborg University software license, and can be downloaded for free from the ASGARD website [2].

15.3.2 Hardware CR Platforms

The software CR platforms introduced in the previous subsections must be combined with a suitable hardware platform, providing analog RF frontend to be driven by the code running in the software.
There are numerous examples of hardware platforms available for developing complete CR systems.

- The *USRP*, standing for Universal Software Radio Peripheral. USRP is a widely popular platform; it was developed by Ettus Research as a low cost hardware platform for software defined radios, and is currently owned by National Instruments. Three families of USRP are currently available:
  - the Networked (N) family, offered in the N200 and N210 models, provides maximum performance, but relies on network connection with an external host PC. Both models sport 14-bits Analog-to-Digital Converters (ADC) and 16-bits Digital-to-Analog Converters (DAC), and mainly differ in the size and performance of the onboard FPGA; furthermore, both models support MIMO operation by means of a dedicated synchronization cable between devices.
  - the Embedded (E) family, also including two models (E100 and E110) that provide performance similar to the N models. The E family adds however a onboard ARM Cortex-A8 processor combined with a DSP, enabling operation without the need of a connection to a host PC, and thus allowing for stand-alone deployments.
  - the Bus (B) family includes the B100 model and the original USRP1, and is the entry level product family for the USRP platform. Smaller FPGAs, lower data rates (8 MS/s versus 64 MS/s for the E and N families) and in the case of the USRP1 no MIMO support make this family only suitable for simple demonstrations and teaching activities.

  As already mentioned, all platform versions are equipped with a FPGA controlling several ADCs and DACs. It is designed for RF applications from DC to 6 GHz, by means of dedicated daughterboards that, combined with suitable antennas, enable operation over different frequency bands.
  One of the main advantages of the USRP platform is the availability of a well maintained and well documented USRP Hardware Driver (UHD), available for all

major Operative systems. The availability of UHD led to a wide support for USRP hardware by many different software tools and development toolkits. Software development platforms supporting the USRP include:

- GNU radio, IRIS and ASGARD software CR and SDR platforms, described in the previous subsection;
- MatLab and Simulink, developed by MathWorks, Inc;
- LabView, developed by National Instruments.

Some of the proposed applications for USRP are: white spaces, mobile phones, public safety, spectrum monitoring, radio networking, cognitive radio, satellite navigation, and amateur radio [16].

- The *BEE2*, standing for Berkeley Emulation Engine 2. The BEE2 system is a FPGA-based computing platform designed to be modular and scalable. Its software design methodology targets a wide range of high performance applications one of which is Cognitive Radio Systems [3]. The FPGAs employed are Xilinx FPGAs (Virtex II Pro 70). The platform has been used to practically carry out a number of scenarios at the physical and link layer levels in [29]. The motivation was to validate various sensing algorithms to prove non-interference to licensed users while evaluating their performance. For trials, the 2.4 GHz spectrum was chosen since a wide range off-the-shelf radio equipment is available. More recently, the development led to the creation of a dedicated spin-off designing, implementing and distributing the latest version of the platform, the BEEcube [4].
- The *VESNA* [36] platform is a modular and fully flexible platform employed for the development of wireless sensor networks. It is based on a high-performance microcontroller with ARM Cortex-M3 core and radio interface spanning over multiple ISM frequency bands. VESNA is a core module and a set of special feature modules (sensor node radio—SNR, sensor node expansion—SNE, sensor node power—SNP) that are used as/if needed. The platform features various peripherals including UART, I2C, SPI, USB, ADC and DAC. All these allow hosting of different sets of sensors and/or actuators including sensors for temperature, humidity, luminance, color, reflectance, pressure, presence, location, sound, acceleration, gasses, motion, range, and actuators such as motor, relay, servo, alarm.
- *WARP*, standing for Wireless Open-Access Research Platform. WARP is a hardware platform originally designed at Rice University and built by Mango Communications [43]. The latest version of the platform, WARP v3, is based on a Xilinx Virtex-6 FPGA, provides two programmable RF interfaces each capable of operating at both 2.4 and 5 GHz, and supports up to 32 GBs of onboard RAM. The platform also provides ethernet connectivity and several I/O possibilities (serial, LEDs and DIP switches). The WARP website also provides information on software tools required to program the platform.

15.3.3 CR Testbeds

Several CR testbeds have been devised by the research community to carry out both small and large scale experiments focusing on Cognitive Radios, and the most relevant ones are discussed in this section. Most of the following testbeds have been implemented based on various combinations of software and hardware platforms introduced in the previous subsections.

- *DARPA XG Program*. The **XG** program aimed to provide dramatic improvements in military communications in support of a full range of worldwide deployments. This has been achieved with the development of enabling technology and system concepts to dynamically redistribute allocated spectrum. Through this programme, the CORAL Cognitive (Policy) Radio Language (**CORAL**) has evolved. This language expresses various policy engines as well as reasoning techniques developed for the XG program, that might be of inspiration in the design and implementation of a Cognitive Engine. Dedicated radios were built within the XG Program Testbes based on a WiMax physical layer and frequency scaler used to move operation frequency to the 225–600 MHz range [28], and used for the creation of a testbed and execution of field tests focusing however mainly on sensing aspects.
- *CREW* Another testbed developed for CR is the result of the CREW project [18], funded by the European Commission in the framework of the Future Internet Research and Experimentation (FIRE) initiative. The **CREW testbed** is based on the federation of five existing testbeds available at different physical locations such as heterogenous ISM test environment, licensed CR testbed (including TV bands), wireless sensor network testbed, LTE cellular test environment and outdoor heterogeneous ISM/TVWS. The CREW project combined a number of functionalities from the aforementioned platforms such as a common portal, the ability to mix and match different aspects of different testbeds, a benchmarking framework, realistic data sets and performance evaluation of external hardware under controlled test conditions. The testbeds currently federated in CREW are:
  - the *w-iLab.t* testbed, hosted by IBBT in Belgium over two locations, combining devices over respectively 200 and 60 spots, with technologies including Wi-Fi, Bluetooth, sensor nodes and USRP platforms.
  - the *IRIS* testbed, hosted by CTVR in Dublin, Ireland that provides access to 8 host PCs connected to USRP N210 devices, and the IRIS software platform described in the previous subsections. More details on the IRIS testbed is provided later in this section.
  - the *TKN Wireless Indoor Sensor network Testbed (TWIST)* testbed, hosted by TKN in Berlin, Germany, focusing on sensor networking, thanks to more than two hundred TMoteSky [42] and eyesIFXv2 [17] sensor nodes.
  - a *LTE/LTE-A* testbed, hosted by TUD in Dresden, Germany composed of 5 LTE eNBs and 8 UEs.

– a *Versatile Sensor Network* testbed, hosted by JSI in Ljubljana, Slovenia, and composed of tens of sensor nodes based on the VESNA platform previously described.

Moving for the 5 component testbeds defined above, CREW aims at supporting a wide range of experiments that take advantage of multiple testbeds, either by simple comparison, or by combination of hardware coming from different facilities to support complex experiments.

- *VT-CORNET*. Another well known CR testbed is the Virginia Tech Cognitive Radio Network Testbed (CORNET), intended as a tool for performing large scale experiments focusing on Software Defined Radio and Cognitive Radio [45]. CORNET is a large testbed composed by three key elements:

1. software defined radio nodes with RF hardware based on the USRP2 platform and the Virginia-Tech Cognitive Radio Open Source System (VT-CROSS) framework;
2. a large set of servers characterized by high processing capabilities;
3. a web interface available for reserving time on the testbed and monitoring the status of the RF platforms.

The objective is to develop a distributed and modular system that provides portability and interoperability between modules developed in different programming languages. The VT-CROSS framework [11] is composed of up to five components:

1. Cognitive Radio Shell
2. Cognitive Engine
3. Software-Defined Radio Host Platform
4. Policy Engine
5. Service Management Layer

The different modules interact and communicate by means of standard TCP/IP sockets, allowing for independent development of each component (possibly in different programming languages). The *Cognitive Radio Shell (CRS)* has the role of parsing a XML file provided as an input to the framework in order to:

1. determine the radio parameters available for configuration on the host platform and their potential values;
2. define the QoS metrics to be taken into account in the operation of the framework;
3. define the observation variables (e.g.: SNR) and their relationships with the utility functions

The *Cognitive Engine (CE)* is in charge of deciding how to adapt the behaviour of the cognitive radio, and in particular the settings of the transmission parameters. The CE receives as a first input the available radio parameters and the utility functions parsed by the CRS. During cognitive radio operations, the CE receives observations gathered by the Cognitive Radio Shell as well as any previous experience cached in the CRS, and returns a solution consisting in the best possible settings for the available parameters given the utility function(s).

The *Software-Defined Radio Host Platform* is the actual hardware that gathers observations and emits signals with parameters determined by the CE. In the present implementation of VT-CROSS the platform is based on the USRP2 hardware, complemented by a custom RF daughterboard capable of operations between 100 MHz and 4 GHz with 20 MHz instantaneous bandwidth.
The *Policy Engine* has the role of storing and managing policies for spectrum access, and validate the selection of the transmission parameters performed by the CE versus the currently active policy.
The *Service Management Layer* provides a higher layer interface allowing for the definition of "Missions", usually involving the optimization of different utility metrics by different instances of the Cognitive Engine. As an example, a Mission could be "Jam all neighboring devices", which would involve the sensing of existing devices and the optimization of radio parameters to jam detected devices, possibly adhering to different standards.

Processing servers

The network of servers is composed of 48 servers connected to the Cognitive Radio nodes, in charge of performing heavy signal processing tasks that may be required for nodes operation, plus 5 management servers. Servers currently deployed are based on Intel Xeon Quadcore processors and equipped with several GB of RAM.

Web interface

The web interface provides free access to the testbed (via a registration procedure to create an account). Users can submit reservation requests that are examined and approved by Virginia Tech system administrators. Reservations can be submitted for the whole testbed or for portions of it (e.g. nodes on one of the 4 floors) allowing for concurrent operations of different users. When a reservation is accepted and the reserved time slot approaches, all nodes are prepared by uploading the OS selected by the user (presently one of several possible Ubuntu Linux versions) along with the selected SDR and CR framework. Presently the following frameworks are available:

– GNUradio (SDR);
– OSSIE (SDR);
– VT-CROSS (CR) or Virginia Tech Cognitive Radio Open Source System (**CROSS**) [11].

The web interface also allows to monitor the status of each node and enquire if it is available for experiments.

- *VESNA-Testbed*. Another testbed that is worth mentioning is **VESNA testbed** [36]. VESNA, as mentioned before, is also a platform that has been developed at JSI (SI). Nevertheless it is being employed by various research communities such as is the JSI in Slovenia. An example of a VESNA based testbed is LOG-a-TEC that

was developed in the city of Logatec, Slovenia as part of the European funded project CREW [39]. People at JSI, upon agreement with the local authorities, they secured the use of the street lighting infrastructure for mounting and power supply their equipment to approximately 1000 locations, consisting mostly of lamp posts but also occasionally other parts of public infrastructure like switching stations. In this particular scenario, LOG-a-TEC testbed was designed to cover cognitive radio experiments in the 2.4 GHz international ISM band, 868 MHz European short range device band and the white-spaces in the UHF broadcast band. To achieve this, VESNA sensor nodes were equipped with a custom designed expansion that contained spectrum sensing equipment and software-reconfigurable digital transceivers operating on these frequencies. These radios operated independently from the management network and form the experimental part of the testbed.

- *IRIS Test-Bed (CTVR).* The main lab at CTVR implemented a large number of cognitive radio models based on USRP technology and the IRIS CR test-bed evolved from the CREW project. The IRIS test-bed at CTVR is divided into two parts. The first part consists of workstations connected to URSPs (dedicated to CREW project [18]) that can be accessed remotely to perform any sort of tasks. One of them is only equipped with a mouse and a keyboard for general monitoring. This part can be accessed through a gateway. The whole system is operational through IRIS. As introduced in a previous subsection, IRIS is an open source software CR platform, that has been developed and expanded in CTVR. It's a modular platform allowing the development of different components that can be connected to form a complete radio platform. The developed experimental platform has been based on two main engines. The first one handles the physical layer which is composed of a number of components connected in a line such as modulator, demodulator, signal scaler etc. to create an interface with the RF front end of a USRP device. The second part of the test-bed is the stack engine which handles the operation for the upper layers. So in IRIS you can create a full stack of layers (from physical to application layer). One of the advantages of IRIS is that it can be reconfigured on the fly and this is achieved through a controller. The controller is an entity that controls in real time the various components of the transceiver node. An event can trigger the controller to update a number of parameters of the transceiver.
  Several field tests have been carried out using the IRIS test-bed; some details are provided in the following.
  - MAC Layer on URSP100—Several URSP devices with an embedded processor have been used to implement the MAC layer running on IRIS. These URSP units have limited processing capabilities when compared with the aforementioned workstations. However they have been used as self-contained devices with adequate processing power. Due to the processing power limitations, time critical processes such as carrier sensing mechanisms have been assigned on an FPGA, thus splitting some of the MAC functionality on software and some on an FPGA. CR engines employing USRP, USRP2 platforms can make use of these FPGA mechanisms since this is an open source tool.

  - Train to perform handoff employing USRP2 Cognitive Radio Engine—Two USRPs have been used to act as base stations (BSs) and another two were connected on two model trains acting as mobile stations (MSs). The MSs had to detect the two signals broadcasted from the two BSs that use different cyclostationary signatures to tell the mobile to which band they should switch to. So the MS should switch to a different BS as well as different band without any previous knowledge which band they should switch to. The trials were conducted using the TV bands since the institution (CTVR) had license to use them.

- *FP7 SAMURAI project testbed.* The testbed, realised as the official demo of the SAMURAI FP7 project, relies on the ASGARD software platform defined above and combined with several USRP N100 and N200 platforms. Various scenarios that are under consideration as part of the experiments running on the testbed are: distributed interference coordination in femtocells while using an Autonomous Component Carrier Selection (ACCS) algorithm for interference reduction, especially for outage protection. The test bed implemented was comprised of 12 nodes, 6 of which were acting as Access Points (AP) and the other half were acting as User Equipment. Affiliation was one UE per cell in a Closed Subscriber Group scenario [7].

15.4 Implementation Issues

The numerous test-beds developed to perform experiments in the cognitive radio field have brought to evidence several potential issues concerning the design and implementation of the cognitive engine, of the cognitive radio systems and networks as a whole.

The ultimate goal for a cognitive engine of finding an optimal configuration is pursued through the maximization of a multiple objective *fitness function* [31, 32] that, for the case at hand, quantifies the advantages of choosing a given system (and network) configuration with respect to others. Such fitness function summarizes with a single figure of merit, how close a given system configuration is to achieving its optimum.

In particular, given a *parameter space* $X \ni \mathbf{x} = (x_1, \ldots, x_N)$, where x_i represents the single parameter values (the *knobs* of our reconfigurable radio), we define a set of *objective functions* as

$$y_m = f_m(\mathbf{x}) \tag{15.1}$$

where the objective vector $\mathbf{y} = (y_1, \ldots, y_M)$ lies in a given objective space Y. The fitness function then computes the single figure of merit as a function of the objective vector. In [31], a simple weighted sum has been used:

$$M(\mathbf{x}) = \mathbf{w} \cdot \mathbf{y} \tag{15.2}$$

where $\mathbf{w} = (w_1, \ldots, w_M)$ is a suitable weight vector with $||\mathbf{w}||_1 = 1$. Here, $|| \cdot ||_1$ denotes the ℓ_1 norm of a vector, i.e., the sum of its elements. Now, the optimization of (15.2) over the parameter space X and the choice of the weight vector $\mathbf{w}$ may result very complex due to the high number of parameters and to the multiple possible choices for $\mathbf{w}$. To this purpose, evolutionary algorithms, like the GA mentioned in Sect. 15.2, are often chosen. For a discussion on the choice of the single objective functions, we refer the interested reader to [32].

As an entity capable of acquiring awareness of the surrounding electromagnetic environment and later exploiting this information, the CR engine typically requires the availability of a significantly high amount of memory. One of the most successful methods adopted for the implementation of CR engines is case-based reasoning [31] (CBR). This method consists in storing a history of cases in which the radio has had to operate. Afterwards, the current operating conditions are compared with the recorded history in order to find a best match and to retrieve the transceiver configuration that was used in that case. Each single case may consist of a relatively large amount of data that are needed to accurately characterize the operational context and the parameters. Moreover, finding a best match between the current situation and the previously recorded cases in order to choose a suitable configuration becomes more and more complex as the amount of available records grows. This ultimately results in a further increase of required processing load.

Although it has been extensively proven that these algorithms are capable of finding suitable solutions to the posed problems, fundamental questions concerning how fast these algorithms converge to a solution and whether the provided solutions are optimal or not remain only partially answered so far.

The memory footprint and processing load of these solutions are more typical of a high-level application running on a workstation rather than of an entity that should be executed into a small and power-efficient wireless transceiver or in a distributed fashion over a network.

Despite these concerns, the processing capabilities and memory availability of newest devices is ever growing and recent works [46] demonstrate that cognitive engines can be implemented and executed even on platforms with limited computational capabilities, like some embedded systems. However, it is likely that some significant effort will be put on the development of more efficient and effective optimization algorithms tailored at the specific case of CR engines in order to obtain more compact, less memory- and processor-hungry solutions.

15.4.1 Standardization Challenges and Efforts

The cognitive radio concept is rather widespread and evolving today. CR test beds and CR engine prototypes have been under development for several years and a number of test beds have been used to evaluate their performance. CR technology is now mature enough to be effectively employed in the products we use in our everyday life. New wireless standards in the process of being defined include specific cognitive features,

Table 15.1 Summary of cognitive standards and main characteristics

Standard	Application	Range/scenario	PHY layer	Incumbent protection
ECMA 392	Personal/portable	Indoor	Coded OFDM	Spectrum sensing
IEEE 802.16h	WRAN (Data)	Outdoor	Coded OFDM	Spectrum sensing
IEEE 802.22	WRAN (Data)	Outdoor	Coded OFDM	Spectrum sensing, WSDB
IEEE 802.11af	WLAN (Data)	Indoor, short -range outdoor	Coded OFDM	WSDB
Weightless™	M2M	Mid-range outdoor	Spread-spectrum	WSDB

like opportunistic transmission with interference protection of primary users. Many of these standards target broadband wireless communications. Other standards and proprietary solutions have been developed for machine-to-machine (M2M) and sensor communications.

Current standards do not define any intrinsic characteristic of the cognitive engine. They rather specify the absolute constraints that the system must comply with in order to protect primary users and to coexist with other similar systems minimizing their mutual interference. Procedures like spectrum sensing, minimum detection levels of primary user signals, channel evacuation procedures and time constraints are specified.

A summary of currently active or drafted wireless standards with cognitive features is shown in Table 15.1. In the following, relevant cognitive characteristics and constraints concerning such standards are summarized.

The first standard for broadband wireless communication in the TV white spaces has been defined by the European Computer Manufacturers Association (ECMA) in 2009. Based on the contributions from the Cognitive Networks Alliance (CogNeA) [8, 10], a European consortium of companies operating in the consumer electronics market, ECMA delivered the first cognitive radio standard for TV white spaces, the **ECMA-392**, in 2009 [15]. This standard has been developed with the aim of providing Internet access and high definition multimedia services to residential users.

ECMA-392 networks are organized as master-slave or peer-to-peer. In the first case, a designated master node coordinates network formation. In the second case, network formation is carried out through distributed beaconing and channel reservation mechanisms.

The standard defines mechanisms for the protection of incumbents and for self-coexistence. As for incumbent protection, sensing is performed both in the operating channel and in adjacent channels during properly scheduled *quiet periods*. Sensing in adjacent channels is required as a protection measure to establish a limit on the maximum power level allowed in the operating channel.

Scheduling of quiet intervals occurs periodically and on-demand based on the occurrence of specific conditions at the PHY/MAC layer, such as increased SINR or packet loss.

When a primary signal is detected, a channel evacuation procedure is started. Channel evacuation consists in suspending all data communication and moving all nodes to another channel, if available. This procedure is coordinated by the master node if the ECMA-392 network is of master-slave type and by the reservation owner if the network is peer-to-peer.

Within the IEEE, several standardization efforts have been carried out and are currently ongoing. In 2010, the IEEE released the **IEEE 802.16h** standard for unlicensed WiMAX operation in licensed bands [24]. This standard defines improved coexistence mechanisms for the protection of primary users (uncoordinated coexistence with other licensed WiMAX users) and coordinated coexistence among 802.16h systems. Improved coexistence is achieved through a set of techniques including detection of channel activity by other users, channel reselection, scheduling of channel measurement, requesting and reporting measurements from different nodes.

A significant effort has been devoted for the release, in 2012, of the first IEEE cognitive wireless communication standard: the **IEEE 802.22** standard for fixed wireless regional area networks (WRAN) operating in the TV white spaces. The standard is currently being implemented and compliant devices are being developed.

Just like 802.16, the structure of 802.22 networks includes infrastructure devices (base stations) and client devices. At the physical layer, IEEE 802.22 uses coded OFDM transmission. It shares with the IEEE 802.16 (WiMAX) a subset of channel coding and modulation schemes. Moreover, its OFDMA-based multiple access schemes exhibit similarities with 802.16.

Being a long-range communication standard, protection of primary users in 802.22 systems is very challenging. In fact, both spectrum sensing and access to geo-referenced spectral occupation data bases, or *white spaces data bases* (WSDB), are mandatory methods. Strict requirements on the accuracy of location information of both BS and terminal equipment are defined as well as the detection characteristics of the sensing units.

In order to allow for incumbent detection to be carried out, so-called *quiet periods* of different durations are adaptively scheduled during transmission. During such periods the whole network is muted and spectrum sensing is carried out.

Furthermore, network coexistence among 802.22 systems is supported through beaconing: the Coexistence Beacon Protocol (CBP) is used to inform other systems of the presence of an 802.22 system operating in a given area and thus avoid coexistence issues.

While the 802.22 is devoted to medium- and long-range operation, the recent **IEEE 802.11af** has been developed with the purpose of providing indoor and short-range outdoor wireless access. This new standard (a.k.a. WhiteFI, or Super Wi-Fi) is a recent effort carried out by IEEE to extend the 802.11 PHY and MAC layers with cognitive features in order to allow broadband communication in the sub-GHz channels available as TV white spaces. The standardization process is ongoing and the first release is at the time of writing, expected by early 2014.

Sub-GHz frequencies are highly preferred as long as they enjoy favorable propagation conditions with respect to the 2.4 GHz Industrial, Scientific and Medical (ISM) bands, are typically used by conventional Wi-Fi systems, thus enabling oper-

ation at lower power and achieving longer ranges at the same time. For this reason, the range of 802.11af operating bands includes frequencies below 1 GHz.

Following the typical 802.11 network structure, a 802.11af system consist of access points (AP) and stations (STA). The latter can be dynamically enabled or disabled by other nodes in the network according to a procedure aimed at protecting primary users from secondary user interference. Primary user protection in 802.11af relies on geo-referenced spectral occupation data bases or White Space Data Bases (WSDB). APs query such DBs before sending activation messages. The use of WSDB requires the availability of location information at the enabling STA and/or AP, while spectrum sensing is not required. Each AP may query the DB for channel availability information of its current location and of multiple neighboring locations, thus acquiring knowledge about the available channels in the whole area in which it operates.

The PHY layer of 802.11af [41] is designed to meet strict requirements in terms of adjacent channel leakage ratio (ACLR), support for TV channel widths, support for contiguous and non-contiguous transmission over multiple TV channels. According to [22], the chosen approach was to reuse the 802.11ac PHY in its 40 MHz configuration, reducing the signal bandwidth through sampling frequency scaling. This way, the so-obtained longer OFDM symbols and corresponding guard intervals are able to tolerate the longer channel time dispersion experienced in outdoor contexts. In fact, in addition to the typical indoor (WLAN) deployment scenario, 802.11af is also able to operate in a short-range outdoor configuration.

Coexistence among heterogeneous systems like the 802.22 and 802.11af is a challenging issue being faced in standardization processes like the 802.19.1 and research works [21]. To avoid mutual interference, different competing networks should be forced to use different channels and hence suitable schemes should be adopted. Starting from the observation that both the 802.22 BS and CPE are required to have a secondary sensing antenna which happens to be underutilized, [23] proposes to use such antenna for the transmission of a busy tone that would prevent other systems like 802.11af to start using that channel.

Recently, a new standardization effort carried out by a consortium of companies called **Weightless** (TM) [44] has led to the release of a new cognitive wireless standard for machine-to-machine (M2M) networking. The network structure consists of master nodes (base stations) connected to a high number of slave devices. The use of white spaces results in extended coverage, while the wireless protocol has been designed to be easily implemented in low-power and low-cost devices.

Standards will hopefully develop into devices and then into market products. At the time of writing, we are at a stage where products are still not available on the mass market, although several companies have shown interest and likely will soon be involved in the development of new products.

The prototype radios and software platforms of the Shared Spectrum Company [37] are a first example of cognitive radio products developed for and made available to US government agencies, research organizations and private sector companies.

Few market products are currently available: Neul [30], is delivering a broadband wireless system (infrastructure and terminal devices) that operate in the TV white

spaces. Some Neul products incorporate the Weightless™ air interface. Carlson Wireless [6] proposes RuralConnect™, a white-space system capable of delivering 16 Mbits/s of data rate over ranges of several tens of miles.

Many companies are devoting significant efforts in supporting standardization activities related to cognitive wireless standards, therefore we expect that new products will be available on the market in a short time.

15.5 Conclusions

Cognitive Radio Testbeds provide the means for evaluating CR systems. Designing and implementing a CR testbed can be challenging. It requires well defined requirement analysis of the application is intended for, a robust design based on the best suited platform, a quick yet flexible and upgradeable implementation and finally an exhausive testing phase. Testbeds are consisted of software and hardware CR platforms that, when combined together, can become a powerful development tool. Testbets play possibly the most significant role towards the establishment of cognitive radios since they provide significant findings with respect to the performance of the system and its impact on legacy systems but, most of all, they are contributing towards raising the confidence of regulators to proceed with the legal framework and allow potential use of the spectrum by CR enabled systems. The current research community developments of customised, highly efficient CR engines running on off-the-shelf CR hardware platforms, is effectively driving the hardware industry towards the development of newer, better and cheaper hardware platforms. Although there is still a long way to go to, it is a fact that CR technology has matured enough to be effectively employed in some of the wireless products used in our everyday life. New wireless standards appear to include specific cognitive features, like opportunistic transmission based on interference avoidance to primary users. Nevertheless, standards do not define any intrinsic characteristic of the cognitive engine, but rather specify the absolute constraints that the system must comply within order to protect primary users and to neatly coexist with other similar systems. Further research and development of new CR algorithms and CR platforms along with the development of new CR standards will hopefully materialize into devices and then into market products. Although there are no products available on the mass market, with the technological advances in CR hardware and the continuous development of new CR algorithms, it is most likely that soon the new testbeds will be able to satisfy the regulator's requirements thus fully CR enabled devices will materialize in our daily lives.

References

1. Amanna, A.E., Ali, D., Gadhiok, M., Price, M., Reed, J.H.: Cognitive radio engine parametric optimization utilizing Taguchi analysis. EURASIP J. Wirel. Commun. Netw. 2012, 2012:5, doi:10.1186/1687-1499-2012-5. Available via http://jwcn.eurasipjournals.com/content/pdf/1687-1499-2012-5.pdf. Accessed: 29 May 2013
2. ASGARD. http://asgard.lab.es.aau.dk/joomla/. Accessed 5 June 2013
3. BEE2. http://bee2.eecs.berkeley.edu/. Accessed 26 May 2013
4. BEEcube website. http://beecube.com. Accessed 15 June 2013
5. Cabric, D., Taubenheim, D., Cafaro, G., Farrel, R.: Cognitive radio platforms and testbeds. In: Wyglinski, A.M., Nekovee, M., Hou, T. (eds.) Cognitive Radio Communications and Networks-Principles and Practice, pp. 539–558. Elsevier, Burlington (2010)
6. Carlson Wireless. Online: http://www.carlsonwireless.com. Accessed 15 June 2013
7. Cattoni, A.F., Nguyen, H.T., Duplicy, J., Tandur, D., Badic, B., Balraj, R., Kaltenberger, F., Latif, I., Bhamri, A., Vivier, G., Kovacs, I.J., Horvath, P.: Multi-user MIMO and Carrier Aggregation in 4G systems: The SAMURAI approach. In: Paper Presented at the 2012 IEEE Wireless Communications and Networking Conference (WCNC), pp. 1–4. Paris, France (2012)
8. Challapali, K., Philips Research North America (private communication). FAQ on CogNeA Alliance, K. Kimyacioglu, Philips
9. Clancy, C.T., Hecker, J., Stuntebeck, E., O'Shea, T.: Applications of machine learning to cognitive radio networks. IEEE Wirel. Commun. **14**(4), 47–52 (2007). doi:10.1109/MWC.2007.4300983
10. CogNeA. http://www.cognea.com/. Accessed 7 June 2013
11. CROSS - Wireless @ Virginia Tech. http://cornet.wireless.vt.edu/trac/wiki/Cross. Accessed 2 June 2013
12. Darpa XG Program. http://xg.csl.sri.com/. Accessed 12 May 2013
13. Denker, G., Elenius, D., Senanayake, R., Stehr, M.O., Talcott, C., Wilkins, D.: Cognitive Policy Radio Language (CoRaL) A Language for Spectrum Policies, XG Policy Language. ICS-16763-TR-07-001, SRI, International, April 2007
14. Denker, G., Elenius, D., Senanayake, R., Stehr, M.O., Wilkins, D.: A policy engine for spectrum sharing. In: Paper Presented at the IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN 2007), pp. 17–20. Dublin, Ireland (2007)
15. ECMA-392: MAC and PHY for Operation in TV White Space. Online: http://www.ecma-international.org/publications/standards/Ecma-392.htm. Accessed 16 June 2013
16. ETTUS Research — USRP2. https://www.ettus.com/product. Accessed 4 June 2013
17. eyesIFXv2 node characteristics. Online: http://www.crew-project.eu/content/hardware-and-testbed-lay-out. Accessed 16 June 2013
18. FP7-CREW Project. http://www.crew-project.eu/overview. Accessed 28 May 2013
19. Friend, D.H., EINainay, M.Y., Shi, Y., MacKenzie, A.B.: Architecture and Performance of an Island genetic algorithm-based cognitive network. In: Paper Presented at the 5th IEEE Consumer Communications and Networking Conference, pp. 10–12. Las Vegas, Nevada (2008)
20. GNU Radio. http://gnuradio.org/redmine/projects/gnuradio/wiki. Accessed 7 June 2013
21. IEEE 802.11af PHY overview. Online: http://standards.ieee.org/develop/project/802.11af.html. Accessed 16 June 2013
22. IEEE 802.11af PHY overview. Online: https://mentor.ieee.org/802.11/dcn/12/11-12-0866-02-00af-phy-overview.pptx. Accessed 16 June 2013
23. IEEE 802.19.1. Online: http://www.ieee802.org/19. Accessed 16 June 2013
24. IEEE Standard for Local and metropolitan area networks Part 16. Air Interface for Broadband Wireless Access Systems Amendment 2: Improved Coexistence Mechanisms for License-Exempt Operation. IEEE Std 802.16h-2010 (Amendment to IEEE Std 802.16-2009). July 30 2010.10.1109/IEEESTD.2010.5538195
25. IRIS weblog. http://irissoftwareradio.wordpress.com. Accessed 15 June 2013

26. Le, B., Rondeau, T.W., Maldonado, D., Scaperoth, D., Bostian, C.W.: Signal recognition for cognitive radios. In: Paper Presented at the Software Defined Radio Forum, pp. 13–15. Orlando, Florida (2006)
27. Maldonado, D., Le, B., Hugine, A., Rondeau, T.W., Bostian, C.W.: Cognitive radio applications to dynamic spectrum allocation. In: Paper Presented at the 1st IEEE International Symposium on New Frontiers in Dynamic Spectrum Access (DySPAN 2005), pp. 8–11. Baltimore, Maryland (2005)
28. McHenry, M., Livsics, E., Nguyen, T., Majumdar, N.: XG Dynamic spectrum access field test results. IEEE Commun. Mag. **45**(6), 51–57 (2007). doi:10.1109/MCOM.2007.374432
29. Mishra, S.M., Cabric, D., Chang, C.: A real time cognitive radio testbed for physical and link layer experiments. In: Paper Presented at the 1st IEEE International Symposium on New Frontiers in Dynamic Spectrum Access (DySPAN 2005), pp. 8–11. Baltimore, Maryland (2005)
30. Neul. Online: http://www.neul.com. Accessed 15 June 2013
31. Newman, T.R., Barker, B.A., Wyglinski, A.M., Agah, A., Evans, J.B., Minden, G.J.: Cognitive engine implementation for wireless multicarrier transceivers. Wiley J. Wirel. Commun. Mobile Comput. **7**(9), 1129–1142 (2007)
32. Newman, T.R., Evans, J.B., Wyglinski, A.M.: Reconfiguration, adaptation, and optimization. In: Wyglinski, A.M., Nekovee, M., Hou, T. (eds.) Cognitive Radio Communications and Networks-Principles and Practice, pp. 177–198. Elsevier, Burlington (2010)
33. Robert, M., Sayed, S., Aguayo, C., Menon, R., Channak, K., Valk, C.V., Neely, C., Tsou, T., Mandeville, J., Reed, J.H.: OSSIE: Open source SCA for researchers. In: Paper Presented at the Software Defined Radio Forum, Phoenix (2004)
34. Rondeau, T.W., Le, B., Maldonado, D., Scaperoth, D., Bostian, C.W.: Cognitive radio formulation and implementation. In: Paper Presented at the 1st International Conference on Cognitive Radio Oriented Wireless Networks and Communications, pp. 8–10. Mykonos Island, Greece (2006)
35. Rondeau, T.W.: Application of artificial intelligence to wireless communications, Ph.D. dissertation, Virginia Polytechnic Institute and State University, Blacksburg (2007)
36. SensorLab — "VESNA". http://sensorlab.ijs.si/hardware.html. Accessed 5 June 2013
37. Shared Spectrum Company. Online: http://www.sharedspectrum.com/. Accessed 16 June 2013
38. Si C, Wyglinski, A.M.: Cognitive radio-enabled distributed cross-layer optimization via genetic algorithms. In: Paper Presented at the 4th International Conference on Cognitive Radio Oriented Wireless Networks and Communications, pp. 22–24. Hannover, Germany (2009)
39. Solc, T., Padrah, Z.: Network design for the LOG-a-TEC outdoor testbed. In: Paper Presented at the 2nd International Workshop on Measurement-based Experimental Research, Methodology and Tools (MERMAT 2013), Dublin, Ireland (2013)
40. Stuntebeck, E., O'Shea, T., Hecker, J., Clancy, C.T.: Architecture for an open-source cognitive radio. In: Paper presented at the Software Defined Radio Forum, pp. 13–15. Orlando, Florida (2006)
41. Supported GNU Radio hardware. http://comsec.com/wiki?GnuRadioHardware. Accessed 15 June 2013
42. TMote Sky Data sheet. Online: http://www.crew-project.eu/sites/default/files/tmote-sky-datasheet.pdf. Accessed 16 June 2013
43. WARP website. http://warpproject.org. Accessed 15 June 2013
44. Weightless. Online: http://www.weightless.org. Accessed 15 June 2013
45. Wireless @ Virginia Tech. http://wireless.vt.edu/. Accessed 2 June 2013
46. Young, A.R., Kaminski, N.J., Fayez, A., Bostian, C.W.: CSERE (Cognitive System Enabling Radio Evolution): A modular and user-friendly cognitive engine. In: Paper presented at the 6th IEEE International Symposium on New Frontiers in Dynamic Spectrum Access (DySPAN 2012), pp. 16–19. Bellevue, Washington, Baltimore, (2012)
47. Zhao, Z., Peng, Z., Zheng, S., Shang, J.: Cognitive radio spectrum allocation using evolutionary algorithms. IEEE Trans. Wirel. Commun. **8**(9), 4421–4425 (2009). doi:10.1109/TWC.2009.080939